FOUR-WAY STOP CONTRASTS IN HINDI: AN ACOUSTIC STUDY OF VOICING, FUNDAMENTAL FREQUENCY AND SPECTRAL TILT

BY

INDRANIL DUTTA

B.A., Jawaharlal Nehru University, New Delhi, India, 1995
M.A., Jawaharlal Nehru University, New Delhi, India, 1997
M.A., University of Illinois at Urbana-Champaign, 2001

DISSENTATION

Submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Linguistics in the Graduate College of the University of Illinois at Urbana-Champaign, 2007

Urbana, Illinois
Abstract

In this dissertation, results from an acoustic phonetic study of the four stops types in Hindi: voiced stops (VS), voiced aspirated stops (VAS), voiceless stops (VLS) and voiceless aspirated stops (VLAS) are reported. The “Standard View” on the distinction between VS and VAS proposes that the voiced aspirated stops are VS with a breathy murmured release and this release feature is sufficient to make the contrast between the VS and VAS. Evidence from studies on the duration of voicing and effect of manner of articulation on the fundamental frequency ($f_0$) of the following vowel in Hindi questions the characterization proposed by the standard view. This study through an examination of durational properties of stop closure, voicing during closure and aspiration following these stops provides evidence against the standard view.

Both VAS and VS have been shown to lower $f_0$ of the following vowel. It has also been shown that the VAS lower $f_0$ even further. This evidence suggests that $f_0$ perturbations can be reliable acoustic cues for stop identification. The goal of this dissertation is to understand not only the magnitude of the $f_0$ perturbations but also the extent of this effect in the following vowel.

Spectral intensity analysis of contrasting breathy and modal vowels in Gujarati, !Xóö and languages which make use of the breathy and modal phonation type as contrastive features provide a background against which spectral analysis of the breathy/murmured release following VAS can be conducted to test the assumptions of the standard view. Spectral analysis based on four measures of spectral intensity of the vowel following the stops indicate that the breathiness following the VAS permeates into a sizeable portion of the vowel. Com-
parisons between durations of breathiness spread and voiceless aspiration also show that voiceless aspiration is shorter in duration than the duration of breathiness characterized by the difference in spectral intensity between the VAS and the unaspirated stops (VS,VLS).

Based on these analyses, I argue that the stop distinctions in Hindi are best understood as a cumulative effect of several acoustic cues, in contradistinction to previous accounts, including the standard view.
To Tulsi, my love and life...
First and foremost, my thanks and gratitude are due to Professor Hans Henrich Hock, my advisor, mentor, Doktorvater and above all an exceptional human being. This project and all of my research is due to him and his constant encouragement, advice, help and support. Hans’ wisdom and intellect is visible in whatever is exceptional in this thesis. Through the years, I learned to appreciate and attempted to emulate Hans’ keen eye for detail and transparency in making arguments. His exceptional intellect, knowledge and enthusiasm led me through the hardest parts of coming up with the research questions, the hypotheses and finally the analyses. Words are not enough to express the gratitude I feel for Hans’ constant support for my dissertation project and my general intellectual growth. Apart from being an exceptional advisor, mentor, source of strength, Hans exhibits a passion for linguistics that is rare and infectious, as much as his sense of humour. I am indebted to him for pointing out so many directions and yet leaving me to choose the way I wanted to go.

Words can also not express entirely, my gratitude for my dissertation committee. Professors Jennifer Cole, José Ignacio Hualde, and Chilin Shih are an extremely supportive group of scholars whose constant encouragement and insight shaped the dissertation. Professor Cole’s keen interest in my dissertation project, her support and advice led me through all of my graduate career. Her valuable insights on formulating relevant research questions and passion for experimentally addressing such questions helped me understand the need for empiricism in linguistics. Right from the very beginning of my graduate career she has been a guide and the rigour she embodies in scientific research has been nothing but a major source of inspiration. Professor Hualde has been a major source of encouragement and knowledge,
especially when it came to understanding the crucial relationship between diachronic and synchronic processes in languages. His unparalleled scholarship, good cheer and sense of humour is worthy of emulation. Professor Shih’s support for my project came by way of her outstanding scholarship and insight on experimental research. Her constant and continuous encouragement, and her enthusiasm for finding a lot more than meets the eye from the data helped shape the observations and results in this work. I am truly indebted to my committee of scholars for their support, encouragement and advice and couldn’t have embarked on this project without their crucial input.

Thanks are due also to my numerous teachers at the Department of Linguistics. Abbas Benmamoun, Eyamba Bokamba, Chin-Chuan Cheng Georgia Green, Molly Mack, Jerry Morgan, Chin-Woo Kim, Daniel Silverman and James Yoon are responsible for the well-rounded education in linguistics that I received here at Illinois. Thanks are also due to the staff of FLB, especially, Mary Ellen Fryer and Pat Gallagher for their passionate commitment that makes life for graduate students easy and stress free.

I am extremely thankful to Parvinder Kaur, Principal and Team Leader of Katha Khazana School for being so helpful during my trip to the field. Her contribution to this thesis is immense, without her help and support, I wouldn’t have been able to collect my data. I am also indebted to the staff of Katha Khazana for making my days at the school comfortable and also to the students for providing general amusement. I am extremely thankful to my subject pool for volunteering to take part in my project without any need for encouragement.

Over the years, the encouragement and good times shared with my friends in Urbana-Champaign have had nothing but an salubrious effect on my general well-being. Countless conversations later, I am glad to say that the following friends have had an extremely positive role to play towards my social and professional life. So, thank you Emin Adas, Somnath Baidya Roy, Himika Bhattacharya, Shefali Chandra, Guha Dharmarajan, Serife Genis, Michael Koliska, Sarasij Majumder, Kelly Maynard, Deepti Misri, Aniruddha Mitra, Mark Nye, Amit Prasad, Srirupa Prasad, Celiany Rivera-Velázquez, Shilpi Sarkar, Anna Schultz,
Debarati Sen and Sunita Singh.

For wonderful times shared with my colleagues at Nuance Communications in Belgium, who quickly became the best friends one can hope for, Duygu Can, Benjamin Freitag, Benedicte Frostad, Elisa Paloni, Elena Sacau, Renaat Seys, Wang-Ju Tsai and Chingwen Tseng, I thank you from the bottom of my heart. Thanks are also due to the folks behind the bar at Manteca, the best watering hole on the other side of the big river.

I am very grateful to my parents, Ratna and Pradip Dutta, for instilling in me a belief in my abilities and for providing the necessary foundation for pursuing life with honesty, courage and strength. I am also very thankful to Geeta and K. Dharmarajan, my parents-in-law, for their support and encouragement. I also want to acknowledge the love and affection of Amamma, Dadi, CIT Thatha and Delhi Thatha, my grand-parents-in-law. From them, I learned the valuable lesson of living with extreme compassion.

My affection and thanks are also reserved for Tashi and Zuri, my adorable puppies, for sharing with me their undying passion for food and a good time. Their love translated into numerous licks made it easy to go through the toughest patches of the dissertation process.

And finally, my thanks, love and gratitude for Tulsi. Her joie de vivre, infectious laugh, lust for life, selfless and caring nature have filled me with a sense of well-being that is unprecedented. Tulsi’s passionate existence is worthy of an entire volume by itself. She has supported me through thick and thin, and there isn’t a word in my vocabulary that can encapsulate my love and affection for her. She is the sole reason that I have been able to complete this project and also the sole reason that I can continue to work towards goals that far exceed my abilities. To her, I dedicate this work with warmth and love.

Quite obviously there are several individuals, entities and institutions who are also responsible for all that is good in this dissertation. As always, the sole responsibility for regrettable omissions and errors rests with me alone.
## Chapter 4  Effects of stop type on fundamental frequency

4.1 Universal tendency for \( f_0 \) lowering following voiced stops .......................... 56
4.2 \( f_0 \) as a function of stop type ........................................ 57
  4.2.1 Subject GA .......................................................... 59
  4.2.2 Subject PB .......................................................... 61
  4.2.3 Subject RM .......................................................... 62
  4.2.4 Subject SD .......................................................... 65
  4.2.5 Subject SV .......................................................... 68
4.3 Discussion ................................................................. 70
4.4 Effect of context on \( f_0 \) ................................................. 72
4.5 Summary and implications ............................................. 72

## Chapter 5  Aspiration and vowel duration

5.1 Outline ................................................................. 74
5.2 Durational properties: Aspiration duration .......................... 75
  5.2.1 Subject GA and PB .................................................. 75
  5.2.2 Subject RM, SD and SV ......................................... 75
5.3 Durational properties: Vowel durations ............................... 77
5.4 Summary of results .................................................... 79

## Chapter 6  Spectral properties of Hindi stops

6.1 Breathiness due to incomplete glottal closure ....................... 81
6.2 The relevance of measures of spectral intensity .................... 82
6.3 Spectral intensity measures .......................................... 83
  6.3.1 \( H_1-H_2 \) ......................................................... 83
  6.3.2 \( H_1-A_1 \) ......................................................... 89
  6.3.3 \( H_1-A_2 \) ......................................................... 94
  6.3.4 \( H_1-A_3 \) ......................................................... 99
6.4 Contributions of the individual spectral intensity measures .... 105
6.5 Summary, discussion and conclusion ................................ 110

## Chapter 7  Conclusions, implications and further research

7.1 Overview ............................................................. 111
7.2 Implications and further research .................................. 113

References .............................................................. 123

Author’s Biography ..................................................... 127
List of Tables

1.1 Sanskrit phonetic and phonological characterization of voice and breath . . . 4
1.2 Hindi stops: Manner and place of articulation . . . . . . . . . . . . . . . . . 5
1.3 Average -VOT values . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 12

3.1 Effect of context on VLT: Post-hoc comparison
based on SNK and Tukey HSD . . . . . . . . . . . . . . . . . . . . . . . . . 50
3.2 Effect of place of articulation on VLT: Post-hoc comparison based on SNK
and Tukey HSD . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 53

4.1 Ordered relations between the stops from 10 to 50 percent of the vowel based
on Tukey HSD comparisons . . . . . . . . . . . . . . . . . . . . . . . . . . . 71

6.1 Maximally distinct distributions and contributing factors towards a distinc-
tion in vowel quality. Means of values at 10 and 30 percent of the vowel. . . 107
6.2 Maximally distinct distributions and contributing factors towards a distinc-
tion in vowel quality. Means of values at
10, 30 and 50 percent of the vowel. . . . . . . . . . . . . . . . . . . . . . . . 107
List of Figures

1.1 Axis of contrast: Voicing and aspiration ........................................ 6
1.2 Laryngeal feature schematic ............................................................. 7
1.3 Lower VLT and f₀ values for voiced aspirated/breathy voiced stops (Speaker HG) ............................................................... 15

2.1 Closure Duration (CD) measurements: Panel A, auditory evaluation confirms long pause between [laːl] and [kʰaːl]. Panel B, CD was annotated between [laːl] and [tʰaːl] .................................................. 26
2.2 VLT measure: From cessation of final [l] in [laːl] to appearance of burst in [gaːl] .................................................................................. 27
2.3 Aspiration duration measure: From burst till appearance of sinusoidal wave-form, [tʰaːl]. Points at which f₀ measures were taken for VLAS are also shown in Tier 2 .......................................................... 27
2.4 Percentage duration of vowel for spectral measurements .................. 30
2.5 Spectral measures: Spectrum, LTAS and LPC analysis ....................... 32

3.1 Effect of place of articulation (POA) on Closure Duration (ms) ............. 37
3.2 Effect of stop type on Closure Duration (ms) ........................................ 38
3.3 Effect of Context on Closure Duration ................................................ 39
3.4 Effect of Stop type on Closure Duration: Excluding velar stops .......... 40
3.5 Effect of Stop type on Closure Duration (with pause): Utterance medial [s] condition; dental POA ......................................................... 43
3.6 Speaker GA: ANOVA results for dependent variable VLT and independent variables, stop type (Panel A), place of articulation (Panel B), context (Panel C), Horizontal lines in boxes (Panel D) represent median values and Summary of effects and interactions (Panel E) .................................................. 45
3.7 Speaker PB: ANOVA results for dependent variable VLT and independent variables, stop type (Panel A), place of articulation (Panel B), context (Panel C), Horizontal lines in boxes (Panel D) represent median values and Summary of effects and interactions (Panel E) .................................................. 46
3.8 Speaker RM: ANOVA results for dependent variable VLT and independent variables, stop type (Panel A), place of articulation (Panel B), context (Panel C), Horizontal lines in boxes (Panel D) represent median values and Summary of effects and interactions (Panel E) .................................................. 47
3.9 Speaker SD: ANOVA results for dependent variable VLT and independent variables, stop type (Panel A), place of articulation (Panel B), context (Panel C). Horizontal lines in boxes (Panel D) represent median values and Summary of effects and interactions (Panel E). ........................................... 48

3.10 Speaker SV: ANOVA results for dependent variable VLT and independent variables, stop type (Panel A), place of articulation (Panel B), context (Panel C). Horizontal lines in boxes (Panel D) represent median values and Summary of effects and interactions (Panel E). ........................................... 49

3.11 Effect of place of articulation on VLT ........................................... 52

4.1 Effect of stop type on $f_0$ (Subject GA): Mean $f_0$ plots ....................... 60

4.2 Effect of stop type on $f_0$ (Subject GA): Variation in $f_0$ values. Horizontal lines within the boxes represent the median $f_0$ values. ........................................... 61

4.3 Effect of stop type on $f_0$ (Subject PB): Mean $f_0$ plots ....................... 63

4.4 Effect of stop type on $f_0$ (Subject PB): Variation in $f_0$ values. Horizontal lines within the boxes represent the median $f_0$ values. ........................................... 64

4.5 Effect of stop type on $f_0$ (Subject RM): Mean $f_0$ plots ....................... 65

4.6 Effect of stop type on $f_0$ (Subject RM): Variation in $f_0$ values. Horizontal lines within the boxes represent the median $f_0$ values. ........................................... 66

4.7 Effect of stop type on $f_0$ (Subject SD): Mean $f_0$ plots ....................... 67

4.8 Effect of stop type on $f_0$ (Subject SD): Variation in $f_0$ values. Horizontal lines within the boxes represent the median $f_0$ values. ........................................... 68

4.9 Effect of stop type on $f_0$ (Subject SV): Mean $f_0$ plots ....................... 69

4.10 Effect of stop type on $f_0$ (Subject SV): Variation in $f_0$ values. Horizontal lines within the boxes represent the median $f_0$ values. ........................................... 70

5.1 Effect of place of articulation (POA) on Aspiration Duration (ms) ............. 76

5.2 Effect of Context on Aspiration Duration (ms) ..................................... 77

5.3 Vowel duration as a function of stop manner ....................................... 78

5.4 Box plot: Percentage of aspiration duration relative to Vowel Duration $Total$ . 79

6.1 Effect of Stop Type on $H_1-H_2$ .................................................... 87

6.2 Box plots show variation in $H_1-H_2$ values and overlapping. .................. 88

6.3 Effect of Stop Type on $H_1-A_1$ .................................................... 92

6.4 Box plots show variation in $H_1-A_1$ values. ..................................... 93

6.5 Effect of Stop Type on $H_1-A_2$ .................................................... 97

6.6 Box plots show variation in $H_1-A_2$ values. ..................................... 98

6.7 Effect of Stop Type on $H_1-A_3$ .................................................... 102

6.8 Box plots show variation in $H_1-A_3$ values. ..................................... 103

6.9 Box plot: Percentage of aspiration duration relative to Vowel Duration $Total$. 105

6.10 Box plot: Effect of place of articulation on $H_1-A_1$. .......................... 106

6.11 Mean spectral intensity for values at 10 and 30 percent of the vowel ........... 108

6.12 Mean spectral intensity for values at 10, 30 and 50 percent of the vowel .... 109

xii
List of Abbreviations

CD  Closure Duration
f₀  Fundamental Frequency
H₁-A₁ Difference between the amplitude of the First Harmonic and the peak amplitude of the first formant
H₁-A₂ Difference between the amplitude of the First Harmonic and the peak amplitude of the second formant
H₁-A₃ Difference between the amplitude of the First Harmonic and the peak amplitude of the third formant
H₁-H₂ Difference between the amplitude of the First Harmonic and Second Harmonic
VAS Voiced Aspirated Stops
VLAS Voiceless Aspirated Stops
VLS Voiceless Stops
VS Voiced Stops
VLT Voice Lead Time
VOT Voice Onset Time
Chapter 1

Introduction

śvāso 'ghoṣāpām
itareśāṁ tu nādāḥ
soṣmoṣmaṇāṁ ghoṣināṁ śvāsanādau

‘Breath’ is emitted for the voiceless sounds but for the others ‘voice’, for the voiced fricative (h) and the voiced aspirates, both breath and voice.

Rk-Prātiṣākhya xiii.4-6

1.1 Aims of this dissertation

This dissertation is an acoustic study of the four stop types in Hindi: voiced stops (VS), voiced aspirated stops (VAS), voiceless stops (VLS) and voiceless aspirated stops (VLAS). The “Standard View” on the distinction between VS and VAS proposes that the voiced aspirated stops are VS with a breathy murmured release and this release feature is sufficient to make the contrast between the VS and VAS (Ladefoged and Maddieson 1996, Dixit 1987a). Evidence from studies on the duration of voicing and effect of manner of articulation on the fundamental frequency (f₀) of the following vowel in Hindi questions the characterization proposed by the standard view. This study through an examination of durational properties of stop closure, voicing during closure and aspiration following these stops provides evidence against the standard view.

Both VAS and VS have been shown to lower f₀ of the following vowel. It has also been
shown that the VAS lower $f_0$ even further. This evidence suggests that $f_0$ perturbations can be reliable acoustic cues for stop identification. The goal of this dissertation is to understand not only the magnitude of the $f_0$ perturbations but also the extent of this effect in the following vowel.

Spectral intensity analyses of contrasting breathy and modal vowels in Gujarati, !Xóô and other languages which make use of the breathy and modal phonation type as contrastive features provide a background against which spectral analysis of the breathy / murmured release following VAS can be conducted to test the assumptions of the standard view. Spectral analyses based on four measures of spectral intensity of the vowel following the stops indicate that the breathiness following the VAS permeates into a sizeable portion of the vowel. Comparisons between durations of breathiness spread and voiceless aspiration also show that voiceless aspiration is shorter in duration than the duration of breathiness characterized by the difference in spectral intensity between the VAS and the unaspirated stops (VS,VLS).

Several studies have shown that prosodic context has an effect on durational properties (Cho and Jun 2000, Cho and McQueen 2005/4, Cole, Kim, Choi and Hasegawa-Johnson 2007). An additional goal of this dissertation is to examine the effect of prosodic context on the durational properties of the four stops in Hindi.

Based on these analyses, I will argue that the stop distinctions in Hindi are best understood as a cumulative effect of several acoustic cues, in contradistinction to previous accounts, including the standard view.¹ Before discussing the motivations for conducting this study, I will provide a brief historical account of the views of the Sanskrit grammarians on the issue of stop contrasts in Sanskrit. This introduction will help situate the arguments that have been made so far, both in the phonetic and phonological literature.

¹Keyser and Stevens (2006) discuss the relevance of acoustic cues that are involved in enhancing contrasts. In chapter 6, I discuss the relevance of a theory of enhancement as proposed in Keyser and Stevens (2006) for a comprehensive understanding of cue interaction in Hindi.
### 1.2 Sanskrit grammarians on voicing and aspiration

Predating the advent of modern phonetic sciences, Sanskrit phoneticians and grammarians understood the complex articulatory dynamics behind the production of voiced aspiration.² In fact one of the most important contributions of the Sanskrit grammarians was identification of the distinction between voiced and voiceless sounds. The Prātiṣākhya clearly states the distinction between sounds that are produced with a closed throat, *ghoṣa* (voice) and those that are produced with an open throat as in a simple breath (voiceless), *śvāsa* (Allen 1953). These lead to further characterization of *madhye hakāraḥ* ‘an intermediate glottal state’ which has both voice and voicelessness (see quote above) *Ibid*, p. 9. This state, where the glottis is understood to produce both ‘breath’ and ‘voice’, was essential in understanding not only the voiced aspirates but also voiced [h]. Most Sanskrit grammarians made the distinction between ‘voice’ and ‘breath’ to be a distinction between closed and open ‘throats’, respectively. William D. Whitney, however, expressed concerns with regards to the characterization of voiced aspirates and voiced [h] by the Sanskrit grammarians in the Prātiṣākhya in the following way:

“The Rk-Prātiṣākhya (xiii.2, r. 6) declares both breath and sound to be present in the sonant aspirates and in *h*, which could not possibly be true of the latter, unless it were composed, like the former, of two separate parts, a sonant and a surd: and this is impossible.” Whitney (1860-1863): 348

Aspiration produced with voice is inconceivable according to Whitney. Despite Whitney’s view, it is clear that the Sanskrit grammarians recognized a clear distinction, first between voiced and voiceless sounds and second between the phonological status of aspiration and the phonetic implementation of ‘both voice and breath’; relevant for the voiced aspirates and the voiced ‘h’. Thus, early on, Sanskrit phoneticians were able to make a distinction between

²In this study, I will reserve the terms ‘voiced aspiration’ to refer to the phonological category and ‘breathy’/‘murmur’ as labels for the phonetic manifestation of the category.
phonological categories and phonetic phenomenon. In Table 1.1, I provide a summary of this distinction.

Table 1.1: Sanskrit phonetic and phonological characterization of voice and breath

<table>
<thead>
<tr>
<th>Phonology</th>
<th>ghośin (=voiced)</th>
<th>d</th>
<th>sośmaṇ (=aspirate)</th>
<th>dh</th>
</tr>
</thead>
<tbody>
<tr>
<td>aghośa (=voiceless)</td>
<td>t</td>
<td>tʰ</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Phonetics</td>
<td>nāda, (+)</td>
<td>+</td>
<td></td>
<td>±</td>
</tr>
<tr>
<td></td>
<td>śvāsa, (-)</td>
<td>-</td>
<td></td>
<td>-</td>
</tr>
</tbody>
</table>

While ghośin (voiced), aghośa (voiceless) and sośmaṇ (aspirate) are used as purely phonological terms, nāda and śvāsa are terms reserved for expressing the phonetic properties of sounds. A four-way stop contrast in Sanskrit, as in Hindi, is asymmetrically addressed in the phonetics, but in the phonology, a completely symmetrical account of the contrast is found. Thus, firstly the contribution of the Sanskrit grammarians begins with the recognition that voicing and voicelessness are phonetically distinct. And secondly, in identifying that the phonetic manifestation of voiced aspiration could be both voiced and voiceless in order to explain the phonological and categorical distinction between the Sanskrit voiceless aspirates and voiced aspirates.

Phonological treatments of aspiration as a feature in this respect and breathiness/murmur as phonetic manifestation of this aspiration contrast find currency even in contemporary analyses of laryngeal phonology and Hindi phonology in particular. In section 1.3, I will provide a brief summary of the nature of the laryngeal stop contrasts in Hindi followed by a discussion of the various feature-based accounts that attempt to express these contrasts through various phonological features. Following this, I will outline the arguments and implications of the standard view on the phonetics of voiced aspiration. In section 1.5, the problems associated with the standard view are discussed. Section 1.6 provides a brief introduction to the findings from previous research that address some of these problems. In
In this section, I will also outline the questions that remain unanswered. Section 1.7 outlines the research questions that will be addressed through this experimental study and provides a brief organization of the following chapters.

1.3 Feature based accounts of the phonology of stop types in Hindi

A large number of modern Indo-Aryan languages exhibit a four-way laryngeal contrast in their stops. While Thai exhibits a three-way contrast, namely, voiceless (VLS), voiceless aspirated (VLAS) and voiced (VS), Hindi possesses a fourth stop type: voiced aspirated stop (VAS). The Hindi stops occur in four places of articulation, and thus can be represented in a 4x4 array (Table 1.2).

Table 1.2: Hindi stops: Manner and place of articulation

<table>
<thead>
<tr>
<th></th>
<th>Labial</th>
<th>Dental</th>
<th>Retroflex</th>
<th>Velar</th>
</tr>
</thead>
<tbody>
<tr>
<td>Voice</td>
<td>b, b^h</td>
<td>d, d^h</td>
<td>d, d^h</td>
<td>g, g^h</td>
</tr>
<tr>
<td>Aspir</td>
<td>p, p^h</td>
<td>t, t^h</td>
<td>t, t^h</td>
<td>k, k^h</td>
</tr>
</tbody>
</table>

The four-way stop contrast in Hindi is conventionally understood to be a laryngeal contrast expressed in two dimensions - voicing and aspiration (Fig. 1.1). This leads to four distinct categories of stops. While voicing requires the vocal folds to be approximated and vibrating, aspiration requires the vocal folds to be abducted. Under this view, a two-dimensional representation establishes unambiguously three of the four stop categories, namely, the voiceless, voiced and voiceless aspirated stops, while failing to address the articulatory mismatch between voicing and aspiration in VAS. VAS, then, pose a difficulty for

---

3 The UPSID survey of languages shows that there are ten languages from six different language families that exhibit a four-way laryngeal contrast.

4 In this array the palatals have not been presented, since they are affricated.
a binary featural analysis of this contrast, in that both voicing and aspiration need to be simultaneously implemented at the release of the stop closure.

Figure 1.1: Axis of contrast: Voicing and aspiration

The focus of research on laryngeal phonology in general and voiced aspiration in particular has not deviated much from the twin poles of simultaneous aspiration and voicing, when characterizing voiced aspirates. Feature based theories of laryngeal phonology have attempted to identify the segmental features that make up the voiced aspirates. These features, be they binary, underspecified or privative, have attempted to express the category of voiced aspirate through an interaction of voicing and aspiration (Avery and Idsardi 2001).

One of the earliest treatments is found in Halle and Stevens (1971), who propose four laryngeal features, [spread glottis](sg), [constricted glottis](cg), [stiff vocal cords] and [slack vocal cords]. In this model, voiced aspirates are characterized as [+sg] and [+slack]. For a recent and complete overview of the feature based accounts see Avery and Idsardi (2001) and Fig. 1.2 below.

As correctly argued by Keating (1988) and Lombardi (1994) these features pose two major problems: First the representation of voicing in this model is not articulatorily accurate, in that not all voiced sounds are produced with slack vocal chords and neither are all voiceless sounds made with stiffening of the vocal cords. Second, the model lacks a single feature that
groups together all voiced sounds. The second problem is crucial in the case of Hindi.

VOT theory (Lisker and Abramson 1964, Abramson and Lisker 1967, Abramson 1977, Poon and Mateer 1985) attempts to remedy these problems by expressing the contrasts in terms of differences in Voice Onset Time (VOT). A major contribution of VOT theory was in identifying the phonetic correlates of voicing. Under this theory, Voice Onset Time (VOT) is a phonetic manifestation of underlying differences in laryngeal timing in relation to the release of the oral occlusion. Lombardi (1994) is correct in stating that VOT theory is insufficient in accounting for the voiced aspirates. In fact, Lisker and Abramson (1964) recognize the difficulty in establishing a distinction between voiced stops and voiced aspirated stops simply on the basis of VOT. The difficulty is brought about by the fact that both voiced aspirated stops and voiced stops are produced with lead voicing or -VOT. These facts make VOT inadequate for making a distinction between these two stop types.

The inadequacy of VOT theory in characterizing the voiced aspirated stops was already pointed out by Schiefer (1986). Lombardi (1994) argues that based on phonetic and phonological data the voiced aspirates in Hindi and other languages are both voiced and aspirated. The bulk of the phonetic evidence that forms the basis for the argument of Lombardi (1994) comes from Dixit (1989), Yadav (1984) and Ingemann and Yadav (1978). Ladefoged (1971)
presents the view that the voiced aspirated stops can not be considered aspirated because the ‘breathy’, ‘murmured’ release of these stops is distinct from voiceless aspiration.

Thus, previous research on the position of voiced aspirates essentially deal with two hypotheses:

- **Hypothesis 1**: Voiced aspiration results from overlap of two independent gestures, voicing and aspiration
- **Hypothesis 2**: Voiced aspiration represents an independent mode of phonation

Research on acoustic phonetics of Hindi voiced aspiration has addressed both of these hypotheses, presenting data that provide evidence in favour of hypothesis 1 (Schiefer 1992) and hypothesis 2 (Ladefoged 1975). Schiefer (1992) finds that listener’s judgements of voiced stops improve for those stimuli that have relatively longer durations of Voicing Lead Time (VLT)\(^5\). Based on the findings from Schiefer (1986), Schiefer (1992) proposes a feature “Lead onset time” with values 1 and 2 for voiced aspirated stops and voiced stops respectively, to account for the difference in preference for VLT. As rightly pointed out by Selkirk (1992), Schiefer’s motivation to posit the feature “Lead onset time” was in effect a way of shifting away from the position that murmured stops represent a third possible state of the vocal chords (Ladefoged 1971). Additionally, width of glottal opening for voiced aspirates is also an important gesture (Hirose 1977, Dixit and MacNeilage 1980, Benguerel and Bhatia 1980). Benguerel and Bhatia (1980) show very different laryngeal timing and opening properties for the voiced aspirated stops as compared to the other stops. Thus, not only do both voiced aspirated stops and voiced stops show closure voicing, the width of glottal opening during the closure for voiced aspirated stops is significantly different from that of the voiced stops.

Thus, in summary, the categorical nature of the four-way contrast expressed along the dimensions of voicing and aspiration have been the primary focus of feature-based accounts of

---

\(^5\)VLT refers to the period of voicing during closure
this contrast. While phonological accounts have attempted to characterize the asymmetrical nature of voicing and aspiration in voiced aspirated stops, phonetic accounts have attempted to provide evidence in favour of either hypothesis 1 or hypothesis 2. Schiefer (1989) argues in favour of Hypothesis 1 and provides evidence to support a gesture-based account, while (Ladefoged 1971) argues in favour of Hypothesis 2 by providing evidence from languages that make contrasts between ‘breathy’, ‘murmured’ and modal vowels. As far as the phonetics of voiced aspirates is concerned, I will consider arguments in favour of Hypothesis 2 to represent the “Standard View” on voiced aspiration. In the following, I will present in greater detail the arguments that form the standard view.

1.4 The standard view on stop types in Hindi

The standard view on the phonetics of Hindi VAS has focused primarily on the post release aspiration phase (Ladefoged and Maddieson 1996, Dixit 1987a). Ladefoged and Maddieson (1996) state that:

“breathy voiced stops in Hindi and many other Indic languages are acoustically distinguished from the plain voiced stops by what happens after the release rather than audible differences during the closure”.

Dixit (1987a) and Dixit (1987b) show that voicing during the closure in VAS is comparable to modal voicing and only at the release of these stops there is turbulent high rate of air flow through the glottis. VOT based theories also assume that the ‘low amplitude buzz’ associated with the release of VAS serves to distinguish the VAS from the VS (Abramson and Lisker 1967, Lisker and Abramson 1964, Abramson 1977). This forms the bulk of the standard view on distinctions between VS and VAS. The standard view also assumes that VAS are VS with breathy / murmured release.

According to Ladefoged (1971) and Ladefoged and Maddieson (1996), the breathy / murmured release following the VAS represents a third state of laryngeal setting to be contrasted with voiceless and modally voiced stops. Arguments have also been made on
the basis of articulatory studies which suggest that only at the release of VAS there is significant difference in glottal states between VS and VAS; while the VS continue to be voiced (modally) the VAS are an admixture of intermittent modal voicing and aspiration at release (Dixit 1982, Dixit 1987a, Dixit 1987b). In addition, VOT has also been used as a measure to distinguish between the four stops. VOT refers to the timing of the onset of vocal fold vibration with reference to the release of a consonant. Abramson and Lisker (1967), Abramson (1977), and Lisker and Abramson (1964) show that in a great number of languages, including Hindi, stops could be discriminated on the basis of VOT. A VOT based theory of stop discrimination, however, is only able to distinguish between voiced stops (VS, VAS), VLS and VLAS based on differences in the timing of the onset of vocal fold vibration. Voiced stops (VS, VAS) are said to have negative VOT values due to the presence of voicing during closure. VLS are considered to be short-lag stops and VLAS stops are considered long-lag stops. The inadequacy of a VOT based distinction between VAS and VS was recognized by the authors of VOT theory. Consider the following quote from Lisker and Abramson (1964):

"The two four-category languages, Hindi and Marathi, present us with our only clear cut cases in which the measure VOT is insufficient for distinguishing among all stop categories of a language. To be sure, the voiced unaspirated and voiced aspirated stops show differences in average values that are almost systematic; nevertheless they occupy ranges that are nearly coextensive. It seems very likely that voiced aspirates are distinguished from the other voiced category by the presence of low amplitude buzz mixed with noise in the interval following the release of the stop." Lisker and Abramson (1964): 403.

VOT based analysis of Nepali, which also exhibits a four-way laryngeal contrast like Hindi, by Poon and Mateer (1985) shows results similar to those of Lisker and Abramson (1964), Abramson and Lisker (1967), and Abramson (1977). Following the standard view,
Davis (1994) shows that Noise Offset Time (NOT) is sufficient to make a distinction between VS and VAS, with breathy release serving to enhance the contrast. Noise Offset Time (NOT) as measured by Davis (1994) is an amended measurement of Lag Time or +VOT. She measures NOT as the time between stop release and appearance of the $F_2$ on the vowel, which signals the cessation of aspiration noise.

The crucial difference between these arguments arises because of the assumptions that - a) breathy / murmured release is a categorically distinct glottal state (Ladefoged 1971, Ladefoged and Maddieson 1996); b) the VS and VAS are both voiced in the closure and thus there is no need to propose a third glottal state (Dixit 1982, Dixit 1987a, Dixit 1987b) or c) VOT differences between VS and VAS are not significant (Lisker and Abramson 1964, Abramson and Lisker 1967, Abramson 1977).

A major assumption of the standard view therefore is:

- The VAS can be characterized as VS with a breathy / murmured release.

This assumption that VAS is VS with a breathy / murmured release has two major implications in relation to the phonetic characterization of the stops in Hindi:

1. VAS and VS are both voiced in the closure and the difference between them arises in the release; this release is characterized by simultaneous voicing and aspiration, which results in a breathy / murmured release.⁶

2. Voicing during closure (VLT) and/or duration of closure (CD) represent non-distinctive differences.

In summary, the major assumption of the standard view implies: (1) the duration of the occlusion or Closure Duration (CD) is not a significant cue for contrasting between the four

---

⁶The terms ‘breathy’ and ‘murmur’ appear interchangeably in the literature on the phonetic manifestation of voicing and aspiration contrast. Research on contrastive non-modal phonation differences in vowels, however, exclusively maintains the term ‘breathy’ to contrast these vowels from modal vowels.
stops in Hindi. (2) the voicing during the stop closure (VLT) is not relevant in maintaining contrast between the voiced stops. (3) the VAS and VS are distinguished only by the breathy / murmured release portion which is associated with the VAS.

1.5 Problems with the standard view

Implications from the assumptions of the standard view notwithstanding, Schiefer (1986) shows that listeners prefer longer VLT durations for unambiguous perception of VS. Similar results for VLT durations (see Table 1.3 below) are also found in the raw data of Lisker and Abramson (1964) and Davis (1994). However, the relevance of VLT towards making a contrast between the VAS and VS does not form the primary focus of these studies.

<table>
<thead>
<tr>
<th></th>
<th>VAS</th>
<th>VS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lisker and Abramson (1964)</td>
<td>-65</td>
<td>-89</td>
</tr>
<tr>
<td>Davis (1994)</td>
<td>-75.22</td>
<td>-89.30</td>
</tr>
</tbody>
</table>

Table 1.3: Average -VOT values

In an acoustic phonetic study, Dutta (2007) argues that there is a significant difference in VLT between VAS and VS in Hindi. VS are produced with significantly longer VLT durations than VS. In this study of initial voiced stops of Hindi, however, the target words were embedded in an utterance medial position such that the preceding environment was voiced, making the closure and voicing durations coterminous. Therefore, from Dutta (2007) the relevance of CD could not be ascertained. These two studies, however, do provide evidence against the assumption that VAS are VS with a breathy / murmured release and the differences in VLT are not significant towards maintaining a contrast between these two stops. The relevance of CD, however, remains to be studied.

Further, voicing is universally expected to lower $f_0$ of the following vowel (House and Fairbanks 1953, Hombert, Ohala and Ewan 1979). Studies have also shown that $f_0$ of the following vowel is reduced in both VAS and VS (Ohala 1979, Schiefer 1986). These two
studies also show that \( f_0 \) following the VAS is lower than the \( f_0 \) following the VS. The VLT durational differences between VAS and VS suggest that VAS are phonetically “less” voiced than the VS. The \( f_0 \) patterns from the VAS and VS however, do not reflect this fact. Contrary to the universal tendency the less voiced VAS reduces \( f_0 \) more than the VS.

Purcell, Villegas and Young (1978) in a study of Panjabi tones suggest that the allotonic low ‘tonal’ characteristic of initial Hindi VAS correlates with the Panjabi phonemic low-rising tone. Aside from the correlation of \( f_0 \), this study shows that Hindi VAS lower \( f_0 \) in the following vowel. However, a limitation of the study by Purcell et al. (1978) was that it did not take into consideration durational aspects of voicing in Hindi, i.e. VLT measurements. They also did not compare the effect of all stop types on \( f_0 \). Schiefer’s study also shows that VAS in Hindi show low ‘tonal’ characteristics.

Both these studies allow us to make the observation that Hindi VAS lower \( f_0 \) even further than the VS. These observations raise important questions about the nature of voicing in VAS and its relation to \( f_0 \) lowering.\(^7\)

The standard assumption that VAS and VS are contrasted by the post-release breathy / murmured portion following VAS and modal phonation following VS has not yet been shown to be an accurate assumption based on acoustic properties (but see Bali (1999) for a study of spectral tilt in Delhi Hindi medial stops). Furthermore, it is unclear from previous studies if indeed voiced and voiceless aspiration can be differentiated not only through spectral differences but also differences in the duration of the aspiration phase for VLAS and the duration of breathy / murmured release following VAS. In the following section, I will outline research findings that address the issue of VLT, \( f_0 \) lowering, and breathy / murmured release following VAS.

\(^7\)Hock (1986/1991) provides an explanation for the greater tonal distinctiveness in Panjabi by suggesting that a process of polarization leads to an enhanced tonal distinction following the loss of aspiration in Panjabi.
1.6 Studies on the effect of stop type on VLT and $f_0$

In Dutta (2007), I have addressed the question of the distinction in voicing durations during closure and the lowering of $f_0$ following VS and its further lowering following VAS. However, in this study, only VLT was studied for VS and VAS. Thus a complete cross comparison between the durations of the occlusions for all the four stop types could not be established. The results from the study show that VAS are produced with shorter durations of VLT than VS (Fig. 1.3 below). Results from the analysis of $f_0$ in Dutta (2007) also show lower $f_0$ values for the VAS. However, this study looked only into the acoustic properties of the voiced stops (VS, VAS) and not the voiceless stops (VLS, VLAS), hence it was not possible to ascertain the relevance of CD and VLT independently. Additionally, the effect of the stops on the initial $f_0$ of the vowel was measured as the average $f_0$ of the first six pitch periods. Methodologically, measuring the $f_0$ over six pitch periods, was a replication of the method for measuring initial $f_0$ perturbations in Schiefer (1986). Thus, it was unclear from the results in Schiefer (1986) and Dutta (2007) what the full extent of the $f_0$ perturbation is in the larger domain of the following vowel. While measuring the average $f_0$ in the first six pitch periods provides an account of the absolute differences in $f_0$ between the VAS and VS, it fails to establish in relative terms the extent of the effect of the stop type on the $f_0$.\(^8\)

Details of absolute and relative $f_0$ measurements that were used in this study are discussed in section 2.3.3.

Another observation that was made in Dutta (2007) was that the breathy portion following the VAS tended to permeate into a sizeable portion of the vowel. This observation was however not instrumentally verified. A mean six-cycle $f_0$ measurement did not adequately correlate in duration with the actual spread of the breathiness. Further, in terms of the phonetic implementation of the aspiration contrast, it would also be interesting to compare and correlate the aspiration durations in VLAS and the breathy portion in VAS.

\(^8\)Purcell et al. (1978) measured $f_0$ in the following vowel at 10 consecutive points in the vowel. See section 2.3.3 for further details.
Spectral intensity analysis by Bali (1999) shows that in Delhi Hindi, intervocalic VAS may be produced without aspiration and that voiced aspirated tokens are produced with a larger open quotient, wider first formant bandwidth and a steeper spectral tilt, while the voiced plosives show the reverse glottal configurations. This goes to show that for discrimination between intervocalic voiced aspirated and voiced stops, spectral intensity measures may be especially relevant as also for distinguishing breathy vowels from modal ones as in Gujarati following Fischer-Jorgensen (1967). However, at the moment there are no studies on the glottal characteristics of word initial voiced aspirated stops in Hindi. Further, the nature and extent of the breathy release following the VAS is not clearly understood.

Figure 1.3: Lower VLT and $f_0$ values for voiced aspirated/breathy voiced stops (Speaker HG)

Further, Ohala (1979), Schiefer (1986), and Dutta (2007) measured $f_0$ in the initial portion of the following vowel and not in the entire vowel (see Purcell et al. (1978) for measurement of $f_0$ contours in the entire vowel). In the case of Schiefer (1986) and Dutta (2007) only the mean $f_0$ of the first six pitch periods was measured. While this particular method provides a fairly accurate account of the effect of stop type on the initial $f_0$ perturbation, it somewhat limits a complete understanding of the nature of the effect of the stops on the $f_0$ of the following vowel. As has been observed in Dutta (2007), the breathy release following the VAS tends to permeate deeper into the following vowel. In order to arrive at a comparison
between the spread of breathiness and lowering of $f_0$, a new method of $f_0$ measurement is proposed in this study (details in section 2.3.3). $f_0$ is measured at 10 equidistant points following the release of the stop. Two major goals are achieved by following this method. First, this method allows us to normalize the $f_0$ contour over variable vowel durations by taking proportional measurements of $f_0$ values and second, the ten $f_0$ values represent the $f_0$ contour in a way that proportional comparisons could be made between stops. In addition, this method also allows us to compare the effects of $f_0$ lowering with the spread of breathiness in the vowel following the VAS. In this respect, this method while facilitating a study of the effect on the initial portion of the vowel also makes it possible to compare normalized $f_0$ contours. Methodologically, this measurement technique is somewhat similar to the one followed by Purcell et al. (1978).

Based on this discussion of previous research, in the following section, I will outline three primary acoustic issues that are key to understanding the nature of the four-way laryngeal contrast in Hindi.

### 1.7 Research questions and outline

The discussion in the previous sections, from the phonetic observations of the Sanskrit grammarians to the body of research on the four-way stop contrast in Hindi allow us to formulate several research questions. As mentioned in section 1.6 above, three outstanding acoustic issues need to be addressed in order to arrive at a comprehensive understanding of the nature of the stop contrasts in Hindi.

First is the issue of VLT and CD. Studies on VLT (Schiefer 1992, Dutta 2007) have shown that VAS are produced with lower VLT than the VS. Previous research on the perception of voicing had also shown that listeners require longer VLT durations for unambiguous perception of VS (Schiefer 1992, Schiefer 1986). One of the primary research questions for this study was to investigate whether a parallel could be found in the duration of closure
(CD) in VLS and VLAS and secondly, it would be necessary to examine the primacy of CD or VLT as features that help in cueing the contrast between unaspirated and aspirated stops. In other words, it needs to be seen whether voicing during closure for voiced stops is an epiphenomenon with the primary contrastive feature being the duration of closure rather than the attendant voicing. In Chapter 3 of this dissertation, I will show that indeed such is the case. VLS stops are produced with longer CD as compared to VLAS. Comparison of the complementary distribution of CD and VLT also leads to the expectation that CD is a relevant feature in cueing the contrast between the aspirated and unaspirated stops. In section 3.5, I address the nature of the complementary distribution of CD and VLT and argue that CD is a primary feature in cueing a contrast between the aspirated and unaspirated stops, with voicing serving to enhance the contrast between the aspirated and unaspirated stops. Here, I also discuss the relevance of a theory of enhancement in possibly explaining the secondary relevance of voicing during closure for cueing the contrast between voiceless and voiced stops.

Second is the issue of f0 lowering following voiced stops. Studies on the f0 in the vowel following the stops have shown the VAS to lower f0 even further than the VS (Ohala 1979, Schiefer 1986, Dutta 2007, Purcell et al. 1978). However, an investigation of the extent of this f0 lowering in the following vowel needs to be conducted to fully understand the effects of all the four stop types on the f0 and not just the voiced stops. In chapter 4 of this dissertation, I address this particular question and show that f0 in VAS is lower till 20-30 percent of the following vowel.

Third is the issue of breathy / murmured release following the VAS. The standard view on the VAS is that the VAS are VS that are produced with a breathy / murmured release. In this respect, one of the goals of this dissertation is to examine the nature and extent of this breathy release. Chapter 6 of this dissertation is a study of four spectral intensity measures that address the question of the nature and extent of the breathy release following VAS. The results from the spectral measures indicate that the reliability of specific measures is
speaker dependent. However, indirect measure of Open Quotient (OQ) is the most reliable measure that helps distinguish between the VAS and the unaspirated stops.

In addition to these three key problems, in this dissertation, I will also look into the duration of aspiration in VLAS, in an attempt to answer the question: What are the differences in duration of aspiration, if any, between VLAS and VAS? In chapter 5, this question is addressed through a discussion of the duration of voiceless aspiration, while chapter 6 discusses the spread of breathiness following the VAS based on the spectral properties of the following vowel.

Much of the previous research on VLT/CD is based on data elicited through carrier sentences. Recent studies have shown that in spontaneous and controlled speech data a consistent effect of prosodic/positional context is found on segmental/durational features (Cho and Jun 2000, Cho and McQueen 2005/4, Cole et al. 2007). Considering the complex articulatory configurations involved in producing the four-way system of stop contrasts in Hindi, it would be of immense interest to compare the prosodic effects of cue strengthening in Hindi with those reported in other languages. Prosodic position is a constant source of variation in not only $f_0$ features, but also durational and spectral features. In this dissertation, I will attempt to look at the scope of this variation and hope to relate the findings with those of Cho and Jun (2000), Cole et al. (2007) amongst others. Barring Shih, Möbius and Narasimhan (1999), who show that preceding context has a stronger effect on segmental features than does the following context, previous studies have overlooked the impact of contextual position. In chapter 3 of this dissertation, I examine the effect of varying prosodic context on VLT and CD. The general consensus on the effect of prosodic context on segmental features is that prosodically stronger positions tend to strengthen segmental features. In the current study the manifestation of prosodically conditioned strengthening is found in the lengthening of VLT/CD durations in the the absolute utterance initial position. All of the above mentioned questions have been addressed through an acoustic phonetic study.

The methodological details of this study follow in chapter 2. In chapter 3, results from the
study of the durational features, VLT and CD are presented. Following that, in chapter 4, results from the study of the effect of stop type on f0 perturbations are presented. Chapter 5 presents the results from aspiration and vowel duration. In chapter 6 results from the effects of stop type on the spectral features of the following vowel are presented. Chapter 7 is a detailed discussion of the conclusions from the studies and the implication of the results from these studies towards a unified account of the contrast between the four stops in Hindi.
Chapter 2
Experimental methodology

2.1 Introduction

In this chapter details of the experimental methodology used to address the issues raised in chapter 1 are presented. In section 2.2, details of the experimental design are presented, followed by a discussion of the methods of acoustic analysis in section 2.3. In section 2.4, various statistical tests and procedures that were employed to test the hypotheses are discussed.

2.2 Experimental design

2.2.1 Material

Recorded material consists of Hindi word-initial voiced aspirated (VAS), voiced (VS), voiceless (VLS) and voiceless aspirated (VLAS) stops in frame sentences (see Appendix 1, for the frame sentences). Frame sentences were of three types, representing three prosodic contexts/positions for the target word:

- Utterance initial [U]
- Phrase initial [I]
- Phrase medial [M]
The phrase medial [M] context consisted of two sub-types, one in which the word preceding the target word ended in [l] and another in which the word preceding the target word ended in [s]. These contexts allow us to not only study the effect of the phrase medial context on the segmental/durational features but also to examine the effect of a preceding voiced and voiceless environment on the voicing of the stops. Additionally, the context with a preceding [s] also allows us to measure closure duration for the stops. All the target words were of the type CVC, where the initial C was the target stop. The vowel following the target stop was an /aː/. The consonants following this vowel were [k, g, t, d, p, b, r, l, ŋ] (see details in Appendix 2). The material was presented to the subjects on index cards and they were asked to read the sentences at a natural speaking pace. Each subject was given a short 5 minute pause between the required 3 recording periods.

2.2.2 Subjects

Five native Hindi speakers were recorded in a quiet environment in New Delhi, India. Fifteen potential subjects were asked to fill out a Language Background Questionnaire (Appendix 3). The LBQ was designed to ascertain the linguistic background, exposure, level of contact, and proficiency in other languages of the subjects. Five speakers were then selected from these fifteen potential subjects based on their level of exposure to other languages. Potential subjects who reported in the LBQ that they had substantial knowledge or contact with Panjabi were not selected for the study. Although a substantial number of Hindi speakers in the vicinity of Delhi also speak Panjabi, they tend to consider themselves to be native Hindi speakers. Since Panjabi does not exhibit the same four way stop contrast as does Hindi (initial VAS in Hindi correspond to VLS with a low rising tone in Panjabi) only subjects who reported Hindi to be their native language and also that of their parents were selected.

1One word for the retroflex voiceless aspirated stop was of the type CVCV since a suitable lexical item with CVC structure is absent from the lexicon.
2.2.3 Recording

All of the recordings were made using a head-mounted AKG C 420 III pp microphone. All of the recordings were made onto a TASCAM DA P1 DAT recorder. The recordings took place in the music room of the Katha Khazana school located at Govindpuri in New Delhi, India. These recordings were then digitized at the Department of Linguistics phonetics laboratory using Computerized Speech Lab (CSL) at 22050Hz.

2.3 Acoustic analysis

Acoustic analysis of the data was done using Praat (version 4.4). The acoustic analysis of the digitized data involved segmentation, annotation and measurement of the relevant acoustic cues. These acoustic cues included:

1. Durational/temporal cues such as Closure Duration (CD), Voicing Lead Time (VLT), aspiration duration and vowel duration (see section 2.3.1 and 2.3.2 for details).

2. Temporally normalized f0 contours (see section 2.3.3 for details)

3. Four measures of spectral tilt (see section 2.3.4 for details)

The segmentation procedures involved identifying the beginning and end of segmental features such as oral closure, closure voicing, duration of aspiration for VLAS and vowel duration. Following the segmentation of the data, segmented portions of the acoustic signal were annotated to identify the segments. These annotations included identifying the place of articulation, stop type and context of the target word. Measurement of the segmented portions was conducted by employing automated Praat scripts.

---

2I am deeply indebted to the Principal of Katha Khazana school for allowing the participants in this study to take time off from their busy teaching schedules and I am grateful for the help and cooperation that was extended to me by all of the participants.

3©1992-2005 by Paul Boersma and David Weenink
Inter/intra-speaker and token variation in \( f_0 \) contours was normalized by taking 10 measures of \( f_0 \) starting at 10 percent of the vowel and ending at 100 percent of the vowel in 10 percent vowel duration increments.\(^4\) In Dutta (2007) \( f_0 \) was measured as an average of the \( f_0 \) in the first six pitch periods following Schiefer (1986). The method used in Dutta (2007) and Schiefer (1986) provides a good estimation of the absolute \( f_0 \) perturbation in the initial six cycles, without providing a complete understanding of the nature of \( f_0 \) perturbation throughout the entire contour in the vowel. Dutta (2007) and Schiefer (1986) provide methods with which segmental effects on initial \( f_0 \) can be measured, however, in the current study \( f_0 \) measurement at 10 consecutive points provides for a normalization of the \( f_0 \) contour over the entire duration of the vowel. The normalization procedure is necessitated by the fact that in this study generalizations about the \( f_0 \) contour and shape are being made over several different prosodic, segmental and vocalic conditions. A similar approach was followed by Purcell et al. (1978), where they measured the \( f_0 \) in the following vowel at 10 consecutive points following the burst so as to provide a relative measure of \( f_0 \) change over the entire stretch of the vowel, starting at 0 percent of the vowel. Due to unstable nature of \( f_0 \) in the periods following the burst, in this study \( f_0 \) perturbations were studied from 10 percent of the vowel and onwards till 100 percent.

The vowel duration over which the spectral measures were taken was also normalized by taking the measurements for the four spectral measures at five points in the vowel, starting at 10 percent and ending at 90 percent of the vowel in 20 percent vowel duration increments. Detailed procedures are presented in section 2.3.4.

\(^{4}\)Variation in the production could be attributed to several causes. Inter / intra speaker variation could be a result of speech rate and token frequency of the target words. In order to be able to make generalizations across speakers and also to minimize the effects of variation in vowel length, normalization of \( f_0 \) was conducted.
2.3.1 Closure duration and VLT

Closure Duration (CD) was measured for the phrase medial [M] and phrase initial [I] conditions for the voiceless stops (VLS and VLAS) and VLT was measured for the voiced stops (VS and VAS). The beginning of the occlusion in the Utterance Initial [U] condition can not be segmented in the acoustics, therefore CD was measured for the voiceless stops only in the phrase medial [M] and phrase initial [I] conditions where the segmental context preceding the target words provide clues towards segmenting the beginning of the occlusion in the target words. In the phrase medial condition with preceding [s] the cessation of frication was considered to be the onset of the CD in cases where no audible pause could be detected. However, a sizeable number of utterances produced by the speakers exhibited pauses between the preceding context and the target word. The pauses could have been caused by two factors. First, a number of nonce words were used in this study to complete a full set of CVC words. Speakers being unfamiliar with these nonce words tended to pause between the context preceding the target word and the target word itself. Second, speakers in some cases paused between the preceding word and the target word due to hesitation, in effect imposing an intonation pattern reserved for use in uttering words in isolation. Auditory judgement on the appropriateness of these utterances was made for these cases. Similar auditory judgements were used for the phrase medial [M] condition with preceding [l] and phrase Initial [I] condition with preceding [e]. Utterances where auditory evaluation led us to confirm that a pausal break existed between the target words and their preceding environments where not annotated with a CD.

In order to confirm that the auditory judgements were robust in identifying naturally intonated sentences without a pausal break, a subset of the data that was not annotated with CD was also examined. This subset included dental stops that occurred in the phrase medial [M] condition with a preceding [s]. The cessation of the frication in the preceding context to the target word provided an adequate environment where the noise in the frication
of [s] contrasted with the absence of signal in the closure and the pausal break. Beginning of the CD was then annotated from the end of the frication in [s] till the visible burst of the target word. This relatively long duration of CD could include the duration of closure and pausal break.

CD annotations are exemplified in 2.1, panels A and B. As can be seen in Panel A in the figure below, V represents the duration of the vowel and the domain over which vowel duration and f0 measurements were taken and H represents the annotated duration over which spectral measures were taken. Tiers 3 and 4 were used to annotate the lexical item and the prosodic context in which the lexical item appeared.

Voicing Lead Time (VLT) measurements involved measuring closure voicing from the initiation to cessation before burst as shown in Fig. 2.2.

### 2.3.2 Aspiration and vowel duration

The duration of aspiration was measured for voiceless aspirated stops (VLAS) from the offset of the burst till the onset of voicing as seen in Fig. 2.3. In most cases this coincided with the appearance of the second formant. However, the cessation of voiceless aspiration was seen in most cases also to coincide with the appearance of sinusoidal waveforms. Therefore, the boundaries of the annotated duration of aspiration were marked by careful observation of both the waveform and the spectrogram. The offset of the burst and the onset of voicing signaled the onset of the vowel for VS, VAS, and VLS, while the cessation of aspiration and onset of voicing signaled the onset of the vowel for VLAS. For VLS, VS, and VAS, vowel duration was segmented from the offset of the burst to the offset of the vowel. For VLAS, the annotation procedure involved segmenting the aspiration portion following the burst and annotating the beginning of sinusoidal waveforms following the aspiration at the beginning of the vowel. The total vowel duration for the VLAS, therefore, included the aspiration duration as well. This procedure was followed so as to be able to examine the relative duration of aspiration as compared to the total vowel duration.

25
Figure 2.1: Closure Duration (CD) measurements: Panel A, auditory evaluation confirms long pause between [lɔːl] and [kʰala]. Panel B, CD was annotated between [lɔːl] and [tʰəla].
Figure 2.2: VLT measure: From cessation of final [l] in [laːl] to appearance of burst in [gaːd]

Figure 2.3: Aspiration duration measure: From burst till appearance of sinusoidal waveform, [tʰaːɾ]. Points at which f₀ measures were taken for VLAS are also shown in Tier 2.
2.3.3  $f_0$ analysis

As discussed in section 1.6 above, effect of stops on the $f_0$ of the following vowel was studied on the initial portion of the vowel following the stop burst (Schiefer 1986, Dutta 2007). Methodologically, both of these studies were similar in measuring the average $f_0$ in the first six pitch periods. This method, while being relatively accurate is not able to demonstrate the full extent of the $f_0$ perturbation which is affected by the stop manner. Further, as stated above, one of the goals of this study was to examine the effect of the stop manner on the $f_0$ curve itself. Therefore, for the purposes of this study the $f_0$ analysis method involved taking $f_0$ measures following the target stops at ten equidistant points after the release of the stop in the following vowel. These ten points correspond to 10, 20, 30, 40, 50, 60, 70, 80, 90 and 100 percent duration of the vowel. This method allows for the possibility of representing the $f_0$ contour as a function of 10 consecutive $f_0$ values and also provides a way of normalizing the $f_0$ contour over varying vowel durations.\(^5\) Additionally, as mentioned in section 1.6, this method allows an indirect comparison of the extent of $f_0$ lowering with the extent of the breathiness spread following the VAS.

While the beginning of the measurements for VS, VAS, and VLS coincided with the cessation of the burst, in the case of VLAS, $f_0$ was measured from the offset of the aspiration and onset of periodic vibrations in the waveform. The rationale behind following this method is the absence of voicing during the aspiration duration in VLAS, which makes it impossible to measure $f_0$ in the aspiration portion.

2.3.4  Spectral measurements

Ideally, glottal airflow measures are taken using direct measures of airflow. However, since articulatory experiments on the glottal airflow could not be conducted, acoustic measures have been used in this study to arrive at indirect measures of estimation of the glottal

\(^5\)In this respect, the method applied here is similar to the method used by Purcell et al. (1978).
characteristics of the stops in Hindi. Spectral intensity measures based on inverse filtering of the acoustic signal provide an indirect method of studying the effect of the glottal source on the acoustic signal (Bickley 1982, Fischer-Jorgensen 1967, Ladefoged and Antonanzas-Barroso 1985). Differences between the amplitudes of $H_1$, $H_2$, $A_1$, $A_2$, and $A_3$ have been shown to correlate with differences in phonation type within vowels (Hanson 1995, Hanson, Stevens, Kuo, Chen and Slifka 2001, Wayland 1998). In this study four measures of spectral tilt were taken that included difference between the amplitudes of $H_1-H_2$, $H_1-A_1$, $H_1-A_2$ and $H_1-A_3$. These spectral intensity measures together provide a measure of spectral tilt.

Due to the high energy in the harmonic component for VAS, it is expected that at least till 20-30 percent of the vowel the difference between $H_1-H_2$ will be sufficiently high when compared to the VS and the VLS. This measure of spectral tilt is also indirectly a measure of Open Quotient (OQ). OQ refers to the ratio between the opening and closing gestures of the vocal folds. in the case of VAS, at least in the initial portion after the stop release, the vocal folds are open for a longer duration than they are closed, leading to a higher OQ.

$H_1-A_1$ refers to an estimation of the first formant bandwidth. in the case of VAS, it is expected that the first formant bandwidth measures based on the spectral intensity measure, $H_1-A_1$, will be higher than the other stops (Bali 1999).

$H_1-A_2$ provides an estimation of the skewness of the glottal source. Typically the opening phase of the vocal folds tend to be longer than the closing phase (Ni Chasaide and Gobl 1997) and acoustic consequence of this is seen in a boosting effect of the lower harmonics if the shape of the glottal pulse is more symmetrical.

$H_1-A_3$ is an acoustic measure of overall spectral tilt. Due to the higher energy in the lower harmonics and dropping off of energy in the higher formants for the VAS, it is expected that in the initial portion of the vowel following the VAS, the overall spectral tilt represented by difference between $H_1$ and $A_3$ should be be fairly large in comparison with the other stops. VLAS stops are aspirated between 10 and 30 percent of the vowel. This implies that there is no fundamental/harmonic component present during this portion of aspiration. Due to
this reason, comparison between the stops on the basis of the spectral tilt measures was restricted to the VLS, VS, and VAS. All of the spectral measures used in this study are direct measurements of spectral intensity differences between the fundamental, the second harmonic, and the amplitudes of the first, second and third formant peaks. Therefore, only comparisons between the VAS and the unaspirated stops are made for all the four spectral measures.

The measurement points for these acoustics measures coincided with 10, 30, 50, 70, and 90 percent duration of the vowel as seen in Fig. 2.4 below. The spectrum was calculated over a 30 ms period. Since the average duration of vowels varied between 200-300 ms, spectral measures were taken over 5 equidistant points rather than 10 points as in the f_0 measurements. Spectral measurements taken at 10 equidistant points with a spectral window of 30 ms would provide values from overlapping consecutive portions of the vowel. In order to get measures that were from non-overlapping portions of the vowel, spectral measurements were taken from 5 equidistant points in the vowel.

![Figure 2.4: Percentage duration of vowel for spectral measurements](image-url)
The measurements were taken with a window length of 0.005 ms for the spectrogram, the F1, F2, and F3 references points were based on the formant values for vowel [a] and these formant values were adjusted for male and female speakers. A spectral window of 30 ms was used for each point where measurements were taken from the spectrum. Spectral peaks were measured by comparing the spectrum, Long Term Average Spectrum (LTAS) and the LPC spectra as shown in Fig. 2.5 below\textsuperscript{6}.

The bottom panel of Fig. 2.5 also shows the peaks at which the amplitude measures were taken. The data was collected by deploying a Praat script, originally developed by Bert Remijsen for capturing spectral moments of steady state vowels. The script was modified to take measures at the five equidistant points. The five values for each of the acoustically determined spectral measures provide an estimation of the glottal characteristics and are perhaps not as accurate as actual articulatory airflow methods, nonetheless, these acoustic measures have been fairly well correlated with articulatory causes (Hanson et al. 2001, Stevens and Hanson 1994).

### 2.4 Statistical analyses

One-way ANOVAs were done to test the significance of independent factors such as Stop Type, Context, Place of Articulation, and Subject. Post-hoc Tukey tests were conducted to help categorize the data into subsets. Multivariate tests provided the results for significant interactions between the various independent variables and also the effect of each of the independent variables on the dependent variables. Post-hoc tests were used to confirm the effect of the independent variables on the outcomes of the dependent variables and also to categorize the patterns into distinct subsets.

---

\textsuperscript{6}LTAS is a representation of the spectral density as a function of frequency, expressed in dB/Hz.
Figure 2.5: Spectral measures: Spectrum, LTAS and LPC analysis
Chapter 3

Closure and voicing in Hindi

3.1 Outline

The goal of this study is to examine the temporal/durational aspects of the stops in Hindi. Studies on VLT (Schiefer 1992, Dutta 2007) have shown that VAS are produced with lower VLT than the VS. One of the primary research questions for this study was to investigate whether a parallel could be found in the duration of closure (CD) in VLS and VLAS. If indeed it is the case that CD and VLT durations show a complementary distribution, then a claim could be made towards the primacy of CD as a defining and contrastive feature. Evidence from this study suggests that VLS are produced with longer CD compared to VLAS. As has been mentioned in section 1.7 of chapter 1, the relevance of VLT and CD need to be looked at comparatively. In this chapter, I will show that VLT and CD in Hindi stops, in their durations, are correlated in such a way that VLT for VS and CD for VLS stops tend to be longer than the VLT in VAS and the CD in VLAS. In the following sections I will discuss the results from this study that show a systematic relationship between CD and VLT.

Closure durations are affected by both the stop type and place of articulation. In the case of VLT, it has been shown that stop type has an effect on VLT durations, with VS showing longer VLT durations than VAS. Further, CD measures for VLS and VLAS stops show that VLS stops have longer CD than VLAS. These observations suggest that in Hindi voiced stops (VS, VAS) and voiceless stops (VLS, VLAS) the duration of voicing during closure and the duration of closure, respectively, show similar patterns. Unaspirated stops
tend to have longer CD and in comparison aspirated stops tend to have shorter CD. These findings suggest that CD shows a parallel or complementary distribution with respect to the stops being aspirated or unaspirated.

These results from CD/VLT provide evidence against the standard view on contrasts in Hindi, as discussed in section 1.4 of chapter 1. The results from the VLT durations in conjunction with the results from CD show that aspirated stops have comparatively shorter closures compared to the unaspirated stops. This observation also provides evidence against the standard assumption that the voicing during closure and closure in and off itself are non-distinctive. Based on the evidence from VLT and CD and Schiefer’s observation that listeners prefer longer VLT durations for the unambiguous perception of VS, it can be concluded that closure durations could play a contrastive role in making the four-way distinctions in Hindi.

This chapter is organized as follows: In section 3.2, I discuss the results from closure durations in VLS and VLAS. Following this, in section 3.3, I present results from the investigation of those tokens that were produced with a pause between the target word and the preceding context. Details of the method employed to measure the CD with pauses is discussed in section 2.3.1. In section 3.4, I present results from the VLT durations for VS and VAS. In section 3.5, I correlate the findings from the CD and VLT durations and propose that duration of closure serves as a contrastive feature. Section 3.6 is a summary of the relevant results and conclusions that can be drawn based on the results.

### 3.2 Durational properties: Closure duration (CD)

As described in Chapter 2, data for this experimental study was collected from three prosodic conditions, namely, utterance initial [U] position, phrase initial [I] and phrase medial position [M]. Additionally, in the phrase medial condition the target words where modified by adjectives, [lə:l] and [kʰə:s]. This environment allows us to contrast a voiceless preceding environment with a voiced environment. Closure duration (CD) data was collected for
voiceless and voiceless aspirated stops. Since closure in voiceless stops cannot be measured at utterance initial position from the speech waveform it was collected from two prosodic contexts, phrase initial [I] and phrase medial [M].

Subject GA was unable to produce a labial voiceless aspirated stop. Synchronically, Hindi is undergoing a change where the canonical labial VLAS is merging with [f], the voiceless labio-dental fricative. This subject reflected the change and only results from dental, retroflex and velar voiceless stops are reported here. Additionally, this subject produced all of the VLAS stops in the phrase medial[s][M] context with considerable pause between the modifier [kʰa:s] and the target words. Due to this, CD measurements in this context could not be annotated and measured.

A univariate ANOVA with CD as the dependent variable and place of articulation (POA), prosodic context and stop type (VS, VLAS) as factors shows no effect of POA on the CD. An F-test reveals that F(2,15)=2.373 has a p=0.188. POA was subsequently removed as a factor from the model. Prosodic context did not have a significant effect on CD, pairwise comparisons, however did reveal a significant difference between the Phrase Medial ([s][M] and [l][M]) and Phrase Initial [I] context (p=0.029 and p=0.046, respectively). Noting that only CD from VLS stops could be measured for the Phrase Medial [s][M], CD for VLS stops was significantly longer in this context than in the other two contexts. There was no significant difference in CD between the Phrase Medial [l][M] and Phrase Initial [I] context. Significant differences in CD obtained between the stop types (VLS and VLAS). VLS had longer closures than VLAS. An F-test reveals that F(1,15)=34.720 is significant at p=0.002. While POA had no effect on the CD, a marginal effect of prosodic context was found in addition to the longer duration of closure for VLS compared to the VLAS.

\footnote{Durational measurements are expressed throughout this study in ms}

\footnote{Only 10 tokens of VLS and 5 of VLAS were measured for this subject, owing to long pausal breaks between the preceding word and the target word. These results, therefore, reflect the low number of tokens measured.}
Subject PB showed a significant effect of POA on CD, $F(3, 145)=26.923$, $p<0.005$. This effect could be attributed to the significant difference in marginal means that obtained between the velar and the three other places of articulation (labial, dental and retroflex, $p<0.005$). The velar stops being produced with significantly short CD. There was no significant difference in CD between the labial, dental and retroflex stops (Fig. 3.1 below). A significant effect of context was seen on CD. The Phrase Initial [I] context resulted in higher CD than the other two contexts (Phrase Medial [s][M] and Phrase Medial [l][M]), $F(2,145)=45.76$ with $p<0.005$ (see 3.3). Significant effect of stop type was found on CD, with VLS having significantly longer closures than VLAS, $F(1,145)=79.36$, $p<0.005$ (see Fig. 3.2). Post-hoc Tukey’s HSD test could only be conducted for POA and prosodic context, since there were fewer than three comparable stop types. The post-hoc tests confirm that these results can be used to significantly categorize the POA differences in mean into two homogenous subsets. Subset 1 consists of velar and Subset 2 consists of labial, dental and retroflex stops. Similarly, for context, subset 1 consists of Phrase Initial [I] and subset 2, Phrase Medial [s][M] and Phrase Medial [l][M] contexts. In summary, for subject PB, velar stops are produced with significantly shorter CD and the Phrase Initial [I] context has an effect on CD, such that CD in this context are significantly longer than the Phrase Medial contexts.

Subject RM showed a significant effect of POA on CD, $F(3, 156)=3.397$, $p<0.05$. Significant difference in marginal means were obtained between the velar and labial stops at $p<0.005$ and dental and labial stops at $p<0.05$ (see Fig. 3.1). Significant difference at the $p<0.05$ level were obtained between the Phrase Medial [s][M] and Phrase Medial [l][M] contexts, in that CD in the Phrase Medial [s][M] context were significantly longer than in the Phrase Medial [l][M] context ($F(2,156)=4.924$). Marginally significant differences were also seen between the Phrase Medial [l][M] and Phrase Initial [I] context at $p<0.05$ level (see 3.3). CD for Phrase Medial [l][M] context were significantly shorter than CD in Phrase Initial [I] context. There were no significant differences between the Phrase Medial [s] and
Phrase Initial [I] context. Stop type had a significant effect on CD, in that VLS stops had significantly longer CD than VLAS stops (F(1,156)=6.366). This difference was significant at the p<0.05 level (see Fig. 3.2). Post-hoc Tukey’s HSD tests showed that velar stops where significantly lower in CD than labial stops, while the mean differences in CD between Phrase Medial [s][M] and Phrase Medial [l][M] contexts were significantly different, with mean CD for Phrase Medial [s][M] context being longer than CD in Phrase Medial [l][M] context. Overall, labial stops had longer CD than velars and in the Phrase Medial[s][M] context the CD tended to be longer than in the Phrase Medial [l][M] context.

Subject SD showed a significant difference between velar and dental stops in CD (overall F(3,154)=2.701). Velar stops showed shorter CD than the dental stops and this difference
Figure 3.2: Effect of stop type on Closure Duration (ms)

was significant at $p<0.005$ level (see Fig. 3.1). The effect of prosodic context was such that the mean differences between the Phrase Medial [s][M] and Phrase Medial [l][M] were significant at the $p<0.005$ level with CD for Phrase Medial [s][M] context being longer than the Phrase Medial [l][M] context ($F(2,154)=5.512$) as can be seen in Fig. 3.3. In addition, significant difference in mean obtained between Phrase Medial [s][M] and Phrase Initial [I] contexts, again, with CD in Phrase Medial [s][M] context being longer. A significant difference in CD was also seen between the stop types at the $p<0.05$ level ($F(1,154)=4.438$, $p=0.037$), with VLS showing longer CD than VLAS (see Fig. 3.1). A post-hoc Tukey’s HSD comparison showed that only the CD differences between the velar and dental stops were significant, while the Phrase Medial [s][M] differed in mean CD from the other two contexts.
Place of articulation had a significant effect on the CD for subject SV, $F(3,143)=10.198$, $p<0.005$ (see Fig. 3.1). The velar stops were significantly shorter than the labial and retroflex stops. Differences in marginal means of CD were significantly different for all contexts. The Phrase Initial [I] context showing longer CD than the Phrase Medial [s][M] context, and the latter being longer than the Phrase Medial [l][M] context ($F(2,143)=24.629$, $p<0.005$), as shown in Fig. 3.3. Post-hoc tests revealed that the differences were categorical between these three contexts. Stop type, like in the case of the other subjects, had a significant effect on the CD. VLS stops showed significantly longer CD than VLAS ($F(1,143)=15.436$, $p<0.005$).

In summary, we observe that the CD measures show that CD for VLS is significantly longer than VLAS (Fig. 3.2). Overall subjects, the velar stops show the shortest CD durations compared to stops from other places of articulation (Fig. 3.1). The effect of context on
CD is less clear, despite the fact that over all subjects the Phrase Initial [I] context tends to show longer CD. Differences between subjects do exist and hence a generalization is difficult to make on the basis of these data.

Velar stops show significantly short CD. In order to test whether these short durations where indeed responsible for the over all short difference in CD between VLAS and VLS, velar stops where excluded from the model. With VLS and VLAS from just three places of articulation significant effect of stop type on CD was obtained (F(1,447)=24.38). This pattern can be seen in 3.4. Excluding subject GA, three speakers show significant difference between the CD for VLS and VLAS.

Figure 3.4: Effect of Stop type on Closure Duration: Excluding velar stops
3.3 Closure duration with pause

As mentioned in section 2.3.1 a sizeable number of utterances produced by the speakers exhibited pauses between the preceding context and the target word. Two probable causes could be responsible for these pauses. One, a number of nonce words were used in this study to complete a full set of CVC words. Unfamiliarity with these words may have caused the speakers to pause between the context preceding the target word and the target word itself. Two, pauses could have been introduced due to speakers exhibiting hesitation in completing the experimental task. This hesitation could have led the speakers to produce an intonation boundary between the preceding context and the target word which in turn would have resulted in an intonation contour reserved for words in isolation. Auditory judgement on the appropriateness of these utterances was made for these cases. Similar auditory judgements were used for the phrase medial [M] condition with preceding [l] and phrase Initial [I] condition with preceding [e]. Utterances where auditory evaluation led us to confirm that a pausal break existed between the target words and their preceding environments where not annotated with a CD.

In order to confirm that the auditory judgements were robust in identifying naturally intonated sentences without a pausal break, a subset of the data that was not annotated with CD was also examined. This subset included dental stops that occurred in the phrase medial [M] condition with a preceding [s]. The cessation of the frication in the preceding context to the target word provided an adequate environment where the noise in the frication of [s] contrasted with the absence of signal in the closure and the pausal break. Beginning of the CD was then annotated from the end of the frication in [s] till the visible burst of the target word. This relatively long duration of CD could include the duration of closure and pausal break. As mentioned in section 2.3.1, auditory judgement was used to identify tokens that didn’t exhibit a pause between the preceding [s] condition and the target word. These measurements will confirm that the auditory judgement used to identify the tokens
which didn’t exhibit a pause between the preceding context and target word, was correct in its estimation.

The general pattern of the distribution of CD (with pause) for VLAS and VLS for three of the five subjects (PB, SD, RM) suggests that CD (with pause) for VLS is longer than the CD (with pause) for VLAS. As can be seen in 3.5, the median CD (with pause) values are consistently longer for VLS compared to VLAS in the phrase medial [s] condition for dental stops. Thus the patterns of CD durations for target words preceded by pauses and target words that are naturally uttered are similar for at least three subjects. This finding confirms that the auditory judgements used to select tokens for which CD measures would be taken are accurate in identifying tokens without pauses.

3.4 Durational properties: Voicing lead time (VLT) durations

All subjects show a significant effect of stop type on VLT durations. VS show significantly longer VLT than the VAS (Figures 3.6-3.10, Panel A) for all speakers. This pattern can be seen in the box plots (Panel D), where the VLT values for VS tend to be higher when compared to the VLT values for VAS. The box plots also show that context has a significant effect on VLT. The Utterance Initial [U] context tends to have higher VLT values than the Phrase medial contexts unambiguously for GA, PB, and SV, while the RM and SD do not show this pattern consistently. GA and PB show higher VLT for the Phrase Initial [I] context than the phrase medial context. Thus initial contexts, like the [U] and [I] conditions tend to have longer VLT than the medial contexts.

Pairwise comparison of means shows that for speaker GA, the mean difference in VLT between the Utterance Initial [U] position and the Phrase medial positions is significant at
Figure 3.5: Effect of Stop type on Closure Duration (with pause): Utterance medial [s] condition; dental POA

p<0.05. The Phrase Initial [I] context significantly differs only from the Phrase Medial [I] context.

Speaker PB shows a significant difference between the Utterance Initial [U] and the Phrase Medial positions; VLT in the Utterance Initial [U] context is higher than in the phrase medial contexts. A similar pattern emerges for the Phrase Initial [I] position and in this position as well, the VLT values are higher than in the phrase medial context. The mean
differences in VLT between the Utterance Initial[U] and Phrase Initial [I] are not significant. The differences within the phrase medial position between [l] and [s] are significant with [l] position showing shorter VLT than the [s] position.
Figure 3.6: Speaker GA: ANOVA results for dependent variable VLT and independent variables, stop type (Panel A), place of articulation (Panel B), context (Panel C), Horizontal lines in boxes (Panel D) represent median values and Summary of effects and interactions (Panel E)
Figure 3.7: Speaker PB: ANOVA results for dependent variable VLT and independent variables, stop type (Panel A), place of articulation (Panel B), context (Panel C), Horizontal lines in boxes (Panel D) represent median values and Summary of effects and interactions (Panel E)
Figure 3.8: Speaker RM: ANOVA results for dependent variable VLT and independent variables, stop type (Panel A), place of articulation (Panel B), context (Panel C), Horizontal lines in boxes (Panel D) represent median values and Summary of effects and interactions (Panel E)
Figure 3.9: Speaker SD: ANOVA results for dependent variable VLT and independent variables, stop type (Panel A), place of articulation (Panel B), context (Panel C), Horizontal lines in boxes (Panel D) represent median values and Summary of effects and interactions (Panel E)
Figure 3.10: Speaker SV: ANOVA results for dependent variable VLT and independent variables, stop type (Panel A), place of articulation (Panel B), context (Panel C). Horizontal lines in boxes (Panel D) represent median values and Summary of effects and interactions (Panel E).
Speaker SV shows a significant difference between the Utterance Initial [U] and phrase
medial contexts, like GA and PB, however in the Phrase Initial [I] position for this speaker,
the VLT is shorter than the Phrase Medial [s] condition.

Speaker RM shows longer VLT in the Utterance Initial [U] position than the phrase
medial position, however only the mean difference between the Utterance Initial [U] and
Phrase Medial [l] context is significant at the p<0.05 level. The difference between the
Phrase Initial [I] and Phrase Medial [l] context is also significant for this speaker with VLT
being longer in the Phrase Initial [I] context. The difference between the [s] and [l] contexts
within the Phrase Medial condition is not significant for this speaker.

Speaker SD shows significant differences between the initial [U,I] and the Phrase Medial
[l] context, and the difference between the Phrase Medial [s] and Phrase Medial [l] contexts
is also significant for this speaker.

Post-hoc comparisons however show that not all the significant results are adequate to
categorize the VLT patterns based on the effect of the context. Table 3.1 details the results
of the post-hoc comparisons for all the subjects.

<table>
<thead>
<tr>
<th>Subject</th>
<th>Post-hoc comparisons</th>
</tr>
</thead>
<tbody>
<tr>
<td>GA</td>
<td>Phrase Medial [l] &lt; Phrase Medial [s], Phrase Initial [I] and Utterance Initial [U]</td>
</tr>
<tr>
<td>PB</td>
<td>Phrase Medial [l] &lt; Phrase Medial [s] &lt; Phrase Initial [I] and Utterance Initial [U]</td>
</tr>
<tr>
<td>RM</td>
<td>Phrase Medial [l] &lt; Phrase Initial [I]</td>
</tr>
<tr>
<td>SD</td>
<td>Phrase Medial [l] &lt; Phrase Medial [s], Phrase Initial [I] and Utterance Initial [U]</td>
</tr>
<tr>
<td>SV</td>
<td>Phrase Medial [l] &lt; Phrase Initial [I], Phrase Medial [s] &lt; Utterance Initial [U]</td>
</tr>
</tbody>
</table>

Table 3.1: Effect of context on VLT: Post-hoc comparison based on SNK and Tukey HSD

The post-hoc comparisons show that the Phrase Medial [l] context where the target word
is preceded by the modifier [laːl] ‘red’, a voiced context, the VLT durations are significantly
lower than in the other contexts for all speakers\(^3\). Based on these results, it can be observed
that a preceding voiced context tends to decrease the VLT duration when compared to a

\(^3\)For speaker RM, VLT for Phrase Medial [s] and the Utterance Initial [l] contexts does not significantly
differ from either the Phrase Medial [l] or Phrase Initial context.
voiceless context (Phrase Medial [s]). Only for two speakers (PB and SV) does the initial context (U and I) have an effect on the VLT, such that the VLT is longer than in the phrase medial context. These results, therefore, indicate that the effect of prosodic context is marginal on VLT. There is a tendency towards longer VLT durations in the initial contexts when compared to the medial contexts, however, a more controlled experimental environment may be able to shed light on this observation.

Except for speaker GA, all of the other four speakers show a significant effect of place of articulation on VLT (Panel C, Figures 3.9-3.13). Within speaker pairwise comparisons show that for speaker PB, significant differences in mean VLT appear between labial and dental, retroflex, and velar places of articulation. The differences in decreasing order of VLT duration are such that: labial and dental > Retroflex > Velar. These differences are significant at p<0.05. In the case of speaker RM the order of the differences is as follows: Labial > Dental > Velar. No significant differences obtain between labial and retroflex places of articulation and between the retroflex and dental, however the retroflex stops are significantly longer in VLT than the velar stops. Speaker SD shows that the difference in VLT due to place of articulation is only significant between the velar and the other three places of articulation, with the velar stops showing significantly lower VLT. Within the labial, dental and retroflex, the differences being not significant. The order of differences between place of articulation for speaker SD is as follows: Labial, dental, Retroflex > Velar. Speaker SV shows that only the labial place of articulation has significantly longer VLT than the other three places of articulation. The order of the differences being: Labial > Dental, Retroflex and Velar. These patterns can be seen in Fig. ?? below.

Post-hoc tests based on multiple comparisons, however, show that only the difference between the velar and the other three places of articulation are significantly different (Table 3.2).

51
Except for speaker RM, all of the other speakers show that there is no interaction between the factors, stop type, place of articulation (POA) and context as far as VLT durations are concerned. Speaker RM shows a marginal effect of interaction between the three factors (p=0.05).

### 3.5 Closure and VLT correlation

One of the primary goals of this study was to examine the hypothesized parallel / complementary distribution between CD and VLT. Based on the results from section 3.2, 3.3 and 3.4 it can be concluded that CD for voiceless stops and VLT for voiced stops show patterns that allow us to draw a parallel between these two proposed categories. CD for VLS and VLT for VS are comparatively longer in duration than CD for VLAS and VLT for VAS,
Table 3.2: Effect of place of articulation on VLT: Post-hoc comparison based on SNK and Tukey HSD

<table>
<thead>
<tr>
<th>Subject</th>
<th>Post-hoc comparisons</th>
</tr>
</thead>
<tbody>
<tr>
<td>PB</td>
<td>Velar &lt; Dental, Retroflex and Labial</td>
</tr>
<tr>
<td>RM</td>
<td>Velar &lt; Dental, Retroflex and Labial</td>
</tr>
<tr>
<td>SD</td>
<td>Velar &lt; Dental, Retroflex and Labial</td>
</tr>
<tr>
<td>SV</td>
<td>Velar, Dental, Retroflex &lt; Labial</td>
</tr>
</tbody>
</table>

respectively. This generalization when correlated with the fact that closure durations for voiced stops are produced with voicing suggests that closure in four-way contrasts can be important in cueing the aspiration contrast. Under this view, the voicing duration during the closure for voiced stops can be understood to be an epiphenomenon. This view is also in keeping with the observations and findings of Schiefer (1986) that listeners perceive tokens with shorter VLT durations to be VAS, all else being equal. These findings suggest that in Hindi production and perception of stops, the duration of the closure is primary and relevant for the stop to be categorized as an unaspirated or aspirated stop. This observation, however, is not a claim towards the redundancy of voicing during closure, on the contrary, voicing during closure, the duration being irrelevant, is an important cue towards distinction between voiceless and voiced stops. Thus, the epiphenomenon of voicing during closure serves to enhance the contrast between voiceless and voiced stops, while the duration of the closure serves to distinguish between the aspirated and unaspirated stops.

Keyser and Stevens (2006) propose a two step process for their model of speech production. The first process involves replacement of universal distinctive features with an appropriate set of motoric instructions. The second process involves a language-specific process that is sensitive to those features that are in danger of losing their perceptual saliency as a consequence of the environment in which they appear. This process is referred to as enhancement, which adds additional motoric instructions to enhance the saliency of the jeopardized features. Closure duration in Hindi can be understood in this context to be subsumed under a language-specific process. While the obliteration of the duration of closure
between aspirated and unaspirated stops puts in jeopardy the contrast between aspirated
and unaspirated stops, motoric instructions to maintain a distinction in duration serve to
salvage the contrast. Similarly, the voicing during closure for VAS and VS serves to enhance
the contrast between the voiced and voiceless stops.

There are two fundamental assumptions behind the model proposed by Keyser and
Stevens (2006). The first assumption is that defining features are universal and the sec-
ond assumption is that defining features need enhancement depending on the contrasts in a
language. In this respect, in Hindi, while the aspiration contrast is a defining feature and is
no different from a distinctive feature that describes aspiration in any other language, the
duration of closure is a feature that serves to enhance the aspiration feature.

However, it is necessary to relate the theory of enhancement with the data in Hindi.
Presumably the motoric instructions that are governed by the language-specific features of
enhancement need to be directed towards either lengthening the closure for unaspirated
stops or shortening the closure for aspirated stops in comparable terms. While adequate
and independent articulatory-aerodynamic grounds exist for shorter closures for aspirated
stops, sufficient grounds are unavailable for lengthening of closures for unaspirated stops.
A manifestation of the theory of enhancement of the aspiration contrast in Hindi therefore,
is a set of motoric instructions that lengthen the duration of closure for unaspirated stops.
Relating this to Schiefer’s findings from perception studies, it becomes clear that enhance-
ment in Hindi results in lengthening of closures for unaspirated stops and not shortening
of closures for aspirated stops. In 6, I discuss in detail the reasons behind the absence of
aerodynamic factors in the spread of breathiness following VAS.

3.6 Summary of results

Closure durations are affected by both the stop type and place of articulation. in the case of
VLT, it has been shown that stop type has an effect on VLT durations. Marginally it has also
been shown that prosodic positions such as the Utterance Initial (U) position tend to lengthen durational features such as closure duration and VLT duration, the phrase medial contexts lead to reduction in duration of segmental features. A complete cross category comparison of closure duration and VLT in the utterance initial position could not be accomplished due to the fact that acoustically it is not possible to segment the beginning of closure in this position.

Further, the effect of segmental contexts on the segmental closure and VLT durations shows that despite the voiceless [s] preceding context, in the phrase medial context VLT durations are longer in this context compared to the phrase medial [l] context.

Place of articulation has expected effects on both closure and VLT durations, in that labials tend to be longer than the velar stops. The articulatory mechanisms that increase the oral tract volume due to the place of the constriction, it seems, not only effect the durations of voicing but also the closure durations. This is unusual in the sense that labials for instance are not only voiced for longer durations, but they tend to have longer closures as well.

Over all the results from VLT and CD provide evidence against the standard view on contrasts in Hindi. As discussed in section 1.4 of chapter 1. The results from the VLT durations in conjunction with the results from CD show that aspirated stops have comparatively shorter closures compared to the unaspirated stops. This observation also provides evidence against the standard assumption that the voicing during closure and closure in and off itself are non-distinctive. Based on the evidence from VLT and CD and Schiefer’s observation that listener’s prefer longer VLT durations for the unambiguous perception of VS, I conclude that closure durations could play a contrastive role in making the four-way distinctions in Hindi.
Chapter 4

Effects of stop type on fundamental frequency

4.1 Universal tendency for $f_0$ lowering following voiced stops

Voiced consonants are known to lower the fundamental frequency in the following vowel (House and Fairbanks 1953, Lehiste and Peterson 1961, Löfqvist 1975, Umeda 1981, Ohde 1984). $f_0$ tends to be higher following voiceless consonants than voiced consonants, an effect that can be found as far as 100 ms from voicing onset (House and Fairbanks 1953, Umeda 1981). This is a well known tendency amongst a large number of languages, irrespective of the nature of laryngeal contrasts that are employed in these languages. That is to say, languages that contrastively use breathy, creaky and modal phonation, tonal contrasts and laryngeal contrasts such as aspiration, all exhibit lowering of $f_0$ following voiced segments (Ohala 1973, Hombert 1978).

Dutta (2007) shows that compared to VS, VAS lower average $f_0$ values in the first six pitch periods. However, this study looked only into the acoustic properties of the voiced stops (VS, VAS) and not the voiceless stops (VLS, VLAS), hence it was not possible to study the effect of stops on the $f_0$ of the following vowel for all the four stop types. Additionally, the effect of the stops on the initial $f_0$ of the vowel was measured as the average $f_0$ of the first six pitch periods. Since the first six pitch periods do not necessarily coincide with 100 ms following the release of the stop it would be interesting to find out how far the effect of the stop type on the $f_0$ persists. While measuring the average $f_0$ in the first six pitch periods
provides an account of the absolute differences in $f_0$ between the VAS and VS, it fails to establish in relative terms the extent of the effect of the stop type on the $f_0$. One of the research goals of this study was to investigate the extent of $f_0$ lowering in the following vowel in order to fully understand the effects of all the four stop types on the $f_0$ and not just the voiced stops. As has been shown in chapter 3, VLT for VAS is significantly lower than VS. Under the assumption that voicing is responsible for lowering $f_0$ universally, a comparatively less voiced VAS tends to lower $f_0$ further than the VS. Therefore it is also important to relate the extent of $f_0$ lowering in VAS with the extent of the spread of breathiness in order to establish whether the $f_0$ lowering and breathiness spread could correlated. The results from the study of $f_0$ lowering show that $f_0$ following VAS is lower compared to the VS, VLAS and VLS till 30 percent of the following vowel. Results from the spectral intensity measures (Chapter 6) suggest that the breathiness spreads till 30-50 percent of the following vowel in VAS. A comparison of the the $f_0$ lowering and breathiness spread durations following VAS suggests a correlation between the nature of phonation following the release of the VAS and the $f_0$ in the following vowel.

This chapter is organized as follows: In the following section, 4.2, I will discuss the effect of the stops on the $f_0$ of the following vowel in Hindi and also present the results from the individual speakers. Section 4.3 is a brief discussion of the results from section 4.2. Section 4.3 is a brief discussion on the effects of prosodic context on the $f_0$ and finally, in section 4.4, I outline the major conclusions from this portion of the study.

### 4.2 $f_0$ as a function of stop type

Stemming from the interest in investigating the consonantal origins of tonal systems, Purcell et al. (1978) is a phonetic study on the correlation of Hindi breathy voiced stops and Panjabi tones. In fact, Ohala (1974) had modeled the Panjabi/Hindi data to account for tonal development in general. A number of studies (House and Fairbanks 1953, Hombert 1978),
to cite a few, started to establish the correlation between voicing and $f_0$. Purcell et al. (1978) in their experimental study established two correspondences between Hindi and Panjabi:

- Panjabi tone originated due to the loss of aspiration
- Panjabi tonal contours and $f_0$ perturbations in Hindi are comparable

They were able to establish the fact that there is a direct correlation between the effects of Hindi VAS on $f_0$ of adjacent vowels and the realizations of Panjabi high and low tones on cognate words. Thus the correlation, as far as low-rising tone in Panjabi following voiceless stops [k̡ar] ‘house’ and Hindi [gʱar] ‘house’ is concerned, it is established without specifically looking into the phonetic manifestation of voicing in Hindi, rather just the phonological contrast. The correlation between the consonantal origins of tones in Panjabi and the synchronic behaviour of $f_0$ in Hindi stops notwithstanding, it is not clear from the study by Purcell et al. (1978), whether the $f_0$ lowering following Hindi VAS could be correlated with the mode of phonation or could it be because of the universal tendency for voiced stops to lower $f_0$.

In this study, the effect of the Hindi stops on the $f_0$ of the following vowel is examined so as to provide an explanation for the above mentioned questions regarding the nature of $f_0$ lowering following VAS. $f_0$ analysis, as mentioned in section 2.3.3 of chapter 2, was conducted on a time normalized contour such as ten $f_0$ values for each token were recorded at 10 equidistant points in the vowel starting at 10 percent of the vowel and ending at 100 percent of the vowel. This way the ten $f_0$ values that are recorded for each token, represent the overall $f_0$ contour for each token. A one-way ANOVA was conducted with $f_0$ at 10 percentage points as the dependent variable with Stop Type (4 levels) as a factor. A multiple pairwise comparison was done to examine differences in $f_0$ as a function of stop type. Tukey post-hoc tests were used to categorize the levels of the stop type according to the effect a particular stop type had on the corresponding $f_0$ values at the different percentage points of the vowel. In what follows, first results from each subject are discussed followed
by a discussion of the most conservative findings that are obtained for all subjects.

4.2.1 Subject GA

The statistical tests indicate that mean $f_0$ differences between the four stops are significant at $p<0.05$ till 40 percent of the vowel. After 50 percent of the vowel the mean $f_0$ differences due to stop type begin to dissipate and after 80 percent till 100 percent of the vowel the differences in $f_0$ between the four stops are no longer statistically significant.\(^1\) The mean differences between $f_0$ at 10, 20, 30 and 40 percent allow us to posit a three category distinction of stops, namely, VS, VAS, and voiceless stops. Voiceless stops here subsumes both VLS and VLAS. The mean difference in $f_0$ between VLAS and VLS is statistically insignificant from 10 percent till 20 percent of the vowel. Significant differences between mean $f_0$ values start at 30 percent and persist till 40 percent of the vowel between all four stops. Subsequently from 50 percent till 60 percent of the vowel, the difference in mean $f_0$ between VAS and VS becomes insignificant, while the distinction between the voiced stops (VAS and VS) and the voiceless stops (VLS and VLAS) and within the voiceless stops (VLS and VLAS) is significant. From 70 till 100 percent of the vowel there is no significant difference in mean $f_0$ between the four stops.

A Tukey post-hoc test was used to confirm these results and also to help categorize the data. The test shows that from 10 till 20 percent of the vowel mean $f_0$ values help categorize the stops into three subsets, VS, VAS, and the voiceless stops. The Tukey test also shows that only the difference in mean $f_0$ between the stops at 30 percent of the vowel can help in categorizing the stops into four subsets. The difference at 40 percent being only sufficient to categorize the stops as voiced, VLS and VLAS. This pattern persists till 80 percent of the vowel after which the difference in $f_0$ between the stops as a factor of stop type remains insignificant till the end of the vowel. These patterns can be seen in the $f_0$ means plots below.

\(^1\)Here and elsewhere, all reported $p$ values are at the level of $p < 0.05$.
in Fig. 4.1. The mean $f_0$ values for VAS are significantly lower than those of the VS till 30 percent of the vowel. While initially at 10 percent there is no significant difference between the VLS and VLAS stops. The boxplots in Fig. 4.2 show the variation in $f_0$ values. The overall range of $f_0$ for VAS between 10 and 40 percent of the vowel tends to be high, however the median values do show a pattern where they are consistently lower than the medians for VS. At 20 and 30 percent of the vowel, however, the VAS show a great degree of variation.

Figure 4.1: Effect of stop type on $f_0$ (Subject GA): Mean $f_0$ plots
Figure 4.2: Effect of stop type on $f_0$ (Subject GA): Variation in $f_0$ values. Horizontal lines within the boxes represent the median $f_0$ values.

4.2.2 Subject PB

Multiple comparisons for the results for speaker PB show that the mean $f_0$ differences are significant between VAS and the other three stops at the $p<0.005$ level till 30 percent of the
vowel. The mean $f_0$ differences are significant between VS and the other three stops only at 20 percent of the vowel. Significant differences between the VLAS and VLS obtain only at 40 percent of the vowel. The mean difference between VAS and the voiceless stops is significant at $p<0.05$ level till 40 percent of the vowel. The mean difference between VAS and VLAS being significant till 80 percent of the vowel. These patterns can be seen in Fig. 4.3. The mean $f_0$ values for VAS are significantly lower than the mean $f_0$ values for VS till 30 percent of the vowel. In this respect the difference in $f_0$ values for speaker PB are similar to that of speaker GA.

A post-hoc Tukey test shows that till 40 percent of the vowel the mean $f_0$ values alone can help categorize the stops into three different subsets: at 10 percent the mean $f_0$ values for VAS<VS=VLAS<VLS, at 20 percent the mean $f_0$ values for VAS<VS<VLS=VLAS, at 30 percent VAS<VS=VLS<VLS=VLAS and at 40 percent, the mean $f_0$ for VAS<VS=VLS<VLAS. Overall the patterns suggest that mean $f_0$ for VAS is lower than the other stops till 40 percent of the vowel. VS also tend to lower $f_0$, however in the case of this subject, only at 20 percent the mean $f_0$ value is significantly lower than that for VLS and VLAS. The boxplots in Fig. 4.4 show variation in $f_0$ values, quite like speaker GA. $f_0$ values for VAS that lie between 10 and 40 percent of the vowel tend to show a great amount of variation. The median values, however, do show a pattern of $f_0$ fall after 10 percent till 30 percent of the vowel.

### 4.2.3 Subject RM

One-way ANOVAs show that speaker RM exhibits similar patterns for $f_0$ values for VAS when compared to speaker GA and PB. However the crucial difference being that the difference in $f_0$ between VAS and VS is significant only till 20 percent of the vowel, with VAS showing
lower mean $f_0$ values than VS ($p=0.007$ and $0.005$ at 10 and 20 percent, respectively). The mean $f_0$ differences between the voiced stops and voiceless stops are significantly different till 60 percent of the vowel.

A post-hoc Tukey test shows that from 10 till 20 percent of the vowel the pattern of $f_0$ values can help categorize the stops into three categories, such that $f_0$ for VAS<VS<VLS=VLAS. These patterns can be seen in Fig. 4.5 below. In Fig. 4.6 we see that for this speaker the variation in the $f_0$ values is reduced, compared to speaker GA and PB. Further, the median
Figure 4.4: Effect of stop type on $f_0$ (Subject PB): Variation in $f_0$ values. Horizontal lines within the boxes represent the median $f_0$ values.

Values show that the VAS have lower $f_0$ values than the voiceless stops as do the VS at least until the initial 10-20 percent of the vowel.
Figure 4.5: Effect of stop type on $f_0$ (Subject RM): Mean $f_0$ plots

4.2.4 Subject SD

Multiple comparisons between mean $f_0$ values for this speaker show patterns similar to those of the other speakers, especially in the initial 30-40 percent of the vowel. For this speaker mean difference in $f_0$ between VAS and VS is statistically significant till 30 percent of the vowel, with the VAS showing lower mean $f_0$ than the VS. The difference in mean $f_0$ between VLS and VLAS stops is significant between 20-50 percent of the vowel, with the VLS showing
lower $f_0$ means. The VS show significantly lower mean $f_0$ than the voiceless stops from 20-70 percent of the vowel. In all of these cases $p<0.005$ for the significant differences in $f_0$ means. These patterns can be seen in Fig. 4.7 below.

A post-hoc Tukey test shows that for this subject at 10 percent of the vowel the $f_0$ means for VAS<$VS=VLS=VLAS$. However, between 20-30 percent of the vowel the $f_0$ means for the VAS<$VS<VLS<VLAS$. Post-hoc tests also confirm that from 80-100 percent of the vowel
the stop type has no significant effect on the $f_0$ of the vowel. The boxplots in Fig. 4.8 show that for this subject also $f_0$ values for VAS vary when compared to the other stops in the beginning 10-30 percent of the vowel. The median values however conform to the pattern seen in the mean $f_0$ differences.

Figure 4.7: Effect of stop type on $f_0$ (Subject SD): Mean $f_0$ plots
4.2.5 Subject SV

Multiple comparisons for this subject show that $f_0$ means for VAS are lower than the other three stops from 10-80 percent of the vowel ($p<0.05$). However the VS for this speaker show $f_0$ values higher than those for the VLAS significantly at 20 percent of the vowel ($p=0.019$). This pattern for the VS is unusual and exceptional when compared to the initial $f_0$ patterns of the other four speakers. While the $f_0$ pattern for the VAS for this speaker like the other
speakers shows a low onset and a steep rise, the VS do not reflect the low onset. These patterns can be seen in Fig. 4.9. The boxplots in Fig. 4.10 show that the variation in $f_0$ values for this speaker are consistent for all the stops. The median values for VAS also show that VAS are produced with lower $f_0$ compared to the other stops.

Figure 4.9: Effect of stop type on $f_0$ (Subject SV): Mean $f_0$ plots
Figure 4.10: Effect of stop type on $f_0$ (Subject SV): Variation in $f_0$ values. Horizontal lines within the boxes represent the median $f_0$ values.

4.3 Discussion

Overall the $f_0$ patterns suggest that the effect of the stop type is reflected in the initial 20-30 percent of the vowel at least for four subjects (GA, PB, RM, and SD). The general pattern for $f_0$ values is such that initially a three-way distinction can be seen, dependent on the stop type. At least for four speakers (GA, PB, RM, and SD), VAS tend to have lower mean $f_0$ values than the VS which in turn have lower $f_0$ than the voiceless stops (both VLS and
VLAS). The initial mean $f_0$ differences between VLS and VLAS is less certain. The relations between the stops based on statistically significant mean $f_0$ differences can be summarized as in Table 4.1 below.

<table>
<thead>
<tr>
<th>Subject</th>
<th>10%</th>
<th>20%</th>
<th>30%</th>
<th>40%</th>
<th>50%</th>
</tr>
</thead>
<tbody>
<tr>
<td>GA</td>
<td>VAS&lt;VS&lt;VLAS&lt;VLS</td>
<td>VAS&lt;VS&lt;VLAS&lt;VLS</td>
<td>VAS&lt;VS&lt;VLAS&lt;VLS</td>
<td>VAS&lt;VS&lt;VLAS&lt;VLS</td>
<td>VAS&lt;VS&lt;VLAS&lt;VLS</td>
</tr>
<tr>
<td>PH</td>
<td>VAS&lt;VS&lt;VLAS&lt;VLS</td>
<td>VAS&lt;VS&lt;VLAS&lt;VLS</td>
<td>VAS&lt;VS&lt;VLAS&lt;VLS</td>
<td>VAS&lt;VS&lt;VLAS&lt;VLS</td>
<td>VAS&lt;VS&lt;VLAS&lt;VLS</td>
</tr>
<tr>
<td>RM</td>
<td>VAS&lt;VS&lt;VLAS&lt;VLS</td>
<td>VAS&lt;VS&lt;VLAS&lt;VLS</td>
<td>VAS&lt;VS&lt;VLAS&lt;VLS</td>
<td>VAS&lt;VS&lt;VLAS&lt;VLS</td>
<td>VAS&lt;VS&lt;VLAS&lt;VLS</td>
</tr>
<tr>
<td>SD</td>
<td>VAS&lt;VS&lt;VLAS&lt;VLS</td>
<td>VAS&lt;VS&lt;VLAS&lt;VLS</td>
<td>VAS&lt;VS&lt;VLAS&lt;VLS</td>
<td>VAS&lt;VS&lt;VLAS&lt;VLS</td>
<td>VAS&lt;VS&lt;VLAS&lt;VLS</td>
</tr>
<tr>
<td>SV</td>
<td>VAS&lt;VS&lt;VLAS&lt;VLS</td>
<td>VAS&lt;VS&lt;VLAS&lt;VLS</td>
<td>VAS&lt;VS&lt;VLAS&lt;VLS</td>
<td>VAS&lt;VS&lt;VLAS&lt;VLS</td>
<td>VAS&lt;VS&lt;VLAS&lt;VLS</td>
</tr>
</tbody>
</table>

Table 4.1: Ordered relations between the stops from 10 to 50 percent of the vowel based on Tukey HSD comparisons

Three observations can be made based on the results of the $f_0$ analysis as well as the summarized table above:

1. In their initial portion, vowels following voiced stops (VS, VAS) show lower mean $f_0$ than the vowels following voiceless stops (VLS, VLAS).

2. VAS<VS

3. The extent of the $f_0$ lowering for vowels following voiced stops is such that this lowering effect can be seen till 30 percent of the vowel (for four speakers).

4. Mean $f_0$ values do not provide a clear understanding of the relation between the VLS and VLAS.

Observation 1 provides evidence for a universal tendency for $f_0$ lowering following voiced stops, however, it does need to be noted based on the results from Chapter 3 on the duration of VLT that VAS show shorter VLT than the VS. Thus, relatively “less” voiced VAS tend to lower $f_0$ more than the VS. Observation 2 provides credence to the claim that $f_0$ lowering for voiced stops can be seen in the first six pitch periods (Schiefer 1986, Dutta 2007) and also shows that the extent of the $f_0$ lowering can extend to about 30 percent of the vowel. Observation 3 claims that based on $f_0$ patterns it is not possible to categorize the voiceless stops unambiguously.
4.4 Effect of context on \( f_0 \)

There was no significant interaction between Stop Type and Context on the mean \( f_0 \) values. However, in all contexts, a low-rising contour was found. This low-rising contour could be an effect of speakers identifying the target words as new or novel instances, which lead the speakers to implement perhaps a ‘focus’ accent in Hindi. It has also been shown by various studies on the manifestation of stress in Hindi that a L*H contour is a close estimation of word stress in Hindi (Dyrud 2001). It is likely that in this study the low rising \( f_0 \) contour could be a manifestation of lexical stress. In that respect, an experiment that is designed specifically to test the effect of prosodic context on the \( f_0 \) of vowels following the stops in Hindi is needed. Unfortunately, the near-laboratory setting under which the recordings were made, made it difficult to obtain natural or near-spontaneous recordings.

4.5 Summary and implications

The results from the effect of the stop types on \( f_0 \) show that VAS and VS tend to lower the \( f_0 \) of the following vowel till about 30 percent. While previous studies had established that \( f_0 \) following the voiced stops and especially the VAS is lower, in this study the extent of this lowering effect has been shown. There are two major implications of these results. First, based on these results it can be said that the \( f_0 \) lowering following voiced stops and especially VAS is not restricted to the first six pitch periods. The lowering effect can be seen till about 30 percent of the vowel, which in most cases will be more than just six pitch periods in the following vowel. Secondly, it can also be said that the VAS for all subjects tend to have lower mean \( f_0 \) in the initial 10-30 percent of the vowel than the VS. In addition to these findings, it can also be said that the universal tendency associated with voiced stops for \( f_0 \) lowering is somewhat supported by these results, however comparison between the proportional duration of \( f_0 \) lowering and breathiness spread (Chapter 6) suggest that breathiness could also be considered a likely reason behind the extent of \( f_0 \) lowering following...
VAS. Languages that contrast between breathy and modal vowels tend to show that breathy vowels are accompanied by low $f_0$ (Esposito 2003, Andruski and Ratcliffe 2000, Thurgood 2004). Voiced stops tend to have lower $f_0$ initially than the voiceless stops (at least for four speakers). However, the VAS, which are “less” voiced (see VLT duration results in Chapter 3) tend to have a further lowering compared to VS. In chapter 6, the nature and extent of breathiness following the VAS is examined based on four spectral intensity measures.
Chapter 5

Aspiration and vowel duration

5.1 Outline

In Dutta (2007) it was observed that the breathy portion following the VAS tended to permeate into a sizeable portion of the vowel. One of the research goals for this study was to be able to compare and correlate the durations of aspiration following VLAS and VAS. This correlation will provide a crucial understanding of the nature and phonetic implementation of the aspiration contrast. In this study, we find that the duration of aspiration following VLAS is comparatively shorter than the duration of breathiness following VAS. Further, we also find that the duration of aspiration is well correlated with the place of articulation of the stop in VLAS. Thus, a further point of occlusion in the oral tract leads to a longer duration of aspiration. The same pattern, however, is not found in breathiness / aspiration following the VAS. This goes to show that aerodynamic factors that are responsible for the varying duration of aspiration following VLAS are not at play in the results of the spread of the breathy portion in VAS. In this respect, supraglottal configurations have no effect on the duration of the breathy portion following the VAS. Breathy release following the VAS are due to particular laryngeal configurations.

This chapter is organized as follows: In section 5.2, I discuss the results from aspiration durations in VLAS. Following this, in section 5.3, I present the results from the vowel durations. In section 5.4, I provide a summary and discussion of the relevant results and conclusions that can be drawn based on the results.
5.2 Durational properties: Aspiration duration

Aspiration duration was measured for the VLAS for all places of articulation and all contexts.\(^1\) In this section, results from the effect of place of articulation and context will be presented for all of the subjects.

5.2.1 Subject GA and PB

Place of articulation had a significant effect on aspiration duration. For subject GA, the absence of VLAS led to only three-way comparisons between the places of articulation. Results indicate a significant effect, F(2,74)=21.046 with p<0.001 (see Fig. 5.1). Velar VLAS are longer in aspiration duration than both dental and retroflex VLAS for GA. Post-hoc comparisons confirm that the difference in mean between dental and retroflex aspiration duration is not significant. Similar results obtain for subject PB, with F(3,114)=26.192 which is significant at the p<0.001 level. For PB as well, the post-hoc comparisons reveal that the velar VLAS stops have longer durations of aspiration compared to the labial, dental and retroflex VLAS (see Fig. 5.1). Subjects GA and PB did not exhibit any significant effect of prosodic context in the duration of aspiration for the VLAS. Hence, prosodic context was not included as a factor in the analysis of aspiration duration.

5.2.2 Subject RM, SD and SV

Results from the effect of place of articulation indicate that subject RM shows longer aspiration duration for velar VLAS and patterns together with subjects GA and PB. SD and SV however pattern together with significant differences between the velar+labial VLAS and the dental+retroflex stops. A consistent finding on the duration of aspiration is the longer

\(^1\)As mentioned above, subject GA could not produce the labial VLAS due to ongoing merger of this stop category with the voiceless labio-dental fricative.
Figure 5.1: Effect of place of articulation (POA) on Aspiration Duration (ms)

duration of aspiration for all velar VLAS (see Fig. 5.1). Subjects RM, SD and SV did show marginal effect of prosodic context on the aspiration duration, however post-hoc tests reveal that the differences in mean are not significant for RM and SD. For SV, prosodic context does have an effect on the aspiration duration, in that, the Phrase Initial [I] (Subset 1) is maximally distinct from the Utterance Initial [U] (Subset 2) and the Phrase Medial [s][M] context (Subset 3) as seen in Fig. 5.2.
5.3 Durational properties: Vowel durations

Vowel duration results show that there is a main effect of stop manner on the vowel duration. Vowels following aspirated stops (VAS and VLAS) are significantly longer than vowels following unaspirated stops (VS and VLS). As can be seen in Fig. 5.3 a difference in vowel duration appears between the unaspirated and aspirated stops. The primary purpose for these measurements was to be able to compare relative durations of aspiration following VLAS with the total vowel durations for these stops. This would help establish the extent of aspiration following the VLAS relative to the total vowel duration.

Vowel Duration\(_{total}\) = Aspiration duration + \(v\)

Here, \(v\) refers to the duration of the vowel following the offset of aspiration.
Comparisons between the relative duration of aspiration and vowel duration show that aspiration duration for VLAS for all speakers persists till about 25 percent of the total vowel duration. This can be seen in the pie diagram in Fig. 5.4. This goes to show that, irrespective of the effects of place of articulation on aspiration duration, the maximal durational effect of aspiration following VLAS is less than the spread of breathiness following VAS (between 30-50 percent). Results from spectral intensity measures were used to arrive at this comparison. Details of this analysis appear in Chapter 6. Suffice it to say, that the effect of aspiration following VLAS is temporally shorter than the voice quality difference between the VAS and the unaspirated stops.
5.4 Summary of results

The results from the duration of aspiration and the vowel durations suggest that the duration of aspiration following VLAS is comparatively shorter than the duration of breathiness following VAS (see 5.4). Further, place of articulation of the stop in VLAS has an effect on the duration of aspiration. Thus, velar stops tend to have longer durations of aspiration compared to the stops from other places of articulation. This goes to show that aerodynamic factors that are responsible for the varying duration of aspiration following VLAS. Thus, supraglottal configurations have an effect on the duration of the aspiration portion following the VLAS. It can be safe to assume, that glottal mechanisms responsible for the breathy release following VAS are not at play either in the manner or spread of aspiration following VLAS. A comparison of these results with the findings from 6 below, suggest that
aspiration as produced following VLAS is acoustically and aerodynamically different from the breathy release following VAS. The duration of aspiration apart from being shorter than the breathy release is also governed by the place of articulation of the stop. Therefore, the manner and duration of aspiration in VLAS is partly due to aerodynamic factors, which due to the shorter duration of closures for velar stops produces longer durations of aspiration. These findings can be directly correlated to supraglottal configurations rather than glottal configurations, as is the case with the breathy portion following VAS.
Chapter 6
Spectral properties of Hindi stops

6.1 Breathiness due to incomplete glottal closure

Two studies by Dixit and MacNeilage (1980) and Benguerel and Bhatia (1980) have reported on the articulatory attributes of breathy voiced stops in Hindi\(^1\). Both these studies present data from post-release measurements of glottal width associated with stops in Hindi. These studies show that the width of the glottal opening in VAS lies in between that of VS and VLAS, suggesting insufficient glottal closure. The acoustic consequence of an articulatory configuration such as insufficient glottal closure is simultaneous periodicity and low frequency noise associated with the release portion of the VAS. A glottal configuration such as this also leads to increased air flow. This has been given impressionistic labels such as ‘breathy’ and ‘murmured’ (Ladefoged (1975)) voicing, both of these labels being used interchangeably. Glottal source signals obtained through inverse filtering typically show more symmetrical opening and closing phases with little or no complete closed phase for Gujarati breathy vowels (Bickley 1982, Fischer-Jorgensen 1967). Fischer-Jorgensen (1967) observes that the high intensity of the first harmonic, H\(_1\), is the most salient spectral feature of Gujarati breathy vowels. Recent spectral analysis by Bali (1999) shows that in Delhi Hindi, intervocalic VAS may be produced without aspiration. However, the H\(_1\)-H\(_2\) measures show that VAS tokens are produced with a larger open quotient and a steeper spectral tilt, while the voiced plosives show the reverse glottal configurations. This goes to say that for discrimination between intervocalic VAS and VS, H\(_1\)-H\(_2\) measures may be especially relevant as also for

\(^1\)See also Kagaya and Hirose (1975).
distinguishing breathy vowels from modal ones as in Gujarati following Fischer-Jorgensen (1967). However, at the moment there no studies on the glottal characteristics of word initial VAS in Hindi.

6.2 The relevance of measures of spectral intensity

Spectral intensity measures provide an indirect method of studying the effect of the glottal source on the acoustic signal (Bickley 1982, Fischer-Jorgensen 1967, Ladefoged and Antonanzas-Barroso 1985). Differences between the amplitudes of $H_1$, $H_2$, $A_1$, $A_2$, and $A_3$ have been shown to correlate with differences in phonation type within vowels (Hanson 1995, Hanson et al. 2001, Wayland 1998). In this study four measures of spectral tilt were taken that included difference between the amplitudes of $H_1$-$H_2$, $H_1$-$A_1$, $H_1$-$A_2$ and $H_1$-$A_3$. These spectral intensity measures together provide a measure of spectral tilt.\(^2\)

One of the observations that had been made in Dutta (2007) was that the breathy release following the VAS tends to permeate deep into the vowel. This observation however, had not been experimentally validated. As has been shown in Chapter 3.3, aspiration in VLAS temporally varies according to the place of articulation. In this context it would also be necessary to examine the effect of place of articulation on the four spectral measures. In this chapter I will show that the VAS can be differentiated from the VLS and VS due to differences in mean $H_1$-$H_2$, $H_1$-$A_1$ and $H_1$-$A_2$ measures. Over all spectral tilt does not have a significant role in distinguishing between the VAS and the unaspirated stops. Further, I will show that the aspiration duration following the VLAS is shorter than the breathy portion of the vowel following the VAS by about 10 percent. Comparing the results from the effect of place of articulation on the duration of aspiration following VLAS with the breathy portion following VAS shows that place of articulation does not have any effect on the spectral

\(^2\)In a recent publication Mikuteit and Reetz (n.d.) suggest a measure of both voiceless and voiced aspiration called Superimposed Aspiration which is based on visual inspection of the waveform and spectrogram.
measures. This chapter is organized as follows. In section 5.3, I discuss the effect of VAS, VLS and VS on the four spectral intensity measures at five points in the vowel. Section 5.4 is a discussion of the contributions of each of the spectral measures in differentiating between the VAS and the unaspirated stops (VLS, VS). Section 5.5 is a summary of the major conclusions.

6.3 Spectral intensity measures

In this section, results from the $H_1-H_2$, $H_1-A_1$, $H_1-A_2$ and $H_1-A_3$ measures will be discussed. VLAS stops are aspirated between 10 and 30 percent of the vowel. This implies that there is no fundamental component present during this portion of aspiration. All of the spectral measures used in this study are direct measurements of spectral intensity differences between the fundamental, the second harmonic and the amplitudes of the first, second and third formant peaks. Therefore in this part of the study only comparisons between the VAS and the unaspirated stops will be made for all the four spectral measures.

6.3.1 $H_1-H_2$

A one-way ANOVA was conducted for all speakers with $H_1-H_2$ values at 10, 30, 50, 70 and 90 percent of the vowel as dependent variable and stop type as an independent factor. For subject GA, significant difference between VAS and the unaspirated stops obtained till 30 percent of the vowel at $p<0.001$. At 10 and 30 percent of the vowel the difference between VLS and VS was found to be not significant. At 50 percent of the vowel the difference between the three stops (VLS, VS and VAS) became insignificant. However, at 70 and 90 percent of the vowel the $H_1-H_2$ means were significantly lower than the VLS and VS put together. Tukey post-hoc comparisons confirm these results. The over all pattern for $H_1-H_2$ means at the 5 points in the vowel are:

1. 10 percent: VAS > VS = VLS
2. 30 percent: VAS > VS = VLS

3. 50 percent: VAS = VS = VLS

4. 70 percent: VS = VLS > VAS

5. 90 percent: VS = VLS > VAS

Subject PB shows a significant effect of VAS on the H₁-H₂ means. ANOVA results show that VAS have significantly higher H₁-H₂ till 70 percent of the vowel at p<0.001 than the VLS and VS put together. The difference between the VLS and VS is not significant till 50 percent of the vowel. At 70 percent of the vowel, the difference between VLS and VS is marginally significant (p=0.028) with VS > VLS. Tukey post-hoc comparisons confirm these results. The overall pattern for H₁-H₂ means at the 5 points in the vowel are:

1. 10 percent: VAS > VS = VLS

2. 30 percent: VAS > VS = VLS

3. 50 percent: VAS > VS = VLS

4. 70 percent: VAS > VS > VLS

5. 90 percent: VAS = VS = VLS

Subject RM also shows a significant effect of VAS on the H₁-H₂. ANOVA results show that VAS have significantly higher H₁-H₂ till 70 percent of the vowel at p<0.001 than the VLS and VS put together. The difference between the VLS and VS is not significant till 50 percent of the vowel. At 70 percent of the vowel, the difference between VLS and VS is marginally significant (p=0.025) with VLS > VS. At 90 percent the difference between VAS and the VLS and VS combined is significant at the p<0.001 with VAS > VLS = VS. Tukey post-hoc comparisons confirm these results. The overall pattern for H₁-H₂ means at the 5 points in the vowel are:
1. 10 percent: VAS > VS = VLS
2. 30 percent: VAS > VS = VLS
3. 50 percent: VAS > VS = VLS
4. 70 percent: VAS > VLS > VS
5. 90 percent: VAS > VLS = VS

Subject SD shows a significant effect of VAS on the $H_1$-$H_2$. ANOVA results show that VAS have significantly higher $H_1$-$H_2$ till 70 percent of the vowel at $p<0.001$ than the VLS and VS put together. The difference between the VLS and VS is not significant throughout the vowel. Tukey post-hoc comparisons confirm these results. The overall pattern for $H_1$-$H_2$ means at the 5 points in the vowel are:

1. 10 percent: VAS > VS = VLS
2. 30 percent: VAS > VS = VLS
3. 50 percent: VAS > VS = VLS
4. 70 percent: VAS > VS = VLS
5. 90 percent: VAS > VS = VLS

Subject SV also shows a significant effect of VAS on the $H_1$-$H_2$. ANOVA results show that VAS have significantly higher $H_1$-$H_2$ till 50 percent of the vowel at $p<0.001$ than the VLS and VS. Between the VLS and VS, the mean difference is significant from 10 till 30 percent at $p<0.05$ and with VS > VLS. At 70 and 90 percent of the vowel the $H_1$-$H_2$ differences between the VAS, VS and VLS are not significant for this speaker. Tukey post-hoc comparisons confirm these results. The overall pattern for $H_1$-$H_2$ means at the 5 points in the vowel are:
1. 10 percent: VAS > VS > VLS
2. 30 percent: VAS > VS > VLS
3. 50 percent: VAS > VS = VLS
4. 70 percent: VAS = VS = VLS
5. 90 percent: VAS = VS = VLS

Several observations can be made about the $H_1-H_2$ differences between the three stops. Except for speaker GA, all other speakers show that the $H_1-H_2$ means for VAS are significantly higher till 50 percent of the vowel than the unaspirated stops. We can also see that for three of the five speakers (PB, RM and SD) the difference between VLS and VS is not significant till 50 percent of the vowel. Based on these results it can be concluded (for three speakers) that the open quotient for VAS stays significantly high till 50 percent of the vowel. The patterns of the mean $H_1-H_2$ differences are shown in Fig. 6.1.

The mean patterns and the conclusions that have been drawn on the basis of the same, however, do not come without qualifications. As can be seen in Fig. 6.2, the data from $H_1-H_2$ show that there is variation between and within speakers. One pattern that can be adduced from Fig. 6.2, however, is the steady nature of the $H_1-H_2$ values for the unaspirated (VLS, VS) stops. This pattern can be seen in speakers PB, SV and RM and to an extent in speaker SD as well till 70 percent of the vowel. The medians and upper and lower quartiles suggest that the VLS and VS pattern together and help establish a baseline for what could be considered $H_1-H_2$ values for ‘modal’ vowels. For speaker PB and SD it is also clear that the $H_1-H_2$ values for VAS do not overlap with the values for the unaspirated stops.
Figure 6.1: Effect of Stop Type on H₁-H₂
Figure 6.2: Box plots show variation in $H_1$-$H_2$ values and overlapping.
6.3.2 $H_1$-$A_1$

A one-way ANOVA was conducted for all speakers with $H_1$-$A_1$ values at 10, 30, 50, 70 and 90 percent of the vowel as dependent variable and stop type as an independent factor.

For subject GA, significant difference between VAS and the unaspirated stops obtained till 90 percent of the vowel at $p<0.001$. At 10 percent of the vowel the difference between VLS and VS was found to be significant with VS > VLS. Following this the difference between VS and VLS was not significant. Tukey post-hoc comparisons confirm these results. The overall pattern for $H_1$-$A_1$ means at the 5 points in the vowel are:

1. 10 percent: VAS > VS > VLS
2. 30 percent: VAS > VS = VLS
3. 50 percent: VAS > VS = VLS
4. 70 percent: VAS > VS = VLS
5. 90 percent: VAS > VS = VLS

Subject PB shows a significant of VAS on the $H_1$-$A_1$ means. ANOVA results show that VAS have significantly higher $H_1$-$A_1$ till 70 percent of the vowel at $p<0.001$ than the VLS and VS. The difference between the VS and VLS is not significant till 30 percent of the vowel such that VS > VLS. At 90 percent of the vowel all the stops behave similarly. Tukey post-hoc comparisons confirm these results. The overall pattern for $H_1$-$A_1$ means at the 5 points in the vowel are:

1. 10 percent: VAS > VS > VLS
2. 30 percent: VAS > VS > VLS
3. 50 percent: VAS > VS = VLS
4. 70 percent: VAS > VS > VLS
5. 90 percent: VAS = VS = VLS

Subject RM also shows a significant effect of VAS on the $H_{1-A_1}$. ANOVA results show that VAS have significantly higher $H_{1-A_1}$ till 90 percent of the vowel at $p<0.001$ than the VLS and VS put together. The difference between the VLS and VS is not significant till 90 percent of the vowel. Tukey post-hoc comparisons confirm these results. The overall pattern for $H_{1-A_1}$ means at the 5 points in the vowel are:

1. 10 percent: VAS > VS = VLS
2. 30 percent: VAS > VS = VLS
3. 50 percent: VAS > VS = VLS
4. 70 percent: VAS > VS = VLS
5. 90 percent: VAS > VLS = VS

Subject SD shows a significant effect of VAS on the $H_{1-A_1}$. ANOVA results show that VAS have significantly higher $H_{1-A_1}$ till 70 percent of the vowel at $p<0.001$ than the VLS and VS put together. The difference between the VLS and VS is significant only at 10 percent of the vowel with the VS < VLS. Tukey post-hoc comparisons confirm these results. The overall pattern for $H_{1-A_1}$ means at the 5 points in the vowel are:

1. 10 percent: VAS > VS > VLS
2. 30 percent: VAS > VS = VLS
3. 50 percent: VAS > VS = VLS
4. 70 percent: VAS > VS = VLS
5. 90 percent: VAS = VS = VLS
Subject SV also shows a significant effect of VAS on the H₁-A₁. ANOVA results show that VAS have significantly higher H₁-A₁ till 50 percent of the vowel at p<0.001 than the VLS and VS. Between the VLS and VS, the mean difference is significant from 10 till 30 percent at p<0.001 and with VS > VLS. At 90 percent of the vowel the H₁-A₁ differences between the VAS and the unaspirated, VS and VLS become significant for this speaker. Tukey post-hoc comparisons confirm these results. The overall pattern for H₁-A₁ means at the 5 points in the vowel are:

1. 10 percent: VAS > VS > VLS
2. 30 percent: VAS > VS > VLS
3. 50 percent: VAS > VS = VLS
4. 70 percent: VAS = VS = VLS
5. 90 percent: VAS > VS = VLS

Two broad observations can be made on the basis of the results from the H₁-A₁ means. First, the H₁-A₁ values tend to be significantly higher for the VAS till about 70 percent of the vowel, which is further than the H₁-H₂ means. However, consonant voicing in VAS and VS tends to have an effect on the peak amplitude of the first formant. The patterns suggest that at least in the initial 10-30 percent of the vowel the VS have significantly higher means than VLS. These patterns can be seen in Fig. 6.3. In Fig. 6.4 the overlapping patterns of the values can be seen. The variation in the H₁-A₁ patterns between the stops shows considerable overlap despite the significant mean values. It is likely that the values at around 50 percent of the vowel, reflect a voice quality that is transient between breathy and modal phonation. Nonetheless, these patterns show that initially (Fig. 6.3) till about 30 percent of the vowel, based on the H₁-A₁ values the vowels can be differentiated successfully.
Figure 6.3: Effect of Stop Type on $H_1$-A$_1$.
Figure 6.4: Box plots show variation in $H_1-A_1$ values.
6.3.3  \(H_1-A_2\)

A one-way ANOVA was conducted for all speakers with \(H_1-A_2\) values at 10, 30, 50, 70 and 90 percent of the vowel as dependent variable and stop type as an independent factor.

For subject GA, significant difference between VAS and the unaspirated stops obtained till 50 percent of the vowel at \(p<0.001\) and at \(p=0.002\) at 70 percent of the vowel. Only 10 percent of the vowel the difference between VLS and VS was found to be significant with VS > VLS. Following this the difference between VS and VLS was not significant. Tukey post-hoc comparisons confirm these results. The over all pattern for \(H_1-A_2\) means at the 5 points in the vowel are:

1. 10 percent: VAS > VS > VLS
2. 30 percent: VAS > VS = VLS
3. 50 percent: VAS > VS = VLS
4. 70 percent: VAS > VS = VLS
5. 90 percent: VAS = VS = VLS

Subject PB shows a significant of VAS on the \(H_1-A_2\) means. ANOVA results show that VAS have significantly higher \(H_1-A_2\) till 70 percent of the vowel at \(p<0.001\) than the VLS and VS. The difference between the VS and VLS is significant till 50 percent of the vowel such that VS > VLS. At 90 percent of the vowel the VAS are significantly higher than the VLS but not the VS. Tukey post-hoc comparisons confirm these results. The over all pattern for \(H_1-A_2\) means at the 5 points in the vowel are:

1. 10 percent: VAS > VS > VLS
2. 30 percent: VAS > VS > VLS
3. 50 percent: VAS > VS > VLS
4. 70 percent: VAS > VS = VLS

5. 90 percent: VAS = VS, VAS > VLS

Subject RM also shows a significant effect of VAS on the H$_1$-A$_2$. ANOVA results show that VAS have significantly higher H$_1$-A$_2$ till 70 percent of the vowel at p<0.001 than the VLS and VS put together. At 90 percent of the vowel the VAS and VLS difference is significant but not the VAS and VLS difference. The difference between the VS and VLS is significant at 10 percent of the vowel. Tukey post-hoc comparisons confirm these results. The overall pattern for H$_1$-A$_2$ means at the 5 points in the vowel are:

1. 10 percent: VAS > VS > VLS
2. 30 percent: VAS > VS = VLS
3. 50 percent: VAS > VS = VLS
4. 70 percent: VAS > VS = VLS
5. 90 percent: VAS = VS, VAS > VLS

Subject SD shows a significant effect of VAS on the H$_1$-A$_2$. ANOVA results show that VAS have significantly higher H$_1$-A$_2$ till 90 percent of the vowel at p<0.001 than the VLS and VS put together. The difference between the VLS and VS is significant till 50 percent of the vowel with the VS > VLS. Tukey post-hoc comparisons confirm these results. The overall pattern for H$_1$-A$_2$ means at the 5 points in the vowel are:

1. 10 percent: VAS > VS > VLS
2. 30 percent: VAS > VS > VLS
3. 50 percent: VAS > VS > VLS
4. 70 percent: VAS > VS = VLS
5. 90 percent: VAS = VS = VLS

Subject SV also shows a significant effect of VAS on the H$_1$-A$_2$. ANOVA results show that VAS have significantly higher H$_1$-A$_2$ till 50 percent of the vowel at p<0.001 than the VLS and VS. Between the VLS and VS, the mean difference is significant from 10 till 70 percent at p<0.001 and with VS > VLS. At 90 percent of the vowel the H$_1$-A$_2$ differences between the VAS and VS are not significant, however both these stops have significantly higher means than the VLS. Tukey post-hoc comparisons confirm these results. The overall pattern for H$_1$-A$_2$ means at the 5 points in the vowel are:

1. 10 percent: VAS > VS > VLS
2. 30 percent: VAS > VS > VLS
3. 50 percent: VAS > VS > VLS
4. 70 percent: VAS = VS > VLS
5. 90 percent: VAS = VS > VLS
Figure 6.5: Effect of Stop Type on $H-A_2$
Figure 6.6: Box plots show variation in $H_1-A_2$ values.
A conservative generalization that can be made on the bases of these results is that the VAS tend to show higher mean $H_1-A_2$ values at least till 50 percent of the vowel for all speakers. Except for Subject SV, all other speakers show a significant difference in mean $H_1-A_2$ values between VS and VLS at 10 percent of the vowel. These patterns can be seen in 6.5. Except for subject SD at 30 percent, all other subjects however, do show considerable overlap in the range and values of $H_1-A_2$ between the three stops despite significant mean differences and high median values for the VAS till about 50 percent of the vowel (see Fig. 6.6). At 90 percent of the vowel the values tend to overlap and become coextensive.

6.3.4 $H_1-A_3$

A one-way ANOVA was conducted for all speakers with $H_1-A_3$ values at 10, 30, 50, 70 and 90 percent of the vowel as dependent variable and stop type as an independent factor.

For subject GA, significant difference between VAS and the unaspirated stops obtained till 30 percent of the vowel at $p<0.001$ and at $p=0.05$ at 70 percent of the vowel. Only at 10 percent of the vowel the difference between VLS and VS was found to be significant with $VS > VLS$. Following this the difference between VS and VLS was not significant. Tukey post-hoc comparisons confirm these results. The overall pattern for $H_1-A_3$ means at the 5 points in the vowel are:

1. 10 percent: VAS > VS > VLS
2. 30 percent: VAS > VS = VLS
3. 50 percent: VAS = VS, VAS > VLS
4. 70 percent: VAS > VS = VLS
5. 90 percent: VAS = VS = VLS

Subject PB shows a significant effect of VAS on the $H_1-A_3$ means. ANOVA results show that VAS have significantly higher mean $H_1-A_3$ till 70 percent of the vowel at $p<0.001$ than
the VLS and VS. The difference between the VS and VLS is significant till 30 percent of the vowel such that VS > VLS. At 90 percent of the vowel the differences in mean H₁-A₃ are not significant between the stops. Tukey post-hoc comparisons confirm these results. The over all pattern for H₁-A₂ means at the 5 points in the vowel are:

1. 10 percent: VAS > VS > VLS
2. 30 percent: VAS > VS > VLS
3. 50 percent: VAS > VS = VLS
4. 70 percent: VAS > VS = VLS
5. 90 percent: VAS = VS = VLS

Subject RM also shows a significant effect of VAS on mean H₁-A₃. ANOVA results show that VAS have significantly higher H₁-A₃ till 50 percent of the vowel at p<0.001 than the VLS and VS put together. At 70 percent of the vowel the VAS and VS difference is significant but not the VAS and VLS difference. At 90 percent all differences in mean are not significant. The difference between the VS and VLS is significant only at 10 percent of the vowel. Tukey post-hoc comparisons confirm these results. The over all pattern for H₁-A₃ means at the 5 points in the vowel are:

1. 10 percent: VAS > VS > VLS
2. 30 percent: VAS > VS = VLS
3. 50 percent: VAS > VS = VLS
4. 70 percent: VAS = VLS, VAS > VS
5. 90 percent: VAS = VS = VLS

100
Subject SD shows a significant effect of VAS on the $H_1$-$A_3$. ANOVA results show that VAS have significantly higher $H_1$-$A_3$ till 70 percent of the vowel at $p<0.001$ than the VLS and VS put together. The difference between the VLS and VS is significant till 50 percent of the vowel with the VS $>$ VLS. At 90 percent of the vowel only the difference in mean for VAS and VS is significant. Tukey post-hoc comparisons confirm these results. The over all pattern for $H_1$-$A_3$ means at the 5 points in the vowel are:

1. 10 percent: VAS $>$ VS $>$ VLS
2. 30 percent: VAS $>$ VS $>$ VLS
3. 50 percent: VAS $>$ VS $>$ VLS
4. 70 percent: VAS $>$ VS $=$ VLS
5. 90 percent: VS $>$ VLS, VAS $=$ VS $=$ VLS

Subject SV also shows a significant effect of VAS on the $H_1$-$A_3$. ANOVA results show that VAS have significantly higher mean $H_1$-$A_3$ till 50 percent of the vowel at $p<0.001$ than the VLS and VS. Between the VLS and VS, the mean difference is significant from 10 till 30 percent at $p<0.001$ and with VS $>$ VLS. From 70 till 90 percent of the vowel the $H_1$-$A_3$ differences between the stops are not significant. Tukey post-hoc comparisons confirm these results. The over all pattern for $H_1$-$A_3$ means at the 5 points in the vowel are:

1. 10 percent: VAS $>$ VS $>$ VLS
2. 30 percent: VAS $>$ VS $>$ VLS
3. 50 percent: VAS $>$ VS $=$ VLS
4. 70 percent: VAS $=$ VS $=$ VLS
5. 90 percent: VAS $=$ VS $=$ VLS
Figure 6.7: Effect of Stop Type on H₁-A₃
Figure 6.8: Box plots show variation in $H_1$-$A_3$ values.
The results from the H$_1$-A$_3$ means show that at least for four subjects the differences in mean between VAS and the unaspirated stops are significant till 50 percent of the vowel. The difference between the means for VLS and VS can be seen at least till 10 percent of the vowel with the VS showing higher means than the VLS. The patterns for the mean of H$_1$-A$_3$ can be seen for the three stops in 6.7. The boxplots in 6.8, however, show considerable overlap in the H$_1$-A$_3$ values between the stops, despite the higher medians for the VAS till 50 percent for four out of the five speakers. At 90 percent of the vowel we also that the values become coextensive for all the stops for all subjects.

The results from the spectral intensity measures show that significant differences do obtain between the means for the VAS and the unaspirated stops till about 50 percent of the vowel. Based on these four measures of spectral intensity and the correlation between these four measures, and voice and vowel quality distinctions it is possible to conclude that the vowel following the VAS is distinct from the vowel following the unaspirated stops (VLS, VS). The vowel following the VAS tends to maintain a ‘breathy’ quality till about 50 percent of the vowel. In comparison to durations of voiceless aspiration as stated in section 3.3. of chapter 3 we can also conclude that the breathy portion following the VAS tends to be longer than the aspiration portion following the VLAS. This trend confirms to an extent the observation made in Dutta (2007) that the breathy release following VAS tends to permeate deep into the vowel compared to the duration of aspiration following VLAS.
Figure 6.9: Box plot: Percentage of aspiration duration relative to Vowel Duration $T_{total}$

6.4 Contributions of the individual spectral intensity measures

As can be seen in Fig. 6.9 the median percentage of aspiration following VLAS is between 20 and 40 percent depending on the subject and also the place of articulation. Compared to the percentage of aspiration in VLAS, for most of the speakers breathy release tends to extend till about 30-50 percent of the vowel. In addition, unlike the effect of place of articulation on aspiration duration following VLAS, no effect of place of articulation was found for all the speakers on any of the four spectral intensity measures (see Fig. 6.10). Fig. 6.10 shows the results for the effect of place of articulation on $H_1-A_1$. The effect of place of articulation on the other three spectral measures were comparable to the effect on
H₁-A₁. This observation allows us to conclude that the volume of the oral cavity has a significant effect on the aspiration duration for VLAS but has no effect on the duration of the breathy release following the VAS.

Although the mean differences for the four spectral measures are significantly higher between the VAS and the unaspirated stops, there is considerable overlap between the values for the three stops. Therefore, it became necessary to ascertain the individual contribution of each spectral intensity measure in comparable terms. In order to accomplish this comparison, first a comparison between the distribution of means for the values at 10 and 30 percent was conducted. Second a comparison between the means for values at 10, 30, and 50 percent was conducted. The distribution of the mean values for these comparisons can be seen in Fig. 6.11 and Fig. 6.12, respectively. The maximally distinct or separated distributions for each spectral measure were considered contributing maximally as well towards a breathy/modal
vowel quality distinction following the stops. These contributing spectral measures and their relation to other measures are summarized in Table 6.1 and Table 6.2. The measures that appear to the left of the column are observed to be better in distinguishing the stops based on comparisons of the boxplots in Fig. 6.11 and Fig. 6.12. Tables 6.1 and 6.2 show that $H_1-H_2$ as a measure of Open Quotient is maximally distinct between VAS and the unaspirated stops for two speakers (PB, SD). $H_1-A_1$, a measure of first formant bandwidth tends to maximally distinguish between the VAS and the unaspirated for two speakers (RM, SV). $H_1-A_2$, a measure of skewness of the glottal pulse, maximally distinguishes between the VAS and the unaspirated stops only for one speaker (GA). These observations allow us to conclude that for distinctions based on the quality of the vowel following the VAS and unaspirated stops, the difference in the amplitudes of the first and second harmonic, first harmonic and the peak amplitude of the first formant, and second formant are relevant acoustic features.

<table>
<thead>
<tr>
<th>Subject</th>
<th>Spectral Measures</th>
</tr>
</thead>
<tbody>
<tr>
<td>GA</td>
<td>$H_1-A_2$, $H_1-A_1$</td>
</tr>
<tr>
<td>PB</td>
<td>$H_1-H_2$, $H_1-A_1$, $H_1-A_2$</td>
</tr>
<tr>
<td>RM</td>
<td>$H_1-A_1$, $H_1-H_2$, $H_1-A_2$</td>
</tr>
<tr>
<td>SD</td>
<td>$H_1-H_2$, $H_1-A_2$, $H_1-A_3$</td>
</tr>
<tr>
<td>SV</td>
<td>$H_1-A_2$</td>
</tr>
</tbody>
</table>

Table 6.1: Maximally distinct distributions and contributing factors towards a distinction in vowel quality. Means of values at 10 and 30 percent of the vowel.

<table>
<thead>
<tr>
<th>Subject</th>
<th>Spectral Measures</th>
</tr>
</thead>
<tbody>
<tr>
<td>GA</td>
<td>$H_1-A_2$, $H_1-A_1$</td>
</tr>
<tr>
<td>PB</td>
<td>$H_1-H_2$, $H_1-A_1$, $H_1-A_2$</td>
</tr>
<tr>
<td>RM</td>
<td>$H_1-A_1$</td>
</tr>
<tr>
<td>SD</td>
<td>$H_1-H_2$, $H_1-A_1$, $H_1-A_2$</td>
</tr>
<tr>
<td>SV</td>
<td>$H_1-A_1$, $H_1-A_2$</td>
</tr>
</tbody>
</table>

Table 6.2: Maximally distinct distributions and contributing factors towards a distinction in vowel quality. Means of values at 10, 30 and 50 percent of the vowel.
Figure 6.11: Mean spectral intensity for values at 10 and 30 percent of the vowel
Figure 6.12: Mean spectral intensity for values at 10, 30 and 50 percent of the vowel
6.5 Summary, discussion and conclusion

The results from the study of spectral intensity measures suggest the VAS tend to be produced with greater difference in the amplitudes of the first and second harmonic compared to the unaspirated stops. VAS are 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 between the four measures it can also be concluded that the breathy portion following the VAS tends to permeate till about 30-50 percent of the following vowel. In comparison with the duration of aspiration following VLAS, the breathy release portion is longer. Overall spectral tilt measured by way of the difference in the amplitude of the first harmonic and the peak amplitude of the third formant does not help distinguish the VAS from the unaspirated stops as well as the other three measures. Further, we have shown that the aspiration duration following the VLAS tends to vary as a function of the place of articulation of the stops. This is due to variable oral tract length and consequently volume during the closure portion of the VLAS. Place of articulation does not, however, have any effect on the breathy portion of the VAS. This goes to show that aerodynamic factors that are responsible for the varying duration of aspiration following VLAS are not at play in the results of the spread of the breathy portion in VAS. In this respect, supraglottal configurations have no effect on the duration of the breathy portion following the VAS. Breathy release following the VAS are due to particular laryngeal configurations. The articulatory correlates of the spectral measures found to be most responsible for distinctions between VAS and the unaspirated stops suggest Hindi VAS are produced with a larger Open Quotient. This implies that the vocal folds during the initial 30-50 percent of the vowel are open for longer durations compared to unaspirated stops. The results from the H₁-A₂ measures also suggest that articulatorily Hindi VAS are produced with comparatively more abrupt closing of the vocal folds than the unaspirated stops.
Chapter 7

Conclusions, implications and further research

7.1 Overview

In this study we have been able to show that several acoustic distinctions between the four stops in Hindi can be seen at the durational, $f_0$, and spectral level. Durationally, VLT differences between the voiced stops (VAS and VS) have been shown to be statistically significant. While Closure Duration data could not be collected in the utterance initial position [U], it has been shown that VLAS are produced with shorter closures in comparison to VLS. Studies on VLT (Schiefer 1992, Dutta 2007) have shown that VAS are produced with lower VLT than the VS. One of the primary research questions for this study was to investigate whether a parallel could be found in the duration of closure (CD) in VLS and VLAS. Temporally, voiceless and voiced stops in Hindi show stop manner dependent duration patterns of closure and closure voicing (VLT). Based on the manner of articulation, the duration of the CD becomes predictable irrespective of whether these stops are voiced or voiceless. The duration patterns for CD suggest that there is a parallel distribution of CD dependent on the manner of articulation. Aspirated stops have shorter durations compared to unaspirated stops.

The duration of aspiration for VLAS is such that the aspiration portion is nearly 20-30 percent of the vowel and is dependent on the place of articulation of the stops in VLAS. Compared to this result, the duration of breathy/murmur following the VAS is relatively longer and is not dependent on the place of articulation of the preceding stop. These results confirm the initial observations from Dutta (2007) that the breathy/murmured portion fol-
lowing the VAS permeate for a longer duration than the voiceless aspiration following the VLAS. Marginal effects of the context on the durations have also been shown.

The $f_0$ contours of the vowels following the four stops in Hindi show that for a majority of the speakers in this study, VAS have lower mean $f_0$ values till about 20-30 percent of the vowel. VS tend to have comparably higher $f_0$ values than the VAS in the beginning of the vowels. VLAS and VLS stops tend to have higher mean $f_0$ values initially when compared to the voiced stops. In this study, effect of these $f_0$ perturbation patterns that are dependent on stop type have been shown to persist till about 30 percent of the vowel, at least for VAS. For four of the five speakers it has been shown that a three way categorical distinction can be made on the basis of the $f_0$ patterns in the beginning 20-30 percent of the vowel.

Spectral measures based on measurement of relative amplitudes of $H_1$, $H_2$, $A_1$, $A_2$ and $A_3$ show that stop type has a significant effect on measures of spectral tilt. Amongst the four measures that were used in this study, $H_1$-$H_2$, an indirect measure of Open Quotient, $H_1$-$A_1$, a measure of the first formant bandwidth, and $H_1$-$A_2$, an indirect measure of the abruptness of the vocal fold closure or skewness are the most reliable measures that help distinguish the VAS from the unaspirated stops (VS, VLS). $H_1$-$A_3$, a measure of the over all spectral tilt is not a reliable cue for discrimination in this data. The spectral measures also indicate that till 30-50 percent of the vowel the VAS can be categorized as having breathy phonation. Comparing the duration of aspiration in VLAS and the breathy portion following the VAS, we can conclude that the breathy phonation is comparably longer. In addition, we have shown that the aspiration portion following the VLAS is dependent on the place of articulation of the stop, with velar VLAS tending to have longer durations of aspiration compared to other places of articulation. Place of articulation does not have an effect on the breathy portion following the VAS. This suggests that place of articulation dependent supraglottal configurations do not have an effect on the breathy portion following the VAS. These findings suggest that while phonologically VAS could be categorized as aspirated stops, where the particular feature that describes this distinctive release is phonetically realized
quite differently from the aspiration following the VLAS. In order to capture the spread of the breathy/murmured release, the aspiration following the VAS would have to be laryngeally specified to be either delayed onset of modal voicing or as a distinctive release following the stop. The durational differences in the phonetic outcome of this phonologically distinctive feature can be better explained if all of the cues responsible in making the four-way stops contrasts are understood in part to function as enhancing cues (Keyser and Stevens 2006). In the following section, I discuss the relevance of the results from this study for a complete understanding of the four-way stop contrasts in Hindi. I also discuss the implications of these findings towards a phonological account of these stop contrasts that takes into close consideration the phonetic facts. Further, I develop the idea that complex cue interaction and correlation leads to cues acting in favor of each other in order to preserve the existing contrasts.

7.2 Implications and further research

Based on the results from the durational features such as VLT and CD it can be said that the standard view of stop production in Hindi needs to be amended in order to better explain the durational patterns for VLT and CD. The CD and VLT durations vary according to stop manner in Hindi, with aspirated stops tending to have shorter closures compared to unaspirated stops. The standard view suggests that the distinction between VAS and VS is primarily a distinction in the breathy release following the VAS. The durational features that have been examined in this study support the view that differences in the duration of closure between the unaspirated and aspirated stops could indeed be relevant for making a contrast between the stop types. Results from the effect of the stop type on $f_0$ also show that $f_0$ following VAS tends to be lower compared to the voiceless stops till about 30 percent of the vowel. These results confirm findings from previous research on the effect of stops on $f_0$ (Schiefer 1986, Ohala 1979). VS also tend to lower $f_0$ as compared to voiceless
stops. These findings suggest that $f_0$ can also act as a valuable cue for stop identification. Based on the differences in VLT between VAS and VS it can also be said the correlation between voicing and $f_0$ lowering is not entirely accurate. VLT for VAS is consistently shorter than VLT for VS. In that respect, a comparatively less voiced VAS tends to lower $f_0$ more than the VS. Hence it can be argued that the causes for $f_0$ lowering following VAS are not entirely due to the universal tendency for $f_0$ lowering following voiced stops. Rather, the breathy/murmured portion following the VAS is also correlated with the duration of the $f_0$ lowering following these stops. It can, hence, be argued that breathy/murmured portion is the primary reason behind the $f_0$ lowering patterns, rather than the universal tendency for voiced stops to lower $f_0$. An alternative analysis that can be formulated based on these results is also dependent on the severity of the $f_0$ lowering following the VAS. Noting that the $f_0$ lowering is significantly lower following VAS than VS it is probable that both the universal tendency and the breathy/murmured mode of phonation are partly responsible for the further lowering of the $f_0$ following the VAS. The spectral intensity measures show that the VAS are distinct in that they are produced with a larger Open Quotient, first formant bandwidth and skewness of the glottal pulse. It has also been shown that the breathy portion following the VAS tends to be longer than the aspiration portion following the VLAS. In terms of aspiration as a contrastive feature, this result confirms Ladefoged’s (Ladefoged 1971) earlier view that the Hindi VAS are produced with a breathy release that is distinct from voiceless aspiration.

These results suggest that the stop distinctions in Hindi are a result of several acoustic features, including durational as well as spectral features. In addition, $f_0$ also plays a crucial role in the acoustic distinction between the voiced stops and the voiceless stops. The results from this study also provide a basis for understanding the complexity of cue interaction in phonological systems with more than three-way contrasts. In light of the results from this study, we can see that a four-way phonological contrast can lead up to a complex interaction between the acoustic cues that could be eventually responsible for either defining a contrast
or maintaining a contrast. Given the level of interaction that can be seen from these results, it can be well-motivated to suggest that perhaps the latter is the case. In section 3.5, I have discussed the relevance of the theory of enhancement towards a better understanding of acoustic cue interaction in Hindi. I have argued that stop distinctions in Hindi could be a cumulative result of the several cues functioning together. However, in order to be able to provide evidence in favour of a theory of enhancement it is essential first to identify those features that threaten to obliterate a contrast. As can be seen from the CD results, there is a parallel distribution of closure durations dependent on the manner of articulation; aspirated stops have shorter closures than unaspirated stops. The voicing during closure for VS and VAS in this respect, is a defining feature in terms of maintaining a contrast between voiced and voiceless stops. Voicing, as is universally attested can be affected by contextual and coarticulatory effects. These effects can in turn result in the obliteration of contrasts between voiced and voiceless stops. In the context of Hindi, then, as specified by Keyser and Stevens (2006), certain language specific features can be argued to be in place to enhance the voicing contrast, for instance. Thus, compensatory mechanisms could be responsible for the acoustic outcomes of the phonological contrasts.

Further research in this direction will have to involve conducting perception experiments that would provide us with an insight into the contribution of the various acoustic features in making the stop distinctions possible in Hindi. This being an acoustic study, CD measurements could not be made for Utterance Initial [U] stops. Further research that examines the relevance of CD in this context will provide important insights into the role of CD in stop distinctions in Hindi. The laboratory environment and the speech material that was recorded did not lend itself to an understanding of the effect prosodic context on the durational and f0 features. Future research in this direction will help understand in greater detail the effect of varying prosodic contexts.

The results from the spectral intensity study also suggest that disparate articulatory causes produce the singular perceptual effect of breathiness in the speech stream. This can
form the basis for conducting a study that can examine the effect of linguistic background on the categorical nature of perceptual voice quality judgements. If indeed speakers’ linguistic background has an effect on their perceptual judgements of voice quality, we should expect American English listeners to behave differently from the Hindi and Gujarati listeners. Further, since Gujarati also has a contrast between modal and breathy vowels, in addition to the breathy and modal stops, we can expect Gujarati listeners to behave differently from Hindi as well as the American English listeners. The results from this study will have several implications, both for a better understanding of the use of spectral cues for making contrasts and for an accurate estimation of voice quality judgements for speakers who do not employ breathiness contrastively. First, this study will show whether speakers whose languages employ breathiness contrastively will differ in their breathiness judgements from American English speakers. Secondly, it will also provide us with insights into the variety of perceptual ratings for breathiness that speakers from all three languages might exhibit. This in turn will prove to be crucial from a psychoacoustic and speech therapy perspective, since critical distinctions between normal and disordered speech could be understood on the basis of a set of parameters that govern speaker judgements of voice quality. Acoustic techniques have benefited our understanding of disordered breathiness as in Hillenbrand, Cleveland and Erickson (1994), the same techniques have also proven to be relevant for studying breathy voicing in vowels, stops and affricates giving us an insight into the production of contrastive breathiness (Bali 1999, Bickley 1982, Blankenship 1998). Results from this study will contribute further towards the utilization of acoustic phonetic techniques to study both normal and disordered speech.
Appendix 1: Frame Sentences

Utterance initial [U]

(1) **Target Word** ka: mətəb vo somk₇:a: nəfi
**Target Word** GEN meaning he understand NEG
He didn’t understand the meaning of (the) **Target Word**

Phrase initial [I]

(2) sunije, **Target Word** ka: mətəb bətaːje ?
Excuse me, **Target Word** GEN meaning tell
Excuse me, could (you) tell (me) the meaning of the **Target Word**

Phrase medial [M]

1. (3) məːne mofiːn ko kʰaːːs Target Word bolte hue suna:
   I Mohan ACC **special** Target Word tell heard say
   I heard Mohan say “**special** Target Word”

2. (4) məːne mofiːn ko laːl Target Word bolte hue suna:
   I Mohan ACC **red** Target Word tell heard say
   I heard Mohan say “**red** Target Word”
# Appendix 2: Word List

<table>
<thead>
<tr>
<th></th>
<th>Voiceless</th>
<th></th>
<th>Voiced</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Plain</td>
<td>Aspirated</td>
<td>Plain</td>
<td>Aspirated</td>
</tr>
<tr>
<td>Labial</td>
<td>pa:r, pa:g, pa:k</td>
<td>pʰa:g, pʰa:l, pʰa:t</td>
<td>ba:l, ba:g, ba:t</td>
<td>bʰa:r, bʰa:g, bʰa:p</td>
</tr>
<tr>
<td>Target Word</td>
<td>Meaning</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-------------</td>
<td>---------</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ka:l</td>
<td>NM. Time; period; age; era</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ka:k</td>
<td>NM. Crow</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ka:g</td>
<td>NM. Cork</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ka:al</td>
<td>NF. Skin; hide</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ka:ad</td>
<td>NF. Manure; fertilizer</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ka:at</td>
<td>NF. Cot; Bedstead</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>gal</td>
<td>NM. A cheek</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>gad</td>
<td>NF. Sediment; dregs</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>gat</td>
<td>NM. Body; person</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>g^h:ag</td>
<td>NM; Adj. Cunning, shrewd (person)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>g^h:at</td>
<td>NM. River bank</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>g^h:al</td>
<td>Nonce</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>tagl</td>
<td>NM. 1. A lake. 2. rhythm</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>tag:b</td>
<td>Nonce</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>tag:al</td>
<td>NM. Metallic plate</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>da:l</td>
<td>NF. Pulse; Lentils</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>da:g</td>
<td>NM. A speck; stain</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>da:k</td>
<td>Nonce</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>d^h:ar</td>
<td>NF. An edge; sharp edge</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>d^h:ak</td>
<td>NF. Commanding/overwhelming influence; sway</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>d^h:ag</td>
<td>Nonce</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>tanl</td>
<td>NF. 1. A stock 2. Prevarication</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>təp</td>
<td>NF. Hoof</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>təb</td>
<td>Nonce</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>tʰæːt</td>
<td>NM. Pomp, splendor</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>tʰæla</td>
<td>Adj. Idle NM. Scarcity</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>tʰæp</td>
<td>Nonce</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>dəl</td>
<td>NF. Branch (of a tree)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>dək</td>
<td>NF. Mail; post</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>dəg</td>
<td>Nonce</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>dʰ aːl</td>
<td>NF. 1. Shield 2. Slope</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>dʰ aːg</td>
<td>Nonce</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>dʰ ək</td>
<td>NM. 1. A particular tree 2. Sticking to an unwelcome convention/custom</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>pəːr</td>
<td>NM. 1. The other coast/bank 2. Adv. Across; on the other side</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>pəːg</td>
<td>NM. 1. Anything boiled in sugar syrup 2. NF. A long winding turban</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>pəːk</td>
<td>Adj. Holy, sacred; pure</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>pʰəːg</td>
<td>NM. A typical song sung during Holi festival</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>pʰəːl</td>
<td>NM. A blade; ploughshare</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>pʰəːt</td>
<td>NM. A division of land</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>bəːl</td>
<td>NM. Hair</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>bəːɡ</td>
<td>NN. A garden; park</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>bəːt</td>
<td>NF. Talk</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>bʰəːr</td>
<td>NM. Load; weight; burden</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>bʰəːɡ</td>
<td>NM. Portion; part</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>bʰəːp</td>
<td>NF. Steam; vapour</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Appendix 3: Language Background Questionnaire

Note: The original language background questionnaire was typed in Devnagari script and potential subjects were asked to fill in the details. The following is a translation from the original Hindi language background questionnaire.

1. Name —__________

2. Age —__________

3. Gender (Please circle one) Female/Male

4. Native language ————

5. Ability in native language (Please say ‘yes’ or ‘no’) speak — read — write —

6. How many years have you learned your mother tongue in school? ————

7. Please write the names of other languages that you can speak, read and write and also indicate your ability in these languages.

<table>
<thead>
<tr>
<th>Language</th>
<th>Speak</th>
<th>Read</th>
<th>Write</th>
</tr>
</thead>
</table>

8. Mother’s native language ————

9. Your mother’s ability in her native language (Please say ‘yes’ or ‘no’) speak — read — write —
10. Please write the names of other languages that your mother can speak, read and write also indicate her ability in these languages.

<table>
<thead>
<tr>
<th>Language</th>
<th>Speak</th>
<th>Read</th>
<th>Write</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

11. Father’s native language

12. Your father’s ability in his native language (Please say ‘yes’ or ‘no’) speak — read — write —

13. Please write the names of other languages that your father can speak, read and write and also indicate his ability in these languages.

<table>
<thead>
<tr>
<th>Language</th>
<th>Speak</th>
<th>Read</th>
<th>Write</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
References


Dutta, I.: 2007, Correlation between voicing lead time (VLT) and $f_0$, *Proceedings of the 39th annual meeting of the Chicago Linguistics Society*.


Stevens, K. and Hanson, H.: 1994, Classification of glottal vibration from acoustic measurements, *Paper presented at the 8th Vocal Fold Physiology conference in Kurume, Japan, April 7-9*.


Indranil Dutta was born in Maligaon, Assam, in India on April 13, 1973. He graduated from the Jawaharlal Nehru University in 1997 with an B.A. degree in German and an M.A. in Linguistics. Following that he came to the University of Illinois and received a second M.A. degree and a Ph.D. in Linguistics in 2007. In between October 2006 and February 2007, he worked on Text-to-Speech research and development for Nuance Communications in Belgium. His research interests lie within the broad fields of acoustic and articulatory phonetics, speech perception, phonological theory, and speech synthesis and technologies. Aside from these interests, he is also inclined towards computational, and statistical speech and articulatory modeling, computational modeling of pronunciation, historical linguistics, language contact and general linguistic theory.

Following the completion of his Ph.D. he will be teaching at Rice University at Houston, Texas.