DZONGKHA SEGMENTS AND TONES: A PHONETIC AND PHONOLOGICAL INVESTIGATION

KINLEY DORJEE

Supervisor
PROFESSOR K.G. VIJAYAKRISHNAN
Department of Linguistics and Contemporary English
Hyderabad

Co-supervisor
Dr. T. Temsunungsang
The English and Foreign Languages University
Shillong Campus

A dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Linguistics
SCHOOL OF LANGUAGE SCIENCES

THE ENGLISH AND FOREIGN LANGUAGES UNIVERSITY
HYDERABAD 500605, INDIA

JULY 2011
To my mother
ABSTRACT

In this thesis, we make, for the first time, an acoustic investigation of supposedly unique phonemic contrasts: a four-way stop phonation contrast (voiceless, voiceless aspirated, voiced and devoiced), a three-way fricative contrast (voiceless, voiced and devoiced) and a two-way sonorant contrast (voiced and voiceless) in Dzongkha, a Tibeto-Burman language spoken in Western Bhutan. Paying special attention to the ‘Devoiced’ (as recorded in the literature) obstruent and the ‘Voiceless’ sonorants, we examine the durational and spectral characteristics, including the vowel quality (following the initial consonant types), in comparison with four other languages, viz., Hindi, Korean (for obstruents), Mizo and Tenyidie (for sonorants).

While the ‘devoiced’ phonation type in Dzongkha is not attested in any language in the region, we show that the devoiced type is very different from the ‘breathy’ phonation type, found in Hindi. However, when compared to the three-way voiceless stop phonation types (Tense, Lax and Aspirated) in Korean, we find striking similarities in the way the two stops (‘Devoiced’ and ‘Lax’) employ their acoustic correlates. We extend our analysis of stops to fricatives, and analyse the three fricatives in Dzongkha as: Tense, Lax and Voiced.

In sonorants (voiced-voiceless distinctions), we show that while the voiceless nasals in Dzongkha pattern with voiceless nasals in Mizo, as pre-aspirated types, the voiceless laterals pattern with voiceless laterals in Tenyidie, as post-aspirated. We also show that the voiced-voiceless distinction in sonorants can be made with two acoustic cues, viz., the sonorant duration and the different phonetic composites of the voiceless sonorants.

Having established the segmental contrasts, we finally make an acoustic study of the tones in Dzongkha. We show Dzongkha to be an incipient, lexically bitonal language, supported by the lexical tone specification on the syllables with voiced sonorant onsets. In obstruents, the ‘Tense’ group occurs in the High Register while the ‘Lax’ group occurs in the Low Register.
ACKNOWLEDGEMENTS

I remain obliged to a number of people in the course of completing this thesis. First and foremost, I thank Prof. K.G. Vijayakrishnan, my guide and mentor, for his constant encouragement, advice and support. Vijayakrishnan’s intellect and wisdom is seen in whatever is exceptional in this thesis. His exceptional intellect and wisdom prodded me through the hardest parts of framing the thesis topic and research questions, and finally the analyses. Apart from being an exceptional guide, Vijayakrishnan’s passion for linguistics is outstanding. I am indebted to him for showing me so many directions and yet leaving me to choose the way I want to go. And above all, I will always admire and appreciate him for not only being an exceptional teacher and guide but also for being a very reasonable human being.

I will never be able to thank my co-supervisor, mentor and friend, Dr. Temsunungsang (Temsu) enough for his help and support right from my day one in Hyderabad till the day I submitted my thesis. He has been there in whatever help I needed during my stay in Hyderabad, both academics and personal (I shall forever remain grateful to you for that). Temsu has tirelessly and meticulously read all the chapters of my thesis (more as a friend than a co-supervisor) over and over. His invaluable comments and feedback contributed enormously in building this thesis (Thanks, Temsu).

I am also indebted to a number of professors at the EFL University for preparing me for my thesis through various coursework modules: Prof. Hemalatha Nagarajan (in Phonology), Prof. Komali Prakash (in Articulatory Phonetics), Prof. Meena Debashish (in Acoustic Phonetics) and Prof. Rahul Balusu (in Research Methodology and Praat software). Prof. Rahul Balusu has also been extremely generous not only in solving my technical impediments with my data, but also for allowing me to use his recording system for the entire period of my data collection in Bhutan (Thanks you, Rahul).

Thanks are also due to linguistics research scholars who have always been around sharing their wisdom and knowledge over coffee and snacks. special thanks to Mukul, Srinivas, Srivatsa, Rajiv, and Masrook.
Over the years, the encouragement and good times I have had with my friends had a very vigorous effect on my general wellbeing, and therefore added vigour to my research. Special thanks are due to Dhammapal (for always being concerned more about my health and wellbeing than his own), Suchi (for always being there to listen to, and make delicious tea for me), Mrinal and Geetashri (for always being true friends, and lightening up my disposition) and Shardool (who would veer any conversation into something humorous and have me laugh, making my days enjoyable ones). Thanks, friends.

My days in Hyderabad would have been difficult if I did not have my country fellows around, always inviting me over their places for delicious home-food; thanks to Tandin, Rinchen, Bharat and Sonam for giving me good company and feeding me with home-food (making a good break from the hostel mess' messy food!). My gratitude to Chimi, Doctor, Bijay, Kunga and all others at Survey of India for the memorable weekends I have had with you all. I will always remain grateful to all the BSA (Hyderabad) members for the wonderful parties we have had at Celebrity Club and other places.

In the course of collecting data from Bhutan, numerous people were involved; thanks to Chechey, Deki, Sangay Om, Tenzin Pem, Tshering Lham, Kinley Tshering, Tshering Penjor, Chencho, Chimi Tsewang (also for arranging informants from Shari Higher Secondary School, Paro; I also thank you for making calls from Bhutan every now and then to make sure I was doing fine), Kunzang, Tandin, (Sister) Cheche and Kinley Situp for willingly sparing your precious time for the recordings.

And finally, a very special thanks to my brother Kinley Tshering, who did not mind sponsoring huge amount of money for my train and air tickets during vacations, of course, with the condition that I worked hard towards my PhD project. Thank you, bro!

While obviously, several individuals, entities and institutions are also responsible for all that is good in this thesis, I hold sole responsibility for all the shortcomings in this thesis
LIST OF ABBREVIATIONS

F  Formant
F0 Fundamental Frequency
H1-H2 Difference between the amplitude of the First Harmonic
      and the Second Harmonic
F  Female
Hon Honorific
IMP Imperative (mood)
M  Male
N  Noun
INTR Intransitive (verb)
TR Transitive (verb)
VOT Voice Onset Time
V  Verb
VLT Voiceless Tense
VLL Voiceless Lax
VLAS Voiceless Aspirated
VD Voiced
LIST OF TABLES

1.1 Languages spoken in Bhutan ........................................ 7
2.1 Consonant phonemes (Mazaudon 1985) .......................... 16
2.2 Consonant phonemes (Mazaudon & Michailovsky 1988) ...... 17
2.3 Consonant phonemes (Van Driem 1992) ......................... 18
2.4 Consonant phonemes (Watters 2002) ............................ 19
2.5 Dzongkha consonant phonemes .................................... 23
2.6 Formant (F1, F2 and F3) frequencies for Dzongkha vowels .. 34
2.7 Classification of Dzongkha vowels ................................. 36
2.8 Dzongkha vowel length contrast in open monosyllabic words 37
2.9 Dzongkha vowel length contrast in closed monosyllabic words 37
2.10 Dzongkha vowel nasalization contrast ............................ 38
2.11 length contrast in nasalized vowels ............................... 38
2.12 Dzongkha diphthongs ............................................... 39
2.13 Dzongkha vowel phonemes ........................................ 40
3.1 English voiced stop phonation ...................................... 49
3.2 Hindi breathy stop phonation ....................................... 50
3.3 Creaky nasal phonation in Kwakw’ala ............................ 51
3.4 Glottal stop contrast in Lawa ....................................... 51
3.5 Glottal stop contrast in Ao ......................................... 51
3.6 Tense stop contrast in Korean ...................................... 52
3.7 Stop inventories of some modern Tibeto-Burman languages 56
3.8 Average VOT values of stop phonation types .................... 60
3.9 Average stop burst energy of Dzongkha stops .................. 63
3.10 Mean closure duration of stops ................................... 66
<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.11</td>
<td>Manipulated F0 of Dzongkha stops 78</td>
</tr>
<tr>
<td>3.12</td>
<td>Interaction of acoustic correlates for Dzongkha stops 93</td>
</tr>
<tr>
<td>3.13</td>
<td>Dzongkha fricative duration 101</td>
</tr>
<tr>
<td>3.14</td>
<td>Comparison of acoustic cues in Dzongkha and Korean 108</td>
</tr>
<tr>
<td>4.1</td>
<td>Phonemic contrast of sonorants in Dzongkha, Mizo and Tenyidie 117</td>
</tr>
<tr>
<td>4.2</td>
<td>Typology of nasals and laterals in Tibeto-Burman languages 122</td>
</tr>
<tr>
<td>4.3</td>
<td>Nasal duration in Dzongkha 129</td>
</tr>
<tr>
<td>4.4</td>
<td>Nasal duration in Mizo 133</td>
</tr>
<tr>
<td>4.5</td>
<td>Mean duration of laterals in Dzongkha 138</td>
</tr>
<tr>
<td>4.6</td>
<td>Duration of laterals in Mizo 141</td>
</tr>
<tr>
<td>5.1</td>
<td>Significance of phonetic pitch contour 183</td>
</tr>
<tr>
<td>5.2</td>
<td>Vowel length of ‘quarrel’, ‘east’, ‘paddy’ and ‘summer’ 188</td>
</tr>
<tr>
<td>5.3</td>
<td>Pitch height significance in the high register 191</td>
</tr>
<tr>
<td>5.4</td>
<td>Pitch height significance in the low register 193</td>
</tr>
</tbody>
</table>
LIST OF FIGURES

1.1 Map of Bhutan 4
1.2 Linguistic map of Tibeto-Burman languages 5
1.3 Political map of Bhutan and her neighbouring states 5
1.4 Classification of Dzongkha in Tibeto-Burman 9
1.5 Dzongkha orthography system 11
2.1 Vowel phonemes (Mazaudon 1985) 31
2.2 Vowel phonemes (Mazaudon & Michailovsky 1988) 31
2.3 Vowel phonemes (Van Driem 1992) 31
2.4 Vowel phonemes (Watters 2002) 32
2.5 F1 – F2 Plot of Dzongkha vowels 35
2.6 Continuum of phonation types 47
2.7 Degree of vocal fold vibration in phonation types 48
3.1 Histogram of VOT differences in Dzongkha stops 60
3.2 Histogram of VOT differences in Dzongkha stops 61
3.3 Spectrogram and waveforms of VOT differences in stop types 61
3.4 Histogram of stop burst in Dzongkha 64
3.5 Histogram of closure duration of Dzongkha stops 66
3.6 H1-H2 in the vowel following four stop types 68
3.7 H1-H2 measurement and waveforms following stop types 70
3.8 Manipulated F0 for the stops 77
3.10a Histogram showing result (in %) of the perception study 78
3.10b Histogram showing result of perception of stimuli 79
3.11a Standard laryngeal feature 90
3.11b Laryngeal feature after Avery and Idsardi (2001) 92
3.12 Histogram of fricative duration difference 101
3.13 Spectrogram and waveform of fricatives showing durational differences

3.14 H1-H2 of the vowel following fricative types

3.15 Waveform of H1-H2 differences in the vowel following the fricatives

4.1a Histogram of the duration of nasals in Dzongkha

4.1b Bar chart of the duration of [ŋ], [h] and [n] in Dzongkha

4.1c Spectrogram and waveform of the nasal types in Dzongkha

4.1d Average percentage of [ŋ], [h] and [n] in /ŋ/ in Dzongkha

4.2a Bar chart of the duration of /n/ and /ŋ/ in Mizo

4.2b Waveform and spectrogram of the nasal types in Mizo

4.2c Bar chart of the duration of [ŋ] and [n] in /ŋ/ in Mizo

4.2e Average % of [ŋ] and [n] in /ŋ/ in Mizo

4.3a Bar chart of the mean duration of laterals in Dzongkha

4.3b Normalized mean duration of laterals in Dzongkha

4.3c Phonetic composites of voiceless laterals in Dzongkha

4.3d Waveform and spectrogram of laterals in Dzongkha

4.4a Bar chart showing duration of laterals in Mizo

4.4b Mean duration of [l] and [l] in Mizo

4.4c Waveform and spectrogram of lateral types in Mizo

5.1 Pitch pattern of the vowel following initial stop types

5.2 Pitch pattern of the vowel following fricative types

5.3 Pitch pattern of the vowel following lateral types

5.2A Pitch of /aː/ 'quarrel' in Dzongkha

5.2B Pitch of /aː/ 'east' in Dzongkha

5.2C Pitch of /bəʔ/ 'paddy' in Dzongkha
5.2D Pitch of /bza:/ ‘summer’ in Dzongkha
5.3 Glottal stop in ‘quarrel’ and ‘paddy’ in Dzongkha
5.4 Vowel length difference in Dzongkha
5.5 With-in register pitch height difference in Dzongkha
TABLE OF CONTENTS

ABSTRACT i
ACKNOWLEDGEMENT ii
LIST OF ABBREVIATIONS iv
LIST OF TABLES v
LIST OF FIGURES vii
TABLE OF CONTENT x

Chapter 1: Introduction

1.1 The language, its location and other languages of the Dzongkha group
   1.1.1 The language 2
   1.1.2 The location 4
   1.1.3 Other languages of the Dzongkha group in Bhutan 6
1.2 Genetic affiliation 8
1.3 Dzongkha script and orthography 9
1.4 Previous research 11
1.5 Methodology and data 12
1.6 Organization of the thesis 13

Chapter 2: Segmental inventory: a review

2.1 Introduction 15
2.2 Consonant phonemes 16
   2.2.1 Differences in consonant inventories 20
2.2.2 Stops

2.2.3 Affricates

2.2.4 Fricatives

2.2.5 Nasals

2.2.6 Approximants

2.3 Consonant Phonemes: the consolidated inventory

2.3.1 Stops

2.3.2 Affricates

2.3.3 Fricatives

2.3.4 Nasals

2.3.5 Liquids

2.3.6 Glides

2.4 Vowel Phonemes

2.5 Vowel phonemes: the consolidated inventory

2.5.1 Vowel formants: facts about vowel formants and vowel quality

2.5.2 Formant plot experiment

2.6 Vowel length and nasalization

2.6.1 Vowel length

2.6.2 Nasalized vowels

2.7 Diphthongs

2.8 On the size of Dzongkha segment inventory

2.9 Conclusion
# Chapter 3: Obstruent phonation types

<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.1</td>
<td>Introduction</td>
<td>43</td>
</tr>
<tr>
<td>3.2</td>
<td>Explaining the term ‘Phonation’</td>
<td>45</td>
</tr>
<tr>
<td>3.2.1</td>
<td>Phonation types</td>
<td>46</td>
</tr>
<tr>
<td>3.2.2</td>
<td>Linguistic use of phonation types</td>
<td>48</td>
</tr>
<tr>
<td>3.3</td>
<td>Dzongkha obstruent phonation: VLT*, VLL, VLAS and VD</td>
<td>52</td>
</tr>
<tr>
<td>3.3.1</td>
<td>Literature on stop phonation types in Dzongkha</td>
<td>53</td>
</tr>
<tr>
<td>3.3.2</td>
<td>Typology of stop phonation types in Tibeto-Burman languages</td>
<td>55</td>
</tr>
<tr>
<td>3.4</td>
<td>Production study of Dzongkha stop phonation</td>
<td>57</td>
</tr>
<tr>
<td>3.4.1</td>
<td>Voice Onset Time (VOT): a brief review</td>
<td>58</td>
</tr>
<tr>
<td>3.4.2</td>
<td>VOT in Dzongkha</td>
<td>59</td>
</tr>
<tr>
<td>3.5</td>
<td>Stop burst in Dzongkha</td>
<td>63</td>
</tr>
<tr>
<td>3.6</td>
<td>Closure duration in Dzongkha</td>
<td>65</td>
</tr>
<tr>
<td>3.7</td>
<td>Spectral tilt (H1-H2)</td>
<td>67</td>
</tr>
<tr>
<td>3.7.1</td>
<td>H1-H2 study</td>
<td>68</td>
</tr>
<tr>
<td>3.7.2</td>
<td>F0 perturbation theory</td>
<td>72</td>
</tr>
<tr>
<td>3.7.3</td>
<td>Summary of Dzongkha stop phonation study</td>
<td>74</td>
</tr>
<tr>
<td>3.7.4</td>
<td>Perception study of the stop phonation types in Dzongkha</td>
<td>75</td>
</tr>
<tr>
<td>3.8</td>
<td>Summary of stop phonation type in Hindi</td>
<td>80</td>
</tr>
<tr>
<td>3.8.1</td>
<td>Comparison of Dzongkha voiceless lax stop with Hindi voiced aspirated stop</td>
<td>83</td>
</tr>
<tr>
<td>3.9</td>
<td>Summary of Korean stop phonation</td>
<td>84</td>
</tr>
</tbody>
</table>
Chapter 4: Sonorant phonation types

4.1 Introduction 114

4.2 Tenyidie and Mizo: Language background 116

4.2.1 Phonological contrast of nasals and laterals in Dzongkha, Mizo and Tenyidie 117

4.3 Cross-linguistic typology of nasals and laterals 199

4.3.1 Nasals 199

4.3.2 Laterals 120

4.4 Typology of sonorants in Tibeto-Burman languages 121

4.5 Background literature on sonorant phonation types 125

4.6 Acoustic study of nasals 128
4.6.1 Nasal duration in Dzongkha 128
4.6.2 Nasal duration in Mizo 132

4.7 Acoustic study of laterals 137
  4.7.1 Dzongkha lateral duration 137
  4.7.2 Mizo lateral duration 141
  4.7.3 Comparison of sonorant phonation types in Dzongkha and Mizo 144

4.8 Phonological representation of Dzongkha nasals and laterals 146
  4.8.1 Typological overview 146
  4.8.2 Theoretical preliminary 147
  4.8.3 Feature representation of Dzongkha nasals and laterals 149

4.9 Conclusion 150

Chapter 5: Acoustic study of tones
5.1 Introduction 154

5.2 Earlier analysis of Dzongkha tone system 156

5.3 Dzongkha initial consonants and the Register contrast 158
  5.3.1 Stop initials 158
  5.3.2 Fricative initials 162
  5.3.3 Sonorant (lateral) initials 164
  5.3.4 Discussion on the acoustic study 167
  5.3.5 Initial phonation types and Register tone: a historical perspective 167

5.4 Pitch pattern study 171
  5.4.1 Acoustic study of the pitch pattern 172
5.4.2 Glottal stop /ʔ/ in Dzongkha
5.4.3 Vowel duration
5.4.4 Phonetic pitch height contrast within register
5.4.5 Historical perspective on the development of Intra-Register contrasts

5.5 Phonological analysis of Dzongkha tones
5.5.1 Phonological interpretation of the phonetic tones in Dzongkha

5.6 Conclusion and looking forward

Bibliography
CHAPTER ONE

INTRODUCTION

The study of linguistic sound patterns is traditionally divided into two sub-fields: 'Phonetics' and 'Phonology'. The distinction usually made between phonetics and phonology is that phonetics concerns the study of the physical or physiological aspects of speech, including its articulatory, aerodynamic, acoustic, auditory, and perceptual aspects; whereas phonology is concerned with accounting for the variation in speech sounds in different but related languages and dialects, and within a given language in the environment of different morphemes, different positions within an utterance, word, or other speech sounds (Ohala 2005). While traditionally, the two disciplines were regarded as quite distinct, there has been a relatively improved acceptance amongst linguists regarding the role of phonetics in the theory of phonological assumptions. One such contribution of phonetics in phonology has been in the formulation of Distinctive Feature Theory. While feature theory since Chomsky & Halle (1968) is primarily guided by the articulatory perspective, there has been a number of contributions in the recent past in which the acoustic (or auditory) and perceptual perspective of distinctive features is explored further (for example, Kingston & Diehl 1994, Harris & Lindsey 1995, Traill 1995, Silverman 1997, Boersma 1998, Jessen 2001). Phonologists and phoneticians alike recognize the importance of acoustic evidence as a prerequisite to the study of acoustic features and feature correlates. "It is on the basis of reliable acoustic data that one can proceed with auditory modelling, categorical perception experiments, and related methods" Jessen (2001). There is no doubt that phonological analysis requires information on phonetic categories, including segments as well as prosodic units such as syllables and tones. According to Ladefoged (2004), "Phonetics and phonology meet most obviously in the definition of the set of features used to describe phonological processes". Then the most significant aspect of integrating phonetics and phonology, perhaps, is that the accounts given for speech sound behaviour are empirically testable, and acoustic phonetic findings can be progressively...
employed in testing and claiming phonological theories, such as distinctive feature theory.

In this thesis, we examine the acoustic cues of segments (consonants only) and tones in Dzongkha, the national language of Bhutan, a Tibeto-Burman language spoken in western Bhutan. Of the various regional dialects of Dzongkha, we focus our study in one dialect, Paro dialect, spoken in Paro district (see map of Bhutan in figure 1.1). We make an acoustic study of the initial consonant types (obstruents and sonorants) and extend our findings in the acoustic analysis of the tones in Dzongkha. We also propose, based on our acoustic correlates, a new analysis of the segment types, especially the three voiceless stops (the existing ‘voiceless’, ‘voiceless aspirated’ and ‘devoiced’) along with the voiced one, and analyse the segments in terms of distinctive feature theory. The primary perspective of this thesis, thus, is a pragmatic one, i.e. a focus on contrasts, in the sense of phonemic oppositions.

1.1 The language, its location and other languages of the Dzongkha group

1.1.1 The language

Dzongkha ‘dzon (fortress) k’a (language)’ - the fortress language2 - is the national language of Bhutan. It furthermore serves as lingua franca and official language throughout the kingdom. An essential trait which Dzongkha shares with national languages of other modern countries is a rich literary tradition of great antiquity. Dzongkha derives from Classical Tibetan through many centuries of independent linguistic evolution on Bhutanese soil. Linguistically, Dzongkha can be qualified as the natural modern descendant of Classical

---

1 In the later discussion in chapter 3, we postulate the ‘devoiced’ stops as ‘voiceless lax’ stop (VLL), and the other two as ‘voiceless tense’ (VLT*) and ‘voiceless aspirated’ (VLAS) stops.

2 Literally, Dzongkha means the ‘k’a’ or ‘language’ spoken in the ‘dzon’ or ‘fortress’, the mighty fortresses which dominate the mountainous landscape of Bhutan from east to west. These fortresses have traditionally been both centres of military and political power as well as centres of learning. Dzongkha is thus the cultivated form of the native language of western Bhutan, the inhabitants of which as well their language have traditionally been known as Ngalong /galon/ (meaning, ‘the first to rise’ to the teachings of Buddhism in the land).
Tibetan, the language in which sacred Buddhist texts, medical and scientific treatises and, indeed, all learned works have been written (Van Driem 1992). Because of its historical role Dzongkha has been for centuries the most important language in Bhutan. The status of Dzongkha as the national language of the royal court, the military elite, educated nobility, government and administration is firmly rooted in Bhutanese history at least as far back as the twelfth century.

Dzongkha is the native language of eight of the twenty districts, with 1,60,000 speakers (Van Driem 1992), of Bhutan, viz. Thimphu, Paro, Punakha, Gasa, Wangdue Phodrang, Chhukha, Ha and Dagana (see shaded portion in figure 1.1). Dzongkha has several regional dialects; Dzongkha dialectal variation, in general, can be specific to a certain region or valley or even district, and one dialect can share various similarities with its immediate neighbouring valley's language or stand as a divergent dialect. Thus the dialects differ between the far north and far west, including the dialect of the alpine yak-herd community of Lingshi in the northernmost part of Thimphu district, which, according to Van Driem (1992), much of the dialects of these regions result from structural and lexical similarities with Tibetan. The Dzongkha spoken in Ha has a character of its own, a lot similar to Dranjongke (in Sikkim). Mazaudon & Michailovsky (1988) found some segmental differences between the three dialects of Dzongkha: Paro, Thimphu and Chapcha (in Chhukha district) dialects, while they found that the tones in Thimphu and Chapcha dialects are same, tones in Paro dialect are predicted to be different from the other two. The standard dialect of Dzongkha is believed to be spoken by people in Wang /wɔŋ/ (in Thimphu valley) and The /tʰe/ (in Punakha valley).

---

1 We may note here that while five districts (towards the north) have a complete homogeneous Dzongkha speaking population, the southern belts of Chhukha and Dagana districts have a fair number of Lhotsamkha (Nepali) speaking population (which is not indicated in the map).
1.1.2 The location

Geographically, Bhutan is located in between the two gigantic nations, China (Tibet) in the north and India in the east, west and south; for this reason, the country is quite often referred to as 'the Sandwich' between the two giants. Linguistically, Bhutan lies right at the foot of Tibetan, and its various dialects, enclosed by various other Tibeto-Burman languages of Arunachal Pradesh in the east, Assam in the south and Sikkim in the west, except in the south-west belt, which shares its border (West Bengal, India) with Indo-Aryan language (see figure 1.2 & 1.3). The geographical size of Bhutan is 38,394 square kilometres (National Statistical Bureau 2005). Other countries neighbouring Bhutan without sharing a direct border are Nepal in the west and Bangladesh and Burma in the south and south-east respectively.

---

Figure 1.2: Linguistic map of Tibeto-Burman languages

Figure 1.3: Political map of Bhutan and her neighbouring states

Source: Van Driem (2008)
Source: http://www.ultrapop.org/blog/category/philanthropy/
1.1.3 Other languages of the Dzongkha group in Bhutan

Although this thesis does not concern any perspective on sociolinguistics or dialectology of the languages and dialects of Bhutan, it is worthwhile enumerating the languages, which are closely related to Dzongkha, spoken in Bhutan for two reasons. Firstly, other Tibeto-Burman languages in Bhutan (except few studies in Dzongkha) are scarcely studied, and secondly, besides Dzongkha (which gained both linguists' and bureaucrats' attention as the national language of the Kingdom) other Tibeto-Burman languages are not made known to linguists around the world.

Van Driem (1998:2) enumerates a total of nineteen distinct languages, including Dzongkha, divided into four linguistic sub-groups in Bhutan. The nineteen languages are presented in table 1.1, with their estimated number of speakers and the regions they are spoken (for the location of the region(s) see map of Bhutan in figure 1.1), following Van Driem 1998:2.

The predominant language in Western Bhutan is Dzongkha. The Dzongkha language area covers well over a third of the country's surface area. The predominant language in Eastern Bhutan is Tsangla. The predominant language in Central Bhutan is Bumthangkha, in Bumthang, with the closely related Khengkha and Kurtoebkha. In the southern belt, the predominant language is Lhotsamkha (Nepali). Other important regional languages include Dzalakha in northeastern Bhutan and Nyenkha in Trongsa, Shengang and Wangdi Phodrang, and Cho-ca-nga-cha, a sister language of Dzongkha, spoken in Mongar and Lhuntse.

Amongst the Tibeto-Burman languages listed in table 1.1, only Dzongkha has a written script and is taught in school. The remaining seventeen Tibeto-Burman languages neither have written scripts nor are taught in school. Considering this fact, and the fact that six languages (Brokpake, Bokkat, Boekha, Chalikha, Monpa and Dakpakha) have just a thousand (or less) speakers, these languages may be considered as endangered languages, if we consider the number of speakers as an important criterion for endangered languages.
Table 1.1: Languages spoken in Bhutan

<table>
<thead>
<tr>
<th>Languages spoken in Bhutan</th>
<th>Central Bodish languages</th>
<th>East Bodish languages</th>
<th>Other Bodic languages</th>
<th>Indo-Aryan language</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dzongkha</td>
<td>1,60,000</td>
<td>Thimphu, Paro, Chhukha, Punakha, Ha, Gasa, Dagana and Wangdi Phodrang</td>
<td>Bumthangkha 30,000</td>
<td>Lhotsamkha (Nepali) 1,56,000</td>
</tr>
<tr>
<td>Cho-can-ga-cha</td>
<td>20,000</td>
<td>Mongar &amp; Lhuntse</td>
<td>Khengkha 40,000</td>
<td>Chhukha, Samtse, Dagana, Sarpang &amp; Tsirang</td>
</tr>
<tr>
<td>Brokpake</td>
<td>5,000</td>
<td>Trashigang</td>
<td>Kurtoepkha 10,000</td>
<td></td>
</tr>
<tr>
<td>Brokkat</td>
<td>300</td>
<td>Northern Bumthang</td>
<td>Nyenkha 10,000</td>
<td></td>
</tr>
<tr>
<td>Lakha</td>
<td>8,000</td>
<td>Wangdue Phodrang</td>
<td>Chalikha 1,000</td>
<td></td>
</tr>
<tr>
<td>Boekha (Tibetan)</td>
<td>1,000</td>
<td>Northern Trashi Yangtse</td>
<td>Dzalakha 15,000</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Monpa (Olekha) 1,000</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Dakpakha 1,000</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Eastern languages

- Bumthangkha 30,000
- Khengkha 40,000
- Kurtoepkha 10,000
- Nyenkha 10,000
- Chalikha 1,000
- Dzalakha 15,000
- Monpa (Olekha) 1,000
- Dakpakha 1,000

Other Bodic languages

- Bumthangkha 30,000
- Khengkha 40,000
- Kurtoepkha 10,000
- Nyenkha 10,000
- Chalikha 1,000
- Dzalakha 15,000
- Monpa (Olekha) 1,000
- Dakpakha 1,000

Central Bodish languages

- Central Bodish languages
- Dzongkha 1,60,000
- Cho-can-ga-cha 20,000
- Brokpake 5,000
- Brokkat 300
- Lakha 8,000
- Boekha (Tibetan) 1,000

East Bodish languages

- East Bodish languages
- Bumthangkha 30,000
- Khengkha 40,000
- Kurtoepkha 10,000
- Nyenkha 10,000
- Chalikha 1,000
- Dzalakha 15,000
- Monpa (Olekha) 1,000
- Dakpakha 1,000

Other Bodic languages

- Other Bodic languages
- Tsangla 1,38,000
- Lhokpu 2,500
- Gongduk 2,000
- Lepcha 2,000

Indo-Aryan language

- Indo-Aryan language
- Lhotsamkha (Nepali) 1,56,000
- Chhukha, Samtse, Dagana, Sarpang & Tsirang

Notes:

- **Central Bodish languages**: Dzongkha, Cho-can-ga-cha, Brokpake, Brokkat, Lakha, Boekha (Tibetan)
- **East Bodish languages**: Bumthangkha, Khengkha, Kurtoepkha, Nyenkha, Chalikha, Dzalakha, Monpa (Olekha), Dakpakha
- **Other Bodic languages**: Tsangla, Lhokpu, Gongduk, Lepcha
- **Indo-Aryan language**: Lhotsamkha (Nepali)
1.2 Genetic affiliation

Classification of Tibeto-Burman languages has been not an easy job for linguists, owing to the lack of linguistic data in many languages in the family. Despite this difficulty, linguists have made different attempts to classify the languages affiliated to the family.

Shafer (1966) classifies Sino-Tibetan language family into six main divisions: Sinitic (Chinese), Daic (Thai), Bodic (Tibetan, and other languages related to it), Burmic (Burmese, and other languages related to it), Baric (Bodo, etc) and Karenic (Karen). Of these six divisions, Bodic, Burmic and Baric comprise of the Tibeto-Burman languages, with further sub-divisions.

Bradley (1997) divides Tibeto-Burman into four major groups: North-Eastern India (or Sal group); Western (or Bodic group); South-Eastern, including Burmic and Kuki-Chin-Naga, and North-Eastern group.

In one perspective, Thurgood (2003) classifies the Bodish group into four subgroups called Tibetan, Tamang-Gurung-Thakali-Manang languages, Takpa or Northern Monpa and Others. The Tibetan dialects are further divided into five groups: Classical Tibetan, Central or U-Tsang (Modern Lhasa, Shigatse, Sherpa, Jirel, Lhomi, Nyamkad, Jad, Kagate, etc), Western Archaic (Balti, Ladakhi, Purik/Purki), Western Innovative (Lahul, Spiti, Tod), and Southern dialects of Sikkim and Bhutan (Khams, Amdo, Sikkimese, Dzongkha).

However, in a more recent classification by Van Driem (2001), following Shafer (1966), he brings together the earlier Bodish grouping with the later ‘Tibetan dialects’ group: Central Bodish with Central, Eastern and North Eastern dialects, south Bodish with Southern dialects, West Bodish with Western Archaic and Western Innovative group, the East Bodish, and the standard literary language or Classical Tibetan. In his sub-grouping of the Southern subgroup or South Bodish, he has four languages, viz. Dzongkha.
(Bhutan), Dranjong (Sikkim), Jumowa or Groma of Chumbi Valley (Tibet) and Choca-nga-ca-kha (East Bhutan); this is presented in figure 1.4 below.

Figure 1.4: Classification of Dzongkha in Tibeto-Burman (Van Driem 2001)

The classification and sub-classification of languages under Tibeto-Burman, especially Bodish where Dzongkha is placed, was based on grammatical features widely shared by the languages in the group. The Bodish subgroup shares the ‘*-s’ ablative/ergative suffix on nouns (LaPolla 2003). The Tibetan dialects, a subgroup under the Bodish, share the second person pronoun ‘*khyo’ ‘you’ and a third person singular, roughly *kho ‘he’ (Thurgood 2003). Dzongkha, in written, it’s orthography accepts the –s ablative/ergative suffix, keeping with the literary writing tradition, but the suffix is slowly disappearing from the commonly spoken Dzongkha, where it has become homophonous with the genitive case –gi or –ki without overtly pronouncing the –s suffix (Thinley 2008).

1.3 Dzongkha script and orthography

A brief mention of the script and orthography system of Dzongkha is necessary since the core basis of our data on tonogenesis (in chapter 5) is based on the Dzongkha (or Tibetan) orthographic system. Dzongkha uses the same script and orthographic system as Tibetan (with few spelling modifications). It has
thirty consonant and four vowel characters and is popularly called ‘Ucen’ Script. Ucen script was adapted from the Gupta script of India in the mid 7th century by the Tibetan Minister named Thumi Sambotra (Van Driem 1998). Ucen script seems to be the origin of various cursive and longhand varieties developed later. Three varieties of the Ucen script, viz. ‘Tshum’, ‘Jotshum’ and ‘Joyig’ were developed and are employed to write documents (van Driem 1998). The Joyig, written as ‘rgyug-yig’ (Aris 1979) literally meaning ‘fast-letter’, is a native cursive variety used in Bhutan and is the most widely used script when people write their personal letters and documents. ‘Joyig’ is also employed in formal letters and documents that need to be written quickly by hand. The choice of the script employed to write depends on the occasion. The ‘Tshum’ script, most closely resembling Ucen script in calligraphy, is the most formal form and is used to write religious texts and other important documents.

As in other Indic scripts, each consonant letter includes an inherent /a/. However, a unique aspect of the Dzongkha (or Tibetan) writing system is that the consonants can be written either as radicals, or they can be written in other forms, such as ‘superscripts’ and ‘subscripts’ (see figure 1.5 for illustration). The ‘superscript’ position above a radical is reserved for the consonants r, l, and s, while the ‘subscript’ position under a radical is for the consonants y, r, l, and w. Besides being written as subscripts and superscripts, some consonants can also be placed in ‘prescript’, ‘postscript’, or post-postscript positions. For instance, the consonants g, d, b, m, 'a (a-chung) can be used in the prescript position to the left of other radicals, while the position after a radical (the postscript position), can be held by the ten consonants: g, n, b, d, m, 'a, r, n, s, and l. The third position, the post-postscript position, is solely for the consonants d and s. A sample of the Dzongkha orthography system is illustrated in figure 1.5 in the word: ‘bsgrubs’ pronounced as /dʒup/ (which roughly means ‘to accomplish or build’). It may be remembered that Proto Tibetan was assumed to be pronounced as it was written, thus making it a more phonetic language than it is now.
1.4 Previous research

Literary work in Dzongkha probably began only after His Late Majesty (the third king of Bhutan) Jigmi Dorji Wangchuk decreed Dzongkha to be the official language of the Kingdom of Bhutan in 1971 (linguistic work began much later). The king’s decree led to systematic efforts to modernize the language as well as providing a firm linguistic foundation for ensuring the preservation of the Kingdom’s traditional and cultural values. From 1971 onwards, Dzongkha has been used as the medium of instruction in the educational system, and Dzongkha has since been undergoing continuous development as the written language.

In 1971, Lopen (title) Nado, Sangay Tenzin and Pemala did a study in Dzongkha on the differences between the liturgical language Choeke (Classical Tibetan) and written Dzongkha (Van Driem 1992). In 1997, Michalovsky and Mazaudon, along with some Bhutanese scholars published a book named, ‘Introduction to Dzongkha’: followed by it, a ‘Dzongkha Handbook’ was written by Dorji Chadro in 1986. Both booklets contain a brief introduction to Dzongkha pronunciation and script in English, vocabulary and example
sentences, but neither attempts to provide any explanation of the grammar of the language (Van Driem 1992). Since 1960 the Dzongkha Development Commission (DDC) has been systematically producing Dzongkha language materials for the instruction of native speakers in primary and secondary schools. The DDC published the first Dzongkha Dictionary in 1986. Since 1986, the work for the advancement of the national language of Bhutan has been set forth by the DDC of the Royal Government of Bhutan. In 1986, Machailovsky, at that time as an external consultant in the Department of Education of the Royal Government of Bhutan, wrote a ‘Report on Dzongkha Development’, which included a useful, first study of Dzongkha phonology. In 1988, Michailovsky and Mazaudon published the article ‘Lost syllables and tone contour in Dzongkha’, recapitulating their pioneering work in Dzongkha phonology and providing diachronic explanation for the Dzongkha contour tone. In 1990, the DDC produced the ‘Dzongkha Rabsel Lamzang’, which was a modernised and much expanded version of ‘An Introduction to Dzongkha (1997)’, written in English for foreign learners of Dzongkha. In 1990, Van Driem conducted preliminary linguistic survey activities in Bhutan. And perhaps, Watters (2002) is the latest proper linguistic work done in Dzongkha (to the best of our knowledge).

1.5 Methodology and data

The methodology adopted in this thesis is a descriptive one. As mentioned earlier, this thesis is an acoustic study of segments and tones in Dzongkha with an aim of establishing phonological contrasts. We describe the acoustic correlates of the consonant phonemes and tonal patterns of Dzongkha. In doing so, we also compare the acoustic correlates of Dzongkha segments with other languages, especially with Hindi and Korean (for obstruents), and Mizo and Tenyidie (for sonorants). However, a descriptive study without any theoretical proposition does not make much sense, as almost all linguistic theory has its origin in descriptive work. We, therefore, extend our acoustic findings of the segments and tones and analyse them in line with distinctive feature theory.
The data in this thesis is restricted to phonemes in minimal pairs in the initial position. As mentioned earlier, Dzongkha has various regional dialects; our focus is on the Paro dialect, spoken in Paro District. Initially, we recorded fourteen speakers, all teacher trainees of Paro College of Education, Royal University of Bhutan. Upon preliminary analysis of the data, we came across some dialectal differences, especially in sonorant segments (voiced-voiceless distinctions) and tonal patterns. This narrows down our informants to four (two male, two female), all in the age group of 20 to 30. All the informants live in and around Paro town, born and brought up there; and they speak at least two more languages (English and Lhotsamkha [Nepali]) besides Dzongkha. Dzongkha dictionaries (mostly the ‘English – Dzongkha Dictionary’ [DDC 2006]) were also used for our data; they were of great help, especially, for our historical data for the illustration of tonogenetic processes (in chapter 5).

1.6 Organization of the thesis

This thesis is divided into five chapters, in which we make an acoustic and phonological study of Dzongkha segment types and tonal patterns.

In chapter two, we review the phonemic inventory of Dzongkha, based on the earlier literature, in conjunction with present findings, and propose the number of phonemes in Dzongkha as 83. The major difference between our proposal and the earlier analyses is in the vowel phonemes; unlike earlier analyses, we propose additional vowel phonemes, the contrast between clear vowel (v) and nasalised vowel (\(\tilde{v}\)), with length distinction.

In chapter three, we study the acoustic correlates of Dzongkha four-way phonation types in stops (VLT*, VLAS, VLL and VD) and three-way phonation types in fricatives (VLT*, VLL and VD). We compare the four phonation types in stops with the four phonation types in Hindi and show that the fourth type, the VLL type in Dzongkha is quite distinct from the Hindi fourth stop type, the voiced aspirated (or the breathy) type. We then compare the three voiceless stop types in Dzongkha with Korean three-way voiceless
stop types, and show remarkable similarities in the acoustic correlates, especially between the voiceless tense and voiceless lax types in Dzongkha and Korean. We conclude that the fourth stop type (VLL) in Dzongkha is analogous to Korean voiceless lax type, and analyse the four stops in Dzongkha as: voiceless tense (VLT*), voiceless aspirated (VLAS), voiceless lax (VLL) and voiced (VD) in line with distinctive feature theory for their phonemic contrast. We propose new features for the obstruents, following Avery & Idsardi (2001), as: VLT* [stiff], VLAS [spread] [stiff], VLL [Ø] (Default) and VD [slack]. We also compare the three fricative types (VLT*, VLL and VD) in Dzongkha with Korean voiceless tense and voiceless lax fricatives, and extend our distinctive feature theory of stops to fricatives.

In chapter four, we turn to the study of sonorant phonation types in Dzongkha. We compare the voiceless sonorants in Dzongkha with voiceless sonorants in Mizo and Tenyidie, and show that while the voiceless nasals are similar to voiceless nasals in Mizo, the voiceless laterals pattern with voiceless laterals in Tenyidie. Based on our acoustic findings, we analyse voiceless sonorants as [stiff] [spread], and voiced as [Ø] (Default) following Avery & Idsardi (2001).

And finally, in chapter five, we make an acoustic study of the tones in Dzongkha. We show that Dzongkha has four phonetic tones triggered by the shape of the syllables, but the language has only two phonemic tones, High and Low, with a default Mid tone. We show that while voiced sonorant onsets co-occur with contrastive tones, the obstruent initials trigger a register system in Dzongkha, in that the VLT* and VLAS (as ‘Tense’ group) segment types correlate with the High Register and the VLL and VD types (as ‘Lax’ group) with the Low Register. We then analyse tones in Dzongkha as bitonal register system. We conclude the chapter, and the thesis, by identifying future areas of research in Dzongkha.
CHAPTER TWO
SEGMENTAL INVENTORY: A REVIEW

2.1 Introduction

In investigating a language, one of the first tasks of a linguist is to establish the phonemic inventory of the language. Dzongkha has been studied by a number of linguists, though not many, for example, by Mazaudon (1985), Mazaudon & Michailovsky (1988) Van Driem (1992) and Watters (2002). The phonemic inventory of the language differs (slightly) in all the four linguists. Though all the four works put forward their inventories for standard Dzongkha, their language informants are speakers of other regional varieties. Mazaudon (1985) and Mazaudon & Michailovsky (1988) have their informants from Chapcha\(^7\), Watters's (2002) informant is from Pasakha\(^8\) and Van Driem (1992) does not mention about the background of his informants, but his work, being a grammar on Dzongkha, is presumed to be on the standard variety spoken in Thimphu. So except Van Driem (1992), all others describe the regional dialects of Chukha District.

Therefore there is a need, firstly, to establish the segmental inventory of the dialect of Paro, which is our subject of study.

In this chapter, we first review the phonemic inventory of Dzongkha put forward by the four linguists who have worked on the language, and propose a revised one with illustrative examples showing the phonemic contrasts. We review the consonant inventories of the four linguists in table 2.1, 2.2, 2.3 and 2.4, and then propose a consolidated one. (We retain the original phoneme symbols by the authors in the review section and use appropriate (revised) IPA symbols in our consolidated proposal).

\(^7\) Chapcha is a village under Chukha District (southern Bhutan), but the place is approximately 50 kilometers south of Thimphu.

\(^8\) Pasakha is a village in the extreme south of Chukha District, bordering West Bengal, India.
2.2 Consonant phonemes

In this section we review the consonant phoneme inventories by the four linguists mentioned above. It is noteworthy to mention here that there is no detailed phonetic and phonological study done on phonemes of the language. The focus of all the four linguists has been on other aspects of the language. Mazaudon (1985) is a study on Dzongkha number systems; Mazaudon and Michailovsky (1988) is a diachronic study of Dzongkha level tone and falling melodic contour in the perspective of their historical origin; Van Driem (1992) is a descriptive grammar of Dzongkha, with a pedagogical goal and Watters (2002) is a study of sounds and tones of Dzongkha, but in combination with four other languages (Lhomi, Sherpa, Dolpo Tibetan and Mugom Tibetan). He has just one inventory for all the five languages; besides, his subject of study in Dzongkha is the regional dialect of Pasakha (a place in southern Bhutan - Indo-Bhutan border).

Thus, there is not yet a comprehensive study done in the segmental aspects of Dzongkha variety spoken in Paro.

We first present the consonant inventories of all the four linguists in table 2.1, 2.2, 2.3 and 2.4.

Table 2.1: Consonant phonemes of Dzongkha after Mazaudon (1985)

<p>| | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Stop</td>
<td>p</td>
<td>b</td>
<td>t</td>
</tr>
<tr>
<td></td>
<td>d</td>
<td>l</td>
<td>ʃ</td>
</tr>
<tr>
<td></td>
<td>dʒ</td>
<td>c</td>
<td>k</td>
</tr>
<tr>
<td></td>
<td>g</td>
<td>ph</td>
<td>bh</td>
</tr>
<tr>
<td></td>
<td>th</td>
<td>dh</td>
<td>ʈʰ</td>
</tr>
<tr>
<td></td>
<td>ɖʰ</td>
<td>ch</td>
<td>jh</td>
</tr>
<tr>
<td></td>
<td>kh</td>
<td>gh</td>
<td></td>
</tr>
<tr>
<td>Affricate</td>
<td>pʃ</td>
<td>bʃ</td>
<td>ts</td>
</tr>
<tr>
<td></td>
<td>dz</td>
<td>pʃh</td>
<td>bʃh</td>
</tr>
<tr>
<td></td>
<td>tʃh</td>
<td>dzh</td>
<td></td>
</tr>
<tr>
<td>Fricative</td>
<td>s</td>
<td>z</td>
<td>ɬ</td>
</tr>
<tr>
<td></td>
<td>ʒ</td>
<td>zh</td>
<td></td>
</tr>
<tr>
<td>Nasal</td>
<td>m</td>
<td>n</td>
<td>ɲ</td>
</tr>
<tr>
<td></td>
<td>ŋ</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Approximant</td>
<td>w</td>
<td>hl</td>
<td>l</td>
</tr>
<tr>
<td></td>
<td>r</td>
<td>j</td>
<td></td>
</tr>
</tbody>
</table>

9 Until we propose our new obstruent phonation types as: voiceless tense (VLT*), voiceless lax (VLL), voiceless aspirated (VLAS) and voiced (VD), we continue to use the existing term 'devoiced' for the VLL consonants. We also maintain the phonemic symbols used by the respective authors.

10 We postpone our discussion on the effect of consonant and prosodic structure until chapter 3 and 4.
As shown in table 2.1, Mazaudon (1985) has 43 consonants: 20 stops with a four-way phonation distinction for stops and affricates across five places (bilabial, dental, retroflex, palatal and velar) and a three-way phonation distinction for fricatives across two places (dental and palatal). He does not have the voiceless glottal fricative /h/ in his inventory. In his inventory the palatal affricates are analyzed as palatal stops. There is no voicing contrast in all four nasals and the dental lateral has a voiced – voiceless (l - hl) distinction. The fourth phonation type (devoiced) in stops and affricates (bh, dh, dzh, bj, and dh) and the third phonation type in fricatives (zh) and (zh) are not studied thoroughly, thus we do not have any phonetic information about their contrast with their counterparts. He analyzes them either as plain voiced with low tone merging with the voiced series or as voiceless aspirates like the regular voiceless aspirated series but occurring with the low tone. However, he mentions dialectal variation as the source of confusion for the phonation types.

**Table 2.2: Consonant phonemes of Dzongkha after Mazaudon and Michailovsky (1988)**

<table>
<thead>
<tr>
<th>Stop</th>
<th>p</th>
<th>b</th>
<th>t</th>
<th>d</th>
<th>tr</th>
<th>dr</th>
<th>k</th>
<th>g</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>ph</td>
<td>bh</td>
<td>th</td>
<td>dh</td>
<td>thr</td>
<td>dhr</td>
<td>kh</td>
<td>gh</td>
</tr>
<tr>
<td>Affricate</td>
<td>pc</td>
<td>bj</td>
<td>ts</td>
<td>dz</td>
<td>cz</td>
<td>j</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>pch</td>
<td>bjh</td>
<td>tsh</td>
<td>dzh</td>
<td>ch</td>
<td>jh</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fricative</td>
<td>s</td>
<td>z</td>
<td>c</td>
<td>z</td>
<td>h</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nasal</td>
<td>hm</td>
<td>m</td>
<td>hn</td>
<td>n</td>
<td>ny</td>
<td>ng</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Approximant</td>
<td>hl</td>
<td>l</td>
<td>r</td>
<td>y</td>
<td>w</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Mazaudon and Machailovsky (1988) have 46 consonants in their inventory: 16 stops and 12 affricates across five places (bilabial, dental, retroflex, palatal and velar) with four phonation types (voiceless, aspirated.

---

11 Mazaudon and Michailovsky (1988) do not specify the place of articulation for the consonants. The places specified in the table are assumed by the author.
voiced and devoiced) and seven fricatives across three places (dental, palatal and velar) with three phonation distinctions (voiceless, voiced and devoiced), except for the voiceless glottal fricative /h/. Of the four places for nasals two of them (bilabial /m/ and dental /n/) make voicing contrasts, and among approximants the lateral /l/ has a voicing contrast. Consonants are not specified for place of articulation in the inventory.

The voiceless and aspirated series are associated with the high register, and the voiced and devoiced with the low.

Table 2.3: Consonant phonemes of Dzongkha after Van Driem (1992)  

<table>
<thead>
<tr>
<th>Stop</th>
<th>p</th>
<th>b</th>
<th>t</th>
<th>d</th>
<th>tr</th>
<th>dr</th>
<th>k</th>
<th>g</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>ph</td>
<td>b*</td>
<td>th</td>
<td>d'</td>
<td>thr</td>
<td>dr'</td>
<td>kh</td>
<td>g'</td>
</tr>
<tr>
<td>Affricate</td>
<td>pc</td>
<td>bj</td>
<td>ts</td>
<td>dz</td>
<td>c</td>
<td>j</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>pch</td>
<td>bj'</td>
<td>tsh</td>
<td></td>
<td>ch</td>
<td>j'</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fricative</td>
<td>s</td>
<td>z</td>
<td>sh</td>
<td>zh</td>
<td>h</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>z'</td>
<td></td>
<td>zh'</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nasal</td>
<td>m</td>
<td>n</td>
<td>ny</td>
<td>ng</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Approximant</td>
<td>hl</td>
<td>l</td>
<td>j</td>
<td>w</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>hr</td>
<td>r</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Van Driem (1992) identifies 44 consonants: 16 stops and 11 affricates across six places (bilabial, alveolar, dental, retroflex, palatal and velar) with four phonation distinctions (voiceless, voiceless aspirated, voiced and devoiced) except the alveolar series of affricates, which is devoid of the devoiced phonation, and seven fricatives across three places (alveolar, palatal and velar) with three phonation distinctions (voiceless, voiced and devoiced), except for the voiceless glottal fricative /h/. According to Van Driem (1992) the ‘devoiced’ consonants are called so because historically they are derived from voiced consonants. In articulatory terms, the devoiced consonants are unvoiced, but in contrast to the voiceless consonants, and they are followed by a murmured or breathy voiced vowel with the low register tone. He also

---

12 The diacritic (') is used for the devoiced phonation type in Van Driem (1992)
observes that for speakers other than from Western Bhutan (Thimphu, Paro, Punakha, Haa, Wangdi Phodrang, Chukha) there is no distinction between the devoiced and voiced. This can be put otherwise. For all other speakers of Dzongkha, there is always a first language; Dzongkha is a second language for them, while for speakers from western Bhutan, Dzongkha is their first language.

Voiceless and voiceless aspirated obstruents occur in the high register tone while voiced and devoiced occur in the low. Sonorants can occur in both the registers.

Table 2.4: Consonant phonemes after Watters (2002)

<table>
<thead>
<tr>
<th>Stop</th>
<th>p</th>
<th>b</th>
<th>t</th>
<th>d</th>
<th>tr</th>
<th>dr</th>
<th>k</th>
<th>g</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>ph</td>
<td>pb</td>
<td>th</td>
<td>tb</td>
<td>thr</td>
<td>dr'</td>
<td>kh</td>
<td>kb</td>
</tr>
<tr>
<td>Affricate</td>
<td>pʃ</td>
<td>pʃ</td>
<td>ts</td>
<td>dz</td>
<td>tf</td>
<td>dʒ</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>pʃh</td>
<td>pʃb</td>
<td>tʃh</td>
<td>tʃb</td>
<td>tʃh</td>
<td>tʃb</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fricative</td>
<td>s</td>
<td>z</td>
<td>ʃ</td>
<td>ʒ</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nasal</td>
<td>m</td>
<td>n</td>
<td>n̊</td>
<td>n̄</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Approximant</td>
<td>hɭ</td>
<td>l</td>
<td>j</td>
<td>w</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Waters (2002) has 43 consonants: 16 stops and 12 affricates with a four-way phonation distinction (voiceless, voiceless aspirated, devoiced and voiced) across five places (bilabial, alveolar, post-alveolar, retroflex and velar) of articulation. Waters (2002) is the only linguist who investigates the segmental properties of consonants in Dzongkha. He describes the phonetic properties of the four phonation types for stops as:

- Voiceless > voiceless without aspiration
- Devoiced > voiceless sometimes with a small degree of aspiration followed by usually breathy voice (devoiced) characterized by a slightly longer VOT than the voiceless unaspirated series

---

13 The elevated /h/ symbol is used by Watters (2002) for the devoiced consonants to distinguish it from the regular aspirated /h/ consonants
• Aspirated > voiceless with heavy aspiration by contrast with the devoiced series, and

• Voiced > voiceless preceded by voicing\(^1\).\(^5\)

Voiceless and voiceless aspirated obstruents occur with the high register tone while voiced and devoiced occur in the low. Sonorants occur in both the registers.

In the next section, we look at the differences in the phonemic inventories of consonants.

2.2.1 Differences in consonant inventories

In Tables 2.1, 2.2, 2.3 and 2.4, a number of differences can be seen vis-a-vis: 1. number of phonemes; 2. place of articulation and 3. manner of articulation.

As we can see in tables 2.1, 2.2, 2.3 and 2.4, the inventories differ with regard to the number of phonemes: Mazaudon (1985) and Watters (2002) have 43 each, Van Driem (1992) has 44, and Mazaudon & Michailovsky (1988) has 46 phonemes. The best way to present their differences is, perhaps, in terms of the manner of articulation.

2.2.2 Stops

As shown in tables 2.1, 2.2, 2.3 and 2.4, except for Mazaudon (1985), all others make a four way place distinction with a four way phonation distinction. Here again, there is a difference in place for the stops; in Watters (2002), stops are distributed over bilabial, alveolar, retroflex and velar, while

---

\(^1\) Waters (2002) observes that the aspiration is not consistently articulated by all speakers, and it is not phonemic.

\(^5\) Waters's (2002) observes that the voiced stops are not completely voiced during closure. His description of the "voiced" phonation is little awkward because irrespective of whether the closure duration of a stop is completely or partially voiced the sound is a voiced one.
Van Driem (1992) has bilabial, dental, retroflex and velar as the places. Mazaudon (1985) has identified five places (bilabial, dental, retroflex, palatal and velar), and four places in Mazauzon & Michailovsky (1988) are not specified. However, all the four linguists have four phonation types; voiceless unaspirated, voiceless aspirated, voiced and devoiced.\(^{16}\) Mazaudon (1985) analyses the palatal series of affricates as stops.

### 2.2.3 Affricates

Van Driem (1992), Watters (2002), and Mazaudon & Michailovsky (1988) recognize three series of affricates, while Mazaudon (1985) has only two series (he identifies the palatal affricate series as palatal stops as mentioned in section 1.1.1). Like the stops, affricates too have been placed over different places; Van Driem (1992) has bilabial, alveolar and palatal places, while Watters has bilabial, alveolar and post-alveolar, and Mazaudon (1985) has bilabial and dental as the places. The places in Mazaudon & Michailovsky (1988) are again not specified. All four have four types of phonations for affricates (voiceless unaspirated, voiceless aspirated, voiced and devoiced) as in stops, except for the alveolar series in Van Driem (1992), which has only three. He recognizes the fact that the alveolar series has only three phonation types: voiceless unaspirated, voiceless aspirated and voiced.

### 2.2.4 Fricatives

As shown in tables 2.1, 2.2, 2.3 and 2.4, all the linguists have two series of fricatives at different places; Mazaudon (1985) has dental and palatal places while Van Driem (1992) has alveolar and palatal, and Watters (2002) has alveolar and post-alveolar places for the fricatives. Mazaudon & Michailovsky (1988) do not specify the places. Except Watters (2002), who has only two phonation types (voiceless and voiced), all others have a three way phonation

---

\(^{16}\) The phonetic details of the so called ‘devoiced’ stop have never been studied in depth. The term is assigned to the stop based on historical ground; historically this stop is derived from the proto Tibetan voiced stop.
distinction (voiceless, voiced and devoiced). Waters (2002), perhaps, analyses the devoiced phonation type as a segment that differs only in pitch (tone) with the voiceless one (though he does not mention this explicitly in his analysis).

2.2.5 Nasals

Van Driem (1992) and Mazaudon (1985) have identified four nasals (/m/, /n/, /ŋ/ & /ŋ/) while Watters (2002) has five (with one more voiceless /hn/) and Mazaudon and Michailovsky (1988) have six (with two more voiceless /hm/ & /hn/) nasals. Watters (2002) has alveolar place for nasal /n/ while Van Driem (1992) and Mazaudon (1985) have dental as the place. The place for /n/ in Mazaudon & Michailovsky (1988) is not specified. The other places are bilabial, palatal and velar. In Van Driem (1992) and Mazaudon (1985), all series of nasals are (simply) voiced, while the other two linguists distinguish voiced and voiceless nasals. Watters (2002) make a distinction between voiced and voiceless nasal for the alveolar place, and Mazaudon & Michailovsky (1988) has the same distinction for /m/ and /n/.

2.2.6 Approximants

In all the inventories, the glides /j/ or /y/ and /w/ do not differ; /j/ or /y/ is a voiced palatal glide and /w/ is voiced labiovelar glide. But sonorants (/l/ & /r/) differ in both place and phonation types. Van Driem (1992) and Watters (2002) have four (two voiced and two voiceless) across the same places (alveolar). Mazaudon's (1985) inventory has three; a voiced /r/ and lateral /l/ with a voicing distinction, while Mazaudon & Michailovsky (1988) have three glides with a voicing distinction in the lateral.

In the next section we propose our consolidated consonant inventory of Dzongkha.
2.3 Consonant Phonemes: the consolidated inventory\textsuperscript{17}

In this section, we bring together all the four consonant inventories of the four linguists along with our analysis and propose a consolidated consonant inventory; additional examples (minimal pairs) are also provided. We have the consolidated consonant phonemes in table 2.5.

Table 2.5: Dzongkha consonant phonemes\textsuperscript{18}

\begin{tabular}{|c|c|c|c|c|}
\hline
Obstruent & VLT & p\textsuperscript{*} & t\textsuperscript{*} & t\textsuperscript{*} & k\textsuperscript{*} \\
\hline
VLL & p & t & l & k & [?] \textsuperscript{19} \\
VLAS & ph & th & th & kh \\
VD & b & d & d & g \\
\hline
Affricate & VLT & pc\textsuperscript{*} & ts\textsuperscript{*} & tc\textsuperscript{*} \\
\hline
VLL & pc & tc \\
VLAS & pch & tsh & tch \\
VD & bz & dz & dz \\
\hline
Fricative & VLT & s\textsuperscript{*} & ç\textsuperscript{*} \\
\hline
VLL & s & ç & h \\
VD & z & z \\
\hline
Nasal & VL & m & n & ŋ & ŋ & ŋ \\
\hline
VD & m & n & ŋ & ŋ \\
\hline
Liquid & VL & l & r \\
\hline
VD & l & r \\
\hline
Glide & VD & j & w \\
\hline
\end{tabular}

\textsuperscript{17} The phonemic symbols are from IPA (revised to 2005) except the voiceless tense symbols. 
\textsuperscript{18} In absence of a symbol for the feature “tense” in the IPA, we take the liberty of borrowing the voiceless tense diacritic (C\textsuperscript{*}) from Korean after Kim & Duanmu (2004), Silva (2006), Kim (2004) and others. 
\textsuperscript{19} We do not specify the phonation type of the ‘?’; we analyze ‘?’ as a phonemic (or rather sub-phonemic) segment as it contrasts with other stops and nasals (only) in the final position, thus in brackets (we postpone further discussion on the glottal stop until chapter 5).
As shown in table 2.5, Dzongkha consonants are distributed across five places of articulation (labial, dental, retroflex, palatal, and velar) and six manners of articulation (stop, affricate, fricative, nasal, liquid and glide). All phonemes, stops (labial, dental, retroflex and velar), affricates (labial, dental and palatal), fricatives (dental and palatal), nasals (labial, dental, palatal and velar), liquids (dental and retroflex) and glides (labiovelar and palatal) occur in more than one phonation type. Obstruents make a contrast between voiceless tense, voiceless lax, voiceless aspirated and voiced (henceforth, VLT, VLL, VLAS and VO respectively). A three-way voiceless stop phonation type is, so far, believed to be attested only in Korean (Kim & Duanmu 2004, Silva 2006, Kim 2004 and others). We shall postpone our in-depth discussion on the VLT and VLL distinction until chapter 3. For now, the VLT type occurs in the high register and VLL in low.

A two-way register tone system in Dzongkha is a well-established fact. All linguists who have worked in the language, viz. Mazaudon (1985), Mazaudon & Michailovsky (1988), Van Driem (1992) and Watters (2002) have analyzed the tone system in Dzongkha as a register tone system. The VLT and VLAS obstruents occur in the high register and VLL and VO in the low. All sonorants occur in both the registers.

Next section is our analysis of stop consonants in Dzongkha.

2.3.1 Stops

We analyse stops with four phonation types, three voiceless (VLT, VLL, and VLAS) as in Korean and an additional voiced (VD) one, as mentioned in table 2.5. We analyse stop places as Mazaudon (1985) and Mazaudon & Machailovsky (1988) do, with labial, dental, retroflex and velar. The VLT and VLAS occur in the high register and VLL and VO occur in the low register (see chapter 5 [on Dzongkha tone] for a detailed discussion). All stops occur initially and medially but the occurrence of stops in the final
position is restricted to voiceless (tense or lax not specified) velar /k/ and labial /p/\(^{20}\). Word medially and in intervocalic positions, the VLL stops become voiced. The phonemic contrasts of the stops across all places are given in 1. In all the examples, the low tone is not marked and high tone is marked by the high diacritic (').

<table>
<thead>
<tr>
<th>Place</th>
<th>Symbol</th>
<th>Example</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>Labial</td>
<td>p*</td>
<td>p*āʔ</td>
<td>'a slice of cooked meat'</td>
</tr>
<tr>
<td></td>
<td>p</td>
<td>pana</td>
<td>'in the middle'</td>
</tr>
<tr>
<td></td>
<td>ph</td>
<td>pháʔ</td>
<td>'pig (astrology)'</td>
</tr>
<tr>
<td></td>
<td>b</td>
<td>baʔ</td>
<td>'bamboo mat'</td>
</tr>
<tr>
<td>Dental</td>
<td>t*</td>
<td>t*āʔ</td>
<td>'tiger'</td>
</tr>
<tr>
<td></td>
<td>t</td>
<td>taʔ</td>
<td>'flag, ceremonial scarf'</td>
</tr>
<tr>
<td></td>
<td>th</td>
<td>tháʔ</td>
<td>'loom'</td>
</tr>
<tr>
<td></td>
<td>d</td>
<td>daʔ</td>
<td>'lick (V)'</td>
</tr>
<tr>
<td>Retroflex</td>
<td>t*</td>
<td>t*ēm</td>
<td>'to spread around'</td>
</tr>
<tr>
<td></td>
<td>t</td>
<td>t̂</td>
<td>'to make into slices'</td>
</tr>
<tr>
<td></td>
<td>th</td>
<td>thē</td>
<td>'pattern'</td>
</tr>
<tr>
<td></td>
<td>dē</td>
<td>dē</td>
<td>'enemy, sameness'</td>
</tr>
<tr>
<td>Velar</td>
<td>k*</td>
<td>k*ē</td>
<td>'voice (of timbre)'</td>
</tr>
<tr>
<td></td>
<td>k</td>
<td>ke</td>
<td>'line, queue'</td>
</tr>
<tr>
<td></td>
<td>kh</td>
<td>khēʔ</td>
<td>'clever, intelligent'</td>
</tr>
<tr>
<td></td>
<td>g</td>
<td>gep</td>
<td>'remuneration for monk'</td>
</tr>
</tbody>
</table>

\(^{20}\) /k/ occurs in some regional varieties instead of /p/, and sometimes both occur in free variation.
2.3.2 Affricates

Dzongkha has eleven affricates spread across three places. Affricates are spread across labial, dental and palatal places of articulation. All series of affricates have four phonation contrasts: VLT, VLL, VLAS and VD, except the dental series, which has only three phonation types (VLT, VLAS and VD). The palatal affricates occur in four phonation types as in stops while the VLL is not attested in the dental place as Van Driem (1992) correctly observes; this fact could have been an oversight in the other authors' analyses. As already mentioned in section 1.1.1, the labio-palatal series in Mazaudon (1985) are analysed as palatal stops, we analyse them as affricates as analysed by others. All affricates occur in initial and medial positions and, like the stops, affricates too are voiced word medially and in intervocalic position. None of the affricates are attested in the final position. The phonemic contrasts of the affricates are given in (2). VLT and VLAS occur in the high register and VLL and VD in the low (see chapter 5 for a detailed discussion).

(2) Labial  
<table>
<thead>
<tr>
<th>pc*</th>
<th>pc*ē</th>
<th>'monkey'</th>
</tr>
</thead>
<tbody>
<tr>
<td>pc</td>
<td>pcē</td>
<td>'thin, slim'</td>
</tr>
<tr>
<td>pch</td>
<td>pche</td>
<td>'bird'</td>
</tr>
<tr>
<td>bz</td>
<td>bze</td>
<td>'cow shed'</td>
</tr>
</tbody>
</table>

Dental  
<table>
<thead>
<tr>
<th>ts</th>
<th>tsēm</th>
<th>'fodder'</th>
</tr>
</thead>
<tbody>
<tr>
<td>tsh</td>
<td>tshēm</td>
<td>'boundary, niece'</td>
</tr>
<tr>
<td>dz</td>
<td>dzem</td>
<td>'pot, chop (V)'</td>
</tr>
</tbody>
</table>

Palatal  
<table>
<thead>
<tr>
<th>te*</th>
<th>te*ē</th>
<th>'hair'</th>
</tr>
</thead>
<tbody>
<tr>
<td>tc</td>
<td>tce</td>
<td>'tea'</td>
</tr>
<tr>
<td>tch</td>
<td>tchē</td>
<td>'pair'</td>
</tr>
<tr>
<td>dz</td>
<td>dze</td>
<td>'hundred'</td>
</tr>
</tbody>
</table>

__Note:__ We have no data of Dzongkha with VLL dental affricates.
2.3.3 Fricatives

Fricatives occur in three phonation types across three places. In terms of place of articulation we analyse fricatives as alveolar (/s'/, /s/ and /z/), palatal (/ç'/, /ç/ and /z/), and glottal (/h/) with three phonation types: VLT, VLL and VD for dental and palatal, and a voiceless glottal /h/ following Mazaudon and Michailovsky (1988), Van Driem (1992 and Mazaudon (1985). The VLT series occurs in the high register and VLL and VD occur in the low register. Watters (2002) suggests that fricatives occur as alveolar and post-alveolar with voicing distinction (voiceless and voiced) as shown in table 2.1; he probably assumes tonal contrast for the two voiceless fricative). However, we analyze fricatives as three phonemes (VLT*, VLL and VD) following other three linguists (also see chapter 3 for the acoustic cues for the voiceless fricatives). All fricatives occur in word initial and medial position. However, as in the stops and affricates, the VLL series get voiced in word-medial and intervocalic positions. All fricatives occur in the initial position and medial position. The phonemic contrasts of the fricatives are shown in (3).

(3)  Alveolar  s'  s'ë  'soil'
      s      se  'eat (V)'
      z      ze  'day (calendar)'

Palatal  ç'  ç'ë  'quarrel, partition wall'
      ç      ca  'night'
      z      za  'keep (V)'

Velar  h  hëm  'blue, green'

2.3.4 Nasals

Unlike all the authors in table 2.1, 2.2, 2.3, and 2.4, we will propose a two way phonatory contrast (voiceless and voiced) across all four places (labial, dental, palatal and velar). The phonemic contrasts in (4) support our
A proposal for voiceless – voiced contrast. The voiceless nasals occur in the high register and voiced ones can occur in both high and low registers (see chapter 4 for detailed discussion).

<table>
<thead>
<tr>
<th>(4)</th>
<th>m - ñ</th>
<th>më</th>
<th>‘lock’</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>mëp</td>
<td>‘husband’</td>
<td></td>
</tr>
<tr>
<td></td>
<td>mëtê</td>
<td>‘low’</td>
<td></td>
</tr>
<tr>
<td></td>
<td>ma:</td>
<td>‘butter’</td>
<td></td>
</tr>
<tr>
<td></td>
<td>má:</td>
<td>‘war’</td>
<td></td>
</tr>
<tr>
<td></td>
<td>më:</td>
<td>‘a kind of bamboo’</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>n - ñ</th>
<th>neni</th>
<th>‘to get sick’</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>nép</td>
<td>‘black’</td>
</tr>
<tr>
<td></td>
<td>nëp</td>
<td>‘snot’</td>
</tr>
<tr>
<td></td>
<td>nê</td>
<td>‘promise (N)’</td>
</tr>
<tr>
<td></td>
<td>nê</td>
<td>‘curse’</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>ñ - ñ</th>
<th>ñe</th>
<th>‘fish’</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>ñëci</td>
<td>‘yoke’</td>
</tr>
<tr>
<td></td>
<td>ñëp</td>
<td>‘reach for something (with hand)’</td>
</tr>
<tr>
<td></td>
<td>ñë</td>
<td>‘sleep’</td>
</tr>
<tr>
<td></td>
<td>ñë</td>
<td>‘site of pilgrimage’</td>
</tr>
<tr>
<td></td>
<td>ñë</td>
<td>‘injure (in an accident)’</td>
</tr>
</tbody>
</table>
2.3.5 Liquids

Our analysis of liquids conform to the analysis of Watters (2002) and Van Driem (1992) in both place (dental\(^{22}\) for /l/ and retroflex for /r/) and phonation type (voiceless and voiced), the other two linguists do not include the voiceless retroflex /r/ in their inventories. The voiceless – voiced phonemic contrast of the liquid consonants are presented in (5).

(5)  

\[ \eta - \eta \]  
\[ \eta a \]  
\[ 'a kind of fasting' \]

\[ \eta á \]  
\[ 'a kind of healing chant' \]

\[ \eta á \]  
\[ 'pillow' \]

\[ \eta e \]  
\[ 'I' \]

\[ \eta ê \]  
\[ 'five, drum' \]

\[ \eta ê \]  
\[ 'early' \]

\[ l - r \]  
\[ lê \]  
\[ 'mountain' \]

\[ lê \]  
\[ 'payment, wage' \]

\[ lê \]  
\[ 'god, deity' \]

\[ lu \]  
\[ 'sheep' \]

\[ lu \]  
\[ 'pour' \]

\[ lû \]  
\[ 'dismantle' \]

\[ r - ð \]  
\[ rëm \]  
\[ 'dismantle (of stone fence) (INTR)' \]

---

\(^{22}\) Articulatorily, the tongue touches both teeth and alveolar ridge, but we choose to specify 'dental' as the place feature.
2.3.6 Glides

The two glides (labio-velar /w/ and palatal /j/) are the only consonants that correspond across all the four inventories. The phonemic contrasts of glides are given in (6). Both the glides occur in both high and low Register.

(6)  

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Pronunciation</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>w</td>
<td>wąŋ</td>
<td>'name of a place'</td>
</tr>
<tr>
<td></td>
<td>wąŋ</td>
<td>'blessing from a monk'</td>
</tr>
<tr>
<td>j</td>
<td>jʊ</td>
<td>'one of a pair'</td>
</tr>
<tr>
<td></td>
<td>jʊ</td>
<td>'itch'</td>
</tr>
</tbody>
</table>

We now turn to vowels.

2.4 Vowel Phonemes

In this section, we review the vowel phonemes by the four linguists. The vowel inventories do not differ as much as the consonants do; yet, each inventory differs from the other in the analysis of vowel phonemes. We shall first present the four inventories by the four linguists and identify the areas in which they differ. As in the consonant inventories, we shall again retain the respective authors’ original vowel symbols in the review section and then use IPA symbols in our consolidated inventory. The four vowel inventories are presented below after their respective authors.
In figure 2.1, Mazaudon (1985) has eight vowels (three front unrounded /i/, /ɛ/ and /æ/; two front rounded /y/ and /o/; two back rounded /u/ and /o/ and an open unrounded /a/). And there is no further discussion on the vowels by the author.

In figure 2.2, Mazaudon & Michailovsky (1988) have eight vowels (three front unrounded /i/, /ɛ/ and /æ/; two front rounded /ue/ and /oe/; two back rounded /u/ and /o/ and an open vowel /a/). They claim that there is no opposition between the front unrounded /ɛ/ and front unrounded /æ/ in short, open monosyllables, and suggest that the realization is rather open. They also claim that vowel length is phonologically distinctive, but there are no examples in support of the contrasts.
In figure 2.3, Van Driem (1992) has eight vowels: two front unrounded /i/ and /e/; two front rounded /ɪ/ and /ɛ/; two back rounded /u/ and /o/ and two low open /æ/ and /a/. He analyses the length of five vowels as phonologically distinctive, the vowels are /æ/, /e/, /i/, /o/ and /u/, while /ɪ/, /ɛ/ and /ɑː/ as always long.

(7) | **Short** | **Long** |
---|---|---|
**a - a:** | máp ‘husband’ | máːp ‘red’ |
**e - e:** | sep ‘stallion’ | sep ‘yellow’ |
**i - i:** | tsip ‘stone wall’ | tsɪp ‘astrologer’ |
**o - o:** | phop ‘put something down’ | phoːp ‘cup’ |
**u - u:** | ku ‘body part (Hon)’ | kuː ‘scoop’ |

However, he notes that the difference between the long and the short vowels are of timbre and vowel quality, besides length.

![Figure 2.4: Watters (2002)](image)

In figure 2.4, in Watters’ (2002) inventory, the front and central vowels which are in square brackets [ ] are problematic vowels. He does not identify if the difference is of vowel duration or quality in both the pairs (/ɛ/ vs /ɛ/ and /ə/ vs /ʊ/). The front and back vowels which are in parenthesis ( ) are supposed to be having allophonic relationships, but he does not specify which is an
allophone of which in both the pairs (front and back). This leaves the number of vowels in his inventory uncertain; however he calls for further phonetic study on the vowels, especially the front vowels to further establish the vowel contrasts.

In the next section, we present our consolidated inventory of Dzongkha vowels.

2.5 Vowel phonemes: the consolidated inventory

Dzongkha has, perhaps, one of the richest vowel inventories among the Tibetan dialects. Besides ten monophthongs (pure vowels) it has seven diphthongs. The monophthongs further have secondary features - length and nasalization contrasts.

In this section, we review all the previous vowel inventories and then propose a consolidated one with acoustic evidence. The formant (F1, F2 and F3) matching experiment is a well-established phonetic paradigm in research on the distinction of vowel quality. We shall establish vowel contrast with vowel formant values. First, we shall present formant readings of a male speaker to support our analysis of the vowel contrasts in Dzongkha. Then further vowel qualities such as length and nasalization will also be discussed in this section. We first present the vowel formants.

2.5.1 Vowel formants: facts about vowel formants and vowel quality

Johnson (2000) suggests that the distinctive features of vowels, in other words the vowel contrasts are better analyzed with acoustic properties, rather than articulatory properties. We have the summary of the correlation between

---

25 All phoneme symbols are from IPA (revised to 2005)
the two formant frequencies (F1 and F2)\textsuperscript{24} and the vowel quality (Ladefoged 1981) listed as:

I. The vowel distinction is best noticeable in the frequencies of F1 and F2
II. Vowel height correlates with the frequency of the first formant (F1), i.e. lower vowels have higher F1, and higher vowels have lower F1
III. Vowel frontness correlates with the second formant (F2), i.e. front vowels have higher F2 and back vowels have lower F2.

2.5.2 Formant plot experiment

Ten words, which are minimal pairs, of a male speaker are recorded (three tokens) with all the ten vowels preceded by the voiceless palatal fricative consonant /\textipa{q}/. Then F1, F2 and F3 values of all the three tokens are measured and their averages listed in table 2.6.

Table 2.6: Average formant (F1, F2 and F3) (Hz) for Dzongkha vowels

<table>
<thead>
<tr>
<th>i</th>
<th>1</th>
<th>ci</th>
<th>'lice'</th>
<th>289</th>
<th>2044</th>
<th>2721</th>
</tr>
</thead>
<tbody>
<tr>
<td>e</td>
<td>2</td>
<td>ce</td>
<td>'know'</td>
<td>344</td>
<td>1861</td>
<td>2611</td>
</tr>
<tr>
<td>e</td>
<td>3</td>
<td>ce</td>
<td>'strength'</td>
<td>435</td>
<td>1916</td>
<td>2574</td>
</tr>
<tr>
<td>y</td>
<td>4</td>
<td>cy</td>
<td>'reminiscence'</td>
<td>289</td>
<td>1755</td>
<td>2117</td>
</tr>
<tr>
<td>e</td>
<td>5</td>
<td>ca</td>
<td>'part of plough'</td>
<td>326</td>
<td>1624</td>
<td>2117</td>
</tr>
<tr>
<td>u</td>
<td>6</td>
<td>cu</td>
<td>'meat'</td>
<td>472</td>
<td>1331</td>
<td>2172</td>
</tr>
<tr>
<td>a</td>
<td>7</td>
<td>ca</td>
<td>'quarrel'</td>
<td>600</td>
<td>1112</td>
<td>2245</td>
</tr>
<tr>
<td>o</td>
<td>8</td>
<td>co</td>
<td>'dice'</td>
<td>399</td>
<td>765</td>
<td>2373</td>
</tr>
<tr>
<td>o</td>
<td>9</td>
<td>co</td>
<td>'come'</td>
<td>472</td>
<td>783</td>
<td>2355</td>
</tr>
<tr>
<td>u</td>
<td>10</td>
<td>cu</td>
<td>'force, strength'</td>
<td>326</td>
<td>801</td>
<td>2154</td>
</tr>
</tbody>
</table>

\textsuperscript{24} We chose to neglect the values of F3 for our analysis since our main concern is with the vowel 'height' and 'frontness'
The mean formant values for the first two formants (F1 and F2) are plotted in figure 2.5 using JPlotformant software (Billerey - Mosier 2002) for the male speaker and we get the results in figure 2.5. (For lack of IPA symbol options in the software, vowel numbers are used instead of vowel symbols, [see table 2.6 for vowel equivalents]).

Figure 2.5: F1 – F2 Plot of Dzongkha vowels

Looking at figure 2.5 and the vowel contrasts in table 2.6, it is apparent that Dzongkha has ten vowel contrasts, and the vowel inventory quite well shapes with the IPA vowel plotting. We have three front unrounded (/i/, /e/ and /æ/), three back rounded (/u/, /o/ and /ɔ/), two front rounded /y/ and /ɛ/, a low unrounded central vowel /a/ and a central open unrounded /a/.

The front open unrounded /i/ and rounded /y/ have exactly the same F1 value (289 Hz), while the back rounded /u/ is slightly lower (by 37 Hz) than its front counterparts. The high mid front rounded /ɛ/ is slightly higher (by 18 Hz) than its front unrounded counterpart /e/, and the front unrounded /æ/ is lower than /ɛ/ by approximately 100 Hz. The front unrounded low vowel [æ] occurs as an allophone of /ɛ/. The back rounded vowels /u/, /o/ and /ɔ/ are more or less consistently spaced in the vowel continuum.
The low open vowel /a/ is lower than the central vowel /e/, and it is closer to back vowels than front vowels, and thus, we may analyze it as a back vowel.

Thus, Dzongkha vowel inventory can be classified as shown in table 2.7.

Table 2.7: Classification of Dzongkha vowels

<table>
<thead>
<tr>
<th>Vowel</th>
<th>Height</th>
<th>Front/Back</th>
<th>Rounded/Unrounded</th>
</tr>
</thead>
<tbody>
<tr>
<td>i</td>
<td>High</td>
<td>Front</td>
<td>Unrounded</td>
</tr>
<tr>
<td>e</td>
<td>Higher mid</td>
<td>Front</td>
<td>Unrounded</td>
</tr>
<tr>
<td>e</td>
<td>Lower mid</td>
<td>Front</td>
<td>Unrounded</td>
</tr>
<tr>
<td>y</td>
<td>High</td>
<td>Front</td>
<td>Rounded</td>
</tr>
<tr>
<td>o</td>
<td>Higher mid</td>
<td>Front</td>
<td>Rounded</td>
</tr>
<tr>
<td>v</td>
<td>Lower mid</td>
<td>Centre</td>
<td>Unrounded</td>
</tr>
<tr>
<td>a</td>
<td>Low</td>
<td>Back</td>
<td>Unrounded</td>
</tr>
<tr>
<td>o</td>
<td>Higher mid</td>
<td>Back</td>
<td>Rounded</td>
</tr>
<tr>
<td>o</td>
<td>Lower mid</td>
<td>Back</td>
<td>Rounded</td>
</tr>
<tr>
<td>u</td>
<td>High</td>
<td>Back</td>
<td>Rounded</td>
</tr>
</tbody>
</table>

In the next section, we look at the secondary features such as vowel length and nasalization contrasts.

2.6 Vowel length and nasalization

2.6.1 Vowel length

Vowel length in Dzongkha is contrastive, except the central vowel /e/; this vowel is always short. Difference in vowel length also has implication on prosodic structure in Dzongkha. We will, however, postpone our discussion on ‘vowel length and prosody’ until chapter five on Dzongkha Tones. The minimal pairs in table 2.8 and 2.9 show vowel length contrast in Dzongkha.
Table 2.8: Dzongkha vowel length contrast in open monosyllabic words

<table>
<thead>
<tr>
<th>Vowel</th>
<th>Word 1</th>
<th>Meaning 1</th>
<th>Word 2</th>
<th>Meaning 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>i</td>
<td>sí</td>
<td>'shake (V)'</td>
<td>sí:</td>
<td>'dry (in air) (V)'</td>
</tr>
<tr>
<td>e</td>
<td>pé</td>
<td>'example'</td>
<td>pé:</td>
<td>'shuffle (V)'</td>
</tr>
<tr>
<td>ē</td>
<td>cé</td>
<td>'strength'</td>
<td>cé:</td>
<td>'wander (V)'</td>
</tr>
<tr>
<td>y</td>
<td>phy</td>
<td>'offering to deity'</td>
<td>phy:</td>
<td>'push (V)'</td>
</tr>
<tr>
<td>o</td>
<td>tché</td>
<td>'you'</td>
<td>tché:</td>
<td>'go mad'</td>
</tr>
<tr>
<td>a</td>
<td>phá</td>
<td>'pig (astrology)'</td>
<td>phá:</td>
<td>'there'</td>
</tr>
<tr>
<td>o</td>
<td>phó</td>
<td>'yeast'</td>
<td>phó:</td>
<td>'stomach'</td>
</tr>
<tr>
<td>o</td>
<td>tó</td>
<td>'food'</td>
<td>tó:</td>
<td>'release (V)'</td>
</tr>
<tr>
<td>u</td>
<td>òú</td>
<td>'force, strength'</td>
<td>òú:</td>
<td>'boil'</td>
</tr>
</tbody>
</table>

Vowel length is more apparent in open monosyllabic words than in closed syllables. However, there are examples of length contrasts in closed syllables as well. Few examples are given in table 2.9.

Table 2.9: Dzongkha vowel length contrast in closed monosyllabic words

<table>
<thead>
<tr>
<th>Vowel</th>
<th>Word 1</th>
<th>Meaning 1</th>
<th>Word 2</th>
<th>Meaning 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>tsíp</td>
<td>'stone wall'</td>
<td>tsí:p</td>
<td>'astrologer'</td>
<td></td>
</tr>
<tr>
<td>dzím</td>
<td>'tongs'</td>
<td>dzí:m</td>
<td>'eyebrow'</td>
<td></td>
</tr>
<tr>
<td>phó:p</td>
<td>'put something down'</td>
<td>phó:p</td>
<td>'bowl, cup'</td>
<td></td>
</tr>
<tr>
<td>sém</td>
<td>'mind, soul'</td>
<td>sém</td>
<td>'pea'</td>
<td></td>
</tr>
<tr>
<td>thé:p</td>
<td>'forehead'</td>
<td>thé:p</td>
<td>'absolute, complete'</td>
<td></td>
</tr>
<tr>
<td>gé:p</td>
<td>'remuneration for monks'</td>
<td>gé:p</td>
<td>'king'</td>
<td></td>
</tr>
</tbody>
</table>
2.6.2 Nasalized vowels

Nasalization is distinctive, at least with four vowels, three back vowels ū, ō and ā and the front high unrounded ī in Dzongkha. Nasalized vowels are attested only in open (both short and long) syllables. Table 2.10 lists minimal pairs of nasalized vowels.

<table>
<thead>
<tr>
<th>Vowel</th>
<th>ī</th>
<th>ī:</th>
<th>ū</th>
<th>ē</th>
<th>ō</th>
<th>ō:</th>
<th>a</th>
<th>ā</th>
</tr>
</thead>
<tbody>
<tr>
<td>'lice'</td>
<td>ī</td>
<td>'tree, firewood'</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>'tip, terminal'</td>
<td>ī:</td>
<td>'glue'</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>'incense'</td>
<td>ū</td>
<td>'tell, say (Hon)'</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>'water'</td>
<td>ē</td>
<td>'basket'</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>'rub (V)'</td>
<td>ō</td>
<td>'stores (N)'</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>'he'</td>
<td>ō:</td>
<td>'freeze (V)'</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>'teeth'</td>
<td>ō</td>
<td>'go (IMP)'</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>'support'</td>
<td>ō</td>
<td>'penalty'</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>'meat'</td>
<td>a</td>
<td>'grass lawn'</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>'tent'</td>
<td>ā</td>
<td>'hill'</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>'loom'</td>
<td>ā:</td>
<td>'play ground'</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Length in nasalized vowels is also contrastive. This is shown in table 2.11.

<table>
<thead>
<tr>
<th>Vowel</th>
<th>ī</th>
<th>ī:</th>
<th>ū</th>
<th>ē</th>
<th>ō</th>
<th>ō:</th>
<th>a</th>
<th>ā</th>
</tr>
</thead>
<tbody>
<tr>
<td>'indifferences'</td>
<td>ī</td>
<td>'pond, fight (V)'</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>'story'</td>
<td>ī:</td>
<td>'say, tell (Hon)'</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>'gravy (of curry)'</td>
<td>ū</td>
<td>'they'</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>'incense leaves'</td>
<td>ū:</td>
<td>'balance (N)'</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
2.7 Diphthongs

Along with the ten monophthongs, there are seven diphthongs as well. Diphthongs in Dzongkha occur only in open syllable. Dzongkha diphthongs are listed in Table 2.12 with examples.

Table 2.12: Dzongkha diphthongs

<table>
<thead>
<tr>
<th>Diphthong</th>
<th>Short Form</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>'au'</td>
<td>'au'</td>
<td>'thief'</td>
</tr>
<tr>
<td></td>
<td>'khau'</td>
<td>'stick'</td>
</tr>
<tr>
<td>'iu'</td>
<td>'tiu'</td>
<td>'pigeon'</td>
</tr>
<tr>
<td></td>
<td>'diu'</td>
<td>'bullet, stupid'</td>
</tr>
<tr>
<td>'eu'</td>
<td>'teu'</td>
<td>'navel'</td>
</tr>
<tr>
<td></td>
<td>'seu'</td>
<td>'hailstorm'</td>
</tr>
<tr>
<td>'eyu'</td>
<td>'theu'</td>
<td>'ash'</td>
</tr>
<tr>
<td></td>
<td>'peeu'</td>
<td>'forehead'</td>
</tr>
<tr>
<td>'ou'</td>
<td>'thou'</td>
<td>'hammer'</td>
</tr>
<tr>
<td></td>
<td>'kou'</td>
<td>'hide'</td>
</tr>
<tr>
<td>'ei'</td>
<td>'ei'</td>
<td>'mother'</td>
</tr>
<tr>
<td></td>
<td>'khei'</td>
<td>'yuck!'</td>
</tr>
</tbody>
</table>

Thus, Dzongkha has 34 vowels: 10 short monophthongs; 9 long (\(\i\)) monophthongs; 4 short nasalized (\(\i\)) monophthongs; 4 long nasalized (\(\i\)) monophthongs and 7 diphthongs. All 34 vowels in Dzongkha are listed in Table 2.13.
Table 2.13: Dzongkha vowel phonemes

<table>
<thead>
<tr>
<th></th>
<th>i</th>
<th>e</th>
<th>y</th>
<th>e</th>
<th>a</th>
<th>o</th>
<th>u</th>
</tr>
</thead>
<tbody>
<tr>
<td>Short</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Long</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Short nasalized</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Long nasalized</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Diphthongs</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

2.8 On the size of Dzongkha segmental inventory

It is worthwhile reviewing the segmental size of Dzongkha before we conclude this chapter, because the size (83 phonemes) does draw our attention. Including length distinction in vowels (both non-nasalised and nasalised) we have claimed that Dzongkha has 83 phonemes, which includes 49 consonants. The question that arises now is, are these many, 83 phonemes, usual in a language? The answer to this question necessitates an areal survey.

The number of segments in languages seems to vary widely. According to the ULCA UPSID\textsuperscript{25} survey, the smallest inventories included only 11 segments as in Rotokas\textsuperscript{26} and Mura\textsuperscript{27}, and the largest has 141 segments as in !Xu\textsuperscript{28} (Maddieson 1984). However, the study also claims that the typical size of an inventory lies between 20 and 37 segments, but only 7% of the languages in the survey fall within these limits. The mean number of segments per language is believed to be little over 31 according to the survey.

Then, the number of segments in Dzongkha seems to be little more than usually expected in languages in general; it exceeds more than double the size of the average number of segments in languages. However, a quick survey on the number of phonemes in peripheral Tibeto-Burman languages saves us from our disbelief. The size of phonemic inventories tends to vary considerably

\textsuperscript{25} (University of California, Los Angeles) Phonological Segment Inventory Database
\textsuperscript{26} Rotokas (East Papuan) is spoken in Vougainville, an island in Papua New Guinea
\textsuperscript{27} Mura (Amazonas) is spoken in Brazil
\textsuperscript{28} !Xu (Khoisan) is a language spoken in Southern Africa
across Tibeto-Burman languages. Ao (Chungli)\textsuperscript{29} (Temsunungsang 2009) has just 19 segments (4 vowels and 15 consonants) at one extreme, on the other, Zhongu Tibetan\textsuperscript{30} (Sun 2003) too, like Dzongkha, has as many as 84 phonemes (42 simple consonants, 33 complex consonants, and 9 vowels), which is two more than number of segments in Dzongkha. In fact, quite a number of dialects of Tibetan seem to be bigger-sized in terms of their segmental inventory; Kham Tibetan\textsuperscript{31} (Olson n.d) has 64 segments (57 consonants and 7 vowels), and Rgyalthang Tibetan\textsuperscript{32} (Xiaosong 1996) has 73 segments (42 consonants and 31 vowels, including diphthongs and nasalised vowels). Thus, the segment size of Dzongkha (83 phonemes) seems to be quite usual by Tibetan standard.

2.9 Conclusion

In this chapter we began by recognizing the need to review the phonemic inventory of Dzongkha; we reviewed the phonemic inventories of four linguists: Mazaudon and Michailovsky (1988), Watters (2002), Van Driem (1992) and Mazaudon (1985). We have shown that Dzongkha has 83 phonemes (including vowel length and nasalization): 48 consonants across five places (labial, dental, retroflex, palatal and velar) with four phonatory distinctions (VLT, VLL VLAS and VD) for obstruents (except fricatives); 10 monophthongs (with length distinction on 9 vowels and nasalization and length distinction on four) and 7 diphthongs.

In section 2.2, we have compared the consonant phonemes of the earlier researchers and proposed a consolidated consonant inventory in section 2.3. The significant alteration we brought about in this section is the recategorisation of earlier devoiced series of consonants as a new feature Voiceless Lax. This makes the claim that in fact, Dzongkha has three voiceless stops and affricates:

---
\textsuperscript{29} Ao (Tibeto-Burman) spoken by Ao tribe in North-East India.
\textsuperscript{30} Zhongu (Tibeto-Burman) is a variety of Tibetan spoken in the Zhongu valley of Songpan County in Northern Sichuan, a borderland between Tibetan and Qiang.
\textsuperscript{31} Kham Tibetan (Tibeto-Bunnan) is a variety of Tibetan spoken in Tibet (Sikang, Szechwan and Yunnan provinces of China).
\textsuperscript{32} Rgyalthang Tibetan is spoken in Zhongdian County, Diqing Tibetan autonomous prefecture in China.
voiceless tense, voiceless lax and voiceless aspirated in addition to voiced, and three types of fricatives: voiceless lax, voiceless tense and voiced.

In section 2.4, we compared the vowel inventories of the earlier researchers, and proposed a consolidated inventory in section 2.5 through section 2.7. We have also supported our claim with acoustic evidence. The average F1 and F2 formant values of the male speaker further maintains our ten vowel contrasts. In addition to 7 diphthongs and 10 monophthongs we have also demonstrated the length and nasalisation distinction in monophthongs.

Then, in section 2.8, we did a quick survey on the number of segments in Dzongkha in comparison with other Tibeto-Burman languages, including three other dialects of Tibetan, and found the segment size of Dzongkha fairly usual by the standard of Tibetan dialects.

In the next chapter, we start with our phonation study of obstruents in Dzongkha: a four-way phonation for stops.
CHAPTER THREE
OBSTRUENT PHONATION TYPES

3.1 Introduction

Consonants in world languages can be distinguished based on three parameters: Manner of articulation, Place of articulation and Phonation type. We devote this chapter to the last parameter, the phonation type (the term warrants an elaborate discussion, see section 3.2) in Dzongkha obstruents; how stops and fricatives employ different phonation configurations in making phonemic contrasts. Extensive phonetic studies in Tibeto-Burman languages in general and Dzongkha in particular is comparatively scarce. The majority of phonetic studies were particularly centered around a few languages, particularly Hindi (Lisker & Abramson 1964, Kagaya & Hirose 1975, Ingemann & Yadav 1978, Ohala 1979, Benguerel & Bhatia 1980, Vaux 1988, Dixit 1989, Davis 1994, Bali 1999, Dutta 2007, among others) and Korean (Lisker & Abramson 1964, Hirose, Lee & Ushijima 1973, Vaux 1996, Cho & Keating 2001, Park 2000, Cho, Jun & Ladefoged 2001, Ahn & Iverson 2003, Kim 2004, Kim & Duanmu 2004, Chang 2006, Silva 2006, Wright 2007, among others). These two languages gained substantial attention from phoneticians and linguists alike for their (presumably) unique phonation systems: Hindi, for its four-way (voiceless unaspirated, voiceless aspirated, voiced unaspirated and voiced aspirated) contrast in stops, and Korean, for its three-way voiceless contrast (tense, lax and aspirated) in stops and two-way voiceless contrast (tense and lax) in fricatives (In fact, throughout our study of Dzongkha phonation system in this chapter, we make reference to Hindi and Korean as precursor to obstruent phonation systems). However, such phonetic studies in Dzongkha have not been done extensively so far. In this chapter, we make a comprehensive study, for the first time, of the acoustic characteristics of the phonation types in Dzongkha. In doing so, we look at acoustic properties (temporal and spectral) such as: VOT, closure duration, stop burst, H1-H2 and fricative duration of Dzongkha obstruents (coronal [dental] stops and fricatives). We will, in principal be, governed by two approaches to linguistic studies: 'experimental'
and ‘acoustic’. Though some linguists tend to use the terms (almost) synonymously, we make a slight difference between the terms: ‘experimental’ part envelops around the frequency of occurrence of certain phonetic and phonological features in the number of speakers, and ‘acoustic’ entails the part that surrounds the physical (visual) characteristics of the sounds in terms of phonetic properties mentioned above. This way we hope to make our phonological claims, in terms of phonemic contrasts, more convincing.

In our segmental inventory of consonants in the previous chapter, we mentioned that Dzongkha has, in fact, a four-way phonation distinction in stops and affricates, with three voiceless, (tense, lax and aspirated, henceforth VLT*, VLL, VLAS) and a voiced (henceforth VD), and a three-way contrast in fricatives with two voiceless (tense and lax) and one voiced (see chapter 2, section 2.3 for lexical contrasts in all the consonant types). These classifications have been put forward as: voiceless unaspirated (for VLT*), voiceless aspirated (for VLAS), voiced (for VD) and devoiced (for VLL) by earlier linguists who have worked on Dzongkha, for example, Van Driem (1992) and Watters (2002). In other words, the two (phonetically) voiceless unaspirated stops and fricatives were classified as ‘voiceless’ and ‘devoiced.’ Both the authors based their analysis on articulatory approach rather than acoustic33. Phonologically, we see a problem with the, so called, ‘devoiced’ series of obstruents in their analysis (see section 3.3.1 for discussion on the problem). In this chapter, we further our new proposal of classification of Dzongkha obstruents as: VLT*, VLL, VLAS and VD with acoustic analysis of the data collected from native speakers of Dzongkha. Besides making a detailed acoustic study of Dzongkha obstruents we also compare the phonetic correlates of Dzongkha with two other languages, Hindi and Korean, in our study. Our interest in these two languages grows from the fact that despite different historical affiliations theses two languages share striking similarities in the consonant (especially, the stops) phonation system. We compare the stop phonation system in Dzongkha and Hindi as, perhaps, the two of the rare languages with a four-way phonation distinction in stops, and show that the two languages employ different acoustic cues in distinguishing their four-way

33 Waters (2001) however mentions about differences in VOT of the stops.
stop phonation contrasts; in that, Dzongkha has three voiceless (tense, lax and aspirated) and a voiced stop, while Hindi has two voiceless (aspirated and unaspirated) and two voiced (aspirated and unaspirated). We then compare the three-way voiceless phonation contrast in Dzongkha with a three-way phonation system in Korean stops, and show striking parallels in the way these two languages contrast their voiceless stops (especially the Tense – Lax distinction).

In section 3.2, we review the literature on ‘phonation types’ and types of phonations languages employ to make phonemic contrasts citing examples from different languages for different types of phonations. In section 3.3, we present what little work exists so far in Dzongkha phonation. This will be followed by a brief typological survey of phonation types in other Tibeto-Burman languages. Section 3.4 is our production study of Dzongkha phonation types in stops. Starting section 3.4 through section 3.7.2, we present our acoustic analysis of Dzongkha stops using VOT (in section 3.4.1 and 3.4.2), stop burst (in section 3.5), closure duration (in section 3.6) and H1-H2 in section 3.7 through 3.7.2. This will be followed by our perception study (using VOT and F0 as the stimuli) in section 3.7.4. In section 3.8, we summarize Hindi phonation (as a four-way phonation system) types in stops and compare the voiced aspirated type phonation with the VLL type in Dzongkha. In section 3.9, we summarize Korean stop phonation type and compare the Tense – Lax phonation types with Dzongkha in section 3.9.1. In section 3.10 through 3.10.2, following Avery & Idsardi’s (2001) modified feature theory, we propose our phonological feature for the four Dzongkha stops. And we conclude this chapter by extending our distinctive features of stops (tense-lax) to Dzongkha fricatives (tense-lax) in section 3.11 through 3.11.6.

3.2 Explaining the term ‘Phonation’

Catford (1977) defines the term ‘phonation’ as: ‘Any vocal activity in the larynx whose role is one neither of initiation nor of articulation.’ Thus, phonation is the various kinds of vocal-fold vibration (or voicing), and the
study of phonation type is aimed at accounting for the various laryngeal possibilities, such as ‘breathy’ and ‘creaky’ voice, among others. Phonation includes laryngeal activities which provide the voicing of sounds, but excludes the articulation of glottal stops or the initiation of a glottalic airstream for ejective plosives. ‘Voiceless’ and ‘whisper’ are also considered to be types of phonation in this analysis. Phonation requires the initiation of egressive or ingressive airflow across the larynx. With most languages it generally involves a pulmonic egressive airstream mechanism as speech sounds produced by a pulmonic ingressive airstream mechanism are only rarely observed. In addition, speech sounds produced by glottalic or velaric airstream mechanism, such as ‘clicks’ and ‘ejectives’, seldom show distinct phonation types. As a result, discussion of phonation type tends to be restricted to speech sounds produced by a pulmonic egressive airstream mechanism.

Configurations of distinct glottal states are sometimes sustained throughout the production of specific speech sounds but usually alternate over time. Glottal gestures, in a sense, can be described as the alternation of vocal fold vibration and its absence. Furthermore, coordination of glottal and supraglottal gestures leads to distinct consonant types. Supraglottal gestures are similar to glottal gestures in that obstruction or constriction alternates with a more sonorant state. The production of a consonant can thus be divided into two phases: pre-release and post-release. Phonetic characteristics of a consonant type are not only represented in terms of pre- and post-release phases but also in terms of voiced and voiceless phases. The relative timing between the two gestures leads to different consonant types. When phonation type is confined to the voiced phase, we can also observe different voice qualities, such as ‘modal’, ‘creaky’, and ‘breathy’ voice.

3.2.1 Phonation types

Phonation types may be linguistically employed to constitute a phonological contrast, so that a difference in phonation type is associated with a difference in lexical meaning. Various phonation types that are linguistically irrelevant or paralinguistic are also observed to occur. When linguistic relevance is taken
into account, phonation type can be narrowly defined as modes of laryngeal settings in normal speech production. Coordination of laryngeal muscle movements constitutes a continuum of glottal width from complete closure to the widest opening of the glottis. Ladefoged (1971) suggests that there might be a continuum of phonation types, defined in terms of the aperture between the arytenoid cartilages, ranging from voiceless (furthest apart), through breathy voiced, to regular, modal voicing, and then on through creaky voice to glottal closure (closest together). Figure 3.1 is an oversimplified schematic representation of the phonation type continuum by Ladefoged (1971).

Figure 3.1: Continuum of phonation types (modified after Ladefoged 1971)

Ladefoged (1971) focuses on a model of possible phonation types based on the glottal constriction continuum as shown in Figure 3.1. In this model, the size of the glottis, which depends on the distance between the vocal folds, can vary from zero, without phonation (for a glottal stop), to that used for voicelessness, again without phonation. The range of glottal positions that allows phonation is divided into three categories, corresponding to the three most common contrasting phonation types: creaky voice, produced with a constricted glottis; breathy voice, produced with a more open glottis; and modal voice, in between these two. Furthermore, since glottal constriction is a continuum, there are degrees of creakiness and of breathiness, and indeed the modal voice category itself varies from more constricted to more open. Ladefoged (1971) also stresses that these categories, like other phonetic categories, are not absolute. Not only might they differ somewhat across languages (so that what counts as breathy voice in one language might count as modal in another), but they are sure to differ somewhat across speakers (so that
what counts as breathy voice for one speaker might count as modal for another, within the same language).

While Ladefoged’s (1971) representation (figure 3.1) is based on the degree of glottal aperture, we can also use an analogous comparison for the degree of vibration the different phonation types make as in figure 3.2.

Figure 3.2: Degree of vocal fold vibration in phonation types

As we can see in figure 3.2, when the degree of vocal fold vibration is considered Ladefoged’s (1971) aperture gives us a different ordering. The glottal stop has the least or no vibration at all through voiceless, creaky, breathy, and the modal (voiced) as the most vibrant phonation type.

In the next section we make a brief review of the linguistic use of phonation types.

3.2.2. Linguistic use of phonation types

Phonation type is linguistically employed to constitute a phonological contrast. For example, voicing contrasts have been examined for a number of languages, and majority of the languages in the world use voicing to contrast their consonants. The UCLA UPSID languages survey (Maddieson 1984) lists 291 (91.8%) languages with plain voiceless phonation and 212 (66.9%) plain voiced stop phonation. This contrast is particularly common among obstruents (sonorants too contrast their voicing but to a lesser degree, for example, Burmese contrasts voiceless and voiced nasals (Ladefoged & Maddieson 1996)) and is employed in a number of widely-spoken languages, such as
English, Japanese, Arabic and Russian (Gordon & Ladefoged 1999). For example, English (British and American) makes stop voicing contrast at three places as shown in table 3.1.

Table 3.1: English voiced stop phonation

<table>
<thead>
<tr>
<th>Place</th>
<th>Sound</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>Labial</td>
<td>pɔ:rˈt</td>
<td>'port'</td>
</tr>
<tr>
<td></td>
<td>bɔ:t</td>
<td>'bought'</td>
</tr>
<tr>
<td>Alveolar</td>
<td>tu:</td>
<td>'two,too'</td>
</tr>
<tr>
<td></td>
<td>dɯ:</td>
<td>'do'</td>
</tr>
<tr>
<td>Velar</td>
<td>kud</td>
<td>'could'</td>
</tr>
<tr>
<td></td>
<td>gud</td>
<td>'good'</td>
</tr>
</tbody>
</table>

A lesser frequent phonation type is the breathy voiced type. Breathy phonation is characterized by vocal cords that are fairly abducted (relative to modal and creaky voice) and have little longitudinal tension (Ladefoged 1971, Laver 1980); this results in some turbulent airflow through the glottis and the auditory impression of 'voice mixed in with breath' (Catford 1977). A good number of languages in India seem to have the breathy stop in their phonemic inventory. However, the occurrence of breathy sounds seems to be infrequent in non-Indian languages. According to Maddieson (1984) the UPSID languages survey lists only seven (2.2%) languages with breathy voiced stops and one (0.2%) with breathy voiced lateral. Hindi (Ladefoged 2001) lists four phonation types, contrasting modal voice with breathy voiced phonation as shown in table 3.2.

---

34 Out of 69 sound inventories of languages of India, 26 have voiced aspirated (breathy) stops, and all have plain voiced stops and voiceless aspirates as well (Pandey 2003).
Table 3.2: Hindi breathy stop phonation (after Ladefoged 2001)

<table>
<thead>
<tr>
<th></th>
<th>Labial</th>
<th>pal</th>
<th>phal</th>
<th>bal</th>
<th>bhial</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>'take care of'</td>
<td>'knife blade'</td>
<td>'hair'</td>
<td>'forehead'</td>
</tr>
<tr>
<td></td>
<td>Dental</td>
<td>tal</td>
<td>thal</td>
<td>dal</td>
<td>dřlar</td>
</tr>
<tr>
<td></td>
<td></td>
<td>'beat'</td>
<td>'plate'</td>
<td>'lentil'</td>
<td>'knife'</td>
</tr>
<tr>
<td></td>
<td>Post-alveolar</td>
<td>tål</td>
<td>tňal</td>
<td>qal</td>
<td>qďlal</td>
</tr>
<tr>
<td></td>
<td></td>
<td>'postpone'</td>
<td>'wood shop'</td>
<td>'branch'</td>
<td>'sheild'</td>
</tr>
<tr>
<td></td>
<td>Velar</td>
<td>kal</td>
<td>khal</td>
<td>gal</td>
<td>gďal</td>
</tr>
<tr>
<td></td>
<td></td>
<td>'era'</td>
<td>'skin'</td>
<td>'cheek'</td>
<td>'confusion'</td>
</tr>
</tbody>
</table>

Creaky phonation seems to be relatively rare in languages. The UCLA UPSID languages survey (Maddieson 1984) does not list a single language with creaky phonation in consonants. However, some languages do contrast creaky and modal voicing among their sonorants. This type of contrast is particularly common in Northwest American Indian languages, e.g., Kwak'wala, Montana Salish, Hupa, and Kashaya Pomo (Gordon & Ladefoged 1999), among others. Creaky phonation is typically associated with vocal folds that are tightly adducted but open enough along a portion of their length to allow for voicing (Ladefoged 1971, Laver 1980). The acoustic result of this laryngeal setting is a series of irregularly spaced vocal pulses that give the auditory impression of a 'rapid series of taps, like a stick being run along a railing' (Catford 1964). Like the contrast between breathy and modal voiced among obstruents, contrasts between creaky and modal voice are also relatively rare in obstruents, though Hausa and (certain other) Chadic (Gordon & Ladefoged 1999) languages make such a contrast for stops. A good example of contrast between modal voice and creaky voice is Kwak'wala (Gordon & Ladefoged 1999, following Boas 1947). Kwak'wala contrasts modal voiced nasals with creaky voiced as in table 3.3.
Table 3.3: Creaky nasal phonation in Kwakw'ala

<table>
<thead>
<tr>
<th>Nam</th>
<th>‘one’</th>
<th>Ṇạṇạna</th>
<th>‘nine’</th>
</tr>
</thead>
<tbody>
<tr>
<td>Naka</td>
<td>‘drinking’</td>
<td>Ṇạla</td>
<td>‘day’</td>
</tr>
</tbody>
</table>

In Ladefoged’s (1971) continuum of phonation types, the last one is the most closed, the phonation for glottal stop /ʔ/, which entails an absence of vocal fold vibration. Often, phonemic glottal stops are realized as creaky phonation on neighboring sounds rather than with complete glottal closure (Ladefoged & Maddieson 1996). Not many languages have glottal stop as a contrastive function, however there are languages such as Lawa (Schlatter 1976) and Ao (Temsunungsang 2009) which make partial contrast of final glottal stop with open syllable and final voiceless velar stop with final glottal stop respectively as shown in table 3.4 and 3.5.

Table 3.4: Glottal stop (partial) contrast in Lawa

<table>
<thead>
<tr>
<th></th>
<th>Open syllable</th>
<th>Partial contrast</th>
</tr>
</thead>
<tbody>
<tr>
<td>So</td>
<td>‘wild cat’</td>
<td>So?</td>
</tr>
<tr>
<td>So</td>
<td>‘to water’</td>
<td>So?</td>
</tr>
<tr>
<td>Te</td>
<td>‘plunder’</td>
<td>Te?</td>
</tr>
<tr>
<td>Pi</td>
<td>‘tired’</td>
<td>Pi?</td>
</tr>
</tbody>
</table>

Table 3.5: Glottal stop contrast in Ao

<table>
<thead>
<tr>
<th></th>
<th>Hatch’</th>
<th>Mu?</th>
<th>‘bite’</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sak</td>
<td>‘scratch’</td>
<td>Sa?</td>
<td>‘measure’</td>
</tr>
<tr>
<td>Mwsak</td>
<td>‘itch’</td>
<td>Mwsa?</td>
<td>‘ask’</td>
</tr>
<tr>
<td>Cak</td>
<td>‘break’</td>
<td>Ca?</td>
<td>‘boil’</td>
</tr>
</tbody>
</table>
There is yet another phonation type which is not discussed in Ladefoged’s (1971) phonation continuum (because his continuum is based on the glottal aperture): the tense (based on the glottal tension) phonation type (as in consonants, and not in vowels) is so far believed to be attested only in Korean (see the section 3.9). Korean makes a three-way phonation distinction with obstruents: voiceless tense, voiceless lax and voiceless aspirated. A number of linguists have studied the tense and lax phonation distinction in Korean; for example, Abramson & Lisker (1964) Hirose, Lee & Ushijima (1973), Cho & Keating (2001), Ahn & Iverson (2003), Kim & Duanmu (2004), Kim (2004), Silva (2006), Kenstowicz & Park (2006), Chang (2006) among others. Korean has three voiceless (tense, lax and aspirated) stops across three places (labial, alveolar and velar). Table 3.6 shows Korean tense stops contrasting with lax and aspirated, following Ahn & Iverson (2003).

Table 3.6: Tense stop contrast in Korean

<table>
<thead>
<tr>
<th>Phonation Type</th>
<th>Labial</th>
<th>Alveolar</th>
<th>Velar</th>
</tr>
</thead>
<tbody>
<tr>
<td>Labial</td>
<td>p*ul</td>
<td>pul</td>
<td>phul</td>
</tr>
<tr>
<td></td>
<td>‘horn’</td>
<td>‘fire’</td>
<td>‘grass’</td>
</tr>
<tr>
<td>Alveolar</td>
<td>t*al</td>
<td>tal</td>
<td>thal</td>
</tr>
<tr>
<td></td>
<td>‘daughter’</td>
<td>‘moon’</td>
<td>‘mask’</td>
</tr>
<tr>
<td>Velar</td>
<td>k*in</td>
<td>kin’</td>
<td>khin</td>
</tr>
<tr>
<td></td>
<td>‘rope’</td>
<td>‘weight unit’</td>
<td>‘large’</td>
</tr>
</tbody>
</table>

We postpone the detailed discussion of Korean stop phonation until section 3.9. In the next section, we begin our discussion of Dzongkha obstruent phonation types.

3.3 Dzongkha obstruent phonation: VLT*, VLL, VLAS and VD

In chapter 2, we proposed a new phonation type for the obstruents; the ‘voiceless tense’ (VLT*) (see chapter 2, section 2.3.1 for the four-way
phonation contrast across all four places). This section will further our deliberation on the four-way phonation distinction in Dzongkha.

3.3.1 Literature on the stop phonation types in Dzongkha

As we have stated in chapter 2 and in the beginning of this chapter as well, there is no comprehensive segmental study of Dzongkha so far. In the phonemic inventory of Mazaudon (1985), he lists four classes of stop phonation types, but phonation specifications are not mentioned.

Mazaudon & Michailovsky (1988) portrays the four phonation types in stops as: voiceless, aspirated, voiced and devoiced. He associates the voiceless and aspirated with the high register and voiced and devoiced with the low register.

Van Driem (1992) goes a step further in describing the stop phonation types in Dzongkha. He portrays the four-way distinction as: voiceless, aspirated, voiced and devoiced and further draws attention to the distinction between the voiced and the devoiced phonation types in standard Dzongkha, in that he states that the distinction is not observed in some regional dialects, especially in the eastern dialects. According to him, the devoiced types are called so because they are historically derived from voiced stops. He states that the devoiced stops are unvoiced, but are in contrast with the voiceless stops. The devoiced stop is followed by murmured or breathy voiced vowels in the low register tone, while the vowels following voiced stops are clear vowels. Both voiced and devoiced stops correlate with the low register tone.

Till date, the most recent and extensive study on phonation types in Dzongkha is Watters (2002). He makes a comprehensive phonetic study of the stop phonation types. According to him, the four phonation types are: voiceless (without aspiration), devoiced (voiceless sometimes with slight aspiration, which is sometimes voiced and followed usually by breathy voice),

35 Watters' (2002) analysis of stop phonation types is in Sherpa, but he claims that the findings are the same even in case of Dzongkha, Mugom, Lhorni and Dolpo Tibetan.
aspirated (voiceless with heavy aspiration (followed by modal voice)), and voiced (voiceless preceded by voicing\(^{36}\)). He observes that the devoiced series is characterised by a slightly longer VOT than the voiceless unaspirated series, and is accompanied sometimes by a small degree of aspiration, but the aspiration is not consistent, and thus not phonemic according to him. “On a scale of time from long to short, voiceless aspirated sounds have the longest period of voicelessness, followed by fully devoiced, voiceless unaspirated, and initially voiced” (Waters 2002). Voiceless and voiceless aspirated stops occur in the high register tone and voiced and devoiced in the low register tone in his analysis as well.

In three, out of the four linguists, Mazaudon & Michailovsky (1988), Van Driem (1992) and Watters (2002) the stop phonation types are portrayed as: voiceless, voiceless aspirated, voiced and devoiced. But phonologically, there is a problem with the devoiced phonation type. Two questions arise about the devoiced type: What is ‘devoiced’ phonologically? (There is no record of languages in literature employing the term ‘devoiced’ as phonological) And is this feature attested in other languages as well? We need to find answers to these questions first.

Crystal (2008) defines the term ‘devoiced’ as: “A sound which is normally voiced, but which in a particular phonetic environment is produced with less voice than elsewhere, or with no voice at all”. By Crystal’s definition, devoiced is not a phonological feature, but a phonetic one. There is a difference between phonetic and phonological features; distinctive features are phonological features and phonetic features are surface realisations of underlying phonological features, which may be realised by more than one phonetic feature. Phonology, on the other hand, is not just concerned with categories such as, consonants, vowels, phonemes, allophones, etc. but is also crucially about relations (for instance, the relationship between stop types - for our current interest – the categories: VLT* vs. VLAS vs. VLL vs. VD). And this is exactly where the problem of Dzongkha phonation devoiced comes in.

\(^{36}\) Watters describes the voiced stop having two phonetic composites: a voiced composite followed by a voiceless one. In our tokens we observe that this phenomenon is not consistent, though few tokens show voicing undershoot before the stop closure is released.
In the sound system of any one language we typically find a number of
common sounds plus a number of rare sounds, and the so called 'devoiced'
(VLL) sound may be considered rare in languages\textsuperscript{37}. Considerations of such
rare sounds may be attributed to their historical origins in the language, but
synchronically it lacks proper phonological categorization in Dzongkha. Thus,
in the remaining sections of the chapter we study the acoustic cues in
Dzongkha obstruent phonation contrasts: four-way for stops (VLT*, VLAS,
VLL and VD) and three-way for fricatives (VLT*, VLL and VD).

Before we discuss our stop phonation types in detail in section 3.4, it is
worthwhile making a quick survey of the phonation patterns in some other
Tibeto-Burman languages. In the following section we take a quick survey of
stop phonation types in some Tibeto-Burman languages.

3.3.2 Typology of stop phonation types in Tibeto-Burman languages

As in other languages of the world, the Tibeto-Burman languages too exhibit a
rich array of stop consonants in their phonemic inventories. The minimum
phonatory contrast is (simply) non-contrastive and the maximum is a four-way
contrast. The stop inventories of some of the modern Tibeto-Burman languages
after Michailovsky (1993) give the following inventories as in table 3.7.

\textsuperscript{37} So far only Korean is supposed to have a three-way voiceless stop contrast, including tense
and lax types.
Table 3.7: Stop inventories of some of the modern Tibeto-Burman languages (Michailovsky 1993)

<table>
<thead>
<tr>
<th>Language</th>
<th>Stop Inventories</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hayu</td>
<td>k kh g, ts tsh dz, t th d, p ph b</td>
</tr>
<tr>
<td>Bahing</td>
<td>k kh g, gh, ts dz dzh, t th d, p ph b, bh, ?b</td>
</tr>
<tr>
<td>Sunwar</td>
<td>k kh g, c j, t th t th d, p ph b, ?b</td>
</tr>
<tr>
<td>Dumi</td>
<td>k kh g, gh, ts dz, t th d, p ph b, bh</td>
</tr>
<tr>
<td>Thulung</td>
<td>k kh g, gh, c ch j, jh, t th d, t th d, t th d, d b, p ph b, bh</td>
</tr>
<tr>
<td>Khaling</td>
<td>k kh g, gh, c ch j, jh, t th d, d b, p ph b, bh</td>
</tr>
<tr>
<td>Kulung</td>
<td>k kh g, c ch j, t th d, d b, p ph b, bh</td>
</tr>
<tr>
<td>Chamling</td>
<td>k kh, k th, t sh, t th d, d b, p ph b, bh</td>
</tr>
<tr>
<td>Bantawa</td>
<td>k kh, c ch, t th d, d b, p ph b, bh</td>
</tr>
<tr>
<td>Limbu</td>
<td>k kh, ts (s), t th, p ph</td>
</tr>
</tbody>
</table>

As we notice in table 3.7, Michailovsky (1993) does not have a single language that does not contrast stops in phonation types; every language makes a minimum of two-way (for instance, voiceless and voiceless aspirated stops in Limbu) phonation contrast. We then assume that a non-contrastive stop phonation in Tibeto-Burman languages is very rare as it is the case in other language families. However, it is a very atypical case with the Chungli dialect of Ao (Temsunungsang 2009) of North-East India, which has four voiceless series of stops (/p/, /t/, /k/ and /ʔ/) without voicing contrasts. This is the result of an ongoing sound change involving de-aspiration (Temsunungsang 2009). A two-way contrast language is exemplified by Limbu, where there is a contrast between voiceless unaspirated and voiceless aspirated stops through four places of articulation. We may note here the total lack of voicing distinction in Limbu, and its lacking in the velar series in Bantawa and Chamling as well. In the three-way distinction, Hayu exhibits an array of 3x3 stop contrasts with their places of articulation at labial, dental and velar. On the other hand, Sunwar contrasts a three-way distinction at labial, dental and velar places of articulation, with a two-way contrast at two places of articulation (palatal and retroflex) making it a five place distinction. Finally, the four-way phonation contrast of stops in Tibeto-Burman languages is exemplified by Bahing, Dumi, Bantawa, and Chamling.

\[26\] In Limbu, [tsh] is found as an allophone of 's'.

56
Thulung, Khaling, Kulung, Chamling and Bantawa. A good number of Tibeto-Burman languages seem to have four-way phonation contrasts as shown in table 3.7.

According to Matisoff (2003), in many Himalayish languages mentioned in table 3.7, and including Chepang and Newar, there is a four-way phonation contrast with stops. He believes that the fourth phonation type, the breathy voiced or the voiced aspirated type is due to the Indospheric influence, first confined to borrowings from Indo-Aryan, but now occurring in native Tibeto-Burman vocabulary as well. Following Benedict (1972), the Proto-Tibeto-Burman (PTB) stop is reconstructed as having a simple two-way distinction, voiced and voiceless. According to him, aspiration is a secondary development related to the loss of prefixes. This further validates the belief that the presence of voiced aspirated series is an influence of borrowing from Indo-Aryan languages that are in contact with the Tibeto-Burman languages of the region. We then propose that the four-way phonation system in Dzongkha is unique to the language and also in Tibeto-Burman languages, and it does not seem to be a borrowed feature from any Indo-Aryan languages.

In the next section, we discuss the phonetic features of the four-way phonation distinction in Dzongkha stops (dental), and put forward the fourth phonation type, the devoiced, as voiceless lax (VLL), thus making a three-way voiceless distinctions in stops: voiceless tense (VLT), voiceless lax (VLL), voiceless aspirated (VLAS), in addition to a voiced (VD) one.

3.4 The production study of Dzongkha stop phonation

In this section, we explore the potential phonetic cues that make the four-way phonation distinctions in stops: VLT, VLL, VLAS and VD. In doing so, we make a production study of the four-way phonation types in dental stops with data collected from native speakers. To contrast the four stops we look at four phonetic cues, viz. VOT, stop intensity burst, closure duration and F0 perturbation in the following vowel of the stops. Although closure duration is used by linguists (Weismer 1980, Maddieson 1997a, Cho & Ladefoged 1999,
Dutta 2007) as an acoustic cue in distinguishing stops, measurement of closure duration appears to be a problematic task for accuracy purposes due to the unpredictable pauses made by speakers (in the initial position) (Dutta 2007) (He measures closure duration in the medial positions), but in the medial position the VLL series become voiced in Dzongkha. So, the only option left for us is to measure closure duration in the medial position in nonsensical words.

However, we may take note of what Positional Faithfulness theorists (Beckman 1988) have to suggest in regard to phonetic and phonological characteristics of segments in different prosodic positions. Beckman (1988) observed that in the perceptually or psycholinguistically prominent positions (roots, root-initial syllables, stressed syllables, and syllable onsets) segmental or featural contrasts are often maintained, though they may be neutralized in non-prominent positions. It has been observed in the literature that certain positions and segment types, which may be considered perceptually prominent, are more resistant to phonological changes, compared to less prominent counterparts. Furthermore, utterance-initial portions make better cues for word recognition and lexical retrieval than either final or medial portions (Horowitz et al. 1968 & Nooteboom 1981) following Beckman (1988). Thus, we choose to consider phonetic features of our data in the word-initial position.

We look at our first acoustic cue of the stops, the VOT, in the next section.

3.4.1 Voice Onset Time (VOT): a brief review

A brief mention of general assumptions made on VOT theory is necessary before we embark on VOT characteristics in Dzongkha. Voice Onset Time (henceforth VOT) is the duration of the period of time between the release of a plosive and the beginning of vocal fold vibration. It is a useful means to distinguish stops. At least three types of stops are said to be distinguished with differences in VOT: Zero VOT for voiceless unaspirated stops - where the onset of vocal fold vibration coincides (approximately) with the plosive
release; Positive VOT for aspirated stops - where there is a delay in the onset of vocal fold vibration after the plosive release and; Negative VOT for voiced stops - where the onset of vocal fold vibration precedes the plosive release.

Phonetic studies on VOT have been conducted in languages as early as 1960s. Since the seminal study by Lisker & Abramson (1964), VOT has been used as a successful acoustic measure in comparing stop categories across many languages. The VOT Theory (Lister & Abramsom 1964, Abramson & Lisker 1967, Abramson 1977, Poon & Mateer 1985, Cho & Ladefoged 1999, Ladefoged & Cho 2001, Gordon & Ladefoged 2001, Kim 2008, Kong 2009, Chen, Chao & Mikuteit 2009, among others) attempts to construct phonation contrasts in terms of differences in VOT. A major contribution of VOT was in identifying the phonetic correlates of voicing. Under this theory, VOT is a phonetic manifestation of underlying differences in laryngeal timing in relation to the release of oral occlusion. Lisker & Abramson (1964), Abramson & Lisker (1967) and Abramson (1977) show that stops in a great number of languages could be discriminated on the basis of differences in VOT. A VOT based theory of stop discrimination, however, is only able to distinguish between voiced stops, voiceless stops and voiceless aspirated stops based on differences in the timing of the onset of the vocal fold vibration. Voiced stops are said to have negative VOT values due to the presence of voicing during closure. Voiceless stops are considered to be short-lag stops and voiceless aspirated stops are considered long-lag stops. In a four-way contrasts language like Hindi (Dutta 2007) along with VOT the language employs closure duration and F0 perturbation in the following vowel to make a four-way distinction.

3.4.2 VOT in Dzongkha stops

In our study of Dzongkha stops three tokens of VOT values of 4 native speakers (2 female and 2 male) were taken in word-isolation. The words were:

---

39 The Language examples in Lisker & Abramson (1964) are: Dutch, Spanish, Hungarian, Tamil, Cantonese, English, Armenian and Thai.
Acoustic measures were taken using Praat (Boersma & Weenink 2006) speech analysis software. VOT values were measured from the release burst of the stop to the onset of voicing in the waveform and spectrogram, as seen in figure 3.3. The results of their average VOT values are presented in table 3.8 and figure 3.3.

**Table 3.8: Average VOT values (msec) of stop phonation types**  
(F-Female, M-Male)

<table>
<thead>
<tr>
<th></th>
<th>VLT*</th>
<th>VLL</th>
<th>VLAS</th>
<th>VD</th>
</tr>
</thead>
<tbody>
<tr>
<td>F1</td>
<td>8.33</td>
<td>26.33</td>
<td>85</td>
<td>-105</td>
</tr>
<tr>
<td>F2</td>
<td>8.66</td>
<td>20.66</td>
<td>62</td>
<td>-105.67</td>
</tr>
<tr>
<td>M1</td>
<td>9.33</td>
<td>15.66</td>
<td>84</td>
<td>-88.67</td>
</tr>
<tr>
<td>M2</td>
<td>9.66</td>
<td>15.33</td>
<td>69.66</td>
<td>-153.67</td>
</tr>
<tr>
<td>Mean</td>
<td>8.9</td>
<td>19.49</td>
<td>75.17</td>
<td>-113.25</td>
</tr>
</tbody>
</table>

**Figure 3.3: Histogram of VOT (msec) differences in Dzongkha stops**
As shown in table 3.8 and figure 3.3, we see the VOT of Dzongkha stops can distinguish at least three types of stops: VLT*, VLAS and VD; VD stop has negative VOT (ranging from -154 to -89msec), VLAS has positive VOT (ranging from 62 to 85msec) and VLT* has the shortest VOT (ranging from 8 to 10msec) and VLL has an intermediate VOT between VLT* and VLAS stops, which may not be as significant as the difference in others stop types. The overall average VOT of all the four speakers is as follows: VLT* is 9msec, VLL is 20msec, VLAS is 75msec and VD is -113msec. The VOT differences of one of the speakers (F1) is presented in figure 3.4 a, b, c and d.

Figure 3.4: Spectrogram and waveform displays of VOT differences in VLT, VLL, VLAS and VD phonation types

(a) /t*ó/ 'food'

(b) /to/ 'two'
As we can see in figure 3.4 a, b, c and d, the VOT duration of the VLT* stop is (approximately) 9msec, the VLL stop is (approximately) 23msec, the VLAS is (approximately) 75msec, and the VD stop is (approximately) -98msec. The average VOT difference range (in msec) for all stop type is as follows:

- VLT*: 8 to 10
- VLL: 15 to 26
- VLAS: 62 to 85
- VD: -154 to -89

A hasty look at our production results seem to suggest that the VOT differences may be a good enough cue to distinguish all the four stops in
Dzongkha, but as we will see in our perception study in section 3.7.4, VOT is able to distinguish only the VLAS and the VD stops with a 100% perception results (see section 3.7.4 for the perception study result). Along with VOT there are, perhaps, other acoustic cues responsible for distinguishing the VLT* and the VLL stops in Dzongkha. Dzongkha is then anomalous in having a three-way contrast in voiceless category. Hence, there is a need to look for other parameters of acoustic cues to keep VLT and VLL distinct. Thus, we look at stop burst cues in all the four stops in the next section.

3.5 Stop burst in Dzongkha stops

Stop Burst\(^{40}\) is another acoustic cue linguists employ in distinguishing stop phonation, for example, Cho & Keating (2001). Cho & Keating (2001) observe that stop burst energy and voicing into closure, were found to vary with prosodic position and, in some cases, to correlate with linguopalatal contact, and thus potentially provide acoustic cues to listeners about prosodic structure. In our study, if systematic correlation between stops and their energy bursts at release is shown it would help us contrast the stops (especially the VLT* - VLL types).

In our study, average stop burst energy (intensity) of three tokens of the same four speakers were taken (precisely at the stop burst) for the same words we used for VOT, and the results are presented in table 3.9 and figure 3.5.

<table>
<thead>
<tr>
<th>Table 3.9: Average stop burst energy (in dB) of Dzongkha stops</th>
</tr>
</thead>
<tbody>
<tr>
<td>Female 1</td>
</tr>
<tr>
<td>Female 2</td>
</tr>
<tr>
<td>Male 1</td>
</tr>
<tr>
<td>Male 2</td>
</tr>
<tr>
<td>Average</td>
</tr>
</tbody>
</table>

\(^{40}\) The amplitude of the burst, which occurs within the first 10-15 ms following consonant release (Odhe & Stevens 1983)
In our results in table 3.9 and figure 3.5, it is hard to say if there is any systematic correlation between the mean stop burst values and the VD and VLAS stop types. The burst value for the VLAS stop is the highest in Female 1 and Male 1, while it is the lowest in Female 2, and in Male 2 it is slightly higher than the VLL stop. A similar picture emerges in the VD type as well, in that, Female 2 and Male 2 have the highest value for the VD type while in Female 1 and Male 1 the stop burst values (more or less) correlate with the VLT* stop. The VLAS and the VD stop types, then, seem to employ VOT as the primary phonetic cue for their contrast as illustrated in section 3.4.2. The mean stop burst values for the VLT*, VLAS and VD types are more or less same: VLT* is 63.25 dB, VLAS is 62.25 dB and VD is 63.5 dB.

However, the stop burst values for the VLL stop type is the lowest in all the speakers (except that in Female 2 it is more or less same as the VLAS type, but still the lowest). Leaving aside the VLT* and VD stop types (since they employ VOT as the main cue), the stop burst values for the VLT* and VLL stop types are significantly apart. This is shown in a tabular form below.
<table>
<thead>
<tr>
<th></th>
<th>VLT* (dB)</th>
<th>VLL (dB)</th>
<th>Difference (dB)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Female 1</td>
<td>64</td>
<td>58</td>
<td>6</td>
</tr>
<tr>
<td>Female 2</td>
<td>65</td>
<td>58</td>
<td>7</td>
</tr>
<tr>
<td>Male 1</td>
<td>58</td>
<td>50</td>
<td>8</td>
</tr>
<tr>
<td>Male 2</td>
<td>66</td>
<td>59</td>
<td>7</td>
</tr>
<tr>
<td>Mean</td>
<td>63.25</td>
<td>56.25</td>
<td>7</td>
</tr>
</tbody>
</table>

On an average, the stop burst of the VLT* stop is higher than the VLL type by 7 dB, which is significant, as shown above. Thus, we show that the VLT* and VLL stops employ stop burst as one of the cues for their contrast\(^{41}\).

In the next section, we look at Closure Duration of the stops.

### 3.6 Closure duration in Dzongkha stops

For closure duration, we look at nonce words containing all the stop types in the intervocalic position (aCa) of the same four speakers we have used for VOT and stop burst. The words were nonce words but the stop sounds are the letter sounds of the Dzongkha alphabet. We study closure duration in the intervocalic position because the acoustic demarcation of closure duration in the initial position is either not reliable or it is not traceable in isolated words. With this technique we could have our speakers focus on the letters (stop sounds) rather than lexical words. The mean values of the four tokens recorded in isolation give us the results in table 3.10 and figure 3.6.

\(^{41}\) Interestingly, in traditional Dzongkha phonology, there is a (though, paradoxical) distinction made between the consonants which are distinguished by their degree of phonation intensity. Van Driem (1992) mentions about (paradoxically) a distinction between (Classical Tibetan terms) dra-dzampa (‘soft’ or tender sound) and dra-takpa (‘hard’ or severe sound). But the paradox (as noted by Van Driem) is that the VLL (the devoiced, following earlier linguists) stops, along with sibilants as well as low register voiced liquids and nasals are termed as ‘soft’ sounds, whereas the VD stops and sibilants as well as high register voiced liquids and nasals are termed ‘hard’. In the traditional analysis, all voiced sibilants and sonorants which occur in the high register are also termed as hard sounds, thus the paradox.
Table 3.10: Mean closure duration (msec) of stops

<table>
<thead>
<tr>
<th></th>
<th>VLT*</th>
<th>VLL</th>
<th>VLAS</th>
<th>VD</th>
</tr>
</thead>
<tbody>
<tr>
<td>F 1</td>
<td>152</td>
<td>116</td>
<td>114</td>
<td>119</td>
</tr>
<tr>
<td>F 2</td>
<td>165</td>
<td>121</td>
<td>102</td>
<td>125</td>
</tr>
<tr>
<td>M1</td>
<td>104</td>
<td>75</td>
<td>58</td>
<td>86</td>
</tr>
<tr>
<td>M2</td>
<td>139</td>
<td>125</td>
<td>84</td>
<td>126</td>
</tr>
<tr>
<td>Mean</td>
<td>140</td>
<td>109.25</td>
<td>89.5</td>
<td>114</td>
</tr>
</tbody>
</table>

Figure 3.6: Histogram of mean closure duration (msec) of Dzongkha stops

In our results in table 3.10 and figure 3.6, we see three correlations emerging: the VLT* has the longest closure duration; the VLAS has the shortest closure duration (though not significantly shorter in F1) and the VLL and the VD have (more or less) same closure duration. We, then, see three types of correlations in the stop types – closure duration: the VLT* stop type has the longest closure duration; the VLAS stop type has the shortest closure duration (excepting F1) and the VLL and VD stop types have (more or less) an intermediate closure duration between the VLT* and the VLAS stop types. On an average, the VLT* is 140 msec, the VLAS is 89.5 msec and the VD and the VLL types are 114 msec and 109.25 msec respectively.

We can, then, assume that closure duration is able to distinguish stop types into three categories: VLT* (longest closure duration), VD and VLL (intermediate closure duration) and VD as one category; and VLAS (shortest closure duration).
In the following section, we study the spectral tilt (H1 – H2) of the stops.

### 3.7 Spectral tilt (H1-H2)

Many linguists, for example, Andruski & Ratliff 200; Wayland & Jongman 2003; Esposito 2003, 2005, 2006; Keating & Esposito 2007, have focused on the acoustic measure H1-H2, the difference between the amplitudes of H1 (the first harmonic, i.e. the fundamental) and H2 (the second) for the phonetic contrast of breathy and modal vowels. The Open Quotient is the proportion of the glottal period where vocal folds are not in contact. This comprises the portion of the glottal wave between the abduction of the vocal folds at their upper margin until adduction occurs along their lower or central margins (Rothenberg 1981, Titze & Talkin 1981, Childers & Krishnamurthy 1985, Childers & Lee 1991, Michaud 2004). The impedance of the vocal folds is directly related to the Closed Quotient, where an increase due to strong adductive tension and medial compression of the vocal folds leads to longer closed periods in the glottal cycle. Subglottal pressure builds up over a longer duration in these cases. Conversely, low impedance on glottal airflow due to weak adductive tension, as found in modal (or tense) phonation, causes an excitement of higher harmonics; slower vocal fold closure which occurs with less adductive vocal fold tension, as found in breathy phonation, does not excite the upper harmonics and causes a lowering of the harmonics' amplitude (Ni Chasaide & Gobl 1997, Ladefoged et al. 1988, Pennington 2005). Thus, one measures the amplitude of higher harmonics to see how ‘Tense’ or ‘Lax’ the vocal folds are during phonation in vowel.

Spectral tilt measures in vowel may be divided into those which compare low-range, mid-range and high-range regions of the spectrum. Low range measures like H1-H2 have a close correlation to Open Quotient values and are therefore good measures of the degree of glottal tension present in different phonation types (Holmberg et al. 1995, Stevens & Hanson 1995). H1-H2 is an acoustic measure well-suited to distinguish modal (or tense) and
breathy (or lax) phonation types, and has been applied to many languages, for example, !Xoo (Traill & Jackson 1987), Guajarati (Fischer-Jorgenson 1967) among others.

3.7.1 H1-H2 study

Keating & Esposito (2006) measures H1-H2 at (randomly) six time-points of the vowel duration. Their main aim of the study was to correlate H1-H2 values with the four tones in Mandarin. Dutta (2007) measures H1-H2 following Hindi stop types at five points (divided by percentage: 10%, 30%, 50%, 70% and 90%); his aim of the study was to check the extent of the breathiness into the vowel. Our main aim of this particular study is to show the correlation between the stop types and the vowel phonation types (modal vs. breathy) following the stops. Thus, we measure the H1-H2 of the vowel (of one token each of three speakers42) for the five consecutive wave periods exactly in the mid-point of the vowel (as shown in figure 3.8), which should give us fairly adequate amount of phonetic information about the voice quality of the vowel following the four stop types. The results are presented in figure 3.7.

Figure 3.7: H1-H2 in the vowel following four stop types

---

42 The quality of the waveforms for Male 2 is not good, therefore we do not include for our study.
In figure 3.7, we make the following observations; the values of the H1-H2 are (more or less) in this order: VLL>VD>VLAS>VLT*, which goes on to suggest that there is a glottal opening (Open Quotient) continuum, narrowest in the VLT* stop type and widest in the VLL type with VLAS and VD types in the middle. This continuum gives us the degree of breathiness in the vowel following the four stop types; a higher value of H1-H2 is found to be correlated with greater glottal Open Quotient; thus, higher the value of H1-H2, the more the degree of breathiness in the vowel and lower the H1-H2 value lesser the degree of breathiness in the vowel (Ladefoged 1971; Holmberg et al. 1995; Stevens & Hanson 1995; Sundberg et al. 1999; DiCanio 2009; Gordon & Ladefoged 2001; Garellek 2010). The degree of the continuum differs significantly between the VLT*/VLAS and VLL/VD types as shown in all the three figures in 3.7. There is a minimum of 5 dB and a maximum of 11 dB
between the high and the low. The H1-H2 values of the VLT* and VLAS stops cross-cut at some point of the vowel and the same phenomenon is seen in the VLL and VD stops as well (as shown in figure 3.7).

The characteristics of the waveform following the four stop types are further illustrated in one of the speakers (Female 1) in figure 3.8.

Figure 3.8: H1-H2 measurement and waveforms following stop types

a) VLT* stop

b) VLAS stop
In figure 3.8, we see that the waveforms in a) (for VLT*) and b) (for VLAS) have smoother pattern, showing that the airflow is not turbulent in these two stop types. But in c) (for VD) and d) (for VLL) we see the waveforms becoming comparatively jagged, showing the turbulent nature of the airflow for the breathy vowel (note that the waveform for the VLL is the most turbulent one, thus the breathiest vowel quality). The H1-H2 values for these waveforms (as seen in the figures) are as follows: VLT* is 8 dB in (a), VLAS is 11 dB in (b), VD is 18 dB in (c), and VLL is 20 dB in (d).

Thus, our H1-H2 study clearly shows that the vowel following the VLT* and VLAS stop types have modal phonation, while the vowel following the
VD and VLL stop types have breathy phonation. This is a property of the initial stops and thus phonetic; phonologically, Dzongkha does not contrast modal and breathy vowels. The physiological difference between the production of the VLT*/VLAS and the VD /VLL stop types is that the former group is produced with the impediment of the vocal folds, which is directly related to the Closed Quotient, where an increased muscular tension is created, due to strong adductive tension and medial compression of the vocal folds leading to higher subglottal pressure. While the latter group is produced with vocal folds that are fairly abducted (relative to the former group) and have lesser longitudinal tension (Ladefoged 1971, Laver 1980, Ni Chasaide & Gobl 1995); this results in some turbulent airflow through the glottis and the auditory impression of “voice mixed in with breath” (Catford, 1977). Based on the physiological process of the production of the two groups we may now term the two groups as Tense Group (VLT and VLAS) and Lax Group (VD and VLL). The Tense group stop types are followed by Modal vowel and the Lax group stop types with breathy vowel, and these characteristics in the vowel are the co-articulation effects of the initial stop types.

3.7.2 F0 perturbation theory

F0 perturbation is the effect of the consonants in the following vowel’s fundamental frequency (F0) or pitch. It is a well-established fact in the literature that the preceding consonant phonation types perturb the following vowel consistently in different ways in both tonal and non-tonal languages. Hombert, Ohala & Ewen (1979) suggest that if articulatory constraints cause the post-consonantal F0 variations which are implicated in the development of tones, it should be possible to find such variations in the speech of anyone producing the relevant consonant types, whether they speak a tone language or not. In other words, all languages that contrastively use breathy, creaky and modal phonation, tonal contrasts and laryngeal contrasts such as aspiration, all exhibit lowering of F0 following voiced segments (Ohala 1973, Hombert 1978). The major perturbation effect is felt by vowels followed by voiced and voiceless stops; voiced stops tend to lower F0 while voiceless stops raise it.
Ladefoged (1967), based on the phonetic basis of pitch perturbation, classifies two types of hypotheses pertaining to F0 perturbation: 'aerodynamic' and 'vocal cord tension' hypothesis. The aerodynamic hypothesis runs as follows, following Ladefoged (1967):

"During a voiced stop, oral pressure gradually builds up, thus decreasing the pressure drop across the vocal cords 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 their vibration at vowel onset. As the airflow returns to normal, the F0 will too. Thus, after voiceless stops, the F0 contour will be initially high and falling."

The basic assumption of the vocal cord tension hypothesis is that, in the course of making the voiced versus voiceless distinction in stops, vocal cord tension is changed so as to affect the F0 of adjacent vowels. Halle & Stevens (1971) suggest that these intrinsic variations are the result of horizontal vocal cord tension:

"The vocal cords are presumably slack in order to facilitate voicing during voiced stops, and stiff in order to inhibit voicing during voiceless stops; and these vocal-cord states spread to adjacent vowels, affecting their F0."

They further claim that the F0 should be perturbed on the pre-consonantal as well as post-consonantal vowels.

The hypotheses are supported by empirical data from different languages. In a tonal language, Gandour (1974) (in his investigation of Thai tones) found that a shorter part of the vowel was affected by the preceding consonant (about 30 ms for voiceless consonants and about 50 ms for voiced ones). In Punjabi, breathy voiced consonants became voiceless unaspirated,
leaving a low tone on the following vowel (Gill & Gleason 1969, 1972, Haudricourt 1972a, b). Glover (1970) too, in his data on Tibeto-Burman languages observes that breathy voiced consonants are stronger F0 depressors than (simple) voiced obstruents. In Ndebele, breathy voiced consonants pattern with voiced obstruents in lowering the pitch on following vowels (Ladefoged 1971).

In a non-tonal language, Dixit (1975) found that the activity in the cricothyroid muscle during the production of Hindi voiceless stops was consistently higher as opposed to voiced ones. Dutta (2007) too observes that in the initial portion (30% of the vowel), vowels following voiced and voiced aspirated stops show lower mean F0 than the vowels following voiceless and voiceless aspirated stops. However, the relation between the voiceless and voiceless aspirated stops is not established, but he further observes that the voiced aspirated tends to further lower the F0.

In Dzongkha, the F0 of the Tense group (VLT* and VLAS) is higher than the Lax group (VD and VLL), however, we cannot use F0 as a phonetic cue for the stop contrasts since F0 is already phonologized in Dzongkha, thus it is a tonal feature (see chapter 5 for a detailed discussion).

3.7.3 Summary of Dzongkha stop phonation study

Thus far, in our production study, we have successfully shown that the four stops (VLT*, VLL, VLAS and VD) in Dzongkha use different parameters of phonetic cues for their contrasts. The VOT is a primary cue in distinguishing the VLAS (longest VOT) and VD (negative VOT) stops from the VLT* and VLL stops. The VLT* and the VLL types are distinguishes by their stop burst and closure duration cues; VLT* stop types have comparatively larger values of these cues than VLL stop types. And finally, H1-H2 study showed that the vowel following the VLT* and VLAS stop types are with modal phonation while vowel following the VD and the VLL stop types have breathy phonation, which categorises the VLT* and VLAS as the Tense group and the VD and the
VLL types as Lax group. In a nutshell, our results of the production study can be summarised as follows:

- **VOT:** \( 
\text{VD} < \text{VLT*}/\text{VLL} < \text{VLAS}. \)
- **Stop burst:** \( \text{VLT*} > \text{VLAS}/\text{VLL}/\text{VD}. \)
- **Closure duration:** \( \text{VLT*} > \text{VLL}/\text{VLAS}/\text{VD}. \)
- **H1-H2:** \( (\text{VLT*} < \text{VLAS}) < (\text{VLL} > \text{VD}). \)

Based on the production study, we propose that while VLAS and VD stops are distinguished by their difference in VOT, the VLT* and the VLL stops cannot be distinguished exclusively by the difference in their VOT (this is also evidenced by the perception study in the next section). The VLT* and VLL stop types are distinguished by other cues such as, stop burst and closure duration, and the VLAS stop is distinguished by VOT from the other two voiceless stop types (VLT* and VLL). The H1-H2 study distinguishes the four stop types into two categories: the Tense and Lax groups; the VLT* and VLAS stops types are Tense and the VD and VLL stop types are Lax.

In the next section, we present the results of perception study on Dzongkha stop phonation types.

### 3.7.4 Perception study of the stop phonation types in Dzongkha

It has proven practical over a long history of research on language sound systems to rationalize phonological units and processes in terms of speech production. It is arguably the case that this early and prolonged emphasis on the articulatory foundations of sound systems was due to the fact that the articulators are open to observation in production. We can observe the movements of the lips, jaw, and the tongue, and the availability of such observations provided an important point of reference for theories of phonology by making available a set of explanatory mechanisms that can be applied to phonological patterns. However, rationalization of language sound systems from the point of view of the listener has had a more spotted history. Some of the more obvious auditory properties have been noted (for example,
sonority, Sievers, 1881), but it was only later, after the development of the sound spectrograph, that a comprehensive approach to language sound structure in terms of acoustic/auditory properties was attempted (Jakobson, Fant & Halle, 1952). Much more recent work in phonological theory has highlighted the role that perception plays in phonological processes. The perceptual basis of contrasts and features has been explored by Flemming (1997, 2002), Gordon (1999), and Kirchner (1997), among others. It is significant, therefore, that the role of speech perception in language sound systems has recently seen a revival of interest among both phoneticians and phonologists alike. These developments in speech perception and phonological research provide a solid foundation for continued and significant progress in understanding language sound systems.

Speech Perception is then the process by which the sounds of language are heard, interpreted and understood. Studies in speech perception seek to understand how human listeners recognize speech sounds and use this information to understand spoken language. The speech sound signal contains a number of acoustic cues that are used in both speech production and perception. The cues differentiate speech sounds belonging to different phonetic categories. For example, one of the most studied cues in speech is VOT.

In our study, we consider two acoustic cues: VOT and F0 of the same four stops we have used for our production study in an attempt to determine which cue plays primary role over the other in contrasting the stops. For technical purposes, the VOT could not be manipulated using speech synthesis, thus we manipulate F0 for our experiment, and presume that the stimuli not affected by F0 manipulation are affected by VOT (and of course, by other cues as well). VOT is not a homogeneous period, whose spectral character changes constantly. Therefore, we cannot cut off or add certain portion to control its duration. Nor can we compress or expand the duration, which is the method of manipulating duration in Praat. Instead of manipulating VOT, we use the VOT range (as shown in table 3.11) of one of the male speakers.
F0 of all the stops of a male speaker were manipulated, using Praat pitch manipulation synthesis, in such a way that stops with higher F0 were manipulated to have lower F0 and stops with lower F0 were manipulated to have higher F0 as follows (the manipulation is done in such a way that the high F0 for the VLT* and VLAS are lowered to the F0 of the VLL and VD, and vice versa):

Figure 3.9: Manipulated F0 for the stops
Dotted/black – original pitch
Solid/green – manipulated pitch

The pitch values were measured both in the onset and offset of the vowel, and their values reversed as shown in table 3.11 (Original Pitch is the un-manipulated pitch and Manipulated Pitch is after the manipulation). The pitch difference between the high F0 and low F0 is approximately 50 Hz in all the stops, and this difference is maintained in the manipulated stimuli as well. Table 3.11 shows the details of the pitch manipulation, including (unaltered) VOT (in milliseconds) differences of the stops.
Table 3.11: Manipulated F0 (in Hz) of Dzongkha stops

<table>
<thead>
<tr>
<th>Stimuli</th>
<th>Onset - Offset</th>
<th>Onset - Offset</th>
</tr>
</thead>
<tbody>
<tr>
<td>t'o 'food'</td>
<td>160 - 111</td>
<td>110 - 71</td>
</tr>
<tr>
<td>to 'two'</td>
<td>130 - 111</td>
<td>181 - 161</td>
</tr>
<tr>
<td>tho 'measurement'</td>
<td>153 - 100</td>
<td>105 - 50</td>
</tr>
<tr>
<td>do 'stone'</td>
<td>128 - 90</td>
<td>181 - 140</td>
</tr>
</tbody>
</table>

The manipulated stimuli were tested on 20 native speakers and their results are presented in figure 3.10 (a) and (b)\(^3\).

Figure 3.10 a: Histogram showing result (in %) of the perception study

The result shows 100% stimuli perception accuracy for the manipulated stimuli of VLAS and VD stops, which probably informs us about the primacy of VOT cue over F0 in the perception of the manipulated pitch for their contrast. 100% of all VLAS and VD stimuli were not influenced by F0 manipulations. It is probably quite valid to draw inferences, from an experiment like this, about how our listeners use one acoustic cue over another (along with other cues) in the perception of our manipulated stimuli. With a

\(^3\) We may note here that the result achieved in our experiment is a 'close to' or 'near to' rather than categorical perception. All the listeners could not make categorical perception with any of the stimuli, which suggests participation of other cues in play. Listeners were instructed to associate their stimuli 'close to' one of the words used as stimuli.
manipulation (lowering) of pitch by 50Hz the listeners could still perceive the stimuli as the VD (with negative VOT) and VLAS (with positive VOT or Aspiration) stops. So there is a primacy of VOT cues over F0 in VLAS and VD stops.

On the other hand, our result for the VLT* and VLL stops are quite different. VLT* stop has 70% of the stimuli perception correct while VLL has only 30% stimuli perception correct. The remaining 30% listeners of the VLT* stimulus perceived it either as VLL or VD stops (20% as VLL and 10% as VD) using F0 cue as shown in figure 3.10 (b).

Figure 3.10 (b) Histogram showing result of perception of stimuli:
't*o' as (> ) either 'to' or 'do'
'to' as (> ) 't*o'

In the VLL stimuli, all the other 70% perceived their stimuli as VLT* (using F0 cue). The result further goes on to suggest that the VLL stop tend to use F0 cue more often than the VLT* stop, considering that 70% of the VLL stimuli were perceived as VLT* stop on the one hand and only 30% of VLT* as either VLL or VD on the other. Our result, then suggests a phonetic trading relation between VOT and F0 (perhaps, with other cues). The VD and VLAS stops seem to use VOT as the primary cue for voicing and aspiration respectively, while the results for VLT* and VLL stops are suggestive of using F0 cues (perhaps, along with VOT, stop burst, closure duration, and breathiness in the vowel). Thus, from the results of our production and perception study we propose that:
VL*, VLL, VLAS and VD stops in Dzongkha initial position are distinguished with the correlation of VOT and F0 as primary cues (along with Closure Duration, Stop Burst and Breathiness in the vowel); neither one by itself seems able to differentiate all types of stops.

There is a phonetic trade-off between VOT and F0 for the VL* and VLL stops.

There is a primacy of VOT over F0 in the distinction of VLAS and VD stops, while in the VL* and VLL stops F0 possibly takes over as the primary cue over VOT (along with other cues).

The VL* and VLL stops employ F0 as the primary cue along with Closure Duration, Release Burst and Breathiness in the vowel.

The phonetic properties in both the pre- and post-voicing portions of the post release phase of stops are relevant to the differentiation of (especially for the VLL) stop types.

Thus, it is clear from our production and perception studies that Dzongkha makes a four way phonation contrast in the stops. While the three stop types: voiceless unaspirated (VL*), voiceless aspirated (VLAS) and voiced (VD) types are attested in many languages, the fourth type (VLL) seems to be rarely attested in languages. We further note that languages making a four-way phonatory contrast are rare as well. However, a number of Indic languages (Hindi, Marathi, Sanskrit etc.) closer home are said to have a four-way phonation contrast. We need to compare the fourth type of stop in Dzongkha with Hindi voiced aspirated or breathy type (Hindi has four stop types: voiceless unaspirated, voiceless aspirated, voiced unaspirated and voiced aspirated or breathy). In the next section, we explore the phonation type of the voiced aspirated or breathy stop in Hindi.

3.8. Summary of stop phonation type in Hindi

To the best of our knowledge, the earliest phonetic study on Hindi stop phonation is by Lisker and Abramson (1964), in which the authors do a cross-
linguistic comparative (acoustic) study of eleven languages\textsuperscript{44} including Hindi. Since then, a number of other linguists have shown curiosity in Hindi phonation types, for example, Ohala (1974), Kagaya & Hirose (1975), Purcell (1978), Benguerel & Bhatia (1980), Dixit & MacNeilage (1980), Schiefer (1992), Bali (1999) & Dutta (2007), among others. Dutta’s (2007) acoustic study perhaps is the most recent and comprehensive study in Hindi stop phonation types. In this section, we will review the findings of Hindi stop phonation types by earlier linguists, paying special attention to Dutta (2007) and Kagaya & Hirose (1975)\textsuperscript{45}.

Kagaya & Hirose (1975) look at VOT, F0 and closure duration, while Dutta (2007) looks at closure duration, voice lead time duration\textsuperscript{46}, F0, aspiration and vowel duration and spectral tilt\textsuperscript{47} of the four Hindi stop types. The consolidated findings (as in both works) of Hindi stop phonation may be summarised as follow:

\begin{itemize}
  \item The VOT of the voiceless aspirated type is longer (on the average) than the voiceless unaspirated, while the voiced types (aspirated and unaspirated) have vocal folds vibration throughout the articulatory closure period. These results are comparable to those reported by Lisker and Abramson (1964).
  \item The results from the voice lead time durations (for voiced) in conjunction with the results from closure duration (for voiceless) show that aspirated stops have comparatively shorter closures compared to the unaspirated stops\textsuperscript{48}.
  \item In the initial portion, vowels following voiceless (aspirated and unaspirated) stops have higher mean F0 than the vowels following voiced (aspirated and unaspirated) stops. Dutta (2002) further observes that the effect of the stop types permeates into the initial 20 to 30\% of the vowel
\end{itemize}

\textsuperscript{44} Dutch, Puerto Rican Spanish, Hungarian, Tamil, Cantonese, English, Eastern Armenian, Thai, Korean, Marathi and Hindi.
\textsuperscript{45} We prefer to review Kagaya and Hirose (1975) and Dutta (2007) because to my view theirs are the most comprehensive studies.
\textsuperscript{46} Voice Lead Time (VLT) refers to the period of voicing during closure.
\textsuperscript{47} Spectral Tilt is a means to measure phonation type differences by comparing H1 and F1 or H1 and F3. These are measures of the slope of the spectrum indicating how much of the energy is in the fundamental frequency as compared with higher frequencies (Ladefoged 2003).
\textsuperscript{48} Dutta (2007) does not consider VOT, he rather studies the correlation of closure duration, voice lead time and aspiration, and vowel length.
and the mean F0 of the breathy type have further F0 lowering effect (up to 50% of the vowel), while in Kagaya and Hirose (1975) the effect of initial stop is in the onset of the vowel.

On the aspiration and duration of vowels, Dutta (2007) observes that the duration of aspiration following voiceless aspirated is comparatively shorter than the duration of breathiness following voiced aspirated stop. He further observes that vowels following aspirated stops (voiced and voiceless) are significantly longer than vowels following unaspirated stops (voiced and voiceless).

On spectral tilt, Dutta (2007) observes that the voiced aspirated stop is produced with greater difference in the amplitudes of the first and second harmonic compared to the unaspirated stops. It is also produced with greater differences in the amplitudes of the first harmonic and the peak amplitude of the first formant. In addition, based on the mean differences he concludes that the breathy portion following the VAS tends to permeate till about 30-50 percent of the following vowel, and that the voiced aspirated stop is produced with a larger open quotient.

Based on the acoustic analysis, the four types of stops in Hindi are distinguishable by VOT, presence and absence of aspiration, F0 after the articulatory release, and spectral intensity. The voiceless unaspirated type is characterized by an open glottis, which facilitates voicelessness. The slightly higher vocal fold tension may also be related to the acoustic characteristics. The voiced aspirated type is characterized by the closed glottis mostly during the articulation of closure period and by the open glottis immediately before and after the release. The former glottal condition facilitates vocal fold vibration during the closure period and the later aspiration. Voicing is maintained through the aspiration period because of the absence of supraglottal constriction. The voiceless aspirated type is characterized by wide-open glottis during the entire consonantal portion, which facilitates voicelessness and gives rise to a considerable period of aspiration.
3.8.1 Comparison of Dzongkha voiceless lax stop with Hindi voiced aspirated stop

In this section, we compare the fourth type of stop in Dzongkha with the Hindi voiced aspirated stop to see if the Dzongkha stop type is similar to Hindi breathy stop type. We present their comparison in a tabular form below.

<table>
<thead>
<tr>
<th>Phonetic cues</th>
<th>Dzongkha (VLL stop)</th>
<th>Hindi (Breathy stop)</th>
</tr>
</thead>
<tbody>
<tr>
<td>VOT</td>
<td>positive</td>
<td>negative</td>
</tr>
<tr>
<td>F0</td>
<td>low</td>
<td>low&lt;sup&gt;49&lt;/sup&gt;</td>
</tr>
<tr>
<td>Stop Burst</td>
<td>low</td>
<td>?&lt;sup&gt;50&lt;/sup&gt;</td>
</tr>
<tr>
<td>Closure Duration</td>
<td>ambiguous</td>
<td>short</td>
</tr>
<tr>
<td>H1-H2</td>
<td>high</td>
<td>high</td>
</tr>
</tbody>
</table>

While the voiceless stop types (voiceless unaspirated and voiceless aspirated) and the voiced unaspirated types (more or less) in both the languages have the same phonetic characteristics, the VLL in Dzongkha and the voiced aspirated type in Hindi are distinct. The table above shows that the two stop types in Dzongkha and Hindi has two phonetic characteristics in common: low F0 and high H1-H2; however, F0 is a tonal property in Dzongkha, and that leaves us with just H1-H2 as the common phonetic characteristic. We further notice that the VOT in the two stop types are very distinct, in that the VLL type in Dzongkha has a positive VOT while the voiced aspirated or the breathy type in Hindi has a negative VOT (with voicing throughout the closure duration). We then conclude that the VLL stop in Dzongkha is very distinct from the

<sup>49</sup> We conducted an acoustic pilot study of the four dental stops (t,th,d,di) in Hindi with speakers from different regions (West Bengal, Orissa, Maharashtra and Bihar) and results seemed to suggest that while VOT cues conform with earlier findings, our preliminary review of speakers with diverse language backgrounds like Hindi, Bangla, Marathi seem to suggest that speakers may perhaps have diverse mechanisms for the phonological voiced aspirated type. This is hardly surprising as many features like tongue height, lip rounding may have different correlates across language groups/speakers etc. Speakers of Hindi from different regions and with different linguistic backgrounds seem to use F0 quite idiosyncratically. While the F0 of the voiced aspirated (or breathy) type was found consistently and significantly lower in all the speakers the F0 of the Voiced type is not significantly distinct from the voiceless types. However, we stick to the findings of Dutta (2007) for F0 in Hindi.

<sup>50</sup> As far as we know, there is no study done on the stop burst in Hindi stop types.
voiced aspirated or breathy stop in Hindi; it is clearly not the voiced aspirated type in Hindi.

Next, we wonder if the fourth stop type in Dzongkha is similar to the Korean voiceless lax stop (Korean is reported to have three-way voiceless stops: voiceless tense, voiceless lax and voiceless aspirated) without a voiced one. In the next section we review the literature in Korean and compare the fourth type of Dzongkha stop to Korean voiceless lax type.

3.9. Summary of Korean stop phonation

Korean is considered unusual among the major languages of East Asia in regard to its stop phonation types. According to Kim (1965) and many others, Korean has a three-way contrast: tense\(^{51}\) (p* t* k*), lax (p t k), and aspirated (ph th kh)) in stops (including affricates). The stops are distinguished at three places of articulation: bilabial, denti-alveolar and velar (see section 3.6 for contrast examples) in syllable onset position. The aspirated and tense stops are voiceless both initially and medially. Lax stops are voiceless initially and voiced medially when they occur between two voiced sounds. In final positions, however, all obstruents are neutralized to unreleased plain stops, such that /ph, p, p*/ is realized as [p], /th, t, t*/ as [t], and /kh, k, k*/ as [k] (Kim and Jongman 1996, Kim 1998). And there is a two-way contrast: tense and lax in denti-alveolar fricatives (we will postpone our discussion on fricatives until section 3.11).


\(^{51}\) The symbol (*) stands for Tense as in Kim (1965) and Ladefoged and Maddieson (1996).
(2002), Ahn and Iverson (2003), Kim (2004), Kim and Duanmu (2004), Silva (2006), among others. The temporal characteristics of Korean phonation types have been studied in terms of VOT (Lisker and Abramson 1964, Kim 1965, Hardcastle 1973, Cho 1996, Shimizu 1996, Ahn 1999, Cho et al. 2002) and duration of the closure (Silva 1992, Han 1996, Cho 1996). All studies of VOT agree that aspirated stops have a longer VOT than lax stops, which in turn have a longer VOT than tense stop. It was noted that there is a substantial difference in VOT between aspirated stop and the other two but no substantial difference between lax and tense stops (Lisker & Abramson 1964). Studies of the duration closure have shown that lax stops have a shorter closure than aspirated or tense stops. Cho (1996) shows that the vowel is longest after tense stops, intermediate after lax ones, and shortest in aspirated ones.

Kim (1965) demonstrated that aspirated and tense stops have a higher level of intensity of the release burst than lax stops. Cho et al. (2002) also showed that (in Seoul dialect) the percentage of the burst energy at the release to that at the midpoint of the vowel was significantly higher in aspirated stop than in the other two. Kim (1965) observed that the aspiration period of aspirated stops consistently have a greater distribution of energy than that of lax stops.

All studies on fundamental frequency agree that F0 is significantly higher at the onset of the vowel after the aspirated and tense stops than at the onset of the vowel after the lax stops, with there being no substantial difference between the onsets of the vowels after the aspirated and tense stops.

Jun (1993, 1995) found that the voicing of lax stops (in intervocalic position) depends on the closure duration of the stop (the longer the closure, the less likely it is voiced). Cho & Keating (2001) have also considered stop closure duration as an important indication of the lax and tense distinction in stop consonants. They observed that the aspirated and tense stops have longer closure duration, which were measured both articulatorily and acoustically. A clear division, however, between the aspirated and tense stops was not made, but tense stops tend to be slightly greater in terms of these measures. In
general, they found that the closure duration is longer for tense stops than for lax stops in both acoustic and articulatory duration.

The difference between lax and tense stops is reflected even in the vowels that follow the consonants. Abberton (1972) found that the onset of the vowel after tense stop has a creaky voice quality, and Han (1998) found that the onset of the vowel after lax stop has a breathy voice quality (i.e., amplitude of $H_1$ (the first harmonic) is larger than that of $H_2$ (the second harmonic)). Cho et al. (2000) reaffirmed these differences in phonation types of the following vowels. They found systematic differences in $H_1$-$H_2$ values of the onset of vowels that follow tense and lax stops. $H_1$-$H_2$ had positive values for lax, which indicates the breathiness of vowels, and negative values for tense, which indicates the creakiness of vowels.

Korean has three voiceless (tense, lax and aspirated) stops which employ at least four acoustic cues for their contrast: VOT, F0, closure duration and stop burst. The facts of Korean stop phonation types can thus be summarised as follows:

\[ \begin{align*} 
\downarrow \text{VOT:} & \quad \text{tense stop} < \text{lax stop} < \text{aspirated stops.} \\
\downarrow \text{Stop release burst:} & \quad \text{aspirated/tense stops} > \text{lax stops.} \\
\downarrow \text{F0:} & \quad \text{aspirated stop} > \text{tense stop} > \text{lax stop.} \\
\downarrow \text{Closure duration:} & \quad \text{aspirated/tense stops} > \text{lax stops.} \\
\downarrow \text{H1-H2:} & \quad \text{lax stop} > \text{aspirated/tense stops.} \\
\downarrow \text{Vowel duration:} & \quad \text{tense stop} > \text{lax stop} > \text{aspirated stop.} 
\end{align*} \]

In the next section, we compare the Dzongkha tense - lax stops with Korean tense - lax stop.

---

52 The difference in amplitude between $H_1$ and $H_2$ has been used to distinguish among modal, breathy and creaky voicing (Bickely 1982, Ladefoged 1983, Huffman 1987, Klatt & Klatt 1990, Blankenship 1997).
3.9.1 Comparison of Dzongkha fourth stop type with Korean voiceless lax stop

In this section, we compare the Dzongkha tense–lax stops with Korean tense–lax stops to see if the two stop types in the two languages have similar phonetic cues for their contrasts. Since aspirated stops can be distinguished from other voiceless stops by the feature [aspirated], the challenge is to distinguish tense and lax stops, which in turn lies in the analysis of the lax stop. We present the comparison in a tabular form below.

<table>
<thead>
<tr>
<th></th>
<th>Tense</th>
<th>Lax</th>
<th>Tense</th>
<th>Lax</th>
</tr>
</thead>
<tbody>
<tr>
<td>VOT</td>
<td>short</td>
<td>long</td>
<td>short</td>
<td>long</td>
</tr>
<tr>
<td>Closure duration</td>
<td>long</td>
<td>short</td>
<td>long</td>
<td>short</td>
</tr>
<tr>
<td>Stop burst</td>
<td>high</td>
<td>low</td>
<td>high</td>
<td>low</td>
</tr>
<tr>
<td>F0</td>
<td>high</td>
<td>low</td>
<td>high</td>
<td>low</td>
</tr>
<tr>
<td>H1-H2</td>
<td>low</td>
<td>high</td>
<td>low</td>
<td>high</td>
</tr>
</tbody>
</table>

The tabular comparison of the two stop types (tense–lax) above suggests that the two stop types in Dzongkha and Korean employ the same phonetic cues for their contrasts. The VOT is shorter in the tense stop than in the lax stop type, closure duration is longer in the tense stop than in the lax stop type, the stop burst and the F0 is higher in tense stop than in the lax stop type, and the H1-H2 is higher in the lax stop than in the tense stop type in both the languages. These correlations strongly suggest two things: 1) both the languages employ the same phonetic cues for the tense–lax contrast and 2) the fourth stop type (VLL) in Dzongkha is similar to Korean voiceless lax stop phonetically. It may be remembered that even phonologically the lax stops behave in a similar manner in both the languages, in that the lax stop becomes voiced in the intervocalic position in both the languages. This phenomenon in Korean persuades Kim & Duanmu (2004) to reanalyse the lax stop as underlyingly voiced and applies devoicing rule in the phrase initial position. This, in fact, is not an attractive analysis; it is counterintuitive since this phenomenon is not attested in any language. It is contrary to the principles of
markedness and positional faithfulness (see Beckman 1998 for positional faithfulness principles).

In the next section, we propose the phonological features for the four-way contrasts of stops in Dzongkha.

3.10 Phonological feature for Dzongkha stop contrast

What do we arrive at finally? Both production and perception study in Dzongkha suggest an acoustic trade-off between VOT and F0 in the contrast of the stops. The results of production study show that VOT is able to contrast only three types of stops: VLT*, VLAS and VD; the VOT of the VLL is not significant from the VLT* stop. The production result is further supported by our perception study, which suggests a robust (100% perception) VOT distinction between the VLAS and VD stops. In case of the contrast between the VLL and VLT* stops both production and perception results suggest that F0 takes over as, perhaps, the primary cue over VOT (along with other cues such as, stop burst, closure duration etc.). This phonetic trade-off is evident in both our production and perception studies. The question that arises then is whether the particular way in which the voicing feature is implemented in Dzongkha is found elsewhere among the languages of the world. As we have already illustrated earlier, Korean tense and lax stops behave in the same way as Dzongkha VLT* and VLL stops do.

In this section, we argue that a close parallel between the Dzongkha phonation system and Korean is found. On the basis of this comparison it is reasonable to characterize the 'devoiced' type of stop found in Dzongkha as voiceless lax. We also argue that the phonetic implementation and phonological status of the VLL stop in Dzongkha is different from that of the Hindi breathy stop.

The contrast between VLT* and VLL stops raises two theoretical issues (segmental and prosodic). Firstly, the segmental issue is that in order to distinguish them, either the contrast in VOT (aspiration) must be increased
from two to three, or a new feature TENSE is needed. The first option (increase in VOT) may not be a plausible solution since the difference in VOT of VLT* and VLL stops is not significant. The only solution we foresee is the feature, TENSE as proposed for Korean. This feature expansion is already proposed in languages, for example, for Korean. But this proposal poses another problem. The initial VLT* and VLAS stops correlate with high tone and VLL and VD with low. We cannot then explain the Dzongkha consonant – tone correlation in the standard theory of tonogenesis (voiceless - high and voiced - low). We postpone our detailed discussion on phonation type – tone (and tonogenesis) until chapter 5. Given the parallel between Dzongkha and Korean in the voice quality of the voiceless tense and lax stops, together with other similarities in the implementation of the voicing features in these two languages, we propose the feature ‘tense’ for the VLT* and VLL distinction in Dzongkha.

3.10.1 Review of distinctive feature theory

Before we propose the feature ‘Tense’ for Dzongkha, it is noteworthy making a quick review of developments in feature theory. The idea that segments are composed of a bundle of features was first thoroughly discussed by Jakobson, Fant & Halle (1952). Later, Chomsky & Halle (1968) (‘Sound Patterns of English’ popularly known as SPE) furthered the theory, which is one of the most influential contributions after Jakobson et al. (1952) in dealing with features from a phonological perspective. Although SPE and Jakobson et al. differ in their approaches in a number of issues, both approaches agree that each feature is defined in terms of some phonetic property. One of the ways SPE and Jakobson et al. systems differed is that while SPE takes the point of view that features are defined solely on articulatory terms, the Jakobsonian approach argued that features have primarily acoustic definitions. In general, phonologists have followed the SPE approach by analysing features articulatorily. In the current standard feature theory, many phonologists including McCarthy (1988), Lombardi (1991) and others put forward the laryngeal feature node as in figure 3.11 (a).
The three laryngeal nodes in figure 3.11(a) efficiently distinguish stops in languages that make contrast in: aspiration (/p/ vs. /ph/ and /b/ vs. /bʱ/ as in Hindi) by the feature [spread], voicing (/p/ vs. /b/ as in English and many other languages) by the feature [voice] and ejectives and implosives (/p/ vs. /pʰ/ and /b/ vs. /bʱ/ as in Uduk (Ladefoged and Maddieson 1996)) by the feature [constricted]. The feature [spread] accounts for the presence vs. absence of aspiration, thus it can account for both voiceless aspirated and voiced aspirated stops in, for example, Hindi. According to Halle and Stevens (1971) [+spread] sounds are produced by a displacement of the arytenoids cartilages creating a wide glottal opening, and [-spread] sounds are produced without this gesture. Halle and Stevens (1971) state that [+constricted] sounds are produced by adduction of the arytenoids cartilages causing the vocal cords to be pressed together and preventing normal vocal cord vibration; [-constricted] sounds are produced without such a gesture. The feature [voice], again accounts for the presence vs. absence of vocal cord vibration.

Thus, aspirated sounds (including breathy sounds in Hindi) are [+spread] and unaspirated [-spread], ejectives and implosives are [+constricted] and other sounds are [-constricted] and voiced sounds are [+voice] and voiceless sounds [-voice].

The laryngeal feature in figure 3.11a can account for laryngeal distinctiveness of sounds in a majority of the world’s languages. However, there are languages which use some additional features besides the three features in figure 3.11, for example, the creaky voice in Guinee and Kwakw’ala, tense voice in Korean (and now in Dzongkha too), the slack voice in Javanese, and pre-aspiration in Icelandic.

---

For lack of space, we skip discussion on the ‘Place’ and ‘Manner’ feature nodes, and restrict our discussion straightaway to ‘Laryngeal’ feature.
However, in the recent past, there has been a trend to allow acoustic or auditory definitions (for example, Flemming 1995, Boersma 1998, Steriade 200, Jessen 2001), though Lahiri & Blumstein recognised the importance of acoustic/auditory in feature theory as early as 1984. Jessen (2001) emphasises on the importance of acoustic phonetics as a prerequisite to the study of auditory/perceptual features and feature correlates as an evidence for featural classification. He believes that it is only on the basis of reliable acoustic data that one can proceed with auditory modelling, categorical perception experiments, and related methods. Jessen (2001) discusses [voice] and [tense] in stop consonants and contributes an important innovation in his distinction between ‘basic correlates’ of any given feature. According to him, ‘basic correlates’ are those correlates which are unique to a particular feature, and which have particularly high contextual stability (that the relevant distinction is expressed by the basic correlate) and perceptual saliency. On the other hand, his ‘non-basic correlates’ have less perceptual saliency and less contextual stability. Jessen (2001) proposes that the non-basic correlates for [voice] are the same as those for [tense].

As an innovative approach to the current (standard) feature theory, a persuasive variation on the privative feature theme has emerged in Avery & Idsardi’s (2001) conception of laryngeal representation under Dimensional Theory. Avery & Idsardi (2001), focussing on laryngeal phonology, propose that the Laryngeal node and its tree dependents in figure 3.11 (a) may be modified as in figure 3.11(b).
Avery & Idsardi (2001) in figure 3.11(b) propose to distinguish laryngeal contrasts according to the three ‘dimensions’ of Glottal Width, Glottal Tension and Larynx Height as an alternative to the three privative features [spread], [constricted] and [voice] of the standard model in figure 3.11(a). In this proposal, ‘dimensions’ imply phonetically opposing ‘gestures’, which are basically the same entities as the phonological features of conventional theories. The gestures are privative and only dimensions can be contrastive in obstruents. They go beyond the previous models in acknowledging explicitly that muscle groups form antagonistic pairs, which constitute the ‘pre-terminal’ organisation of the gestures, a layer they call ‘dimension’. Avery & Idsardi (2001: p 44) put it as:

"In our model the dimensions organise antagonistic pairs. The reciprocal antagonism of the muscles is modelled by restricting the dimensions to be non-branching. Thus, only a single muscle within a dimension can be active in any given speech sound...the gestures, which are the ultimate constituents of the phonological representation, do not bear the major contrastive burden and are in many ways unlike the features in traditional feature theory. They cannot be binary, as the muscle is either activated or inhibited by the activation of its antagonistic partner. For the most part, the phonology begins at one level removed from the gesture – the
dimension, and this is the level that is largely responsible for contrast."

The terminal elements (under ‘Gestures’, in square brackets) are to be interpreted as motor instructions to the articulators. In the new proposal the dependents of Glottal Width [spread] and [constricted] make different predictions about the nature of contrasts; it does not allow contrasts as the earlier standard theory. The dependents of Glottal Tension [stiff] and [slack], being mutually opposing, are not present in the same representation. Avery & Idsardi (2001) rather allow more latitude in the phonetic implementation of Glottal Width. And the dependents of Larynx Height are [raised] and [lowered] (they skip discussion on Larynx Height in their paper).

3.10.2 Phonological representation of Dzongkha stops

The acoustic cues of Dzongkha stops we explored in this chapter were: VOT, closure duration, stop burst and H1-H2. The interaction of phonetic correlates in the distinction of the Dzongkha VLT*, VLL, VLAS and VD stops may be presented as in table 3.12.

Table 3.12: Interaction of acoustic correlates for Dzongkha stops

<table>
<thead>
<tr>
<th>VOT (Asp)</th>
<th>short</th>
<th>short</th>
<th>long</th>
<th>negative</th>
</tr>
</thead>
<tbody>
<tr>
<td>Closure Duration&lt;sup&gt;54&lt;/sup&gt;</td>
<td>long</td>
<td>short</td>
<td>ambiguous</td>
<td>short</td>
</tr>
<tr>
<td>Stop Burst</td>
<td>high</td>
<td>low</td>
<td>ambiguous</td>
<td>ambiguous</td>
</tr>
<tr>
<td>H1-H2</td>
<td>modal</td>
<td>breathy</td>
<td>modal</td>
<td>breathy</td>
</tr>
</tbody>
</table>

<sup>54</sup> There are two opposite ways in which the duration of the stop closure can act as a correlate of [tense]. Firstly, closure duration can be a durational correlate of [tense] that is independent of aspiration as in word-medially in English (Lisker 1986) or German (Jessen 1998). Secondly, closure duration can sacrifice its own potential as a durational correlate of [tense] for the enhancement of aspiration duration (Jessen 2001); aspiration can be lengthened by shortening closure duration has been proposed in association with place-of-articulation differences in closure and aspiration (Maddieson 1997, Kohler 1984). It is then fair to assume that closure duration in the VLAS and VD stop types in Dzongkha may not be ambiguous if aspiration is taken as a dependent feature (which we did not in our acoustic study) of the VLAS stop type.
Based on our production and perception studies, the acoustic correlate for the VD stop is rather a straightforward one. The negative VOT (with voicing throughout the closure duration) can be proposed as the primary correlate; it employs H1-H2 (breathiness [or slackness] in vowel) as a secondary correlate. This is supported by our results of perception study; despite a raise of F0 by 50 Hz, there was 100% perception of the stop type. The fact that F0 (significantly heightened) leads to a 100% perception result motivates us to propose VOT as the primary correlate.

In the three voiceless (VLT*, VLL and VLAS) stops, the VLAS type is, again a straightforward one; it employs positive VOT (aspiration) as the primary correlate, and perhaps H1-H2 as the secondary correlate. This proposal again is supported by both production and perception study. In our production study, the VLAS stop has a significantly longer duration of VOT compared to the other two voiceless stops (VLT* and VLL). And the perception study again shows a 100% perception result despite a significant lowering (50 Hz) of the pitch. This fact inspires us to assume that the VLAS stop too employs VOT (which is positive) as the primary correlate.

The VLT* and the VLL types are rather tricky in employing phonetic correlates, in that we are not sure which, from the two correlates – closure duration and stop burst – the stop types employ for their contrast. VOT is certainly not one of the correlates, as shown by both production and perception studies. Production study showed that the VOT difference in the two stop types was insignificant, further supported by the perception study. However, the two stop types differ significantly in their values in stop burst and closure duration. We thus assume that the two stop types employ closure duration and stop burst either in combination or complementarily for their contrast.

Next, we also need to translate acoustic correlates that would group the VLT* and the VLAS types as ‘Tense’ and the VLL and the VD types as ‘Lax’. One of the primary correlates of ‘Tense’ and ‘Lax’ distinction in stop types is the H1-H2 cue. We have shown (in section 3.7.1) that the H1-H2 values for the VLT* and the VLAS types are significantly lower than the VD and the VLL types; thus the vowel following the former group is a modal (or clear) one and
the latter a breathy (or lax) one, which is a carryover of the initial stop types. Phonetically, the difference between the production of the ‘Tense’ and the ‘Lax’ stop types is that the former group is produced with the impediment of the vocal folds, which is directly related to the Closed Quotient, where an increased muscular tension is created, due to strong adducted tension and medial compression of the vocal folds leading to higher subglottal pressure. On the other hand, the latter group is produced with vocal folds that are fairly abducted (relative to the former group) and have lesser longitudinal tension (Ladefoged 1971, Laver 1980, Ni Chasaide & Gobl 1995); this results in some turbulent airflow through the glottis and the auditory impression of ‘voice mixed in with breath’ (Catford 1977). These features in the vowel, then, are attributes of the initial stop types.

The second correlate - closure duration - though ambiguous in the VLAS type cannot be overruled as a correlate of the Tense-Lax distinction. Languages seem to have a trade off relation between closure duration and aspiration duration in associating the voiceless aspirated stops as [tense]. According to Jessen (2001), aspiration can be lengthened by shortening closure duration to mark the stop type as [tense] (also see footnote 23). Jakobson, Fant & Halle (1952) too remark that if a language uses both [voice] and [tense], [tense] is implemented with aspiration. From the perceptual perspective too, VOT (aspiration) seems to be more important than other durational correlates of [tense]; Silverman (1997) notes that the auditory system is particularly sensitive to events that follow (rather than presence or coincide with) the silence or low amplitude associated with stop closure, and this makes stop release a very salient perceptual event. In Danish, for instance, Hutters (1985) reports that while aspiration is particularly long compared to other languages, closure duration is quite short, to the effect that closure in /p, t, k/ (which are ‘Tense’) is systematically shorter than in /b,d,g/ (which are ‘Lax’). One would, then, land up in problem if in a language like Danish, the closure duration is interpreted as a constant attribute of ‘Tense’ stops. Closure duration as a correlate of [tense] in Dzongkha may be analogous to Danish, in that the short closure duration in the VLAS types may be compensated by the aspiration duration (if aspiration duration is considered as a dependent feature of the
VLAS stop type). It is then fair to propose that in Dzongkha closure duration for the VLAS type may not be ambiguous; the short closure duration in the VLAS type may be a compensatory feature for the aspiration duration. Thus, in Dzongkha, the ‘Tense’ and ‘Lax’ stops are distinguished by two phonetic correlates: H1-H2 (breathiness [or laxness] in the vowel) and closure duration.

Based on the interaction of the phonetic correlates above, we propose phonological contrasts, following Avery & Idsardi (2001), for the four stop types in Dzongkha. According to Avery & Idsardi (2001), the phonological contrasts are always of the type Ø/Marked, so that in any two-term system there is an element that has no marking and an element that has some specification. This means that for any contrasting pair, at the phonological level they differ only by the presence versus the absence of a single node, which makes the ‘Dimensions’ as the primary interface between the phonology and the phonetics. This system gives the prediction that the basic two-way systems will involve a Ø/GT or a Ø/GW contrast, and the dimensions can be combined (as for example, in Hindi as Ø/GT/GW/GT+GW for the four stop types, voiceless, voiceless aspirated, voiced and voiced aspirated).

Thus, we propose the Dzongkha phonological contrast system as:

\[ \text{Ø/GT/GW/GT+GW} \]

Using this phonological system, the VLT* can be classified with gestural completion\(^{55}\) [stiff] using the Glottal Tension (GT) dimension. We combine Glottal Width (GW) and Glottal Tension (GT) (recall that the dimensions can be combined) for the VLAS stops with gestural completion [stiff] and [spread] respectively; the VD stop with Glottal Width (GW) dimension with gestural completion.

\(^{55}\) Avery and Idsardi’s (2001) concept of ‘completion’ is different from ‘enhancement’ in standard theory, in that ‘completion’ merely involves the addition of gestural information to the already present dimensions, while ‘enhancement’ involves the addition of a dimension node.
completion [slack] and VLL is Ø (unmarked)\textsuperscript{56}. This can be presented as follows:

<p>| | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>VLL</td>
<td>/t/</td>
<td>Ø (unmarked)</td>
<td>(Default)</td>
<td></td>
</tr>
<tr>
<td>VLT*</td>
<td>/t*</td>
<td>(GT stop)</td>
<td>[stiff]</td>
<td></td>
</tr>
<tr>
<td>VLAS</td>
<td>/th/</td>
<td>(GW+GT stop)</td>
<td>[spread] [stiff]</td>
<td></td>
</tr>
<tr>
<td>VD</td>
<td>/d/</td>
<td>(GW stop)</td>
<td>[slack]</td>
<td></td>
</tr>
</tbody>
</table>

In Dzongkha, we analyse the Ø (unmarked/default) stop as the VLL\textsuperscript{57}, GT stops are VLT*, VLAS stops have a combined GW+GT dimension, and the GW stops are VD.

In the following section(s), we study the fricative phonation contrast in Dzongkha.

### 3.11 Fricative phonation in Dzongkha

In this section, we first have a quick overview of the distribution of fricatives in Dzongkha (and their historical derivations) and other languages, paying specific focus on the ‘six-fricative languages’ cited in Maddieson (1984). We then have a brief survey of phonation (two-way voiceless) system in Korean because, as mentioned earlier, Korean too makes a two-way voiceless phonation distinction in fricatives. This is followed by an acoustic examination of potential phonetic cues for fricatives in Dzongkha; we examine fricative duration and H1-H2 (along with other corroborative) cues to make a contrast.

\textsuperscript{56} Considering the fact that in most languages with a two-way laryngeal contrast (either plain vs. voiced or a plain vs. aspirated distinction) is found, Avery & Idsardi (2001) propose [slack] as a default completion for GT and [spread] for GW for the Ø (Unmarked) dimension.

\textsuperscript{57} Considering the fact that in most languages with a two-way laryngeal contrast (either plain vs. voiced or a plain vs. aspirated distinction) is found, Avery & Idsardi (2001) propose [slack] as a default completion for GT and [spread] for GW for the Ø (Unmarked) dimension.
between the three fricatives: two voiceless (VLT* and VLL) and a voiced (VD) one, and see the parallels between the Dzongkha and Korean phonation system in fricatives. And we conclude this section on 'fricative phonation' by extending our [tense] feature of stops to fricatives in order to make a three-way phonation contrast (VLT*, VLL and VD) in Dzongkha.

3.11.1 Dzongkha as a 'six-fricative' language

In chapter 2, we proposed seven fricatives in Dzongkha: /s*/, /s/, /z/, /ç*/, /ç/, /ç/ and /h/ (see chapter 2 for phonemic contrasts and places of articulation of fricatives). However, we do not include /h/ as a typical fricative in our analysis here. This is because of the fact that the voiceless glottal fricative /h/ does not fit appropriately by the traditional definition of a fricative sound. Ladefoged (1971) defines fricative sounds as those sounds produced by the narrow approximation of two articulators so as to produce a turbulent airstream. We may note here that the sound /h/ does not fit into the traditional definition of the fricatives because there is no oral constriction in the articulation of the sound, except the glottis. Maddieson (1984), following Pike (1943) and others, further remarks that "... /h/ is normally a voiceless counterpart of an abutting voiced segment (most often a vowel)". Furthermore, as we have shown in our segmental inventory, in chapter 2, that the phonatory behaviour of /h/ is different from the regular fricatives. We do not include /h/ as a typical fricative in our acoustic study here. This leaves us with the classification of Dzongkha as a 'six-fricative' language. According to Maddieson (1984) in the UPSID survey of languages, 'six-fricative' languages contribute to only 8.8% (28 languages) in his survey. The typical structure of fricatives in a 'six-fricative' language has three pairs of fricatives contrasting in voicing, viz. /f/, /v/, /s/, /z/, /ʃ/ and /ʒ/. Maddieson's (1984) survey includes only Korean as the two-way voiceless fricative language, but there is no mention of languages with a three-way phonation distinction as found in Dzongkha. However, the number of fricatives in languages can vary from zero (as in Hawaiian (Hockett 1955)) to
22 (as in Kabardian (Maddieson 1984)). However, an estimate of the average number of fricatives given by Maddieson (1984) is four fricatives.

Mattisof (2003) reconstructs Proto-Tibeto-Burman (PTB) fricatives as *s and *z for dental place, and *ʃ and *ʒ for palatal place. No labio-dental (/f/ or /v/) fricatives are reconstructed for PTB, though many daughter languages have /v/ (for example, Lahu). He, however, states that the contrast between dental and palatal sibilants (including, affricates) is shaky or non-existent in many TB languages (including Burmese and Lahu). Following Mattisof's (2003) reconstruction for PTB fricatives, the VLL dental and palatal fricatives in Dzongkha may be posited as a historical derivation of PTB *z and *ʒ, following the same process of tonogenesis as the VLL stops (see chapter 5 for a detailed discussion).

### 3.11.2 Fricative phonation system in Korean

As stated earlier, Korean has two denti-alveolar fricatives. As in the stops, there is no voicing contrast, but there is a two-way contrast between /s/*/ and /s/: /s/*/ has been called a tense (or fortis) fricative and /s/ as lax (or lenis). The contrast between the two fricatives is shown in the minimal pairs (Ladefoged et al. 2002):

- **Tense** /s*ata/ ‘to wrap’
- **Lax** /sata/ ‘to buy’

Unlike the stops, the phonetic properties of the fricatives have not been well studied, and the categorization of the lax /s/ has been controversial. It is sometimes categorized as Lax and sometimes as aspirated (Kagaya 1974, Iverson 1983, Jun 1993, Park 1999, Yoon 2002).

In Korean orthography, according to Ladefoged et al. (2002), the /s/ is regarded as lax and behaves that way in phonological processes. But its behaviour in phonetic processes and its phonetic realizations are generally believed to be similar to stops in the aspirated category (/s/ becomes tense
when followed by a lenis stop). On the other hand, it is also similar to the aspirated stop, in that it is not likely to become voiced between voiced segments, and it generally triggers a high tone in the beginning of an Accentual Phrase (Jun 1993). Fiberscopic data in Kagaya (1974) and Jun et al. (1998) show that /s/ has a glottal opening configuration similar to aspirated stops, and Jun et al. (1998) further show that /s/ has a larger glottal opening than /s*/. Based on acoustic data, Yoon (1998) and Park (1999) also show that /s/ has some degree of aspiration as in the aspirated stops, and Park (1999) claimed, based on the energy difference between the first and second harmonics at vowel onset, that the lax should be categorized in the aspirated category because the vowel onset after it is breathier than after the tense one. Furthermore, an Electropalatography (EPG) study by Kim (2001) shows that /s/ has a significantly less linguopalatal contact with a wider frication channel than /s*/. As in stop consonants, Cho et al. (2000) find systematic differences in H1-H2 values of the onset of vowels that follow /s/ and /s*/. H1-H2 has positive values for /s/, which indicates the breathiness of vowel, and negative values for /s*/. As in stop consonants, Cho et al. (2000) find systematic differences in H1-H2 values of the onset of vowels that follow /s/ and /s*/. H1-H2 has positive values for /s/, which indicates the breathiness of vowel, and negative values for /s*/, which indicates the modal characteristic of vowel.

The two fricatives also differ in their duration both in word-initial and medial positions. Many linguists, for example, Yoon (1998), Cho et al. (2000), Kim (2001) find that the duration of /s*/ is longer than /s/. They further observe that the duration difference is larger in word-medial position than in word-initial position.

However, our aim here is not to re-examine the facts of Korean fricatives, but to review the facts of the Korean fricative phonation system so that it helps us investigate the phonetic characteristics of the Dzongkha fricative phonation system. We compare the fricative phonation system in Korean and Dzongkha towards the end of this section. Before we compare the Dzongkha and Korean fricative phonation systems we explore the acoustic properties, fricative duration and H1-H2, of fricatives in Dzongkha using the same methods we had used for stops.
3.11.3 Fricative duration

In studying the temporal cues of the fricatives, we measure the duration of the three fricatives in the initial position; the first three tokens were considered for our study. The words were:

VLT* /s*ʃ:/ ‘thumb measurement’

VLL /ʃoː/ ‘sickle’

VD /zɔː/ ‘carpenter’

The results of the fricative duration are presented in table 3.13 and figure 3.12.

Table 3.13: Dzongkha fricative duration

<table>
<thead>
<tr>
<th></th>
<th>VLT*</th>
<th>VLL</th>
<th>VD</th>
</tr>
</thead>
<tbody>
<tr>
<td>F1</td>
<td>207</td>
<td>170.25</td>
<td>165</td>
</tr>
<tr>
<td>F2</td>
<td>184</td>
<td>141</td>
<td>141</td>
</tr>
<tr>
<td>M1</td>
<td>169.75</td>
<td>140.25</td>
<td>110</td>
</tr>
<tr>
<td>M2</td>
<td>149.75</td>
<td>130.75</td>
<td>100</td>
</tr>
</tbody>
</table>

Mean 177.63 145.56 129

Figure 3.12: Histogram of fricative duration difference
From our results in table 3.13 and figure 3.12, we see a clear correlation between the VLT* and the VLL fricatives; the VLT* fricative is significantly longer (on an average by 32 msec) than the VLL fricative. However, we cannot draw any such correlation between the VLL and the VD fricatives, though the results of the two male speakers (M1 and M2) seems to suggest some correlation (a difference of 30 msec in both the speakers), the two female speakers (more or less) have the same duration for the two fricatives, as shown in figure 3.12. The overall mean difference between the three fricatives is: VLT* - 178 msec, VLL - 146 msec and the VD - 129 msec. Though, overall duration seems to suggest a three-way correlation, the VD fricative is distinguished from the other two voiceless fricatives by the presence of voicing throughout the fricative duration. Thus, the VLT* fricative is longer in duration than the VLL and VD fricatives. The spectrograms and waveforms of the first tokens of each fricative of M1 are presented in figure 3.13.

Figure 3.13: Spectrogram and waveform of fricatives showing durational differences

\( \text{a) VLT*} \)
In figure 3.13, we show the durational differences of the three fricative types; the VLT* is 172 msec, the VLL is 156 msec and the VD is 131 msec. Besides the durational differences, we also see that the VD fricative has voicing (indicated by the voice band in the spectrogram and the periodic waveform) throughout the fricative duration as shown in figure (c). On closer look at both the waveform and spectrogram of the VLT* and VLL fricatives, we also notice slight acoustic differences in the two fricative types. Both the waveform and the spectrogram suggest that the VLT* fricative is produced with higher muscular force than the VLL fricative; this is indicated by the thicker energy bands (in the spectrogram) and the higher amplitude in the waveform (especially in the onset of the fricative) in the VLT* type.
Thus, while the VD fricative is distinguished by the presence of voicing, the two voiceless fricatives are distinguished by their duration corroborated by tenseness and laxness of the muscular activities.

In the next section, we study the H1-H2 (breathiness in the following vowel) of the three types of fricatives.

3.11.4 H1-H2 study

We use the same method of studying H1-H2 in fricatives as we have used for the stops; the measurements of the same four speakers were taken from the middle of the vowel (for five wave cycles of one token) following the fricatives. The results are presented in figure 3.14.

Figure 3.14: H1-H2 of the vowel following fricative types
In figure 3.14, we show that the H1-H2 in the vowel following fricatives are same as the stop counterparts, in that the H1-H2 values of the five wave cycles in the middle of the vowel of the VLT* fricative is significantly
higher than the VLL and VD types. There is a minimum difference of 5 dB and a maximum of 9 dB. Based on these results, we assume that the vowel following the VLT* fricative has a modal phonation, while the vowel following the VLL and VD fricatives have breathy phonation as their stop counterparts. The H1-H2 differences of Female 1 are given in figure 3.15 to illustrate the difference.

Figure 3.15: Waveform showing the H1-H2 differences in the vowel following the fricatives

![Waveform showing the H1-H2 differences in the vowel following the fricatives](image)
In figure 3.15, we show the waveforms continuum of the breathiness; the vowel is breathiest after the VLL fricative and the least breathy in the VLT* fricative, with the VD fricative (somewhat) at an intermediate degree. This is evidenced by the fact that the waveform in the VLT* is a smooth one, while the waveforms in the VLL and the VD fricatives are jagged, as in their stop counterparts. As seen in the figures, the H1-H2 difference for the VLT* fricative is 7 dB in (a), the VD fricative is 15 dB in (b) and the VLL fricative is 16 dB in (c).

Thus, our H1-H2 study clearly shows that the vowel following the VLT* type is modal phonation, while the vowel following the VD and VLL types are breathy phonation. This is, again as in the stops, a property of the initial fricatives and thus phonetic; phonologically, Dzongkha does not contrast modal and breathy vowels. The physiological difference between the production of the VLT* and the VD/VLL fricative types is that the former is produced with the impediment of the vocal folds, which is directly related to the Closed Quotient, where an increased muscular Tension is created, due to strong adductive tension and medial compression of the vocal folds leading to higher subglottal pressure. While the latter group is produced with vocal folds that are fairly abducted (relative to the former group) and have lesser longitudinal tension (Ladefoged 1971, Laver 1980, Ni Chasaide & Gobi 1995), which results in some turbulent airflow through the glottis. Based on the acoustic evidence of the H1-H2 results we may now, extending our analysis of
stop counterparts, term the two groups as **Tense** (VLT) and **Lax** (VD and VLL). The Tense type is followed by Modal vowel phonation and the Lax types with breathy vowel, and these characteristics in the vowel are the coarticulation effect of the initial fricative types.

In the next section, we compare the two voiceless fricative phonation systems in Dzongkha and Korean.

### 3.11.5 Comparison of the voiceless fricative phonation system in Dzongkha and Korean

In this section, we compare the two voiceless fricative phonation systems in Dzongkha and Korean. We compare in juxtaposition the three acoustic cues: fricative duration, H1-H2 and (presence or absence of) aspiration, which is presented in a tabular form in table 3.14

<table>
<thead>
<tr>
<th></th>
<th><strong>Dzongkha</strong></th>
<th><strong>Korean</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Tense</strong></td>
<td>long</td>
<td>short</td>
</tr>
<tr>
<td><strong>Lax</strong></td>
<td>short</td>
<td>long</td>
</tr>
<tr>
<td><strong>Tense</strong></td>
<td>long</td>
<td>short</td>
</tr>
<tr>
<td><strong>Lax</strong></td>
<td>long</td>
<td>short</td>
</tr>
<tr>
<td><strong>H1-H2</strong></td>
<td>low</td>
<td>high</td>
</tr>
<tr>
<td><strong>Aspiration</strong></td>
<td>no</td>
<td>no</td>
</tr>
<tr>
<td></td>
<td>no</td>
<td>yes</td>
</tr>
</tbody>
</table>

Table 3.14 shows striking similarities in the way the two voiceless fricatives behave in terms of phonatory characteristics. In these languages, the duration of /s*/ is longer than its counterpart /s/, indicating greater degree of muscular force in the production of /s*/ than in /s/. H1-H2 has higher values for /s*, which indicates the breathiness of vowel, and lower values for /s*/. which indicates the modal characteristic of vowel, in both the languages. However, the two languages differ in one way, in that a number of studies had shown that the /s/ has phonetic aspiration in Korean (Kagaya 1974, Cho et al. 2000, Kim 2001, Ladefoged et al. 2002, Wright 2007, among others), while, as
we can see in figure 3.13 (b), the /s/ in Dzongkha does not have any characteristic of aspiration. However, this should not let us believe that the two fricatives are very different in the two languages. We may remind ourselves that if a language uses both [voice] and [tense] (as Dzongkha does), [tense] is implemented with aspiration (Jakobson, Fant & Halle 1952). Furthermore, Kim (2001) shows that /s/ in Korean has a significantly less linguopalatal contact than /s*/. which is a result similar to the difference in the amount of linguopalatal contact between the lax and the tense stops in Korean.

Phonologically too, the two voiceless lax fricatives behave similarly in both the languages. In Dzongkha, /s/ becomes voiced in the medial position (/sam/ 'bridge' vs. /cazam/ 'iron bridge') like the VLL stops do. In Korean too, according to Ladefoged et al. (2002) 47% of the tokens of their speakers' /s/ were fully voiced in the medial position and about another 40% of the tokens were voiced over more than half the frication period. Since /s/ in Dzongkha is phonetically an unaspirated one, we propose that /s/ in Dzongkha is a lax fricative and not aspirated, which will implicate the inertness of the feature [spread]. In Korean, the /s/ can still be analysed as a lax fricative, as in the stops, in opposition to /s*/.

In the following section, we extend our phonological feature analysis of stops to fricatives.

3.11.6 Phonological representation of Dzongkha fricatives

From our acoustic study of Dzongkha fricatives so far, we have found three phonetic cues, fricative duration, H1-H2 and Voice, that distinguish the three fricatives (VL{T*}, VLL and VD). The phonetic nature of the VD fricative is rather straightforward; it is produced with voicing throughout the fricative duration. The two voiceless fricatives (/s*/ and /s/) employ frication duration and H1-H2 for their contrasts; /s*/ has longer duration and lower H1-H2, while /s/ has shorter duration and higher H1-H2.
Based on the above phonetic assumptions, we extend our feature system for the stops (minus the feature for VLAS stops) to fricatives and propose the features for the fricatives three (VLT*, VLL and VD) in Dzongkha as follows:

$$\emptyset/GT$$

Using this feature system, the VLT* can be classified as [stiff] with GT dimension, VD fricative as [slack] GT fricative, and VLL as $\emptyset$ (unmarked). This can be presented as follows:

<p>| | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>VLL</td>
<td>/s/</td>
<td>$\emptyset$ (unmarked)</td>
<td>(Default)</td>
</tr>
<tr>
<td>VLT*</td>
<td>/s*</td>
<td>Glottal Tension (GT)</td>
<td>[stiff]</td>
</tr>
<tr>
<td>VD</td>
<td>/z/</td>
<td>Glottal Tension (GT)</td>
<td>[slack]</td>
</tr>
</tbody>
</table>

Following the revised model of Avery and Idarsi (2001), we propose the feature for the VLL as unmarked $\emptyset$, for the VLT* we use GT dimension with completion [stiff] and for the VD fricatives we use GT with the completion [slack].

### 3.12 Conclusion

In this chapter, we began by recognising a segmental problem of Dzongkha obstruents in terms of their segmental classification. We pointed out that the four-way phonation types in stops and three-way in fricatives have been inadequately analysed in Dzongkha as: voiceless unaspirated, voiceless aspirated, voiced and devoiced. We rather propose that the three (phonetically) voiceless stops and two fricatives be classified as VLT*, VLL and VLAS for stops, and VLT* and VLL for fricatives, along with their VD counterparts.
In section 3.2 through 3.2.2, we saw the major phonation types languages employ to contrast their consonants, and opened our discussion of tense phonation type in Korean as the only language so far believed to contrast this type of phonation. In the following two sections (3.3 and 3.3.1), we reviewed Dzongkha obstruent phonation in the literature and identified the problem of, the so called 'devoiced' type in the light of distinctive feature theory. This was followed by a general survey of phonation typology in other Tibeto-Burman languages in section 3.3.2; we saw that besides Dzongkha a few other Tibeto-Burman languages have a four-way phonation contrast. But we conclude that the four-way phonation system in other Tibeto-Burman languages, in fact, differ from Dzongkha, in that other Tibeto-Burman languages have four-way contrast (a borrowed feature from Indo-Aryan languages) based on voicing and aspiration as in Hindi and other Indo-Aryan languages, while in Dzongkha the contrast is a three-way voiceless (VLT*, VLL and VLAS) and a voiced (VD) one. Thus, we conclude that the Dzongkha four-way stop phonation system is unique not only in Tibeto-Burman languages, but cross-linguistically, similar to Korean.

In section 3.4, we embarked upon our production study on stop phonation types and presented the findings on Dzongkha stop VOT in section 3.4.2 with mean values and spectrographic support for the difference in VOT. In our production study, the VOT of the VLAS and VD stops are significantly different, while the VLT* and the VLL stop types were found to lack such a difference; this was further supported by perception study. In section 3.5, we studied stop burst in Dzongkha and found that the VLT* and VLL stops have a correlation in stop burst, in that the VLT* has a higher burst energy than its counterpart, but we could not assume any such correlation in the VLAS and VD stops. In the following section (3.6), closure duration of the stops were studied, and found the results similar to stop burst. We conclude that closure duration, like stop burst, has a correlation in only the VLT* and VLL stops, in that VLT* stop has longer closure duration than the VLL stop. There was no systematic correlation in the VLAS and the VD stops. In section 3.7 and 3.7.1, our H1-H2 study showed that the VLT* and the VLAS types were followed by modal vowel phonation, while the VLL and VD types were followed by
breathy vowel phonation, thus dividing the four stop types into two major groups: Tense (VLT* and VLAS) and Lax (VLL and VD) group.

In section 3.8, we summarized the stop phonation system of Hindi, and showed that the four stops in Hindi are distinguished by voicing and aspiration. The fourth type (VLL) in Dzongkha was compared with Hindi breathy type and found that the two stops are significantly different.

In section 3.9, we summarized the three-way voiceless stop phonation system in Korean, and compared the tense-lax stops with Dzongkha VLT*-VLL types; we found that the two voiceless stops in both the languages employ the same phonetic cues for their contrasts.

In section 3.10, after a short review of the laryngeal features discussed in the current literature, we follow Avery and Idsardi's (2001) modified feature theory and proposed the phonological feature, based on three dimensions: glottal width, glottal tension and glottal width+glottal tension, of four stops in Dzongkha with GT [stiff] correlating with ‘Tense’ articulation and GT [slack] with ‘Lax’ articulation:

- VLL /t/ $\emptyset$
- VLT* /t*/ GT (with completion [stiff] gestural specification)
- VLAS /th/ GW+GT (with completion [spread] [stiff] gestural specification)
- VD /d/ GT (with completion [slack] gestural specifications)

The earlier ‘voiceless’ is now ‘voiceless tense’ (VLT*), ‘voiceless aspirated’ and the ‘voiced’ remain unchanged, and the ‘devoiced’ is ‘voiceless lax’ (VLL) in our proposal.

In section 3.11, we studied the fricative phonation system in Dzongkha and extended our stop feature analysis to fricatives. In section 3.11.1 we did a short literature review of the fricatives in languages and placed Dzongkha into the six-fricative languages. In section 3.11.2 through 3.11.5, we studied the
acoustic cues in Dzongkha fricatives. In section 3.11.2, we summarized the fricative phonation system in Korean. In section 3.11.3 and 3.11.4, we studied fricative duration and H1-H2 in Dzongkha and found the results similar to Korean, except that the Korean lax fricative type has phonetic aspiration while the Dzongkha counterpart does not have aspiration.

Finally, in section 3.11.6, we extended our feature classification of stops to fricatives and proposed features for Dzongkha fricatives, based on the dimensions Glottal Tension + Glottal Width, as:

\[
\begin{align*}
\text{VLL } /s/ & \quad \emptyset \\
\text{VLT*}/s*/ & \quad \text{GT (with completion of [stiff] gestural specifications)} \\
\text{VD } /z/ & \quad \text{GT (with completion of [slack] gestural specification)}
\end{align*}
\]

We conclude that, in the absence of any phonetic aspiration in any of the fricatives, the feature [spread] is underspecified in Dzongkha fricatives.
CHAPTER FOUR
SONORANT PHONATION TYPES

4.1 Introduction

One of the key features of Dzongkha phonology is the distinctive voiced - voiceless contrast in its sonorants; nasals across all four places (bilabial, dental, palatal and velar), and in its lateral as well (see table 4.1 for examples of contrasts). Cross-linguistically, using contrastive phonation in sonorant consonants such as the ones found in Dzongkha is rare, but relatively common in Tibeto-Burman languages. The presence and use of contrastive phonation in languages of this region have been well documented in the phonemic inventories (see section 4.4 for examples). However, acoustic studies in the sonorant phonation types to provide a comprehensive description of the distinction between voiced and voiceless sonorants in Tibeto-Burman languages is very limited, and absolutely nil in Dzongkha. The goal of this chapter, therefore, is to examine the acoustic correlates of voiced and voiceless sonorants (nasals and laterals) in comparison with two other Tibeto-Burman languages, Tenyidie54 and Mizo, thus enabling a firmer understanding of the nature of the phonetic properties of the two sonorant phonation types in Dzongkha for their phonemic contrasts.

This chapter has three goals: 1) to study the characteristics of voiceless nasals and laterals in Dzongkha and compare them with Mizo and Tenyidie, 2) to determine if voiceless nasals and laterals are /Ch/ (or /hC/) single voiceless segments or /Ch/ (or /hC/) clusters, and 3) finally, to analyse the segments phonologically in terms of distinctive feature theory.

The organisation of this chapter is as follows. In section 4.2, we present a brief language background of Mizo and Tenyidie. This is followed by examples of sonorant voiced-voiceless contrasts in section 4.2.1 in Dzongkha.

54Ladefoged & Bhaskararao (1993) refers to the language as ‘Angami’; we choose to call it ‘Tenyidie’ as our speaker for this study tells us that ‘Angami’ refers to a tribe name and ‘Tenyidie’ to the name of the language different tribes speak, including the Angami tribe.
In section 4.3 through 4.3.2, we review a typological distribution of nasals and laterals (emphasising on the voiceless types) in the UPSID database, and show that the voiceless types of sonorants are relatively marked than the voiced ones. This is followed by another typological survey of the distribution of the voiceless-voiceless sonorant types in twenty Tibeto-Burman languages in section 4.4, and show that the same markedness typology of the voiceless types of sonorants is true in Tibeto-Burman languages as well. Section 4.5 is a brief literature review of nasal and lateral phonation types in languages, including the breathy voiced types in Sumi and some other Indic languages. We then acknowledge the fact that Sumi (despite being a Tibeto-Burman language) has the breathy voiced nasals, as in the Indic languages.

In section 4.6 through 4.6.2, we embark upon our acoustic study of nasal duration and show that the duration of voiceless nasals in all the three languages is longer than the voiced ones. Besides examining the overall duration of the nasals, we also examine different phonetic composites (voiceless composite [η], voiced composite [n] and aspiration [h]) of the voiceless nasals, and show that Dzongkha and Mizo nasals may be classified as pre-aspirated and Tenyidie as post-aspirated.

In section 4.6 through 4.6.6, we study the acoustic characteristics of the laterals in the three languages. In section 4.6.1 through 4.6.3, we show that the duration of the voiceless laterals in all the three languages is longer than the voiced laterals. We also show that, unlike the voiceless nasal, the voiceless lateral in Dzongkha does not have the phonetic voiced composite [l], patterning with the Tenyidie voiceless nasal and lateral as a post-aspirated type. In Mizo in section 4.7.2, we show that the phonetic characteristics of the voiceless lateral patterns with the voiceless nasal, in that the /l/ in Mizo has a phonetic voiced composite [l].

In section 4.7.3, we compare the acoustic characteristics of sonorants in the three languages. We compare the duration of the sonorants and show that the duration of the voiceless sonorants is longer than their voiced counterparts across all three languages. We also compare the different phonetic composites
of the voiceless sonorants and show that while the characteristics of the voiceless nasal in Dzongkha pattern with Mizo, the voiceless lateral patterns with Tenyidie.

And finally, in section 4.8 through section 4.8.3, we represent, following Avery & Iodsardi's (2001) model, sonorants in Dzongkha in terms of distinctive feature theory for their phonemic contrasts. In section 4.8.1, we show, once more, that typologically the voiceless sonorants are marked types in languages, including Tibeto-Burman, which motivates us to posit the voiced types as the unmarked segments. In section 4.8.2, we present a brief review of the theoretical preliminaries of voiceless sonorants and present the problems of representing the voiceless sonorants using standard model feature theory. And finally, in section 4.8.3, we represent the sonorants in Dzongkha, following Avery and Iodsardi (2001).

The next section presents a short language background of Mizo and Tenyidie.

4.2 Tenyidie and Mizo: Language background

Tenyidie is a Tibeto-Burman tonal language, belonging to the Naga group of languages, spoken in the Naga Hills of Nagaland, in the north-eastern part of India. According to Ladefoged & Bhaskararao (1993), there are several dialects of Tenyidie, the most prominent being Chokri, Khonoma and Kohima. The Kohima variety is considered the standard variety for publications and is taught in schools. The dialects of Tenyidie are mutually intelligible but differ in tonal and segmental inventory. Ladefoged & Bhaskararao (1993) estimate the number of speaker for Khonoma dialect at 4,000 people in the extreme west of the Angami region. In our study, we refer to the language as ‘Tenyidie' as suggested by our language informant for this study. Ladefoged & Bhaskararao (1993) report four tones, though Burling (1960) and Ravindran (1974) (reported in Ladefoged & Bhaskararao (1993)) claim that there are five.
Mizo, also a tonal language, belongs to the Kuki-Chin group of Tibeto-Burman family. Mizo speakers are scattered over a vast area comprising of Northern Mizoram, South district of Manipur and parts of North Cachar Hills and Cachar districts in Assam. Like Tenyidie, Mizo too has a number of dialectal variations. Different sub-tribes and clans speak different varieties. Of these, the more prominent dialects are Khosak, Thiek, Hrangkhól, Bieté, Fāihriem, and Sakachap. The differences among them however, are marginal. Among these dialects, Khosak dialect, literally, the dialect of eastern group has been adopted as the common and standard language and this has been reduced to a written language during the thirties. Chhangte (1986) and Fanai (1992) report four tones, which are confirmed by an acoustic study by Sarma & Wiltshire (2010).

4.2.1 Phonological contrast of sonorants in Dzongkha, Mizo and Tenyidie

In Mizo and Tenyidie, the phonological information about nasals and laterals (sonorants in general) is scarce. We, however, could elicit a list of minimal pairs that contrast the sonorants (nasals and laterals) from our informants. Table 4.1 shows the minimal pairs of the three languages under study. We do not have the tonal details of the words in Mizo and Tenyidie, thus the vowels are left unmarked for tones. In Dzongkha for now, as mentioned in chapter 2, we assume that the voiced nasals and laterals occur in both the registers, while voiceless ones occur only in the high register.

Table 4.1: Phonemic contrast of nasals and laterals in Dzongkha, Mizo and Tenyidie

<table>
<thead>
<tr>
<th>Dzongkha</th>
<th>Labial</th>
<th>भा: वार ‘war’</th>
<th>भा: वो ‘low’</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dental</td>
<td>वेप ‘promise’</td>
<td>वेप ‘curse’</td>
<td></td>
</tr>
<tr>
<td>Palatal</td>
<td>वेप ‘site of pilgrimage’</td>
<td>वेप ‘injure’</td>
<td></td>
</tr>
<tr>
<td>Velar</td>
<td>वेप ‘five, drum’</td>
<td>वेप ‘early’</td>
<td></td>
</tr>
</tbody>
</table>

59 For the sake of uniformity we use uniform symbols न and फ for all the voiceless sonorants, though Ladefoged et al. (1993) use different symbols (N and l).
4.3 Cross-linguistic typology of nasals and laterals

4.3.1 Nasals

In this section, we make a brief summary of the occurrence of nasals and laterals in the UCLA Phonological Segment Inventory Database (UPSID) of Maddieson (1984). Of the 1057 nasals in the UPSID file, a great majority (934 or 88.4%) are simple plain voiced nasals, out of which, 50 nasals are plain voiced but have distinctive length or a secondary articulation. The voiceless nasals consist of only 36 or 3.4%, a number which is almost equal to the laryngealized nasals (34 or 3.2%). There are also three breathy nasals reported in Hindi-Urdu and !Xü. Esposito et al. (2005) also report of the presence of breathy nasals in Bengali and Marathi. The authors also suggest that the breathy nasals in Hindi appear to be more like segment clusters (N+h) rather than single segments (N₀) based on their electroglottographic study. Cross-linguistically, languages with contrasting breathy and modal voiced sounds are rare, but those that make such a contrast do so in stops, nasals, laterals and vowels (Ladefoged & Maddieson 1996).

Maddieson’s (1984) analysis of nasals is primarily based on the work of Ferguson, Hyman & Ohala (1975). Based on the assumptions of Ferguson et al. (1975) he restates his assumptions about the typology of nasals as follows:\footnote{We include only the assumptions that pertain to the voiced, voiceless and breathy nasals.}

- Most languages have at least one nasal (96.8%).
- If a segment is a nasal, it is voiced (93.1%).
- A voiceless nasal is more likely to have a bilabial place of articulation than any other place (30.6%).
- The presence of a voiceless, laryngealized or breathy voiced nasal implies the presence of a plain voiced nasal with the same place of articulation (100%).
- The presence of a nasal with a secondary articulation implies the presence of a simple nasal at the same place of articulation (92.3%).
In Dzongkha, as we have stated in chapter 2, the nasals make a voiced–voiceless distinction at four places (labial, dental, palatal, velar), along with the lateral. In Mizo, nasals make the same distinction at three places (labial, dental, velar), along with the lateral. And in Tenyidie, though nasals are attested at four places (labial, dental, palatal and velar) the voiced–voiceless distinction is attested only at three places (labial, dental and palatal); the velar nasal does not have the voiceless counterpart.

In the next section, we present a typological survey of the sonorant types in the world languages.

---

60 The minimal pairs are elicited from a female informant of Mizo, who is a Lushai speaker. We do not have tonal information of the words, thus we do not mark the words for their tones.
61 After Ladefoged & Bhaskarao. (1993)
Ohala (1975), reported in (Maddieson 1984), states that nasals are rarely voiceless. He further observes that voiceless nasals can barely be distinguished from each other on the basis of their noise spectrum, and that in languages which have voiceless nasals, there is a voiced portion before the vowel begins and it is that portion which makes them recognisable. However, Ladefoged & Bhaskararao (1993) report that in Khonoma dialect of Angami (Tenyidie) the voiceless nasals remain voiceless throughout the nasal articulation and even beyond the release; they observe that the voicing of the following vowel begins well after the articulatory stricture has been released.

4.3.2 Laterals

Maddieson (1984) categorises laterals under four broad headings: lateral approximants, taps/flaps, fricatives and affricates. We replicate the frequency of occurrence of different lateral types in Maddieson (1984) below.

*Approximant lateral types*

<table>
<thead>
<tr>
<th>Type</th>
<th>Number</th>
<th>% of laterals</th>
</tr>
</thead>
<tbody>
<tr>
<td>Plain voiced</td>
<td>313</td>
<td>74.7%</td>
</tr>
<tr>
<td>Plain voiceless</td>
<td>11</td>
<td>2.6%</td>
</tr>
<tr>
<td>Laryngealized voiced</td>
<td>8</td>
<td>1.9%</td>
</tr>
<tr>
<td>Breathy voiced</td>
<td>1</td>
<td>0.2%</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>333</strong></td>
<td><strong>79.7%</strong></td>
</tr>
</tbody>
</table>

*Lateral tap/flap*

<table>
<thead>
<tr>
<th>Type</th>
<th>Number</th>
<th>% of laterals</th>
</tr>
</thead>
<tbody>
<tr>
<td>Plain voiced</td>
<td>9</td>
<td>2.2%</td>
</tr>
<tr>
<td>Laryngealized voiced</td>
<td>1</td>
<td>0.2%</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>10</strong></td>
<td><strong>2.4%</strong></td>
</tr>
</tbody>
</table>

*Lateral fricative types*

<table>
<thead>
<tr>
<th>Type</th>
<th>Number</th>
<th>% of laterals</th>
</tr>
</thead>
<tbody>
<tr>
<td>Plain voiceless</td>
<td>34</td>
<td>8.1%</td>
</tr>
<tr>
<td>Plain voiced</td>
<td>9</td>
<td>2.2%</td>
</tr>
<tr>
<td>Ejective voiceless</td>
<td>2</td>
<td>0.5%</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>45</strong></td>
<td><strong>10.8%</strong></td>
</tr>
</tbody>
</table>
*Lateral affricates*

<table>
<thead>
<tr>
<th>Type</th>
<th>Frequency</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ejective voiceless</td>
<td>14</td>
<td>3.3%</td>
</tr>
<tr>
<td>Plain voiceless</td>
<td>9</td>
<td>2.2%</td>
</tr>
<tr>
<td>Plain voiced</td>
<td>4</td>
<td>1.0%</td>
</tr>
<tr>
<td>Aspirated voiceless</td>
<td>2</td>
<td>0.5%</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>29</strong></td>
<td><strong>6.9%</strong></td>
</tr>
</tbody>
</table>

In the above list, we notice that languages prefer approximant laterals to other types of laterals; 333 languages (79.7%) have approximant laterals against 10 languages with lateral tap/flap, 45 with lateral fricative types, and 29 with lateral affricates, making approximately 20% of the total languages in the UPSID. However, interestingly, the frequency of occurrence of the voiceless lateral is comparatively more in lateral fricative types than in the approximant types despite the approximant types having a high degree of frequency (approximately 80%). Unlike voiceless approximants, voiceless lateral fricatives are reported in inventories that contain no voiced lateral approximant (for instance, in Tlinglit, Puget Sound Salish, Chukchi, Kabardian) Maddieson (1984). As the table shows, lateral taps and flaps are reported fairly rarely.

In the next section, we look at the typology of nasals and laterals in some Tibeto-Burman languages.

### 4.4 Typology of sonorants in Tibeto-Burman languages

In this section, we enumerate nasals and laterals in 20 Tibeto-Burman languages mostly from the Indian sub-continent, including Tibet (gLo Tibetan, Zhongu, Rgyalthang, Kham) and Myanmar (Burmese) and compare the typological distribution of nasals and laterals. Except in Tenyidie and Burmese, in all other languages, to the best of our knowledge, there is no acoustic study done on the nasals and laterals. The phonemic symbols are from the phonemic
inventories of the respective languages (see footnote '6' for references). Table 4.2 lists the nasal and lateral inventories of 20 Tibeto-Burman languages.63

Table 4.2: Typology of nasals and laterals in Tibeto-Burman languages

<table>
<thead>
<tr>
<th>#</th>
<th>Language</th>
<th>Voiced</th>
<th>Voiceless</th>
<th>Voiced</th>
<th>Voiceless</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Tamang</td>
<td>m n η</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>Manipuri</td>
<td>m n η</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>Thadou</td>
<td>m n η</td>
<td></td>
<td>l</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>Thulung</td>
<td>m n η</td>
<td></td>
<td>l</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>Garo</td>
<td>m n η</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>Manange</td>
<td>m m₄ m n η η₆</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>Ao (Chungli)</td>
<td>m n η</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>Tshangla</td>
<td>m n η</td>
<td></td>
<td>l</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>kurtoep</td>
<td>m n η</td>
<td></td>
<td>l</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>Nar-Phu</td>
<td>m n η</td>
<td></td>
<td>l</td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>gLo Tibetan</td>
<td>m n η</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>Denjongka</td>
<td>m n η</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>Angami</td>
<td>m n η</td>
<td>ηⁿ ηⁿ˚ ηⁿ˚</td>
<td>l</td>
<td>ηⁿ˚</td>
</tr>
<tr>
<td>14</td>
<td>Lai</td>
<td>m n η</td>
<td></td>
<td></td>
<td>l</td>
</tr>
<tr>
<td>15</td>
<td>Zhongu</td>
<td>m n η</td>
<td></td>
<td></td>
<td>l</td>
</tr>
<tr>
<td>16</td>
<td>Pyen</td>
<td>m n η</td>
<td></td>
<td></td>
<td>l</td>
</tr>
<tr>
<td>17</td>
<td>Rgyalthang</td>
<td>m n η</td>
<td>hm hn hp hn</td>
<td>l</td>
<td></td>
</tr>
<tr>
<td>18</td>
<td>Tibetan</td>
<td>m n η</td>
<td>hm hn hp hn</td>
<td>l</td>
<td>hl</td>
</tr>
<tr>
<td>19</td>
<td>Kham</td>
<td>m n η</td>
<td>mh nh nh nh</td>
<td>l</td>
<td>lh</td>
</tr>
<tr>
<td>20</td>
<td>Burmese</td>
<td>m n η</td>
<td></td>
<td>l</td>
<td></td>
</tr>
</tbody>
</table>

63 We choose to keep the original phoneme symbols (as they appeared in their respective inventories) for the voiceless consonants, and assume that all these consonants have some sort of voicelessness and/or aspiration (which is not breathy voiced) irrespective of whether the voicelessness or aspiration is pre or post.

Table 4.2 calls for a number of observations regarding the typology of nasals and laterals in the twenty Tibeto-Burman languages. We list the observations below.

- Nasals seem to be attested in all Tibeto-Burman languages; there is no language without a nasal (at least we know of no Tibeto-Burman language without a nasal segment). However, we cannot have the same assumption regarding laterals; 5 (Garo) does not have a lateral in table 4.2.

- One of the key features of Tibeto-Burman sonorants is the presence of voiceless nasals and laterals. Matisoff (2003) points out that voiceless nasals are widely distributed in Tibeto-Burman languages.

- The pattern voiced-voiceless seems to be interestingly uniform in both nasals and laterals, i.e., if a language makes voiced-voiceless distinction in nasals it also seems to make the same distinction in the lateral (though three languages do not conform to this pattern [10, 11 and 17]).

- If a language has voiceless nasals or laterals it also have their voiced counterparts, but not vice versa. Eleven languages (1 to 11) do not have voiceless nasals, while nine languages (12 to 20) make voiced-voiceless contrast, and at least one language (12) does not have voiceless counterparts for the labial nasal.

- At least one language (6) has complex nasal segments for labial /mʷ/ and velar /ŋʷ/ places; however complex segments seem to be rare in Tibeto-Burman languages compared to plain segments, while laterals do not have any complex segments.

- None of the languages in the list has the breathy voiced nasal or lateral (however, Sumi (Harris 2009) is reported to have the breathy voiced nasal).

We, as linguists, are more successful in presenting the facts of the kind listed above than in explaining them, but explanation of such facts would throw significant insights into the way languages choose their segments. The very fact that all 20 languages (except Garo for lateral) have nasals and laterals supports the assumption that nasals and laterals are highly distinctive in Tibeto-Burman languages and in languages in general, i.e., they are rarely subject to confusion with other types of consonants and are reliably identified as nasals
and laterals. Maddieson (1984) states that the distinctiveness of nasals as a class indicates that there is value in incorporating such sounds into any language, and that they are likely to be retained over time.

The presence of voiceless nasals and laterals, however, is not easy to argue on the grounds of distinctiveness. Perception studies show that although nasals and laterals as a class are distinct, they are confusing within the class. Malécot (1956) and Nord (1976) show that nasals with different places of articulation are poorly distinguished in terms of the nasal murmur itself (as opposed to in terms of the transitions to adjoining vowels). However, Recasens (1982) reports that, in Catalan, although transition provided more effective cues for place of articulation than murmurs, the murmurs did make a significant contribution to the /n/-/n/ distinction. On the contrary, in an interesting study, Hillenbrand’s (1983) subjects (who were six-month-old infants, who were phonetically trained on a /na/-/ma/ contrast) found that infants did recognize the similarity of nasal consonants sharing place of articulation independent of variation in talker (stimuli) and vowel context. Maddieson (1984) proposes the hierarchy of preferred places of articulation for nasals which runs in descending order of popularity as: bilabial, velar, palatal. Zee (1981) reports that /n/ is most likely to be correctly identified, while House (1957) found that /ɲ/ is more likely to be misidentified as /n/ after /i/.

Maddieson’s (1984) assumptions and the findings of Zee (1981) and House (1957) have some bearing on the pattern of nasals in the Tibeto-Burman languages, in that out of the 20 languages we have enumerated in table 4.2, only 12 languages (6, 8, 9, 10, 11, 13, 15, 16, 17, 18, 19 and 20) have the palatal nasal /ɲ/ while all languages have /m/, /n/ and /ŋ/. This peculiar pattern may be attributed to the poor perceptual characteristics of /ɲ/ compared to the other three nasals (/m/, /n/ and /ŋ/). Despite all the nasals not being equally perceptually distinctive, Tibeto-Burman languages tend to have a minimum of three and maximum of six (in case of Manange [6]) in table 4.2, including voiceless ones.
In the next section, we start with our comparative study of nasals in the three languages; we first present a review of the acoustic characteristics of the voiceless nasals and laterals (including the breathy voiced types) in languages.

4.5 Background literature on voiceless sonorant phonation types

Before we embark on our acoustic study of nasals and laterals in the three languages under study, it is noteworthy to review the existing acoustic literature on the sonorants’ phonation types. As we have mentioned in the introduction the literature on nasals and laterals is comparatively limited in Tibeto-Burman languages. Most of the literature is on nasals (laterals are comparatively rarely studied) in Tibeto-Burman languages (as far as we know), except the breathy type of nasal phonation in Indic languages. However, a few languages have been studied acoustically in terms of their consonants’ phonation types, for instance, Sumi (Harris 2009); Burmese (Dantsuji 1984, 1986; Ladefoged & Bhaskararao 1991); Khonoma dialect of Tenyidie (Ladefoged & Bhaskararao 1993) and Bengali, Hindi & Marathi (Esposito, Khan & Hurst 2005); Hindi-Urdu (Aziz 2002).

There are basically two types of nasal phonation types reported in languages: 1) the voiceless type, and 2) the breathy type. The voiceless types are further (phonetically) categorized as pre-aspirated, as in Burmese (Dantsuji 1984, 1986) and post-aspirated types, as in Khonoma Tenyidie (Ladefoged & Bhaskararao 1993). Both the types are attested in Tibeto-Burman languages, while the breathy phonation type seems to be attested mainly in Indic languages (for example, Hindi, Bengali, Marathi, Urdu) in South East Asia. However, Harris (2009) reports of breathy nasals in Sumi (a Tibeto-Burman language spoken in the North-Eastern part of India) as well. The segmental status of breathy nasals in Indic languages is not well-established; there appears to be disagreement in the literature over the status of breathy nasals in, for example, Hindi. Some linguists, for example, Maddieson (1984), Hinskens & Weijer (2003) treat breathy nasals as single segments, while some other
linguists like Ohala (1983) and Botma (2004) regard them as nasal plus /h/ (N+h) clusters. On the other hand, Esposito, Khan & Hurst (2005) report an inconclusive finding on the breathy nasals in Hindi; in that, according to their finding, while duration indicates that Hindi /Nh/ is more like the single segment, the schwa epenthesis suggests it is more like a cluster, and conclude that the breathy nasals behave like a cluster as well as a single segment. However, they are in favour of treating Hindi breathy nasals as cluster sequences rather than single segments. Esposito et al. (2005) treat breathy nasals in Marathi as proper breathy segments /N^0/, and not as clusters /N+h/.

In Urdu, Aziz (2002) studies word-initial, word-medial, and word-final /N/ + /h/ sequences and report that in word-initial position, spectrographic evidence reveals that speakers consistently insert a schwa between the nasal [N] and the [h]. Thus, there is no breathiness during the nasal. In word-medial positions, speakers were inconsistent in the pronunciation of [h]. When speakers did produce the [h], there was some aspiration, but only in cases of /mh/, and not other places of articulation. In word-final position, speakers nasalized the preceding vowel, deleted the nasal consonant, and were inconsistent in their production of [h], creating [vM(h)] sequences. In other words, in all three positions, there was little evidence for phonetically breathy nasals in Urdu.

In Bengali, Esposito et al. (2005) report that some data in the within-language (comparison of nasal aspiration with stop aspiration) comparison, (such as duration) point to cluster status for the /Nh/, while other data (such as vowel length), suggest the /N^h/ as a singleton. Interestingly, based on their comparative study of Hindi, Bengali and Marathi, Esposito et al. (2005) propose that there might be three different treatments of /Nh/’s in Indic languages: 1) phonemic breathy nasals (as in Marathi), 2) nasal plus /h/ [Nh] clusters (as in Hindi) and 3) a possible transitional state between the two (as in Bengali).

The phonation type (including manner) of laterals, on the other hand, seems to be more intricate than the nasals; in that laterals make more complex phonation types (including manner types) than nasals. There are at least four
types of laterals attested in languages. Maddieson (1984) lists lateral types as: approximants, taps/flaps, fricatives, and affricates. The distinction between voiceless approximant lateral and voiceless fricative laterals still remains uncertain. Many schemes of phonetic classification propose that there is only one manner of articulation for voiceless laterals, namely that they are fricatives. Pike (1943) states that: "Laterals... upon becoming voiceless narrow the opening sufficiently to get local friction." Unlike voiceless approximants, voiceless lateral fricatives are reported in inventories that contain no voiced lateral approximant, for example, in Tlingit, Nookta, Chukchi and Kabardian (Maddieson 1984). Catford (1977) too implies that all phonemic voiceless laterals are invariably fricatives. However, for phoneticians like Ladefoged (1980) and Maddieson & Karen (1984) the voiceless lateral fricatives and approximants are distinct types of sounds; Ladefoged (1980), however, maintains that the difference is not used contrastively in languages. The contrast between lateral approximants and lateral fricatives occurs only among voiceless sounds (for example, in Burmese and Welsh) (Ladefoged 1980).

Phonologically, there seems to be differences in the way languages employ the two types of laterals, namely the voiceless lateral approximant and the voiceless lateral fricative. Maddieson & Karen (1984), in their comparison of 60 languages with these two types of laterals observe that the two types of laterals differ in their phonotactics, allophonic variations and co-occurrence with other lateral in the inventory. They state the most significant differences as follows:

- Voiceless lateral approximants are restricted to syllable initial positions; fricatives are not.
- Voiceless lateral fricatives may have affricate allophones; approximants do not.
- Voiceless lateral approximants always occur together with a voiced lateral approximant in the inventory; voiceless lateral fricatives may occur without a voiced lateral.

In the next section(s), we embark on our acoustic study of voiceless nasals in Dzongkha, Mizo and Tenyidie.
4.6 Acoustic study of nasals

In this section and the sections that follow, we make an acoustic study of the dental nasals /n/ in initial open syllables in Dzongkha, Mizo and Tenyidie. In doing so, we take into account durational and spectral analysis of the voiced and voiceless nasals and laterals in all the three languages. We make an acoustic study of the phonation types of nasals, especially the voiceless nasals, taking into account the durational properties of both the nasals (/n/ and /ŋ/). We consider: 1) the overall duration of the nasal types and 2) the duration of the phonetic composite parts, the voiceless part and voiced part in the voiceless nasals.

4.6.1 Nasal duration in Dzongkha

For the Dzongkha nasal phonation study, data was recorded for three minimal pairs (voiced nasal with high tone and the voiceless nasal) with three native speakers (2 male, 1 female) for the dental nasals. The speakers were asked to make five tokens in word isolation. The words were:

<table>
<thead>
<tr>
<th>Voiced</th>
<th>ná</th>
<th>‘pus’</th>
</tr>
</thead>
<tbody>
<tr>
<td>Voiceless</td>
<td>ŋá</td>
<td>‘to make slice (of meat)’</td>
</tr>
</tbody>
</table>

Using Praat software the duration of both the nasals was taken; the results (mean duration of four tokens) are presented in table 4.3 and figure 4.1. We measure the three composites of the voiceless nasal: the oral closure part (which is completely voiceless), the aspiration part, and the voiced part based on our spectrographic cues. The first half of the table presents the duration of the nasal types and the second half presents the phonetic duration of the different composites (the voiceless composite [ŋ], the aspiration composite [h] and the voiced composite [n]) of the voiceless nasal /ŋ/ as labeled in the table.

---

65 Henceforth, we use the voiceless diacritic (IPA) (₵) for all the voiceless sonorant types in all the three languages uniformly and assume that all such sounds have phonetic aspiration in them.
Table 4.3: Nasal duration (msec) in Dzongkha

<table>
<thead>
<tr>
<th></th>
<th>85</th>
<th>120</th>
<th>71</th>
<th>29</th>
<th>20</th>
</tr>
</thead>
<tbody>
<tr>
<td>F1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M1</td>
<td>76</td>
<td>144</td>
<td>80</td>
<td>42</td>
<td>22</td>
</tr>
<tr>
<td>M2</td>
<td>78</td>
<td>120</td>
<td>70</td>
<td>25</td>
<td>25</td>
</tr>
<tr>
<td>Mean</td>
<td>80</td>
<td>128</td>
<td>74</td>
<td>32</td>
<td>22</td>
</tr>
</tbody>
</table>

Figure 4.1a: Histogram of the duration (msec) of nasals in Dzongkha

In figure 4.1a, we show the durational results of the nasal types in Dzongkha. We see that all speakers have the duration of voiceless nasal /ŋ/ longer than the voiced nasal. This property of voiceless nasals have been observed in other languages too, for example, Khonoma Tenyidie (Ladefoged & Bhaskararao 1993), Burmese (Dantsuji 1984, 1986), including breathy nasals in Sumi (Harris 2009) in Tibeto-Burman languages and for Indic languages (Hindi, Marathi, Urdu, Bengali) as well. The overall mean duration of the voiced type is 80 msec, while the voiceless type is 128 msec, which is longer than its counterpart by 48 msec.

The ratio of the different phonetic composites of the voiceless nasal is presented in figure 4.1b. The first portion (blue) is the nasal oral closure

---

66 Our study showed that the tone does not reflect the duration of the voiced nasal. We postpone nasal types and tonal effect until chapter 5.
composite, followed by the aspiration composite (red) and voiced portion (green). Unlike Khonoma Tenyidie (Ladefoged & Bhaskararao 1993), where authors do not make minute phonetic distinctions in the voiceless nasals (for them the voiceless nasal is phonetically voiceless throughout the nasal portion), the voiceless nasals in Dzongkha have three distinctive composites: the oral closure (which is voiceless), the aspiration portion after the release of oral closure (without voicing) and the voiced portion (which is comparatively shorter) before the onset of the vowel, as shown in figure 4.1b.

Figure 4.1b: Bar chart of the duration (msec) of [ŋ], [h] and [n]

In figure 4.1b, we see the durational ratio of [ŋ], [h] and [n] in /ŋ/. The duration of the nasal oral closure (blue) is longer than the aspiration (red) portion, and the aspiration portion is again longer than the voiced portion (green). These differences are presented in the waveforms and spectrograms of one of the female speakers (F1) in figure 4.1c.
Figure 4.1c: Spectrogram and waveform of the nasal types

/ŋaː/ ‘pus’

In figure 4.1c in ‘pus’, the waveform and spectrogram of the voiced nasal is self-explanatory; the duration (marked with broken lines) is the duration of the voiced nasal followed by approximately 100 msec of the vowel. The voiced nasal in ‘pus’ is 86 msec, while the voiceless nasal in ‘to make into slice’ is 184 msec. As shown in the figure, we have three distinctive parts of the voiceless nasal: the first part, the oral closure portion (marked with thick unbroken arrow) is 121 msec, next portion, the aspiration portion (marked with
thin unbroken arrow) is 37 msec, and the voiced portion (marked with thin broken arrow) is 26 msec. Unlike in the voiced nasals (which have unambiguous spectrographic distinctiveness for the nasals and vowel), the voiced composite [n] in /ŋ/ does not have such distinctive cues; we rely on the formant (for place) transition from nasal to vowel as shown in figure the figure. As shown in the mean durations in table 4.3 and figure 4.1b, the same composite proportion of the voiceless /ŋ/ is shared in all the three speakers. The average percentage of the [ŋ], [h] and [n] in /ŋ/ is presented in figure 4.1d. One advantage of presenting the mean differences of different composites in percentage is that it automatically normalizes the different speech rates of different speakers.

Figure 4.1d: Average percentage of [ŋ], [h] and [n] in /ŋ/

In figure 4.1d, we show the average percentage of different parts of /ŋ/: [ŋ] is 54%, [h] is 27% and [n] is only 19%. In other words, the duration of the three phonetic composites is: [ŋ] > [h] > [n].

In the next section, we do a pilot study of nasal duration in Mizo and compare it with Dzongkha.

4.6.2 Nasal duration in Mizo

In this section, we do a sample study of nasal duration in Mizo; we recorded two native speakers (1 male, 1 female) and the results of the nasal duration are
presented in table 4.4 and figure 4.2a. We follow the same method of measurement for Mizo as we followed for Dzongkha. We do not have tonal information of the words we have recorded, therefore in our comparison of acoustic properties of /n/ and /ŋ/ tones are not marked. The left column shows the duration of phonemic /n/ and /ŋ/ and the right column shows the phonetic composites of /ŋ/. The words are:

<table>
<thead>
<tr>
<th>Voiced</th>
<th>na</th>
<th>‘pain’</th>
</tr>
</thead>
<tbody>
<tr>
<td>Voiceless</td>
<td>ŋa</td>
<td>‘work’</td>
</tr>
</tbody>
</table>

Table 4.4: Nasal duration in Mizo

<table>
<thead>
<tr>
<th></th>
<th>/n/</th>
<th>/ŋ/</th>
<th>[ŋ] ln /ŋ/</th>
<th>[ŋ] ln /ŋ/</th>
</tr>
</thead>
<tbody>
<tr>
<td>F</td>
<td>78</td>
<td>112</td>
<td>81</td>
<td>31</td>
</tr>
<tr>
<td>M</td>
<td>120</td>
<td>148</td>
<td>110</td>
<td>38</td>
</tr>
<tr>
<td>Mean</td>
<td>99</td>
<td>129</td>
<td>95.5</td>
<td>34.5</td>
</tr>
</tbody>
</table>

Figure 4.2a: Bar chart of the duration of /n/ and /ŋ/ in Mizo

As shown in table 4.4 and figure 4.2a, the duration of /ŋ/ is longer than /n/ (by 28 msec in the male speaker and by 25 msec in the female speaker) as shown in figure 4.2a. The durational cue for voiced-voiceless nasal in Mizo conforms to the findings in other languages with voiced-voiceless nasal
distinction. However, we may note that the phonetic property of the voiceless nasal differs slightly from voiceless nasal of Dzongkha, in that in Dzongkha there are three distinct composites of /ŋ/: the voiceless oral closure [ŋ], the (voiceless) aspiration portion [h] and the voiced portion [n], whereas Mizo /ŋ/ has only two distinct composites: the voiceless composite [ŋ] (with a slight aspiration) and the voiced composite [n] as shown in the waveform and spectrogram in figure 4.2b.

Figure 4.2b: Waveform and spectrogram of the nasal types in Mizo

/na/ 'pain'

/ŋa/ 'work'
As shown in figure 4.2b, unlike Dzongkha, the distinction between nasal oral closure and aspiration is not clear from the waveform and spectrogram for /n/ in Mizo. The two portions blend smoothly with a slight aspiration. For this reason we do not have two composites: [ŋ] and [h] in Mizo as we have in Dzongkha (see figure 4.1e for Dzongkha); the whole voiceless composite may be labeled (phonetically) voiceless [ŋ] (in the absence of significant aspiration). We may further note that the voice quality of the voiced composite in Mizo differs slightly from its counterpart in Dzongkha; in that while Dzongkha voiced composite [n], besides being comparatively shorter than in Mizo, is a voiced one (see figure 4.1c) while its counterpart in Mizo is somewhat a devoiced one. This is evident from the waveforms and the spectrograms of the nasal in the two languages. The duration of the voiceless composite [ŋ] is 122 msec and the duration of [n] composite is 49 msec for the single token in ‘work’. The mean duration of the voiceless and the voiced composite is presented in figure 4.2c.

Figure 4.2c: Bar chart of the duration (msec) of [ŋ] and [n] in /ŋ/ in Mizo

In figure 4.2c, the male speaker has an average duration of 120 msec for the voiceless composite [ŋ] and the voiced composite [n] 38 msec; the female speaker has an average of 78 msec for the voiceless composite [ŋ] and the voiced composite [n] is 31 msec. The duration of [n] is proportionate to the duration of their voiceless counterparts. The average percentage of [ŋ] and [n] is 73% : 27% as shown in figure 4.2d.
Figure 4.2d: Average % of [n] and [n] in /n/ in Mizo

Thus, in our acoustic study of nasal duration study, we are able to show that the duration of the voiceless nasal is longer than the voiced one in both Dzongkha and Mizo; the duration of the voiced is slightly longer in Mizo (99 msec) than in Dzongkha (80 msec), while the duration of the voiceless nasal is more or less same (overall average of 128 msec in Dzongkha and 129 msec in Mizo) in both the languages. The acoustic characteristic of the voiced nasal type is a straightforward one in both the languages, i.e. it is produced with voicing throughout the nasal duration, but the voiceless nasal types differ slightly in their acoustic characteristics. The differences can be captured as follows:

- Dzongkha has three phonetic composites for the voiceless nasal: voiceless composite [\( \text{\textalpha} \)], aspiration composite [h] and voiced composite [n], while Mizo voiceless nasal has only two phonetic composites: voiceless composite [n] and voiced composite [n].
- The voiced composite [n] in Dzongkha is a fully voiced one, while its counterpart in Mizo is somewhat breathy voiced.

---

\[67\] In our pilot study of Tenyidie, the voiceless nasal type has two phonetic composites: voiceless composite [\( \text{\textalpha} \)] and aspiration composite [h]; there is no voiced composite [n] as found in Dzongkha and Mizo, this seems to be the main difference between Tenyidie and Dzongkha/Mizo. The duration of the voiceless nasal in Tenyidie is longer than the voiced nasal; however, the duration of the (phonetic) aspiration composite [h] of the voiceless nasal in is longer than the duration of the voiceless composite [\( \text{\textalpha} \)] in Tenyidie (in Dzongkha, the duration of [\( \text{\textalpha} \)] is longer than [h]). However, such phonetic variations, such as, [\( \text{\textalpha} \)] vs. [h] as in Dzongkha and Tenyidie is observed in other consonants as well, for instance, VOT differences between voiceless and voiced stops are found to vary across languages; thus, voiceless [\( \text{\textalpha} \)] and aspiration [h] may vary across languages. But phonemically, there could be just two major types, namely, voiceless [\( \text{\textalpha} \)] as in Mizo and voiceless aspirated [\( \text{\textalpha} \)] as in Dzongkha and Tenyidie.
The duration of the voiced composite [n] in Mizo is longer (overall average 35 msec) than its counterpart in Dzongkha (overall average 22 msec).

However, we assume that languages employ slightly different acoustic characteristics for the contrast of voiced-voiceless nasals as evidenced by Dzongkha and Mizo, thus the acoustic variations are language-specific; we conclude that the voiceless nasals in Dzongkha and Mizo are of the same type; in that they are voiceless segments. The voiceless nasal segments may be more accurately represented as /ŋ/ and /ŋ/ in Dzongkha and Mizo respectively, based on our acoustic findings.

4.7 Acoustic study of the laterals

In this section and the sections that follow, we study the second class of sonorants, the laterals, in Dzongkha, Mizo and Tenyidie. We follow the same acoustic methods we have used for the nasals, in that we look at acoustic properties of laterals in the initial position in open syllables followed by the open vowel /a/. In doing so, we look at two acoustic properties: duration of the lateral types and the phonetic composites of the voiceless laterals. We start with Dzongkha, first.

4.7.1 Dzongkha lateral duration

For Dzongkha lateral duration, we study the duration of the two laterals: voiced lateral /l/ and the voiceless lateral /ʃ/. We recorded three native speakers (one female, two male) for the following words:

Voiced  lá    ‘to waste’
Voiceless ʃá    ‘excess’
The results of our lateral duration study are presented in table 4.5 and figure 4.3. The figures in the left hand column are the duration of lateral types, and on the right column are the phonetic composites of the voiceless lateral.

Table 4.5: Mean duration (msec) of the laterals in Dzongkha

<table>
<thead>
<tr>
<th></th>
<th>I</th>
<th>I</th>
<th>II</th>
<th>[h]</th>
</tr>
</thead>
<tbody>
<tr>
<td>M1</td>
<td>92</td>
<td>146</td>
<td>113</td>
<td>33</td>
</tr>
<tr>
<td>M2</td>
<td>89</td>
<td>160</td>
<td>126</td>
<td>34</td>
</tr>
<tr>
<td>F1</td>
<td>99</td>
<td>147</td>
<td>111</td>
<td>36</td>
</tr>
<tr>
<td>Mean</td>
<td>93.33</td>
<td>151</td>
<td>116.66</td>
<td>34.33</td>
</tr>
</tbody>
</table>

Figure 4.3a: Bar chart of the mean duration of laterals in Dzongkha

In table 4.5 and figure 4.3a, we see that the mean duration of /l/ is longer than /l/ in all the three speakers. The overall average duration of /l/ in is 93.33 msec and /l/ is 151 msec, with a difference of 58 msec. The difference in terms of percentage (the normalized difference) of all the three speakers together is shown in figure 4.3b.
Figure 4.3b: Normalized mean duration of laterals in Dzongkha

Figure 4.3b shows that the mean duration of the two laterals is: 61% : 39%, in other words, the voiceless lateral is longer than the voiced one by 22%.

The phonetic composites of the voiceless lateral are presented in figure 4.3c.

Figure 4.3c: Phonetic composites of voiceless laterals in Dzongkha

The voiceless composite [l] is longer than the aspiration composite [h] in all the speakers. The overall average of the [l] is 117 msec, while the [h] is 34 msec. The duration differences of the lateral types and the phonetic composite differences of the voiceless type are further illustrated in the waveforms and spectrograms of one of the tokens of M2 in figure 4.3d.
In figure 4.3d, the voiced lateral is characterized by voicing throughout the entire lateral portion, and the transition to the following vowel can be seen with the (place) formant transition from lateral to vowel. The duration of /l/ for this token is 96 msec. In the voiceless lateral in 'excess', we see two composites of the voiceless lateral: the voiceless composite [l] and the aspiration composite [h]; there is no voiced composite [l] as we saw in the

*68 The short devoiced portion after the aspiration composite [h] is a devoiced onset of the vowel; this is based on two cues: perception cue and F1 cues. Perceptually, it is a devoiced...
voiceless nasal). The duration of the voiceless lateral in this token is 122 msec, while the duration of the voiced lateral is 96 msec. The duration of the [l] is 80 msec, while the [h] is 42 msec.

### 4.7.2 Mizo lateral duration

For Mizo lateral duration, we study the two lateral types of two speakers (one male, one female) in isolation. The words were:

**Voiced**

|  la | ‘wool’ |

**Voiceless**

|  h  | ‘song’ |

As in the nasals, we do not have tonal information of the words, thus the words are not marked for tone. The results of the mean duration of /l/ and /l/ are presented in table 4.6 and figure 4.4. The first column (to the left) is the duration of /l/ and /l/ and the second column consists of the two phonetic composites ([l]) and [l]) of /l/.

<table>
<thead>
<tr>
<th></th>
<th>l</th>
<th>l</th>
<th>m</th>
<th>m</th>
</tr>
</thead>
<tbody>
<tr>
<td>Female</td>
<td>129</td>
<td>142</td>
<td>95</td>
<td>46</td>
</tr>
<tr>
<td>Male</td>
<td>130</td>
<td>151</td>
<td>107</td>
<td>44</td>
</tr>
<tr>
<td>Mean</td>
<td>129.5</td>
<td>146.5</td>
<td>101</td>
<td>45</td>
</tr>
</tbody>
</table>
In table 4.6 and figure 4.4a, we show the duration of /l/ and /l'/ in Mizo. As in the nasals, the duration of the voiceless /l'/ is longer than /l/ in both the speakers. In M1, the mean duration of /l'/ is longer than /l/ by 21msec, and in F1 the mean difference is 13msec. The overall average duration for /l/ is 129.5msec and /l'/ is 146.5msec.

In the voiceless lateral /l/, Mizo has the same phonetic composites as in the nasals, a voiceless [l] followed by a voiced [l]; these composites are shown in figure 4.4b and c. The voiced composite [l] is slightly longer in the lateral than in the nasal (see table 4.4 and figure 4.2d for the Mizo /n/ composite duration).
Figure 4.4b: Mean duration (msec) of [l] and [r] in Mizo

In figure 4.4b, we see that the duration of [r], in both the speakers, is shorter than [l] by 63 msec in the Male speaker and 49 msec in the Female. The overall average duration of [r] is 101 msec and [l] is 45 msec. The waveforms and spectrograms of the two laterals in Mizo are presented in figure 4.4c and d.

Figure 4.4c: Waveform and spectrogram of lateral types in Mizo

/la/ 'wool'
In figure 4.4c, the voiced /l/ has the usual waveform and spectrogram similar to that of the voiced lateral in Dzongkha; the lateral portion is voiced throughout and the transition to the following vowel is evidenced by the (place) formant transition from lateral to the vowel, (which rises slightly). The duration of /l/ for this token is 146 msec. In 'song', the voiceless /l/ like the voiceless nasal, again, has two phonetic composites: the voiceless composite [I] (which is 111 msec), with a slight aspiration, and the voiced composite [I] (which is 58 msec). As we can see, the duration of [I] is shorter than [I] by 53 msec in this token.

4.7.3 Comparison of sonorant phonation types in Dzongkha and Mizo

In our study of lateral types, we showed that the lateral duration of the voiceless type is longer than the voiced type. However, the voiceless lateral types in these two languages are slightly different, in that the voiceless lateral in Mizo has two composites: voiceless [I] composite (with a slight aspiration)

---

69 In our pilot study of lateral types in Tenyidie, the duration of the voiceless lateral is longer than the voiced one. The voiceless type has two phonetic composites: voiceless composite [I] and aspiration composite [h]; like in the nasal counterpart, the aspiration composite is comparatively longer than its composite counterparts in Dzongkha and Mizo, in fact, the two composites ([I]), 73 msec and [h], 64 msec) are roughly of the same duration.
and voiced [l] composite. In Dzongkha, there are only two composites (unlike its nasal counterpart, which has three composites): voiceless [l] composite and aspiration [h] composite; the offset of the aspiration composite is immediately followed by the onset of the vowel (which is slightly devoiced, roughly for 10 msec; see figure 4.3c). Like in the nasal counterpart, the voiceless composite is longer than the aspiration composite. In Mizo, the voiceless composite is always longer than the voiced one.

Thus, in our durational study of the sonorants (nasals and laterals) in Dzongkha, Mizo and Tenyidie, the voiceless nasals and laterals are longer than their voiced counterparts across all three languages. The voiceless nasals in Dzongkha and Mizo are of the same type, i.e. they are pre-aspirated, as reported in Burmese (Dantsuji 1986), while in Tenyidie, the voiceless nasal is post-aspirated, conforming to the findings of Ladefoged & Bhaskarao (1993)⁷⁰. However, the phonation type of voiceless lateral in Dzongkha patterns with Tenyidie, in that the voiceless types are post-aspirated, while the voiceless lateral in Mizo is pre-aspirated.

We thus conclude that the voiced-voiceless sonorant (n-ŋ and l-ʃ) contrast in Dzongkha and Mizo is made based on two acoustic cues: the sonorant duration and the different composites of the voiceless sonorants. The voiceless sonorants are longer than the voiced sonorants, though there are language-specific quantitative variations. The voiceless sonorants have different phonetic components: voiceless composite, aspiration composite and voiced composite, with, again, language-specific quantitative variations. Thus, phonologically, a pertinent question to ask is: what is the phonological status of these segments? Are they, Clusters? Affricates? Or Singletons? We conclude that the voiceless segments in these languages are singletons, evidenced by the following verification:

⁷⁰ We, however, note that Tenyidie voiceless nasals (including voiceless laterals) are phonetically characterized by a slight aspiration before the onset of the vowel; Ladefoged & Bhaskarao (1993) analyze the voiceless nasals as "voiceless throughout the nasal portion". We assume that the authors presume the 'aspiration' characteristic as a feature of voicelessness.
The voiced composite of the voiceless sonorants are significantly shorter than the voiced sonorants.

The absence of any sort of vowel epenthesis in these voiceless sonorants, as reported in Hindi (Esposito et al. 2005).

At least, in Dzongkha, the phonotactics of the consonants do not favour any sort of consonant clusters except complex segments (affricates).

However, there is a need to do further investigations on the sonorant phonation types in the languages that contrast voiced-voiceless sonorants, especially in Tibeto-Burman languages, to characterize these sound segments.

In the next section, we present the phonological representation of the sonorant types.

4.8 Phonological representation of Dzongkha nasals and laterals

In this section, we propose the distinctive feature for the two types of nasals and laterals in Dzongkha. In doing so, we, first, present an overview of the literature on the featural classification of sonorants in languages with voiced–voiceless distinction. In section 4.8.1, we present a summary of the typological overview of the voiceless sonorants in line with our analysis of distinctive feature for the sonorant types. This is followed by a short summary of the theoretical background on the analysis of the voiceless sonorants that exist in the literature. And finally, we analyze sonorants in Dzongkha, following Avery & Idsardi’s (2001) model of distinctive feature theory.

4.8.1 Typological overview

The voiceless nasals and laterals occur in only 3.4% (nasals) and 2.6% (laterals) of the languages in the UPSID database (recall section 4.3.1 and 4.3.2), which implies that the voiceless nasals and laterals are relatively

71The same proposal can be extended to Mizo and Tenyidie, though our emphasis is on Dzongkha.
marked. Furthermore, the presence of a voiceless nasal or lateral presupposes the presence of the voiced series (recall section 4.3.1 and 4.3.2 for sonorants in the UPSID database, and section 4.4 for typological overview of nasals and laterals in Tibeto-Burman languages). This, firstly, demonstrates the markedness of the voiceless nasals and laterals. Secondly, the maximum number of phonation contrasts in any language in nasals and laterals appears to be two, as is assumed by Maddieson (1984) and also as is shown by our typology of Tibeto-Burman (recall section 4.4). However, the categorization in the UPSID database predicts a maximally three-way distribution, which to the best of our knowledge is unattested in any language. The key problem lies in the categories of voiceless and breathy voiced nasals, for instance, in Hindi and Sumi (recall section 4.5); but to the best of our knowledge, there is no language in which these two types of nasals (voiceless and breathy voiced) contrast. Under these circumstances, we propose that both voiceless types of nasals (pre-aspirated and post-aspirated), including breathy voiced types, comprise a single phonological category, following Botma (2005), relying on the phonetic presence of aspiration [h] in both the types, irrespective of whether aspiration is pre- or post-nasal. The phonetic realization of the voiceless nasal varies between voicelessness and aspiration (whether it is pre or post).

4.8.2 Theoretical preliminary

The voiceless sonorants (nasals and laterals) are recognized in Chomsky & Halle (1968), and are specified as [-voice]. This analysis suffers from two theoretical problems. First, it relies on the binary-valued feature [voice], which fails to express the relative markedness of voiceless sonorants as compared to voiced sonorants. Second, an account in terms of [-voice] fails to offer an adequate characterization of the phonological behaviour of the sonorants involved, which is characteristic of aspiration rather than of voicelessness. The second point is supported by the universal implication that the presence of voiceless nasals in a language implies the presence of aspirated plosives.

72 UPSID contains 24 languages with voiceless nasals; of these only two, Hopi and Aleut, are described as lacking aspirated plosives (Maddieson 1984). However, in these languages, too there are good grounds for recognizing as aspirated series of plosives. According to Voegelin
Phonetically, the relation between voiceless nasals and aspirated plosives is supported by the fact that both types of sounds involve vocal cord abduction. This motivates Halle and Stevens (1971) to specify both voiceless sonorants, including aspirated plosives, in terms of the feature [spread glottis].

There has been some disagreement, however, regarding the treatment of the manner of the voiceless laterals. Many schemes of phonetic classification propose that there is only one manner of articulation for voiceless laterals, namely that they are fricatives. Pike (1943) expresses this as: "Laterals upon becoming voiceless narrow the opening sufficiently to get local friction." Catford (1977) also implies that phonemic voiceless laterals are invariably fricatives. Yet, for other phoneticians, voiceless lateral fricatives and voiceless lateral approximants are distinct types of sounds. This view was taken, for instance, by Ladefoged (1980). In general, linguists working on languages of Africa, the Americas and Europe favour terms that imply that the voiceless laterals are fricatives, whereas those working on Asian languages favour terms and transcription that imply that the voiceless laterals are approximants. We follow the Asian tradition, and categorize our voiceless laterals in all the three languages as approximants. The phonetic characteristic of the voiceless lateral (high frequency noise, implying frication) has also been observed, for instance, in Tenyidie (Ladefoged & Bhaskararao. 1993). We analyze voiceless laterals in Dzongkha as approximants, and not as fricatives, based on the following arguments:

- Despite differences in phonetic characteristics between voiceless lateral approximants/fricatives, the difference between the two types of manners is a subtle one (Maddieson & Emmorey 1984).
- No language makes a phonological contrast between the voiceless lateral approximant and fricative (Ladefoged 1980, Maddieson 1984).
- Voiceless lateral approximants always occur together with a voiced lateral approximant in the inventory; voiceless lateral fricatives may occur without a voiced lateral (Maddieson & Emmorey 1984).

(1956), reported in Botma (2005), those Hopi dialects which have voiceless nasals also have pre-aspirated plosives, the latter contrasting with a series of plain plosives. In Botma (2004), reported in (2005), the author argues that the voiceless plosives of Aleut are phonologically aspirated.
4.8.3 Featural representation of Dzongkha nasals and laterals

We follow Avery & Idsardi (2001) in specifying features for nasals and laterals in Dzongkha. A summary of the current theoretical stand on distinctive feature theory and the theoretical tenets of Avery and Idsardi have already been discussed in chapter 3; we do not repeat the discussion here. As in the obstruents in chapter 3, we restrict our feature representation to the laryngeal dimension only.

Based on the two acoustic cues, voicelessness and aspiration, in the voiceless sonorants, we present our phonological contrast for sonorants, (following Avery & Idsardi 2001) as:

\( \emptyset /GW+GT \)

Using this feature system, the voiced sonorants are unmarked (\( \emptyset \)), and the voiceless sonorants are represented as a combination of the two dimensions, Glottal Width (GW) + Glottal Tension (GT). For the GW dimension, the gestural completion is [spread] and for the GT the completion is [stiff], which has the same feature combination for VLAS stops in chapter 3. This can be represented as follows:

| /n/ /l/ | \( \emptyset \) (Unmarked) (Default) |
| /n/ /l/ | \( GT+GW \) [stiff] [spread] |

We represent the voiced sonorants as the unmarked segment; the voiceless sonorants are more marked than the voiced ones (recall section 4.13.1). Avery & Idsardi's (2001) model has an advantage over Chomsky and Halle (1968) and Halle & Stevens (1971), in that, firstly, the relative markedness of the voiceless sonorants as compared to the voiced sonorants can be correctly expressed in the feature system by representing the voiced sonorants as the unmarked segment, with a default gestural completion.
Secondly, by combining the two dimensions\(^7\) (the GT dimension and the GW dimension) both the acoustic characteristics, viz., voicelessness and aspiration can be represented in the feature system, which was otherwise not possible in Chomsky & Halle's and Halle & Stevens' binary feature system. Our motivation for the GT dimension is evidenced by the fact that the voiceless sonorants in Dzongkha are always aspirated, as in the VLAS stops (which are \([\text{stiff}]\)). The presence of aspiration is then a reliable acoustic cue for tension under the GT dimension, both in sonorants and stops in Dzongkha.

Cross-linguistic evidence further shows that the presence of voicelessness, the relative order of voicing and aspiration (pre- or post-aspiration), and the degree of overlap between the two are a matter of language-particular phonetic implementation. For instance, Smith (2002) notes that the realization of /\text{wh}/ in Scottish English varies between \([\text{h}^{*}]\) and \([\text{h}^{*}\]). The data in the three languages in our study suggest that the range of phonetic variation in aspirated sonorants is considerable, and includes realizations which may be characterized in phonetic terms as voiceless (as in Mizo, with very little aspiration) and aspirated (as in Dzongkha and Tenyidie). However, from the perspective of phonological contrast the important point is that none of the three languages we have studied makes distinctive use of this variation. This supports the hypothesis that we are dealing with one and the same phonological segment types in all the three languages, represented in terms of the dimension GT+GW with the gestural completion \([\text{stiff}]\) and \([\text{spread}]\), respectively. Thus, we can extend the same phonological representation we have used for Dzongkha to Mizo and Tenyidie sonorants as well.

### 4.9 Conclusion

We began this chapter by stating a fact that Dzongkha, like many other Tibeto-Burman languages in south-east Asia, makes a two-way phonatory distinction with its sonorants. We also recognized the fact that while acoustic studies of

\(^7\) In Avery andIdsardi's (2001) model, the dimensions can be combined, but not the gestures.
sonorant phonation types in Tibeto-Burman languages are scarce, in Dzongkha such studies were never done. We then identified three goals of this chapter: 1) to compare sonorant types of Dzongkha with Tenyidie and Mizo, 2) to show if voiceless sonorants are single segments or segment clusters, and 3) to represent sonorants in terms of distinctive feature theory.

In section 4.2, we presented a short language background of Mizo and Tenyidie and illustrated sonorant contrasts in all the three languages in section 4.2.1. In section 4.3.1 and 4.3.2, we reviewed the cross-linguistic typology of nasals and laterals (emphasizing on the voiceless types) and conclude that the voiceless types are relatively marked than the voiced ones in languages. In section 4.4, we made a separate typological review of sonorants in 20 Tibeto-Burman languages and conclude that the voiceless types of sonorants are relatively marked in Tibeto-Burman languages as well. In section 4.5, we reviewed some laryngeally modified sonorants, including the breathy voiced sonorants in Indic languages (Aziz 2002, Esposito et al. 2005), including Sumi (Harris 2009) and reported that the breathy voiced nasals in Indic languages suggested three possible patterns: 1) phonemic breathy nasal /Nʰ/ as in Marathi, 2) nasal plus /h/ clusters /Nh/ as in Hindi and, 3) a possible transitional state as in Bengali, following Esposito et al. (2005). We also reviewed the discrepancies in the analysis of the voiceless lateral types; the voiceless laterals being invariably represented as voiceless fricatives (Pike 1943, Catford 1977) on one hand, and on the other, that the voiceless lateral approximants and fricatives are distinct types of sounds (Ladefoged 1980, Maddieson & Karen 1984).

In section 4.6 through 4.6.2, we studied nasal duration in Dzongkha, Mizo and Tenyidie and conclude that the voiceless nasals are longer than the voiced ones. We also studied the phonetic composites of the voiceless nasals and showed that the voiceless nasals in Dzongkha and Mizo are pre-aspirated, while Tenyidie voiceless nasals may be post-aspirated.

In section 4.7 through 4.7.3, we studied the acoustic characteristics of the lateral types. We showed that the duration of the voiceless laterals in all the three languages is longer than the voiced ones. Unlike the voiceless nasal, we
showed that the voiceless lateral in Dzongkha does not have the phonetic voiced composite [I], patterning with the voiceless lateral of Tenyidie as a post-aspirated type. In Mizo and Tenyidie, we showed that the voiceless types have the same pattern in both nasals and laterals.

In section 4.7.3 we compared the sonorant duration and the phonetic characteristics of the voiceless sonorants and conclude that:

- the voiceless nasals in Dzongkha and Mizo are pre-aspirated
- the voiceless nasals in Tenyidie is post-aspirated
- the voiceless lateral in Mizo is pre-aspirated, and
- the voiceless laterals in Dzongkha and Tenyidie are post-aspirated.

We, then, noted the idiosyncratic behaviour of the voiceless sonorants in Dzongkha, in that the voiceless nasal and voiceless lateral are not of the same type; voiceless nasal was found to be pre-aspirated (patterning with voiceless nasal in Mizo) while the voiceless lateral was post-aspirated (patterning with voiceless lateral in Tenyidie). We do not have any explanation for this phenomenon (as of now); we will have to do a typological study of voiceless laterals in other Tibeto-Burman languages and see if it is a common phenomenon. In fact, our study suggests that there is a need to recognize the phonetic characteristics of the types of voiceless sonorants that are attested in languages that make voiced – voiceless contrast despite the fact that the different phonetic distinctiveness not being contrastive in languages. There seems to be, at least, three ways in which the phonetic characteristics of the voiceless sonorants differ:

- the first type is pre-aspirated (as in Mizo)
- the second type is post-aspirated (as in Tenyidie), and
- the third type is both pre- and post-aspirated (as in Dzongkha; the voiceless nasal is pre-aspirated and the voiceless lateral is post-aspirated).

However, the third type, as we have suggested, will need further typological survey in the languages that make voiced – voiceless sonorant contrasts.
And finally, in section 4.8 through section 4.8.3, we made a featural analysis of the sonorants in Dzongkha, following Avery and Idsardi’s (2001) model of distinctive feature theory and presented the features for sonorants in Dzongkha as:

<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>/n/</td>
<td>/n/</td>
<td>Ø(Unmarked) (Default)</td>
</tr>
<tr>
<td>/ŋ/</td>
<td></td>
<td>GT+GW [stiff] [spread]</td>
</tr>
</tbody>
</table>

The voiced sonorants are analysed as unmarked, with a default gestural completion, and the voiceless sonorants as a combination of two dimensions: the Glottal Tension (GT) and the Glottal Width (GW), with their gestural completions as [stiff] and [spread] respectively. We have also shown that this feature captures both the [voice] and [spread] which was not possible in the standard model of feature theory.
5.1 Introduction

In chapter 3 and 4, we have described the major segmental characteristics of consonant (obstruents and sonorants) phonation types in Dzongkha and concluded that while sonorants can have a two-way phonation distinction (voiced and voiceless [aspirated]), obstruents (stops) can have as many as four-way distinctions (VLT*, VLAS, VLL and VD). In this chapter, we extend our findings of the segmental characteristics to the study of the prosodic structure; more specifically, to the development of register tone in Dzongkha. In other words, the characteristics of consonants will now serve as a basis for the study of register tone in Dzongkha. This chapter has two goals: 1) to validate the effect of initial consonant types on the register system and, 2) to investigate the Intra-Register contrasts in the two syllable types: final _? type and open syllable type.

In section 5.2, we review the earlier literature on Dzongkha tones, especially by Mazaudon & Machailovsky (1988) and Watters (2002) and review Dzongkha as a Register tone system with four tones (two level and two contour tones).

In section 5.3 through 5.3.5, we experiment the effect of initial consonant types and the Register system, and show that, in fact, the initial phonation types do interact with the Register system in Dzongkha; in that the VLT* and VLAS types associate with the High Register and the VLL and VD types with the low Register in stop types in section 5.3.1; the VLT* fricative types associate with High and the VLL and VD fricative types with Low in section 5.3.2, and the voiceless sonorant associate with High and the voiced with Low in section 5.3.3. In Dzongkha, since lexical tonal contrasts can be made only with sonorant initials, we also show the lexical tonal contrast in sonorants with lateral initials in section 5.3.3. We summarise the findings in section 5.3.4. Finally, in section 5.3.5, we present a historical perspective on
the development of the Register system in Dzongkha based on orthographic data and show that the two historical phenomena generated the Register system in Dzongkha: 1) devoicing of the (historically) voiced obstruents and 2) initial cluster simplification, we also illustrate the historical development of voiceless sonorants in this section.

In section 5.4 through 5.4.5, we study the pitch patterns in both the Registers, including Intra-Register pitch contrasts. In 5.4.1, we show that the High Register pitch has a falling pattern while the Low Register pitch has a rising pattern. In 5.4.2, unlike earlier linguists (who posited the glottal stop as a feature of tone) we give sub-phonemic status to glottal stop. We illustrate the historical development of the glottal stop using orthographic data, in which all (except *b) the ancient final obstruents (*d, *g and *s) merged into a final _?. In 5.4.3, we show that the vowel in the open syllable type is significantly longer than the vowel in the syllable with a _? final. In 5.4.4, we study the Intra-Register pitch differences and show that there is a significant phonetic Intra-Register pitch differences triggered by the syllable shapes. We show that the closed syllable type (CV?) is significantly higher than the CVV type in both the High and Low Register. We also show the vowel length differences in the two syllable types. In the following section in 5.4.5, we provide historical data in support of the Intra-Register pitch contrasts.

In 5.5, we analyze Dzongkha as an incipient, bitonal language, with a right-to-left tonal association. We propose two phonological tones [+High], [-High] separated by a [+Upper], [-Upper] Register system, plus a default Mid tone. In 5.5.2, we propose a tonal feature for Dzongkha tones following Yip's (1989) tonal feature for two tone Register system.

And finally in 5.6, we conclude the chapter by saying that our analysis of tones in Dzongkha differ from earlier literature, in that Dzongkha (perhaps, only Paro dialect) has only two tones. In earlier literature, Thimphu, Chapcha and Pasakha dialects were analysed as a four tone languages. We then end the chapter by identifying future areas of research in Dzongkha.
5.2 Earlier analysis of Dzongkha tone system

To the best of our knowledge, tone system in Dzongkha has been studied, so far, by only four linguists, viz., by Machailovsky (1986) (we did not have access to the paper), Mazaudon & Machailovsky (1988), Van Driem (1992) and Watters (2002). In the three studies, linguists agree that Dzongkha has a register tone system, in which the pitch of the vowels following the VLT* and VLAS consonants is high and the VD and VL low. Voiced sonorants occur in both the registers (except for ‘r’, which occurs only in the low register). Mazaudon & Michailovsky’s (1988) study is on two regional dialects, Thimphu and Chapcha and Watters’ (2002) is on a dialect of Pasakha (Chukha District). Van Driem’s (1992) study is (supposedly) on the standard dialect of Thimphu, though he does not make any explicit mention neither of his informants and does nor make a comprehensive description of the tone system in Dzongkha (his work is a pedagogical grammar of Dzongkha). We may note here that Dzongkha has various regional dialects, and each dialect differs from the other with considerable phonological differences, including tone. In fact, Mazaudon & Michailovsky (1988) mention that Paro dialect may not have the contour contrast within the register as is found in Thimphu and Chapcha dialects.

Mazaudon & Michailovsky’s (1988) name the tones in Dzongkha, in the order of their absolute pitch, as: 1 high-level, 2 high-falling, 3 low-level, 4 low-falling. On short open syllables and on monosyllables with final ‘_n’ there is no distinctive contour. According to Mazaudon & Michailovsky, the contour opposition is clearest on long, open syllables: tone 1 and 3 rise slightly and end with a glottal stop, while tone 2 and 4 tend to fall, and end smoothly. They observe that the tone contour before final ‘_m’, tones 1 and 3 are glottalized, in addition to being “somewhat higher” than tones 2 and 4. Before final ‘_p’, the authors note that “the main difference is one of absolute pitch”, and this difference is perceived more clearly in the high than in the low register, with no apparent contour difference.

---

74 We could find no example of ‘r’ in the high register, however, in the low register it makes the same within register distinction as exemplified by the minimal pair: /ryʔ/ ‘flood, land slide’ and /ry:/ ‘to rot’. 

156
Waters (2002) observes that Dzongkha (along with Lhomi, Sherpa, Dolpo Tibetan and Mugom Tibetan) is an "emergent tone" language, accompanied by phonational and length contrasts. A high–low pitch contrast on monosyllables has been noted, which is accompanied among other things by a phonational difference. He refers to this pitch contrasts as high and low registers, and gives the major phonetic features which characterize the high and low registers as:  

**High register 'Tense'**

- Modal voice on the vowel
- Glottal stop on vowel initial syllables
- Strong aspiration
- Plain stops and affricates are voiceless

**Low register 'Lax'**

- Absence of above
- Often breathy voice on the vowel
- Voiceless stops and affricates vary between slight and no aspiration
- Vowels are longer than their high register counterparts; vowel quality contrasts are maintained rather than vowel length contrasts.

Thus, Dzongkha has been analysed by all earlier linguists as a register tone language with a level tone and a falling tone in both the registers. In the following section(s), we posit a slightly different analysis of the register system in Dzongkha (Paro dialect) with acoustic evidence, however maintaining the same two register distinctions as earlier linguists did.

In the next section, we study the interaction of the initial consonants and the register system in Dzongkha.

---

75 Waters' (2002) data is of Lhomi; he extends his analysis to Dzongkha and the other languages under study.
5.3 Dzongkha initial consonants and the Register contrast

In this section, we study the pitch pattern of the vowel following the initial consonant (obstruents and sonorants) types: stops (VLT*, VLAS, VLL and VD), fricatives (VLT*, VD and VLL), and laterals (VLT* and VD). We first present the pitch patterns of all the initial consonant types; make a phonetic description of the pitch patterns of each consonant type (manner) separately and draw a generalisation finally. For all the consonant types, we measure the pitch of the entire vowel following initial types; we take three tokens of four speakers (2 male, 2 female) of all the consonant types and take their pitch values using Praat Automatic Pitch Listing option (this system extracts pitch values of the entire vowel at each pitch point). The values were then transferred to Microsoft Excel for the calculation of the average means. We do not normalize the duration of the vowel (this helps us see the impact of initial types in the length of the vowel, if there exists any). We first start with the stops.

5.3.1 Stop initials

The words we used for the stop study were:

VLT*  t*o  'food'
VLAS  tho  'hand measurement',
VD    do   'stone' and
VLL   to   'two')

The results of the pitch values of the vowel following stop types are presented in figure 5.1.

---

76 Since voiced sonorants (nasal and lateral) have contrastive tones we include the pitch patterns of both High and Low tones following the voiced lateral.
In figure 5.1a, we show the pitch patterns of Male 1, we make several observations in the figure. There is a clear High-Low distinction (henceforth, we refer to this pitch distinction as 'High' and 'Low'), at least, till the middle of the vowel; the pitches following the VLT* and VLAS is high and the pitches following the VLL and VD low. The High has a falling tendency, while the Low has a rising tendency. We may also note here that in the initial portion the pitch of the VLAS and VLL types are slightly lower than their counterparts in High and Low. The High and Low cross-cut and merge after the seventh pitch point (which is approximately after 60% to 70% of the vowel duration). The pitch following the VLT* and VLAS initials (more or less) remain steady in the middle of the vowel, then fall, while the pitch of the vowel following the VLL and VD initials starts with a fall, rise steeply in the middle of the vowel, and fall in the end. We further note that the vowel following the VLT* and VD types are slightly longer than the vowel following VD and VLL types.
In figure 5.1b, Male 2 shares the same acoustic features as in Male 1, except that the pitch of the vowel following the VLAS initial is slightly higher (after the third pitch-point) than the pitch following the VLT* initial in the High, however, in the initial portion it is slightly lower than the vowel following the VLT* type. In the Low Register, the pitch following the VLL initial is slightly lower than the pitch of the vowel following the VD initial throughout the entire vowel duration. The pitches of the vowel following the VLT* and the VD initials cross-cut at the seventh pitch point, which is roughly after the 70% of the vowel. The vowel following the VLT* and the VD types are slightly longer than their counterparts.

c) Female 1
In figure 5.1c, Female 1 has hardly any steady pitch level in the High (as seen in the male speakers); the High falls and the Low is the same as the male speakers. Unlike the male speakers, the pitches following the VLT* and VLAS initials do not cross-cut throughout the entire vowel duration in the High, however in the Low, the pitches following the two initials merge after the third pitch point, till the end of the vowel.

d) Female 2

In figure 5.1d in Female 2, other characteristics remaining (more or less same as other speakers) the pitches of the initial portion of the vowel following VLAS and VLL types are slightly lower than their counterparts in both High and Low. The vowel following the VD type is slightly longer than the other stop types.

Thus, we can generalize the pitch patterns of the four stop types as follows:

► There is a clear High and Low pitch distinction in the vowel following initial stop types; the pitch in the vowel following VLT* and VLAS are High and the pitch following VLL and VD are Low (we refer to this High and Low distinction as Register distinction later in our discussion).

► The High has a falling tendency and the Low has a rising tendency, criss-crossing towards the end of the vowel.

► The pitch of the VLAS and VLL types in the initial portion of the vowel are slightly lower than the VLT* and VLL types. We may associate this
phenomenon to the effect of the presence of aspiration in the initial stop types.

In the next section, we study the pitch pattern of the vowel following fricative types.

5.3.2 Fricative initials

In this section, we study the pitch pattern of the vowel following fricative types (VLT*, VLL and VD) in Dzongkha. We take the dental fricatives, and the words were:

VLT* s*ô: 'thumb measurement, inch'
VLL so: 'sickle'
VD zo: 'carpenter'

The results of the pitch values following the three fricative types are presented in figure 5.2 (as there is not much idiosyncratic differences in the pitch patterns of the speakers we do not present our discussion of the speakers separately).

Figure 5.2: Pitch pattern of the vowel following fricative types

a) Female 1
Examining figure 5.2 a, b, c and d, we make several observations (quite similar to those made on stop types). Like in the stops, there is a clear High and Low pitch division; the pitch of the vowel following the VLT* fricative type is High and the VD and VLL types Low. The High has a falling tendency (sharply in the initial portion of the vowel, except speaker Male 2), and the Low has a rising tendency. The High and the Low criss-cross each other towards the end of the vowel, as in the stops. The pitch of the vowel following the VLT* fricative maintains a, more or less, steady pitch in the major portion of the vowel, until it makes a fall towards the end. The pitches of the vowel following the VD and VLL fricative types make a steady rise in the major portion of the vowel and ends in a fall (except Male 1 77). In the initial portion of the vowel, in the Low Register, the pitch of the VLL type starts at a slightly higher level than the VD type, criss-cross with the VD type, and continues at a slightly lower pitch than the VD type. We may further note that the vowel following the VD fricative is always (slightly) longer than the VLT* and VLL types.

Thus, in this section, we showed that, as in the stops, there is a High-Low Register distinction, generated by the initial fricative types. The pitch of the VLT* fricative patterns with the VLT* and VLAS stops, while the VLL and VD fricatives pattern with the VLL and VD stops.

In the next section, we study the pitch pattern of the vowel following the sonorant types.

### 5.3.3 Sonorant (lateral) initials

In this section, we study the pitch patterns of the vowel following lateral types. Since sonorants are the only initials in Dzongkha that has contrastive pitch (tone) we also illustrate the pitch contrast in the VD laterals. The pitch values of the lateral types in Dzongkha are presented in figure 5.3 (as in the fricative

---

77 The pitch pattern of the vowel following the VLL and VD fricative types of Male 1 rises steeply in the end (which we do not show in figure 5.2c). This may be because of the fact that this speaker has a glottalized vowel.
initials, there is not much idiosyncratic differences in the pitch patterns of the speakers we present a combined discussion of the speakers).

Figure 5.3: Pitch pattern of the vowel following lateral types
In figure 5.3 a, b, c, and d, we make similar observations we made in section 5.3.1, 5.3.2 and 5.3.3, except that the voiced lateral makes a High and Low contrast, with the High patterning with the voiceless lateral. We may also note here that the pitch of the voiceless lateral starts at a slightly lower level (due to the presence of aspiration in the lateral type) than the voiced type (in the High), criss-crosses the pitch of the voiced type and continues to rise before it falls (more or less) with the voiced type. The pitch height difference between the Low and the High of the voiced lateral is 20 to 50 Hz, and the vowel is slightly shorter after the voiceless lateral type.
5.3.4 Discussion on the acoustic study

From our acoustic study of the pitch patterns of the vowel following the consonant initial types in Dzongkha we make the following assumptions:

- As posited by the earlier linguists, there is a clear high-low Register distinction based on the phonation type of the initials; the VLT* and VLAS correlate with High Register and the VD and VLL with the Low; this distinction exists till (roughly) 75% of the vowel.
- The pitch of the High and Low cross-cut after roughly 75% of the vowel duration.
- There is no systematic correlation between the initial types and pitch within the Register.
- The vowel tends to be longer after the VD initial type.

Thus, in this section, among other things, we showed that the two-way Register distinction in Dzongkha is evident, and it is as a result of the initial phonation types. Thus, in Dzongkha, both obstruents and sonorants (both nasals and lateral), including semivowel /j/ and /w/, can be both High and Low. However, obstruents are not lexically specified for tone. The obstruent initials can best be posited as tonogenesis in progress (we illustrate tonogenetic process in the following section). The sonorants, lateral (as illustrated above) and nasals (/nà/ ‘barley’ vs. /ná/ ‘pus’) including semivowels /j/ (/jê/ ‘one of a pair’ vs. /jê/ ‘allergy, itch’) and /w/ (/wân/ ‘name of a place in Bhutan’ vs. /wân/ ‘blessing from a Buddhist priest’) are predictable, and therefore contrastive.

In the next section, we postulate the tonogenetic perspective of the Register system in Dzongkha.

5.3.5 Initial phonation types and Register tone: a historical perspective

In the theory of tonal evolution or tonogenesis, linguists have discovered various influences of consonants and vowel quality on the pitch values of
vowels. At least two of these influences are most-documented and claimed as universal phenomena. The effect from initial consonant voicing on the pitch values of the following vowels has been shown to give rise to tones. Voiceless initial consonants increase the pitch values, while voiced initial consonants decrease the pitch values, resulting in a high pitch and a low pitch, respectively. These findings have been attested in many non-tonal and tonal languages all over the world (e.g., House & Fairbanks 1953, Lehiste & Peterson 1961, Hombert et al. 1979, Thavisak 2000, Teeranon 2006). This leads to the conclusion that voiceless consonants give birth to high tones and voiced consonants give birth to low tones.

The influence of initial voicing distinction on the F0 of the following vowels is one of the most documented areas in acoustic studies for the theory of tonogenesis. Many studies have been carried out to explain the loss of initial consonant voicing distinction as a cause of the development of high and low tones (House & Fairbanks 1953, Haudricourt 1954, Lehiste & Peterson 1961, Lehiste 1970, Gandour 1974, Erickson 1975, Hombert et al. 1979, Maddieson 1984, L-Thongkum 1990, Thurgood 2002, Watkins 2002). The emergence of a high tone was presumably through the loss of voiceless initial consonants, while a low tone was developed through the loss of voiced initial consonants. The most convincing explanation from a physiological basis is the mechanism involving the cricothyroid muscles of the larynx to stiffen the vocal folds for the voiceless state or to slacken the vocal folds for the voiced state (Hirose & Gay 1972, Lofqvist et al. 1989).

The contrast between high and low Register and its relationship to the voicing opposition of Tibetan orthography has been long established (Jaschke 1881; Sprigg 1954, 1955, 1996, 1972; Richter 1964; Chang & Chang 1978). In the investigation of tonogenesis in Asian languages linguists rely heavily on the orthography of the languages for historical reconstruction. According to Glover (2007), "In many orthographic systems in Asia a contrast symbols derived respectively from the Devanagiri voiced and voiceless consonants symbols (used for Sanskrit) is used to mark syllable modifications like tone... Tibetan, Thai, Laotian and Burmese are examples". Thus, Henderson (1952)
notes that the first and second registers in Cambodian are signalled orthographically by consonant symbols corresponding respectively to the symbols for voiceless and voiced consonants in Sanskrit. The historical nature of this correlation is argued for Tai languages by Li (1966) as well.

In Dzongkha, the history of tonogenesis can be attributed to three historical phenomena (relying heavily on the orthography, which is the same as Tibetan orthography): 1) devoicing (followed by splitting) of the (historically) voiced obstruents, as illustrated in (I) below, 2) initial cluster simplification, as illustrated in (II), in the genesis of the low register. We then posit that historically, the phonetic pitch following both voiced and voiceless initials was high before tonogenesis occurred (this is still attested in Amdo Tibetan (Sun 2003)). This is cross-linguistically the best-known tongenic path, and can be partially explained in terms of articulatory and auditory phonetic mechanisms as due to the tendency for prevocalic voiced obstruents to depress pitch height (Hombert et al. 1979).

<table>
<thead>
<tr>
<th>WD\textsuperscript{78}</th>
<th>Pronunciation</th>
<th>Gloss</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>I.</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1. *ba</td>
<td>/pè/</td>
<td>'cow'</td>
</tr>
<tr>
<td>2. *bu</td>
<td>/pù/</td>
<td>'son'</td>
</tr>
<tr>
<td>3. *dar</td>
<td>/tà/</td>
<td>'flag, scarf'</td>
</tr>
<tr>
<td>4. *dug</td>
<td>/tù?/</td>
<td>'poison'</td>
</tr>
<tr>
<td>5. *drim</td>
<td>/ têm/</td>
<td>'smell, odour'</td>
</tr>
<tr>
<td>6. *'bri</td>
<td>/tù/</td>
<td>'write'</td>
</tr>
<tr>
<td>7. *ga</td>
<td>/kà/</td>
<td>'who'</td>
</tr>
<tr>
<td>8. *gjel</td>
<td>/kè/</td>
<td>'line, queue'</td>
</tr>
<tr>
<td><strong>II.</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1. *ma</td>
<td>/mè/</td>
<td>'lock (N)'</td>
</tr>
<tr>
<td>2. *rma</td>
<td>/mè/</td>
<td>'wound'</td>
</tr>
</tbody>
</table>

\textsuperscript{78} 'WD' stands for 'Written Dzongkha', which is more or less the same as Written Tibetan, and the symbol '*' indicates the ancient Tibetan forms.
3. *la /lə/ ‘hills, mountain’
4. *lha /lê/ ‘god, deity’
5. *lham /lám/ ‘shoes’
6. *nas /nâ/ ‘wheat, barley’
7. *rnag /nâʔ/ ‘pus’
8. *sna /nâ/ ‘nose’
9. *snam /núm/ ‘oil’
10. *lo /lò/ ‘year, age’
11. *glo /lô/ ‘cough’

The data in (I) show that the ancient voiced obstruents got devoiced, resulting in the synchronic VLL series of obstruents with a low tone in the low Register. In (II), the first process involved the genesis of tone following sonorants. The first stage of this involved the simplification of complex onsets in which the second member was a sonorant, in case of nasals, and a voiceless velar fricative /h/, in case of lateral. The first member of the onset cluster triggered high tone (in nasals) and the cluster simplified in favour of the sonorant (as in 2, 7, 8, 9 and 10). Whenever the onset cluster consisted of a sonorant (lateral) plus ‘h’ the cluster simplified, again, in favour of the sonorant, devoicing the sonorant and leaving the remnant of aspiration in the sonorant (as in 4 and 5). This process - cluster simplification - also explains the historical development of voiceless sonorants (voiceless nasals and lateral types) in chapter 4. So, the voiceless sonorants are a later development in the phonology of Dzongkha. Unlike in the development of voiceless nasals where only the preceding voiceless consonants seem to have played role in the development of voiceless nasals and tonal split in Dzongkha (and other languages in South East Asia), the following segment in the cluster (voiceless glottal fricative /h/), not the preceding segment, had its influence in the development of the voiceless lateral in Dzongkha. Syllable onsets which
consisted only of a sonorant consonant developed into a low tone on the following vowel without a change in the initial consonant (as in 1, 3, 6 and 9).

In Dzongkha, the most plausible explanation in which tone may shed light on the phonological status of aspirated sonorants is tonogenesis, in that aspirated sonorants may have their provenance to the process of tonogenesis, more specifically tonal split. This has been reported in a number of languages in South-East Asia, including Chinese, Tibeto-Burman, Miao-Yao, Vietnamese, and Kadai languages (Botma 2005). Tone splitting concerns a diachronic process whereby tone contrasts in vowels emerge under the influence of a preceding (and following consonant as well, as in the development of voiceless lateral in Dzongkha) consonant. The typical conditioning factor of tone splitting is voicing, in such a way that after a tone splitting has occurred one tone series is associated with preceding voiceless consonants, and the other with preceding voiced consonants (Botma 2005). Both tonal contrast and the phonological status of voiceless sonorants can be explained with the phenomenon of tonogenesis.

In the next section, we study the phonetics of the Intra-Register distinction.

5.4 Pitch pattern study

As we mentioned, earlier studies in Dzongkha have posited four tones divided into two Registers; Madaudon & Michailovsky’s (1988) name them, in the order of their absolute pitch, as: 1 high-level, 2 high-falling, 3 low-level, 4 low-falling (in Chapcha and Thimphu dialects). According to them, tone 1 and 3 rise slightly and end with a glottal stop and tone 2 and 4 tend to fall and end smoothly. Following Madaudon and Michailovsky, Watters (2002) too, posits the same tonal pattern (for Pasakha dialect). The earlier linguists have also analysed the glottal stop, associated with tone 1 and 3, as a property of tone, and not of segment. However, in our study, we show that firstly, all the four tones in Dzongkha are, in fact, phonetically dynamic tones and not level, and tone 1 is not a rising tone (it falls) and tone 4 is not a falling tone (it rises).
thus, we propose that the four tones in Dzongkha are phonetically contour tones. Secondly, we analyse the glottal stop ‘ʔ’ as a property of segment (providing historical evidence), and not of tone. And thirdly, we show the difference in vowel length (the tone bearing unit [TBU]) is significant (in what Mazaudon and Michailovsky calls tone 1 and 3 versus tone 2 and 4). We first begin with the acoustic study of the general pitch patterns of the four tones in Dzongkha.

5.4.1 Acoustic study of the pitch pattern

In order to determine: 1) if there are, in fact, four tones (including Intra-Register distinctions) and 2) if there exist the level versus falling contrasts, we measured the F0 extracted at each pitch point (using Automatic Pitch Tracking option in Praat) of the following words: (We follow earlier linguists in naming the tones (for now) as: tone 1, 2, 3 and 4; tone 1 as the highest and tone 4 as the lowest, as marked below).

| High Register | falling     | ćåʔ¹ | ‘quarrel’ |
|               | level       | ćå:² | ‘east’    |
| Low Register  | falling     | bzaʔ³ | ‘paddy’  |
|               | level       | bza:⁴ | ‘summer’. |

The same four informants for the experiment in section 5.3 were made to produce the above words four times in isolation; first three iterations were taken for the measurement. The average of their pitch values for three iterations at each pitch point were then collected in an excel file for further analysis. We present the excel analysis of all the four speakers in figure 5.2.

The results suggest four observations:

1) All the pitches are (more or less) dynamic, not level; High Register words fall and Low Register rise.
2) The vowel length is significantly different between the two words in both the Registers.
3) The two words in both the Registers also differ in their syllable shape, in that ‘quarrel’ and ‘paddy’ are closed syllables, with a glottal stop ‘?’ coda and the other two words are open syllables.
4) The pitch pattern within the Register can best be analyzed with pitch height differences rather than level – falling contour.

We illustrate all the four phenomena separately; we first make a description of the dynamic nature of each pitch of each speaker and then draw a generalization of the pitch patterns of each word.

Figure 5.2A: Pitch of /cã/ ‘quarrel’

a) Male 1

In figure 5.2A, we show the pitch patterns of the word /cã/ ‘quarrel’ of all the four speakers. In (a) in Male 1, the pitch starts at 146 Hz, falls rather steeply till the sixth pitch point and remains steady for another three pitch points, and then falls gradually till 119 Hz. There is a difference of 27 Hz between the start-point and the end-point.
In (b) in Male 2, the pitch starts at 188 Hz, begins to fall gradually till (approximately) 70% of the vowel and then falls steeply, and ends with 128 Hz. The difference between the start-point and the end-point is 60 Hz.

In (c) in Female 1, the pitch starts at 285 Hz, falls steeply till the sixth pitch point and then remains (more or less) steady for another three pitch points, and falls steeply and ends at 235 Hz. The difference between the start-point and the end-point is 50 Hz.
In (d) in Female 2, the pitch starts at 265 Hz, falls steeply till the fourth pitch point, then the fall becomes gradual for another four pitch points, and ends with a steep fall. The difference between the start-point and the end-point is 34 Hz.

In 5.2A, in all the four speakers, we observe that the pitch pattern for 'quarrel' has a significant fall with a (slight) steady pattern in the middle of the vowel. The overall average difference between the start-point and the end-point is 42.75 Hz.

In figures 5.2B, we present the pitch patterns of the word, /cā:/ ‘east’ of the four speakers.

Figure 5.2B: Pitch of /cā:/ ‘east’

a) Male 1
In (a) in Male 1, the pitch starts at 140 Hz, falls steeply till the fifth pitch point and remains steady for another four pitch points; it then falls slightly for another four pitch points and remains (more or less) steady for another six pitch points and ends with a steep fall at 114 Hz. The difference between the start-point and the end-point is 26 Hz.

b) Male 2

In (b) in Male 2, the pitch starts at 170 Hz, falls steeply till the fourth pitch point and maintains a (smooth) level pattern for another twenty pitch points and then ends in a steep fall at 100 Hz. The difference between the start-point and the end-point is 70 Hz.

c) Female 1

In (c) in Female 1, the pitch starts at 275 Hz, makes a steep fall till the sixth pitch point; it then maintains a level pattern for another ten pitch points, rises
slightly and then falls in the end (we may note here that the vowel quality of this speaker is a glottalized one in the entire tokens, which is associated with a slight rise in pitch towards the end of the vowel). The difference between the start-point and the end-point of the pitch is 40 Hz.

d) Female 2

In (d) in Female 2, the pitch starts at 250 Hz, falls steeply till the sixth pitch point; it (more or less) maintains a steady pattern for another ten pitch points and then ends in a fall. The difference between the start-point and the end-point is 40 Hz.

In 5.2B, in ‘east’, we observe that the pitch point starts with a high, falls steeply in the beginning (roughly) five pitch points and then maintains a steady (level) pattern for the major portion of the vowel (roughly 60%) and ends in a fall. The overall average difference between the start-point and the end-point is 42.75 Hz, which is exactly same as in ‘quarrel’. Unlike in 5.2A, in 5.2B, the pitch slide differs slightly, in that there is a difference of an immediate fall and a postponed fall. We then propose that the pitch patterns of the two words in the high Register are phonetically a falling one. The longer level pattern in the middle of the vowel in ‘east’ may be attributed to the comparatively longer duration of the tone bearing unit, i.e., the vowel in ‘east’ than in ‘quarrel’.

In figure 5.2C, we present the pitch patterns of the word: /bzaʔ/ ‘paddy’ in the low Register.
Figure 5.2C: Pitch of /bədaʔ/ ‘paddy’

a) Male 1

In (a) in Male 1, the pitch starts with a low at 123 Hz, remains level till the fifth pitch point and makes a steep rise until 128 Hz, and ends in a steep fall. The difference between the start-point and the end-point is 6 Hz.

b) Male 2

In (b) in Male 2, the pitch starts with a low at 122 Hz, rises steeply until the fifth pitch point and then (more or less) remains steady for another ten pitch points, and ends with a fall. The difference between the start-point and the end-point is 24 Hz.
In (c) in Female 1, the pitch starts at a low 227 Hz, remains steady till the sixth pitch point and then rises steeply until 240 Hz and drops steeply in the end. The difference between the lowest point and the highest point is 13 Hz.

In (d) in Female 2, the pitch starts at a low 222 Hz, falls (slightly) to 219 Hz; it remains steady for three pitch points and rises steeply until 232 Hz, and then ends with a fall. The difference between the lowest pitch point and the highest point is 13 Hz.

In 5.2C, in ‘paddy’, the pitch starts at a low, maintains a steady state for 4 to 5 pitch points (except Male 2, which makes a steep rise right from the
start-point79), and rises steeply and then ends with a fall. The overall average between the lowest point and the highest point is 14 Hz. We again observe the same phenomena of immediate rise and postponed rise, as we saw in 5.2A, and 5.2B.

In figure 5.2D, we present the pitch pattern of the word /bza:/ 'summer' of all the four speakers.

Figure 5.2D: Pitch of /bza:/ 'summer'

a) Male 1

In (a) in Male 1, the pitch starts with a low 123 Hz, dips little further (till the third pitch point) and then maintains a steady state for another five pitch points before it rises till 124 Hz; it maintains a (more or less) steady pattern for at least seven pitch points before it falls in the end. The difference between the lowest point and the highest point is 7 Hz.

---

79 Male 2 is the only speaker which differs from other three speakers in age and educational background; all other three speakers are in their early 20s and they are undergraduate students, while Male 2 is an illiterate in his late 40s, thus, perhaps the difference in pitch pattern.
In (b) in Male 2, unlike all others (see footnote '3'), the pitch starts with a rise at 115 Hz, rises steeply till the third pitch point and maintains a steady pattern for another seven pitch points before it gradually rises until 130 Hz, and then makes a steep fall in the end. The difference between the lowest point and the highest pitch point is 15 Hz.

In (c) in Female 1, pitch begins at a low 235 Hz, dips further (steeply) for four pitch points and maintains a (more or less) steady state for another six pitch points before it rises steeply to 252 Hz and ends with a fall. The difference between the lowest point and the highest point is 33 Hz (highest in the low register rise).
In (d) in Female 2, the pitch begins at a low 215 Hz, dips further to 205 Hz till the sixth pitch point and then makes a steep rise for another eleven pitch point until 228 Hz before it makes a steep fall in the end. The pitch difference between the lowest point and the highest pitch point is 23 Hz.

The general pitch pattern for the word ‘summer’ is then a rising one (we ignore the terminal fall; this fall may be a result of list intonation). Ignoring Male 2 (see footnote ‘3’), the pitch for this word begins with a low, rises steeply (at least in the two Female speakers) and then ends with a fall.

Thus, phonetically, the pitch pattern of the four words is significantly dynamic: high Register falls and low Register rises. The pitch slide differences (high to low for High Register and low to high for Low Register) are shown in table 5.1.
In Table 5.1, assuming that a difference of 5 Hz slides (either rise or fall) is significant (see footnote '4') our study shows that all the speakers' pitch slides is significant. The word 'quarrel' has a fall range of 27 to 60 Hz, 'east' has a fall range of 26 to 68 Hz, with an overall average of 42.75 Hz for both the words in the high Register. In the low Register the pitch rise is again significant; the word 'paddy' has a slide range of 7 to 24 Hz and 'east' has 7 to 33 Hz. The overall average slide difference for 'paddy' is 14.25 Hz and 19.5 Hz for 'summer'.

In general, contour tones in tone languages demonstrate a varying degree of variations in terms of the difference the onset and offset of the pitch contour. One the one hand, languages like Thai shows large differences between the onset and offset of contour tones. According to Abramson (1962), the falling tones in Thai show a fall of about 55 Hz, while low (falling) tones show a fall of about 10 Hz; the high rising tones show a rise of about 45 Hz and low (rising) tones show a rise of about 15 Hz. On the other hand, the differences between the onset and offset of contour tones in Taiwanese have rather substantially small differences. Peng (1997) notes that in Taiwanese tones the difference between onset and offset of rising tones can be as small as 10 Hz; while in Kammu (Svantesson & House 2006) the pitch difference between the two tones (high and low) is quite small, with the average ranging between 4 Hz.

---

*Footnote: We assume that a difference of 5 Hz between the start-point and the end-point of the contour makes perceptual significance. According to Hewlett & Beck (2006), within the frequency region of 1000 Hz, a difference of 2 to 3 Hz can be detected (thus perceptible), at least by trained listeners. However, fineness of discrimination depends on the absolute frequencies involved, as well as the interval between them. “So it is possible to detect a difference between 1000 Hz and 1003 Hz but not between 4000 Hz and 4003 Hz” Hewett & Beck (2006).*
to 25 Hz (for male speakers). Thus, the pitch differences between the onset and the offset of the contour tones differ from language to language, and Dzongkha contour average pitch range between onset and offset falls within the ranges attested in literature.

To sum up, we make the following general assumptions about the facts of pitch pattern in Dzongkha, in the within Register study:

- High Register syllables are characterised by a fall, and low Register syllables by a rise in pitch.
- On long syllables, pitch movements are (somewhat) flattened.
- The glottal stop in the coda ‘ʔ’ induces a fall in pitch.
- There is at most a two-way Register contrast, high versus low, on all rhyme types in Dzongkha (\_v, \_vv, \_vp, \_vʔ and \_vN[asals]).

We, then based on the acoustic study of the four speakers, conclude that the pitch patterns of the four words under study in Dzongkha suggest that the pitches are dynamic, and not level.

In the next section, we show evidence of the presence of a glottal stop in ‘quarrel’ and ‘paddy’ and illustrate the historical development of final ʔ.

### 5.4.2 Glottal stop /ʔ/ in Dzongkha

Earlier literature had posited the glottal stop as the property of tone; we posit it as a feature of the segment. Glottal stop in Dzongkha, however, occurs only in the coda (syllable-final along with /p/, /m/, /n/ and /ŋ/) position and not in the onset (word-initial) position. This is true of many other Tibeto-Burman languages in the region, for example, in Mizo (Lalrindiki 1992), Garo (Burling 1992), Ao (Temsunungsang 2009). The presence of glottal stop in the final position is illustrated in the waveforms and spectrograms of one of the female speakers in figure 5.3.
Figure 5.3: glottal stop in 'quarrel' and 'paddy'

a) $\text{ca}? \text{ 'quarrel'}$

b) $\text{ca}: \text{ 'east'}$

c) $\text{bza}? \text{ 'paddy'}$
In figure 5.3 (a) and (c), we make two observations, in the waveforms and spectrograms: 1) the presence of glottal stop /ʔ/ and, 2) the vowel is significantly shorter than their counterparts (we show the significance of vowel length in the next section). We also see that the quality of the vowel, especially towards the end (just before /ʔ/) has a glottalized quality, which is not apparent in ‘east’ and ‘summer’; the vowel rather ends smoothly in these two words.

Historically, we can then posit that all final obstruents (except the labial *b) got merged into a glottal stop; this is illustrated in the following data.

<table>
<thead>
<tr>
<th>WD</th>
<th>Pronunciation</th>
<th>Gloss</th>
</tr>
</thead>
<tbody>
<tr>
<td>*phag-pa</td>
<td>/phaʔ/</td>
<td>‘year (calendar)’</td>
</tr>
<tr>
<td>*tshag</td>
<td>/tshaʔ/</td>
<td>‘a punctuation (full-stop)’</td>
</tr>
<tr>
<td>*shog</td>
<td>/coʔ/</td>
<td>‘come’</td>
</tr>
<tr>
<td>*mdzod</td>
<td>/dzeʔ/</td>
<td>‘treasure, store’</td>
</tr>
<tr>
<td>*tshod</td>
<td>/tshoʔ/</td>
<td>‘disposition’</td>
</tr>
<tr>
<td>*sdod</td>
<td>/deʔ/</td>
<td>‘sit’</td>
</tr>
<tr>
<td>*shes</td>
<td>/ceʔ/</td>
<td>‘knows’</td>
</tr>
<tr>
<td>*'bras</td>
<td>/bzaʔ</td>
<td>‘paddy’</td>
</tr>
<tr>
<td>*rtsis</td>
<td>/tsiʔ/</td>
<td>‘account’</td>
</tr>
</tbody>
</table>
In the above data we show that all ancient final obstruents (*g, *d, *s) have merged into a glottal stop /h/ in the contemporary pronunciation of Dzongkha words. On the other hand, Dzongkha still preserves the ancient labial final (*b [devoiced to /p/]) and nasal finals (*m, *n, *ŋ) as shown in the following data.

<table>
<thead>
<tr>
<th>WD</th>
<th>Pronunciation</th>
<th>Gloss</th>
</tr>
</thead>
<tbody>
<tr>
<td>*skudb</td>
<td>/kyp/</td>
<td>'thread'</td>
</tr>
<tr>
<td>*khab</td>
<td>/khp/</td>
<td>'needle'</td>
</tr>
<tr>
<td>*gsum</td>
<td>/sum/</td>
<td>'three'</td>
</tr>
<tr>
<td>*khjim</td>
<td>/khim/</td>
<td>'house, home'</td>
</tr>
<tr>
<td>*sman</td>
<td>/mén/</td>
<td>'medicine'</td>
</tr>
<tr>
<td>*skjon</td>
<td>/kön/</td>
<td>'defect'</td>
</tr>
<tr>
<td>*shing</td>
<td>/ɕin/</td>
<td>'wood'</td>
</tr>
<tr>
<td>*chang</td>
<td>/tɕan/</td>
<td>'beer'</td>
</tr>
</tbody>
</table>

In the next section, we compare the vowel length significance in ‘quarrel’ versus ‘east’ and ‘paddy’ versus ‘summer’.

5.4.3 Vowel duration

In this section we compare the length of vowel in /caʔ/ ‘quarrel’, /caʃ/ ‘east’, /bzaʔ/ ‘paddy’ and /bzaʃ/ ‘summer’ and show that the vowel is ‘east’ and ‘summer’ are significantly longer than their counterparts. We measure the length of three tokens of the vowel in all the four words and their average values are presented in table 5.2 and figure 5.4 (length differences are shown in

---

81 Final /ŋ/ in Dzongkha is perhaps undergoing change; in some pronunciations the final /ŋ/ is deleted, leaving the vowel preceding it with a nasal quality. Thus, *shing may be pronounced as either /ɕi/ or /ɕin/; we presume, for now, that these forms are in free variation.
percentage; the percentage system normalises the individual length variations [refer table 5.2 for numeric differences]).

Table 5.2: Vowel length (msec) of 'quarrel', 'east', 'paddy' and 'summer'

<table>
<thead>
<tr>
<th></th>
<th>qa?</th>
<th>ca?</th>
<th>hapa?</th>
<th>hapa</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male 1</td>
<td>168</td>
<td>296</td>
<td>215</td>
<td>322</td>
</tr>
<tr>
<td>Male 2</td>
<td>186</td>
<td>338</td>
<td>208</td>
<td>372</td>
</tr>
<tr>
<td>Female 1</td>
<td>168</td>
<td>288</td>
<td>230</td>
<td>342</td>
</tr>
<tr>
<td>Female 2</td>
<td>165</td>
<td>291</td>
<td>210</td>
<td>304</td>
</tr>
<tr>
<td>Average</td>
<td>171.75</td>
<td>303.25</td>
<td>215.75</td>
<td>335</td>
</tr>
</tbody>
</table>

Figure 5.4: Vowel length difference

A number of comments can be made about figure 5.4:

- Vowel in 'east' is significantly longer than its counterpart 'quarrel' in the high Register (with an overall mean difference of 131 msec [12%]).
Vowel in ‘summer’ is also significantly longer than its counterpart ‘paddy’ in the low Register (with an overall mean difference of 120 msec [12%]).

The (overall) duration of vowel is slightly longer in the low Register words than in the high Register.

We show that the vowel in the open syllable words are significantly longer than the syllables closed by glottal stop, thus contrastive (which is again a segmental feature like the glottal stop). Short vowel occurs with glottal stop and long vowel with open syllable; however, there is vowel length contrasts in the syllables with final _p as in: /tsip/ ‘stone wall’ versus /tsi:p/ ‘astrologer’ and final nasals _N (in both the Register) as in: /sem/ ‘mind, soul’ versus /se:m/ ‘pea’; /dzim/ ‘tongs’ versus /dzim/ ‘eyebrows’ (pitch is not constrastive in these words).

In the next section, we show pitch height contrast with-in register.

5.4.4 Phonetic pitch height contrast within register

In this section, we study the pitch height differences within the Register. We compare the pitch height differences in both the Registers for three speakers of the same words we used in section 5.4.1. the Intra-Register pitch differences for the four words are presented in figure 5.5 (A) and (B).

---

82 We do not consider data of Female 1 because of the bad quality of the recordings.
Figure 5.5: within register pitch height difference

A. /əʔ/ 'quarrel (dark-diamond) vs. /əʔ/ 'east' (light-square)

In Male 1, the pitch height difference in the initial portion is 14 Hz (147–133 Hz) followed by a difference of 10 Hz (134–125 Hz) in the midpoint of the vowel and ends with a difference of 6 Hz (128–122 Hz). We may note that the pitch of ‘east’ maintains a slightly level pitch (almost) till the end. This is because the speaker is idiosyncratic in his pronunciation of the word, in that the vowel is slightly glottalized in all the tokens, which is not seen in other speakers.

In Male 2, in the initial portion there is a pitch difference of 32 Hz (192–160 Hz), followed by a difference of 23 Hz (165–142 Hz) in the midpoint of the vowel and ends with a difference of 17 Hz (130–113 Hz).
In Female 2, the two pitches start off with a height difference of 15 Hz (165–150 Hz), followed by a difference of 19 Hz (246–227 Hz) in the mid-point, and ends with a difference of 18 Hz (231 Hz – 213 Hz).

We, thus, see that in all the three speakers the pitch height difference is significant (recall footnote 6 for the significance of pitch height difference) in the words ‘quarrel’ and ‘east’ in the high Register. The pitch height differences are further presented in tabular form in table 5.3.

Table 5.3: Pitch height significance in the high Register

<table>
<thead>
<tr>
<th></th>
<th>Start-point</th>
<th>Mid-point</th>
<th>End-point</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male 1</td>
<td>14</td>
<td>yes</td>
<td>10</td>
</tr>
<tr>
<td>Male 2</td>
<td>32</td>
<td>yes</td>
<td>23</td>
</tr>
<tr>
<td>Female 2</td>
<td>15</td>
<td>yes</td>
<td>19</td>
</tr>
<tr>
<td>Average</td>
<td>20.33</td>
<td>17.33</td>
<td>13.66</td>
</tr>
</tbody>
</table>

In table 5.3, we show that the pitch height difference is significant in all the three speakers across the entire portion of the vowel (start-point, mid-point and end-point). We also note that the pitch height difference is greatest in the
start-point (overall average difference of 20.33 Hz) and least in the end-point (overall average difference of 13.66 Hz) with an intermediate difference in the mid-point (overall average difference of 17.33) in the high Register.

In the low register in figure 5.5 (B) too, we see a similar pattern in the pitch height differences.

In Male 1, the two pitches start off with a difference of 7 Hz (124–117 Hz); it differs by 6Hz (125–119 Hz) in the mid-point, followed by a difference of 10 Hz (126–116 Hz) in the end-point.
In male 2, the pitch height difference in the start-point is 14 Hz (129–115 Hz); it differs by 14 Hz (142–128 Hz) in the mid-point, followed by a difference of 16 Hz (128–112 Hz) in the end-point.

![Female 2](image)

In Female 2, the two pitches start off with a height difference of 15 Hz (265–250 Hz); it differs by 17 Hz (245 Hz–226 Hz) in the mid-point, and ends with a difference of 18 Hz (231–213 Hz).

In general, in all the speakers, the two pitches begin with a smaller difference in the start-point and the differences gradually increase as the two pitches travel in parallel through the vowel and terminate with the greatest difference. The pitch height difference in the low register is shown in a tabular form in Table 5.4.

<table>
<thead>
<tr>
<th></th>
<th>Start-point</th>
<th>Mid-point</th>
<th>End-point</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Male 1</strong></td>
<td>7</td>
<td>Yes</td>
<td>6</td>
</tr>
<tr>
<td><strong>Male 2</strong></td>
<td>14</td>
<td>Yes</td>
<td>14</td>
</tr>
<tr>
<td><strong>Female 2</strong></td>
<td>15</td>
<td>Yes</td>
<td>17</td>
</tr>
<tr>
<td><strong>Average</strong></td>
<td>12</td>
<td>12.33</td>
<td>14.66</td>
</tr>
</tbody>
</table>

Table 5.4: Pitch height significance in the low register
In table 5.4, we show that the pitch height difference for ‘paddy’ and ‘summer’ is significant in all the speakers; the overall average difference is 12Hz in the start-point, slightly widened to 12.33 Hz in the mid-point and terminates with a difference of 14.66 Hz in the end-point.

Thus, along with the Register contrast, Dzongkha also has an Intra-Register contrast (phonetic), which is generated by the syllable types. The closed syllables (CV?) heighten the pitch; while the open long syllables (CVV) lower the pitch in both the High and Low Register. A historical perspective on the development of the two syllable shapes: CV? and CVV, and thus the development of the phonetic Intra-Register contrast, is illustrated in the following section.

5.4.5 Historical perspective on the development of Intra-Register contrasts

In section 5.3.2, we illustrated how the initial consonant types and initial cluster reduction contributed to the high and low Register distinction in Dzongkha. Here we illustrate the contribution of ancient final consonants in the development of within Register distinction in Dzongkha. The Dzongkha Intra-Register contrast, which is phonetic, developed through the loss of syllable-final consonants; this is illustrated in (1) and (2) below. We name the four (phonetic) tones as: tone 1, 2, 3 and 4 (as we did in section 5.4.1).

<table>
<thead>
<tr>
<th>WD</th>
<th>Pronunciation</th>
<th>Gloss</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>*phag-pa</td>
<td>/phaʔ/ ‘pig (calendar)’</td>
</tr>
<tr>
<td></td>
<td>*phar</td>
<td>/phaː/ ‘there’</td>
</tr>
<tr>
<td></td>
<td>*shag</td>
<td>/caʔ/ ‘room’</td>
</tr>
<tr>
<td></td>
<td>*shar</td>
<td>/caː/ ‘east’</td>
</tr>
<tr>
<td></td>
<td>*shed</td>
<td>/ceʔ/ ‘full stop’</td>
</tr>
<tr>
<td></td>
<td>*bshel</td>
<td>/ceː/ ‘roam’</td>
</tr>
<tr>
<td></td>
<td>*bsreg</td>
<td>/seʔ/ ‘fry’</td>
</tr>
</tbody>
</table>
In (1) (for the High Register) and (2) (for the Low Register), we observe that the ancient syllables with final obstruents (*d, *g and *s) have merged into a glottal stop and consequently contributed to a phonetic tonal split, raising the pitch in both High and Low Register. The ancient syllables with final liquids (*l and *r) have disappeared, leaving its remnant in the length of the vowel; this contributed to a phonetic tonal split by lowering the pitch height in both the High and the Low Register.
In the next section, we make a phonological analysis of the tones in Dzongkha.

5.5 Phonological analysis of Dzongkha tones

In this section, we first present a short summary of the phonetic facts of tones in Dzongkha. Based on the phonetic facts, we present our phonological representation of the tones in Dzongkha.

5.5.1 Phonological interpretation of the phonetic tones in Dzongkha

Thus far, we established the following phonetic facts of Dzongkha tones.

1. Dzongkha has two distinct Register contrasts: High and Low Register, generated by initial phonation types; the ‘Tense’ (VLT* and VLAS) correlate with High and the ‘Lax’ (VLL and VD) with Low, in the obstruents; while voiced sonorants allow contrastive tonal specification.
2. Pitch pattern in both the Registers is a dynamic one; High Register words have a falling pitch, while Low Register a rising pitch.
3. There is also an Intra-Register contrast, generated by the syllable shapes (CV? and CVV); closed syllable (CV?) generates higher pitch than open long syllable (CVV) in both the High and Low Register.
4. Vowel duration is slightly longer in the Low Register than in the High Register.

Based on the phonetic characteristics of Dzongkha tones, we argue that both pitch range (height) differences and contour are important cues for discriminating tones. Hence, both averaged pitch and the direction of pitch are important in categorising tones in Dzongkha. The phonetic facts of Dzongkha tones can be interpreted phonologically as follows:
<table>
<thead>
<tr>
<th>Phonetic</th>
<th>Phonological</th>
</tr>
</thead>
<tbody>
<tr>
<td>High fall</td>
<td>High tone (Mid tone)</td>
</tr>
<tr>
<td>Low rise</td>
<td>Low tone (Mid tone)</td>
</tr>
</tbody>
</table>

The phonetic 'High fall' can be interpreted as High tone, with a default Mid tone, phonologically, and the phonetic 'Low rise' as Low tone, with a default Mid tone, phonologically. The difference in pitch range between the CV? and CVV syllables can be interpreted as phonetic implementation due to durational excess (syllable types). In Dzongkha, the tonal range can be divided into two pitch Registers, correlating with initial phonation types, i.e., 'Tense' (correlating with High) and 'Lax' (correlating with Low) phonation types in obstruents, while voiced sonorant onsets allow contrastive tones, with a default Mid tone at the right edge. Dzongkha is then an incipient, lexically bitonal language, supported by the lexical tone specification on the syllables with voiced sonorant onsets. The High in the Upper Register is a phonetic falling tone whereas the Low in the Lower Register is a phonetic rising tone. The assumption is that the grammar contains two tone melodies, High Mid and Low Mid, one of which is part of the representation of any given monosyllabic word. We can propose that every syllable has a phonological tone, and the inventory contrasts two tones: High and Low; both the tones are then phonologically active in Dzongkha. If we add to an underlying High vs. Low system the possibility of syllables being unspecified, and a default process that provides a Mid tone, then we have three phonetic tones (High, Mid and Low), but only two phonological tones (High and Low) as in Yoruba (Pullyblank 1986), where the Mid tone is the default, and there is a High vs. Low vs. Zero underlying system.

Looking at other Tibeto-Burman languages in the region, a similar tonal pattern is also reported in Bodo (Sarmah 2004) and, Rabha and Dimasa (Sarmah 2009). According to Sarmah, while Rabha and Dimasa have three lexical tones (rising, falling and level-mid), Bodo has two lexical tones (high and low) with a Mid default tone. Dzongkha tones are similar to tones in Bodo, except there is a difference of tonal association, in that while Bodo tones follow left-to-right tonal association (Sarmah 2004), the case is just the opposite in Dzongkha; it is
right-to-left. This is to be expected as tone is still largely tonogenetic in Dzongkha arising from the onset phonation types, unlike those languages where lexical tone is distinctive and totally unpredictable.

5.6 Conclusion and looking forward

In this chapter, we did an acoustic study of Dzongkha tones and, the acoustic study suggests that, in fact, there are only two phonological tones: High and Low, separated by a Register system. This, perhaps, confirms Mazaudon and Michailovsky’s (1988) speculation on the absence of Within Register contrasts in Paro dialect. Thus, while we agree with earlier literature in terms of the Register system of tonal analysis, we differ from the earlier literature in the number of tones and the tonal patterns. We also differ in terms of the tone shapes; in earlier literature, tone 1 and 3 were analysed as level tones and tone 2 and 4 as contour tones. However, our study showed that both the tones in the (Paro dialect of) Dzongkha are phonetically contour tones.

However, a lot is yet to be studied regarding segments and tones in Dzongkha. In chapter 3, we studied only the dental stops for our acoustic study; we need to study stops across the remaining three places (labial, retroflex and velar), and see if results conform to the dental stops. It would also be worthwhile comparing the acoustic correlates of stops across different places in Dzongkha, and see if the results are similar to the findings in other languages.

In chapter 4, we compared the sonorant phonation types of Dzongkha with two other Tibeto-Burman languages (Mizo and Tenyidie), and found that all the three languages have slightly different modes of employing acoustic cues for the voiceless sonorants. The voiceless nasals in Dzongkha and Mizo were found to be same (post-aspirated), while the voiceless lateral in Dzongkha patterned with voiceless lateral in Tenyidie (pre-aspirated), and not with Mizo as expected. It would be worthwhile to do a cross-linguistic study of sonorants in the Tibeto-Burman languages (with voiced-voiceless sonorant contrast) and
see if such pattern of acoustic correlates is attested in other Tibeto-Burman languages as well.

And finally, in this chapter, we extended our acoustic findings of segmental studies in chapter 3 and 4 in the study of tones, and analysed Dzongkha as a two toned language. In our Intra-Register study, we could attend to only limited data: short closed syllable (CV?) and long open syllable (CVV). The future study could extend this study to other types of syllables (CVN [nasal], CVp, CV). We feel (if our hunch is correct) that even within Paro dialect, there exists sub-dialectal variations in both segmental and tonal features (owing to the peculiar distribution of the geographical terrain); there is a need to carry out similar studies in different sub-dialects within Paro dialect. In this study we looked at only monosyllables; we need to study the tonal patterns in disyllabic words as well.

Thus, what we have achieved regarding segments and tones in Dzongkha in this study is a preliminary work, which would usher linguists into further research into the linguistic investigation of Dzongkha.
BIBLIOGRAPHY


Dutta, I. (2007). *Correlation between voicing lead time (VLT) and f0.* Proceedings of the 39th annual meeting of the Chicago Linguistics Society, pp. 405-422.


Kim, M. (2004). *Correlation between VOT and F0 in the perception of Korean stops and affricates*. Eighth International Conference on Spoken Language Processing.


