

Phonation Types in Marathi: An Acoustic Investigation

By

Copyright 2013

Kelly Harper Berkson

Submitted to the graduate degree program in Linguistics and the Graduate Faculty of the University of Kansas in partial fulfillment of the requirements for the degree of Doctor of Philosophy.

---

Co-Chair Dr. Jie Zhang

---

Co-Chair Dr. Allard Jongman

---

Dr. Joan Sereno

---

Dr. Harold Torrence

---

Dr. Raymond Pierotti

---

Dr. Geetanjali Tiwari

Date Defended: December 7, 2012

The Dissertation Committee for Kelly Harper Berkson  
certifies that this is the approved version of the following dissertation:

Phonation Types in Marathi: An Acoustic Investigation

---

Co-Chair Dr. Jie Zhang

---

Co-Chair Dr. Allard Jongman

Date Approved: December 7, 2012

## Abstract

This dissertation presents a comprehensive instrumental acoustic analysis of phonation type distinctions in Marathi, an Indic language with numerous breathy voiced sonorants and obstruents. Important new facts about breathy voiced sonorants, which are crosslinguistically rare, are established: male and female speakers cue breathy phonation in sonorants differently, there are an abundance of trading relations, and—critically—phonation type distinctions are not cued as well by sonorants as by obstruents.

Ten native speakers (five male, five female) were recorded producing Marathi words embedded in a carrier sentence. Tokens included plain and breathy voiced stops, affricates, nasals, laterals, rhotics, and approximants before the vowels [a] and [e]. Measures reported for consonants and subsequent vowels include duration, F<sub>0</sub>, Cepstral Peak Prominence (CPP), and corrected H1-H2\*, H1-A1\*, H1-A2\*, and H1-A3\* values. As expected, breathy voice is associated with decreased CPP and increased spectral values.

A strong gender difference is revealed: low-frequency measures like H1-H2\* cue breathy phonation more reliably in male speech, while CPP—which provides information about the aspiration noise included in the signal—is a more reliable cue in female speech. Trading relations are also reported: time and again, where one cue is weak or absent another cue is strong or present, underscoring the importance of including both genders and multiple vowel contexts when testing phonation type differences. Overall, the cues that are present for obstruents are not necessarily mirrored by sonorants.

These findings are interpreted with reference to Dispersion Theory (Flemming 1995; Liljencrants & Lindblom 1972; Lindblom 1986, 1990). While various incarnations of Dispersion Theory focus on different aspects of perceptual and auditory distinctiveness, a basic claim is that

one requirement for phonological contrasts is that they must be perceptually distinct: contrasts that are subject to great confusability are phonologically disfavored. The proposal, then, is that the typology of breathy voiced sonorants is due in part to the fact that they are not well differentiated acoustically. Breathily voiced sonorants are crosslinguistically rare because they do not make for strong phonemic contrasts.



## Acknowledgements

My first thank you goes to the Marathi consultants who were generous with their time, energy, input, and good humor. Anything I learned I owe to them, and I am thankful beyond measure. Special thanks to Namrata Barve and Priyanka Raut, without whom I never would have started working with Marathi. Thank you also to Abhijeet Barve, Narayani Barve, Vijay Barve, and Priyamvada Prabhu.

I am immensely grateful to my co-advisors, Dr. Allard Jongman and Dr. Jie Zhang, for the myriad ways in which they supported me during the writing of this dissertation. Their critiques, commentary, and genuine enthusiasm for the project kept me engaged and carried me forward when my own energy flagged. I owe you both a debt that goes beyond words, and I can only hope that I am able to pay it forward.

Thank you also to my other committee members—Dr. Raymond Pierotti, Dr. Joan Sereno, Dr. Geetanjali Tiwari, and Dr. Harold Torrence—who asked me questions that made me think harder and produce better work. Throughout this process I felt challenged in all the best ways and supported in all the most important ways. It was a true honor and privilege to work with you all, and I only wish that every graduate student could be graced with such a committee.

My colleagues in the Linguistics Department—Rania Agarbeh, I Nyoman Aryawibawa, José Alemán Bañón, Turki Binturki, Kristi Bond, Philip Duncan, Hiba Gharib, Pedro Mateo, Mircea Sauciuc, Khady Tamba, and Mahire Yakup—were with me from the beginning. They, along with Dr. Alison Gabriele and Dr. Sara Rosen, offered critical support along the way. My classmates and colleagues in LING 850 provided important insights as I wrestled with data and with presentations. I am particularly grateful to Goun Lee, Hyunjung Lee, Maite Martínez-García, and Mahire Yakup. Stephen Politzer-Ahles deserves special mention, as well, for

technical support, coding assistance, and general awesomeness. Thank you, all, for unending patience and support.

It should also be said that a dissertation is both a solo and a group endeavor: although I wrote this document, and am solely responsible for any errors, it would not exist without the love and support I received from my friends and family during its creation. Abigail Berkson, Diane Helen Harper, H. Neil Berkson, Heather Bandalo, Ilene Lang, William G. Johnson, my beloved aunts and uncles, and my Art of Living and Lawrence families: thank you, on so many levels.

Finally, I am also grateful to the Marathi language for being my constant companion over these last years. Ashok Kelkar wrote the following dedication in his 1958 dissertation on Marathi: I could never say it better.

“To

The Marathi language  
It was a pleasure to work  
with you, although I must  
confess that your messiness  
sometimes infuriated me.  
There never was, however,  
a dull moment!”

As always: all that is well done benefitted from the help of those I have mentioned. Any mistakes are mine alone.

## Table of Contents

1	Introduction.....	1
1.1	Overview of the present study .....	3
1.2	Major themes that emerge in the present study .....	6
1.3	Significance of the present work.....	8
1.4	Overview of the dissertation .....	8
2	Setting the stage: Background information and previous research.....	10
2.1	An overview of sonorants and obstruents.....	10
2.2	An overview of phonation types.....	12
2.3	Breathy voice in Indic and beyond: A review of previous findings .....	14
2.4	Duration .....	22
2.4.1	Consonant Duration .....	22
2.4.2	Post-release measures: VOT, NOT, ACT, and the PVI.....	24
2.4.3	Vowel Duration.....	35
2.5	Fundamental Frequency.....	36
2.6	Spectral Measures .....	38
2.6.1	H1-H2 .....	43
2.6.2	H1-A1, H1-A2, H1-A3 .....	46
2.6.2.1	H1-A1 .....	47
2.6.2.2	H1-A2 .....	49
2.6.2.3	H1-A3 .....	51
2.7	Cepstral Peak Prominence .....	52
2.8	Broad review .....	56
2.9	An overview of Marathi.....	58
3	Overview of the Present Study – Methods and Procedures.....	62
3.1	Participants.....	62
3.2	Materials .....	63
3.3	Recording and token presentation.....	66
3.4	Data analysis .....	67
3.5	Intervals measured .....	69
3.6	Predictions.....	69
3.7	Statistical analyses .....	71
4	Duration Results.....	73
4.1	Consonant Duration .....	75
4.2	Pre-Vocalic Interval.....	81
4.3	Vowel Duration.....	85
4.4	A brief recap of duration results .....	88
5	Fundamental Frequency Results.....	90
5.1	Fundamental frequency in voiced consonants .....	91
5.2	Fundamental frequency in vowels .....	96
5.3	F0 recap.....	105
6	Spectral Measure Results.....	107
6.1	H1-H2c Results.....	107
6.1.1	H1-H2c – consonant results .....	108
6.1.2	H1-H2c – vowel results .....	115

6.1.3	H1-H2c recap .....	125
6.2	H1-A1c .....	127
6.2.1	H1-A1c – consonant results .....	128
6.2.2	H1-A1c – vowel results .....	134
6.2.3	H1-A1c recap .....	145
6.3	H1-A2c .....	147
6.3.1	H1-A2c – consonant results .....	148
6.3.2	H1-A2c – vowel results .....	153
6.3.3	H1-A2c recap .....	161
6.4	H1-A3c .....	163
6.4.1	H1-A3c – consonant results .....	164
6.4.2	H1-A3c – vowel results .....	169
6.4.3	H1-A3c recap .....	175
6.5	Cepstral Peak Prominence .....	177
6.5.1	CPP – consonant results .....	178
6.5.2	CPP – vowel results .....	187
6.5.3	CPP recap .....	194
7	An overview of all results .....	196
8	Discussion .....	200
8.1	The present results in relation to previous findings .....	202
8.2	Trading relations .....	204
8.3	Gender differences .....	213
8.4	Sonorants and obstruents .....	225
9	Conclusion .....	234
	References .....	237
	Appendix A – Demographic Data .....	244
	Appendix B – Wordlist .....	246

## List of Figures

Figure 1	The Glottal Continuum .....	16
Figure 2	Illustration of breathy voiced obstruent: waveform and spectrogram of the word [b <sup>h</sup> at], ‘rice’ .....	18
Figure 3	Sample waveforms of four stop-types in Marathi.....	20
Figure 4	Illustration of NOT beginning and endpoints in [k <sup>h</sup> as], ‘special’ .....	27
Figure 5	Measurement of ACT in [p <sup>h</sup> aʈa], ‘fork (in the road)’ .....	28
Figure 6	Clear ACT, missing ACT: two waveforms of the [b <sup>h</sup> ] in [b <sup>h</sup> at], ‘rice’ .....	31
Figure 7	Waveform and spectrogram of [b <sup>h</sup> at], ‘rice’ .....	32
Figure 8	Waveform and spectrograms of modal and breathy voiced stops .....	40
Figure 9	Illustration of H1, H2, A1, A2, and A3 on spectra generated from the modal and breathy voice sources shown in Figure 8 .....	41
Figure 10	Illustration of CPP measures from modal and breathy voiced tokens .....	54
Figure 11	Marathi consonant chart.....	59
Figure 12	Marathi consonants recorded in the present study .....	63
Figure 13	Full syllable duration (consonant, PVI, vowel) by phonation type, obstruency, and voicing.....	74
Figure 14	Consonant duration by obstruency, voice, and phonation type .....	77
Figure 15	Duration of sonorants and voiced obstruents in word-initial position .....	79
Figure 16	Medial consonant durations by obstruency, voice, phonation type, and place of articulation .....	80
Figure 17	PVI durations by obstruency, voice, breath, and place of articulation .....	82
Figure 18	PVI durations by obstruency, voice, breath, and vowel context.....	84
Figure 19	Time course of F0 (in Hz) in voiced consonants .....	92
Figure 20	Time course of F0 across voiced consonants by vowel context – male data.....	94
Figure 21	Time course of F0 across voiced consonants by word position.....	95
Figure 22	Time course of F0 across vowels after voiced consonants by obstruency and phonation type.....	98
Figure 23	Time course of F0 across vowels after obstruents by voice (post hoc) .....	99
Figure 24	Time course of F0 across [a] and [e] (post hoc) .....	102
Figure 25	Time course of F0 across vowels by word position (post hoc).....	104
Figure 26	Time course of H1-H2c across sonorants by phonation type and gender.....	109
Figure 27	Time course of H1-H2c across sonorants by breath, place, and gender .....	111
Figure 28	Time course of H1-H2c across sonorants by phonation type and word position.....	113
Figure 29	H1-H2c across syllables by gender (sonorant data only).....	115
Figure 30	H1-H2c across vowels after voiced segments by phonation type and obstruency.....	116
Figure 31	H1-H2c across vowels after obstruents by voice – male data .....	118
Figure 32	H1-H2c across [a] and [e] by phonation type and obstruency .....	121
Figure 33	H1-H2c across vowels after voiced segments by phonation type, obstruency, and word position .....	124
Figure 34	H1-A1c values across sonorants by gender .....	129
Figure 35	H1-A1c of sonorants by phonation type and place of articulation .....	130
Figure 36	H1-A1c values across sonorants by breath, vowel context, and gender .....	132
Figure 37	H1-A1c across syllables by gender (sonorant data only).....	134
Figure 38	H1-A1c in vowels after obstruents by phonation type and voice .....	135
Figure 39	H1-A1c in vowels after voiced consonants by obstruency and breath .....	137

Figure 40	H1-A1c in vowels after voiced consonants by phonation type and obstruency – female data, with error bars .....	138
Figure 41	H1-A1c in [a] and [e] after voiced consonants by phonation type, obstruency, and gender.....	141
Figure 42	H1- A1c after voiced consonants by breath, obstruency, and word position ....	143
Figure 43	H1-A2c across sonorants by consonant phonation type and gender.....	149
Figure 44	H1-A2c across sonorants by consonant phonation type and vowel context.....	151
Figure 45	H1-A2c across syllables by gender (sonorant data only).....	152
Figure 46	H1-A2c across vowels after obstruents by consonant phonation type and voice	154
Figure 47	H1-A2c across vowels after voiced segments by consonant phonation type and obstruency .....	155
Figure 48	H1-A2c across [a] and [e] following voiced segments by consonant phonation type and gender .....	157
Figure 49	H1-A2c values across vowels following voiced segments by consonant phonation type, word position, and gender .....	158
Figure 50	H1-A3c across sonorants by phonation type and gender.....	165
Figure 51	H1-A3c values across sonorants by phonation type, place of articulation, and vowel context.....	166
Figure 52	H1-A3c across syllables by gender (sonorant data only).....	169
Figure 53	H1-A3c across vowels after obstruents by consonant phonation type and voice	170
Figure 54	H1-A3c across vowels after voiced segments by consonant phonation type and obstruency .....	172
Figure 55	H1-A3c values in vowels after voiced segments by consonant phonation type and word position .....	173
Figure 56	H1-A3c in vowels after voiced segments by consonant phonation type, vowel context, and gender .....	174
Figure 57	Time course of CPP across voiced segments by phonation type and obstruency	178
Figure 58	Time course of CPP across voiced obstruents .....	180
Figure 59	Time course of CPP across sonorants by phonation type and gender .....	181
Figure 60	CPP across sonorants ~ vowel context and word position (female data) .....	183
Figure 61	CPP across sonorants by phonation type and place of articulation (female data)	185
Figure 62	CPP across full syllables by obstruency .....	187
Figure 63	CPP in vowels after obstruents by breath, voice, and gender .....	188
Figure 64	CPP in vowels after voiced consonants by breath, obstruency, and gender .....	190
Figure 65	CPP across vowels after voiced consonants ~ vowel contexts .....	191
Figure 66	CPP across vowels after voiced consonants ~ word positions .....	193
Figure 67	F0 across syllables by gender, phonation type, and obstruency .....	216
Figure 68	H1-H2c across syllables by gender (sonorant data only).....	217
Figure 69	H1-A1c across syllables by gender (sonorant data only).....	218
Figure 70	CPP across syllables by gender and obstruency .....	219

## List of Tables

Table 1	Indic Oral Stops: A Typical Array.....	15
Table 2	Phonatory settings: A basic articulatory description of four stop-types prevocally.....	16
Table 3	Average Closure Duration for Intervocalic Bengali and Hindi Stops .....	23
Table 4	Average VOT of velar stops in Hindi and Marathi (in ms).....	26
Table 5	Average NOT of Hindi velar stops (in ms.....	27
Table 6	Average ACT values of East Bengali stops (least square means, in ms) .....	29
Table 7	Summary of Previous Findings.....	57
Table 8	Mean vowel durations by vowel context, obstruency, and breath.....	87
Table 9	F0 of [a] and [e] by obstruency and place of articulation (omnibus).....	97
Table 10	ANOVA results for H1-H2c of vowels following voiced segments by obstruency and window.....	119
Table 11	ANOVA results for H1-H2c of [a] and [e] by obstruency, gender, and window.....	122
Table 12	Standard deviation of H1-A1c values (in dB) in vowels by gender, obstruency, phonation type, and window .....	139
Table 13	Post hoc analyses of H1-A3c differences triggered by phonation in vowels following voiced and voiceless obstruents .....	171
Table 14	Post hoc analyses of H1-A3c in [a] and [e] by consonant phonation type and gender .....	175
Table 15	Success of the measures of duration, F0, H1-H2c, H1-A1c, H1-A2c, H1-A3c, and CPP in distinguishing phonation types in sonorants and obstruents produced by male and female speakers. ....	196
Table 16	Parameters which reflect an overall effect of Phonation type .....	197
Table 17	Breathy voiced sonorants in multiple word positions and vowel contexts in Marathi, Sumi, and Camling.....	209
Table 18	A comparison of the measures that successfully distinguish phonation type in OBSTRUENTS versus SONORANTS .....	226

## 1 Introduction

This dissertation presents an acoustic analysis of phonation type in Marathi sonorants and obstruents. While it is likely that the work will be of interest primarily to linguists, Indic language enthusiasts without formal linguistic training may find it of interest as well, and so brief definitions of key linguistic concepts are provided throughout. Foremost among the terms that must be understood are *phonation*, *obstruent*, and *sonorant*: short descriptions are provided here, and more thorough discussion appears in Chapter 2. *Phonation* refers to the production of sound through vibration of the vocal folds (VF). The VF can be held in a number of different configurations during vibration: breathy voice is a phonation type generally produced with less tension in the folds, which leads to slower vibration and greater airflow. These articulatory facts are relevant because they directly affect the acoustic characteristics associated with breathy phonation. *Obstruent* is a term referring to sounds produced with a significant obstruction in the vocal tract, like the oral stops [p] and [b]. *Sonorants* are sounds produced with a looser constriction than obstruents and relatively free airflow, like the nasal [m] or the lateral [l]. Sonorants typically involve greater acoustic energy than obstruents. These are phonetic—meaning articulatory and acoustic—definitions of obstruents and sonorants: a phonological distinction is drawn between the two as well, however, notably in terms of distinctive features or feature sets à la Chomsky and Halle. This topic will be returned to in Chapter 2.

Phonation is commonly used to create meaningful linguistic differences in the languages of the world. While certain phonation-type are crosslinguistically common—phonemic distinctions between voiced and voiceless consonant sounds, for instance, are found in many languages—other, less common phonation types also exist. Breathily phonation is one example, and has been the subject of ongoing phonetic investigation—particularly with regards to vowels



(Andruski & Ratliff 2000, Blankenship 1997, Esposito 2006, Wayland & Jongman 2003) and Indic obstruents (Davis 1994, Dutta 2007, Mikuteit & Reetz 2007, Ohala & Ohala 1972).<sup>1</sup> Very little research, however, has focused on breathy phonation in sonorants, probably due in part to their crosslinguistic rarity.<sup>2</sup> The UCLA Phonological Segment Inventory Database (UPSID), which indexes 451 languages, reveals that five languages (1.11% of the total) utilize breathy phonation in vowels and 13 (2.9% of the total) in consonants. These 13 languages can be further subdivided: only five of the 13 contain breathy voiced sonorants.

As such, this dissertation provides critical new insights into the many questions about breathy voiced sonorants which remain unanswered by presenting a detailed acoustic analysis of breathy voiced sonorants and obstruents in Marathi. The hope is that by determining the acoustic parameters which cue breathy phonation in sonorants, and investigating the ways in which sonorants either pattern with or diverge from obstruents, we will also be able to shed light on the typology of phonemic breathy voice. The primary questions addressed herein are:

1. What are the acoustic correlates of breathy phonation in sonorants?
2. Is the pattern similar to or different than the pattern found in obstruents?
3. Does the acoustic pattern provide insights into their crosslinguistic rarity?

The dissertation reveals that breathy voice is cued better by obstruents than by sonorants. This result is interpreted with reference to the concept of dispersion originally formulated in the Theory of Adaptive Dispersion (Liljencrants & Lindblom 1972; Lindblom 1986, 1990) and modified in Dispersion Theory (Flemming 1995). There is a growing body of literature providing

---

<sup>1</sup> Breathly voice arising from vocal pathology in speakers of languages that do not feature breathy phonation has also been the subject of investigation. In this dissertation, pathological breathy voice is considered distinct from linguistically-relevant (or phonemic) breathy voice, though the pathological literature will be referred to occasionally, particularly with regards to specific acoustic measures such as cepstral peak prominence (CPP).

<sup>2</sup> One exception is an MA thesis out of Australia focused on breathy voiced nasals and laterals in Sumi, a Tibeto-Burman language (Harris 2009). I will refer to this study repeatedly throughout the dissertation: it represents an important contribution, particularly in light of the dearth of research on breathy voiced sonorants, but is limited in that data came from a single female speaker.

support for the basic claim made by these theories of dispersion, which is that one requirement of phonological contrasts is that they must be easy for listeners to distinguish: contrasts that are subject to great perceptual confusability are phonologically disfavored (Engstrand & Krull 2009, Flemming 1995, Kawahara 2005, Lindblom 1986, Padgett 2003, Padgett & Tabain 2005).

While Liljencrants and Lindblom (1972) at first argued that contrasts must be maximally distinct, Lindblom later modified this by stating that contrasts must at least be sufficiently distinct (Lindblom 1986, Schwartz *et al.* 1997). In working to formalize Adaptive Dispersion Theory into an optimality theoretic framework, Flemming (1995) made explicit reference to the link between distinctness and auditory or acoustic information. Flemming's recast, dubbed Dispersion Theory, posits that linguistically relevant contrasts must be as auditorily distinct as possible, and that acoustic information is directly related to the auditory distinctiveness of sounds. The key idea here is that sounds which are well-differentiated acoustically make for better phonemic contrasts. A phonemic contrast that is poorly cued, on the other hand—as breathy voice is cued more poorly by sonorants than by obstruents in Marathi—fails to meet the requirement that “the auditory distinctiveness of the contrasts should be maximized so that the differences between words can be easily perceived by a listener” (Flemming 1995: 23). With this in mind, I argue that the acoustic profile and the crosslinguistic rarity of breathy voiced sonorants are connected.

## **1.1 Overview of the present study**

This study consists of a detailed acoustic analysis of sonorants and obstruents in Marathi, an Indo-Aryan language which is uniquely suited for investigation of breathy voice in sonorants and obstruents due to its rich phonetic inventory. Spoken by approximately 72 million people

primarily in the state of Maharashtra (2001 India Census), Marathi—in common with its Indic relatives like Hindi and Bengali—features a unique four-way contrast among obstruents composed of plain voiceless, aspirated voiceless, plain voiced, and breathy voiced categories (i.e. /t, t<sup>h</sup>, d, d<sup>h</sup>/). Even more noteworthy than its breathy voiced obstruents, however, are Marathi’s breathy voiced sonorants. Contrasts include plain and breathy voiced distinctions among nasals (/m, m<sup>h</sup>, n, n<sup>h</sup>, ŋ, ŋ<sup>h</sup>/), approximants (/v, v<sup>h</sup>/), laterals (/l, l<sup>h</sup>/), and rhotics (/r, r<sup>h</sup>/).<sup>3</sup> Thus Marathi utilizes a crosslinguistically rare phonation type in an uncommon way across multiple manners of articulation.

In conducting an analysis of the acoustic correlates of phonation types in Marathi, the key comparisons are between plain and breathy phonation, and between obstruents and sonorants. Particular attention is paid to the ways in which the acoustic correlates of breathy voiced sonorants differ from those of breathy voiced obstruents. Data are from ten native speakers of Marathi—five male and five female—and measures reported include consonant duration, the duration of the breathy interval, vowel duration, fundamental frequency, the spectral measures of H1-H2c,<sup>4</sup> H1-A1c, H1-A2c, H1-A3c, and cepstral peak prominence (CPP).<sup>5</sup> Descriptions of these measures, as well as previous findings related to each measure, are provided in the next chapter.

---

<sup>3</sup> This distinction holds in the fricatives /z/ and /z<sup>h</sup>/ as well, but these sounds are largely allophonic rather than phonemic and alternate freely with the affricates /dz/ and /dz<sup>h</sup>/ . They are set aside at present and will be addressed in future work.

<sup>4</sup> H1-H2c means “corrected H1-H2” – this is explained more fully in the following sections, but in short it means that an algorithm is used to correct for the influence of formant frequencies and bandwidths on spectral measures.

<sup>5</sup> There has been some debate regarding measures used to capture breathiness: while the spectral measure H1-H2 has been found to differ significantly from plain to breathy phonation in various languages including Green Hmong (Andruski and Ratliff 2000), Gujarati (Esposito 2006), Hindi (Dutta 2007), and Khmer (Wayland and Jongman 2003), the reliability of this measure has also been called into question (Simpson 2012), and cepstral peak prominence (CPP)—which reflects periodicity, and has been utilized in vocal pathology research for a number of years (Hillenbrand *et al.* 1994, Heman-Ackah *et al.* 2003)—is being used increasingly in linguistic research (Blankenship 2002, Esposito 2012). In other words, there is not yet a clear consensus in the field regarding the measures which should be reported and so the current dissertation reports a range of measures in order to be thorough.

Broadly speaking, this research accomplishes two main goals. One of these is that it is—to the best of my knowledge—the first detailed acoustic analysis of breathy phonation in sonorants, and the most comprehensive comparison of breathy phonation across obstruents and sonorants within a single language. As such, it establishes that the acoustic correlates of breathy phonation in sonorants are in line with the pattern seen in obstruents. Second, while neither breathy voiced obstruents nor sonorants could be called common, sonorants are even more rare than obstruents. Why should this be the case? The acoustic results reported herein move us toward answering this question, for they reveal that the cues associated with sonorants are less robust than the cues associated with obstruents. Simply put, then, breathy phonation is not cued as well by sonorants as it is by obstruents. Thus, according to the principles laid out by the various dispersion theory incarnations (Flemming 1995; Liljencrants & Lindblom 1972; Lindblom 1986, 1990), a distinction between modal and breathy-voiced sonorants makes for a poor phonemic contrast.

Acoustically, this research reveals some consistent patterns. For both genders, and for both obstruents and sonorants, breathy voice is cued by a relatively long breathy interval after the release of the consonant constriction. The presence of this interval is not surprising; it was in fact referred to by Lisker and Abramson in their landmark 1964 article, wherein they state that breathy voiced stops are probably “distinguished from the other voiced category by the presence of low amplitude buzz mixed with noise in the interval following release of the stop” (p. 403). What is new here, however, is that while Lisker and Abramson were focused on obstruents, this work confirms the presence of a comparable breathy interval after sonorants as well. In addition to the breathy interval, reliable differences in H1-A2c and H1-A3c values in subsequent vowels based on consonant phonation type are triggered by both sonorants and obstruents, for both male

and females speakers.

## **1.2 Major themes that emerge in the present study**

Several major lessons are revealed by the results reported herein, and these mainly take the form of gender differences, trading relations, and differences between sonorants and obstruents. In terms of gender—or, more accurately, sex, for I refer not to sociolinguistic differences based on gender but on sexual dimorphism arising from biological sex differences—male and female speakers cue breathy phonation quite differently in Marathi, particularly with regards to sonorants. While an increase in the strength of the first harmonic relative to other low-frequency components of the spectrum (such as the second harmonic) in the vowel following breathy sonorants results in significantly greater H1-H2c values after breathy sonorants than after plain sonorants in male speech, the same is not true in female speech. Instead, breathy phonation in female speech is cued by added noise in the voice source, as indicated by the fact that females—but not males—show consistently lower CPP values within breathy sonorants than within plain sonorants.

As has been mentioned, differences in the way that phonation type is cued by sonorants and by obstruents also arise. Obstruents are consistently cued better than sonorants, and this can be seen even if we consider only those two parameters mentioned in the paragraph above, H1-H2c and CPP. With regards to H1-H2c values in subsequent vowels, both sonorants and obstruents trigger significant differences based on phonation type in at least some contexts for males. For females, however, no differences arise after sonorants but there are at least some contexts after obstruents where differences are significant. Thus while H1-H2c is a good cue for males across the board, it is useful for females only after obstruents. Meanwhile, CPP values

within consonants reliably distinguish between phonation types for both sonorants and obstruents for female speakers; for male speakers, however, CPP values distinguish plain from breathy obstruents, but they do not distinguish plain from breathy sonorants. Like H1-H2c for females, then, when it comes to CPP values in male speech sonorants lack a cue which obstruents have.

Finally, what also emerges is that Marathi phonation type contrasts show numerous examples of trading relations: many times, cues that are strong in Context A are weaker in Context B, while cues that are absent in Context A are more robust in Context B. The cues that work for males are not necessarily the same as those that work for females. The cues that are present for obstruents are not necessarily mirrored by sonorants. What distinguishes plain from breathy phonation word-initially may disappear word-medially, and vice versa. All these differences serve to highlight an important point, however. Phonation type contrasts are associated with multiple cues, and while the availability of these cues shifts from one context to the next the sheer number of them means that in any given context, at least some cues are available.

If there is an overarching lesson that can be gained from this research, it is as follows. The acoustic correlates of breathy voiced phonation are numerous, and they fit together in a complex and shifting way. Acoustic differences triggered by consonant phonation type are often weaker after sonorants than after obstruents, and this presumably underlies—at least partially—the fact that breathy voiced sonorants appear so rarely in the world’s languages: they simply are not cued that well.

### **1.3 Significance of the present work**

This work is significant on a number of levels. Theoretically, the implication of the findings presented herein is that breathy voiced sonorants are typologically uncommon at least in part because they are poorly cued. Time and again, phonation type differences are cued more robustly by obstruents than by sonorants. Beyond the theoretical contribution, this work is important phonetically, because Marathi allows for a thorough analysis of breathy voice across multiple manners of articulation; typologically, because breathy voiced sonorants are incredibly rare, occurring in perhaps 1% of the world's known languages; and descriptively, because Marathi is 15<sup>th</sup> on the list of world languages with the most speakers (meaning that only 14 languages have more speakers than Marathi) but is relatively understudied phonetically and is virtually undocumented acoustically. Taken together, these elements mean that the current work represents a significant contribution to the literature.

### **1.4 Overview of the dissertation**

The remainder of the dissertation reads as follows. Chapter 2 includes background information and a literature review. An overview of phonation types is presented. The four-way contrast observed in Indic obstruents is reviewed, as are basic articulatory facts relevant to phonation. Next, temporal, F0, and spectral measures that have been utilized in previous phonation type analyses are reviewed, and conflicting results are highlighted. An overview of Marathi is then presented. Chapter 0 includes the methodology and procedures utilized in the current study, and concludes with commentary about the statistical analyses performed. Results from the temporal measures appear in Chapter 4, while Chapter 5 presents results related to fundamental frequency. Spectral measure results—H1-H2c, H1-A1c, H1-A2c, H1-A3c, and

CPP—appear in Chapter 6, and a brief overview of all key results appears in Chapter 7. Chapter 8 includes discussion of the findings; it ties results from the separate measures together, and provides a comprehensive overview of the full acoustic picture provided in the results chapters. Topics addressed include the numerous trading relations discovered in the results, a discussion of the pervasive gender-related differences that emerged, and the ways in which breathy phonation is cued differently by obstruents and by sonorants. Chapter 9 presents the overall conclusion, and discusses future directions.



## **2 Setting the stage: Background information and previous research**

The sections in Chapter 2 provide background on several topics related to phonation type differences in Marathi sonorants and obstruents, in order to properly set the stage for the data and discussions that follow. The first sections provide overviews of the distinction between obstruents and sonorants, phonation, and breathy voiced obstruents in Indic languages. Review of measures that have previously been used to assess phonation type differences follows, and the section ends by presenting a phonetic sketch of the Marathi language.

### **2.1 An overview of sonorants and obstruents**

The distinction between obstruents and sonorants can be approached from multiple perspectives. Phonetically, we can consider the articulatory and/or acoustic characteristics of obstruents and sonorants. Phonologically, we can consider the ways in which these groups pattern as natural classes. Typologically, we can consider crosslinguistic trends related to obstruency—which types of sound occur most frequently in consonant inventories? Which types of phonemic contrasts are common, and which types are rare?

In terms of articulation, obstruents are a class of sounds that are produced with a significant obstruction of airflow somewhere in the vocal tract. These include stops, fricatives, and affricates. Sonorants, in contrast, are produced with a relatively unimpeded airflow. The sounds typically associated with the natural class of sonorants are nasals, liquids and glides (or approximants), and vowels. These sounds are not produced with a stricture severe enough to allow for as great a build up of air pressure in the vocal tract as obstruents, and they are associated with greater internal acoustic energy than obstruents.

Phonologically, the division of sounds into obstruents and sonorants is motivated by the fact that these sounds seem to form natural classes: numerous observation can be made about

processes that target obstruents but not sonorants, or sonorants but not obstruents. Thus in traditional feature notation the two are defined in opposition to one another: sonorants are [+sonorant], and obstruents are [-sonorant]. Crosslinguistically, obstruents are more likely than sonorants to show phonemic voicing contrasts, and more likely to undergo phonological processes such as voicing assimilation (Hayes 2009).

In terms of typology, an analysis of the make-up of consonant inventories indexed in UPSID reveals that obstruents tend to account for between 66% and 75% of the individual languages' inventories (Lindblom and Maddieson 1988). There is some variation, of course, but the crosslinguistic trend is that inventories contain more obstruents than sonorants, and this is true across numerous languages, geographic areas, and language families. After making this observation, Lindblom and Maddieson state the following:

Since the constancy of obstruent-sonorant proportions applies cross-linguistically we feel justified in suggesting that it reflects a *phonetic universal*. We interpret this universal to be the physical characteristics of the regions of the 'phonetic space' that obstruent and sonorant consonants [exist] in. We take the obstruent space to be larger, that is perceptually—and perhaps articulatorily—richer, than that of sonorants (cf. Ohala 1983) (p. 66).

This concept of perceptual (and perhaps articulatory) richness is of particular interest, and may be relevant when considering yet another typological trend: Languages that employ phonemic consonant length are more likely to feature geminate obstruents than geminate sonorants (Kawahara 2007). Based on the results of a perception study, Kawahara posits that this may be because geminate sonorants are difficult to perceive—the more sonorous a sound, the more difficult it is for listeners to distinguish segment duration. This is interpreted as providing support for Dispersion Theory: the proposal is that because length contrasts in sonorants are difficult to perceive, geminate sonorants are not auditorily distinct and are therefore crosslinguistically marked.

Recall that the figures in UPSID indicate that the same typological observation Kawahara makes regarding geminacy can be made regarding phonemic breathy voice: languages that employ phonemic breathy phonation are more likely to feature breathy obstruents than breathy sonorants. The question, then, is whether similar findings will emerge regarding the auditory distinctiveness of phonation type contrasts in sonorants. This work does not test perception directly, but instead seeks to establish the acoustic profile of phonation type contrasts in Marathi sonorants and obstruents. This is a critical first step: only once we begin to understand the ways in which modal and breathy voiced sounds are differentiated acoustically can we hope to create well-designed perception experiments.

The commentary above is not meant to represent a comprehensive review of the differences between obstruents and sonorants: rather, it is meant to establish the fact that the distinction between the two manifests in a number of linguistically relevant ways. As such, it is not surprising that the typology of breathy phonation shows a distinction between obstruents and sonorants. Having established this, we turn now to a more thorough discussion of phonation.

## **2.2 An overview of phonation types**

In the sound systems of the world's languages, meaningful differences often hinge on distinctions in phonation type. Production of distinct phonation types is achieved through control of the larynx; to effect various linguistically relevant distinctions we may manipulate, for example, the width of the glottal opening, the distance between the vocal folds, the stiffness or slackness of the folds, and the rate of airflow (Marchal 2009).<sup>6</sup> While voicelessness and voicing,

---

<sup>6</sup> This work focuses on phonation type or voice quality as it pertains to phonemic distinctions, and disregards other arenas in which it may be relevant such as the conveyance of extralinguistic information (i.e. emotion, excitement) or speech pathology.

or modal voice, are widely used to establish phonemic contrasts crosslinguistically, additional phonation types such as creaky voice and breathy voice also exist (Catford 1977, Laver 1980, Gordon and Ladefoged 2001, Marchal 2009).

Previous investigation into the articulatory properties and acoustic correlates of phonation types suggests crosslinguistic consistencies (Gordon and Ladefoged 2001). This is unsurprising given that the human vocal tract does not vary considerably from one language group to another. In addition, laryngeal contrasts (i.e., phonation type contrasts) tend to be preserved in some locations and neutralized in others, and it is theorized that these patterns again tie back into perceptibility. Steriade (1997) argues that the crosslinguistic preservation and neutralization patterns observed arise due to the availability or absence of perceptual cues in various positions; contrasts tend to be preserved in onset and prevocalic positions because these are the positions in which the acoustic cues most relevant for the perception of laryngeal distinctions may be fully realized. Furthermore, as has already been mentioned, typological study of the world's languages reveals that some phonation types are widely utilized for making phonemic contrasts while others are considerably more rare.

Given these facts, a number of points of interest arise when considering phonation types.

Minimally, these include:

- (1) What acoustic correlates are associated with different phonation types both crosslinguistically and language-specifically;
- (2) Why are some phonation types uncommon typologically; and
- (3) Can the acoustic correlates of phonation type shed light on the relative rarity of occurrence of a particular phonation type utilized in a particular way (i.e. breathy voice in sonorants)?

In order to address such issues, it is critical that we develop an understanding of the acoustic correlates of both the more and less common phonation types.

While a number of previous studies have involved detailed analysis of modal and breathy phonation in various languages, they have tended to focus on either vowels (Andruski & Ratliff 2000, Blankenship 1997, Esposito 2006, Wayland & Jongman 2003) or obstruents (Davis 1994, Dutta 2007, Mikuteit & Reetz 2007, Ohala & Ohala 1972). Another line of research has compared phonation types in a particular manner of articulation across multiple languages; Esposito *et al.* (2005), for instance, compared modal and breathy nasals in Bengali, Hindi, and Marathi. Investigation of breathy voice across multiple manners of articulation within a single language is virtually nonexistent, however, and as such the current research fills an existing gap. The hope is that through presenting detailed analysis of breathy phonation in one language, we will move towards a more thorough consideration of the broad questions outlined above.

### **2.3 Breathy voice in Indic and beyond: A review of previous findings**

While phonemic breathy voice is somewhat rare cross-linguistically, Indic languages are relatively well-known for containing breathy voiced obstruents. The four-way contrast among oral stops and affricates shown in Table 1 is typical of Indic phonetic inventories (Masica 1991).

**Table 1 Indic Oral Stops: A Typical Array**

<b>BILABIAL</b>	p	p <sup>h</sup>	b	b <sup>h</sup>
<b>DENTAL</b>	t	t <sup>h</sup>	d	d <sup>h</sup>
<b>RETROFLEX</b>	ɖ	ɖ <sup>h</sup>	ɗ	ɗ <sup>h</sup>
<b>ALVEO-PALATAL</b>	tʃ	tʃ <sup>h</sup>	dʒ	dʒ <sup>h</sup>
<b>VELAR</b>	k	k <sup>h</sup>	g	g <sup>h</sup>

The right-most column of sounds are alternately referred to as voiced aspirated, murmured, or breathy voiced stops (Schiefer 1987, Ladefoged and Maddieson 1996, Dutta 2007). They are characterized by vocal fold vibration during and after the stop closure, and a breathy interval after the release which was referred to as “buzz mixed with noise” by Lisker and Abramson (1964). The sounds are transcribed herein with a raised /h/ to indicate that the breathy release resulting from this continuous voicing differs from voiceless aspiration (Ladefoged and Maddieson 1996, Dutta 2007).

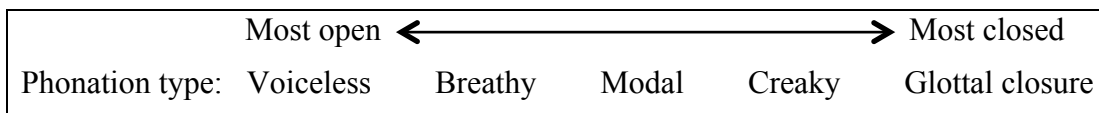
It is useful to consider the articulatory characteristics of the four stop types listed above before reviewing the acoustic measures which have been used to capture differences, because of the obvious connection between articulation and acoustics. A simplified overview, drawn primarily from Marchal (2009: 121-25) and Laver (1980: 132-33), appears in Table 2.

**Table 2 Phonatory settings: A basic articulatory description of four stop-types prevocally**

<b>VOICELESS UNASPIRATED</b>	The vocal folds (VF) are apart and are not vibrating during consonant closure, but vibration for the subsequent vowel begins immediately or almost immediately upon the release of the stop closure.
<b>VOICELESS ASPIRATED</b>	The VF are apart and are not vibrating during the consonant closure. They remain apart for some interval of time after closure release. During this interval, air passes freely through the open vocal folds. This unobstructed airflow is referred to as <i>aspiration</i> , and aspiration occurs for an interval of time before the VF begin vibrating for the subsequent vowel. The temporal window between closure release and onset of VF vibration is significantly longer in voiceless aspirated than in voiceless unaspirated sounds.
<b>MODAL (PLAIN) VOICE</b>	VF vibration occurs during the consonant closure; during the portion of the cycle when the VF are closed, they are pressed together along their full length. In a fully voiced stop, vibration persists throughout the entirety of the stop closure and transition into the subsequent vowel.
<b>BREATHY VOICE</b>	The VFs vibrate while remaining separated slightly. Though held closer together than during the production of voicelessness, they never fully close—they are, according to Catford (1977: 99), merely ‘flapping in the breeze.’ Increased airflow volume and velocity compared with the other settings and the loose vibratory mode persists into some portion of the subsequent vowel.

While the full range of phonation types will not be treated in the present work, they are sometimes conceptualized in terms of a continuum of glottal opening, with voicelessness falling at one end and full glottal closure falling at the other. This is illustrated in Figure 1, recreated from Gordon and Ladefoged (2001: 384, after Ladefoged 1977).

**Figure 1 The Glottal Continuum**



*from Gordon & Ladefoged 2001: 384*

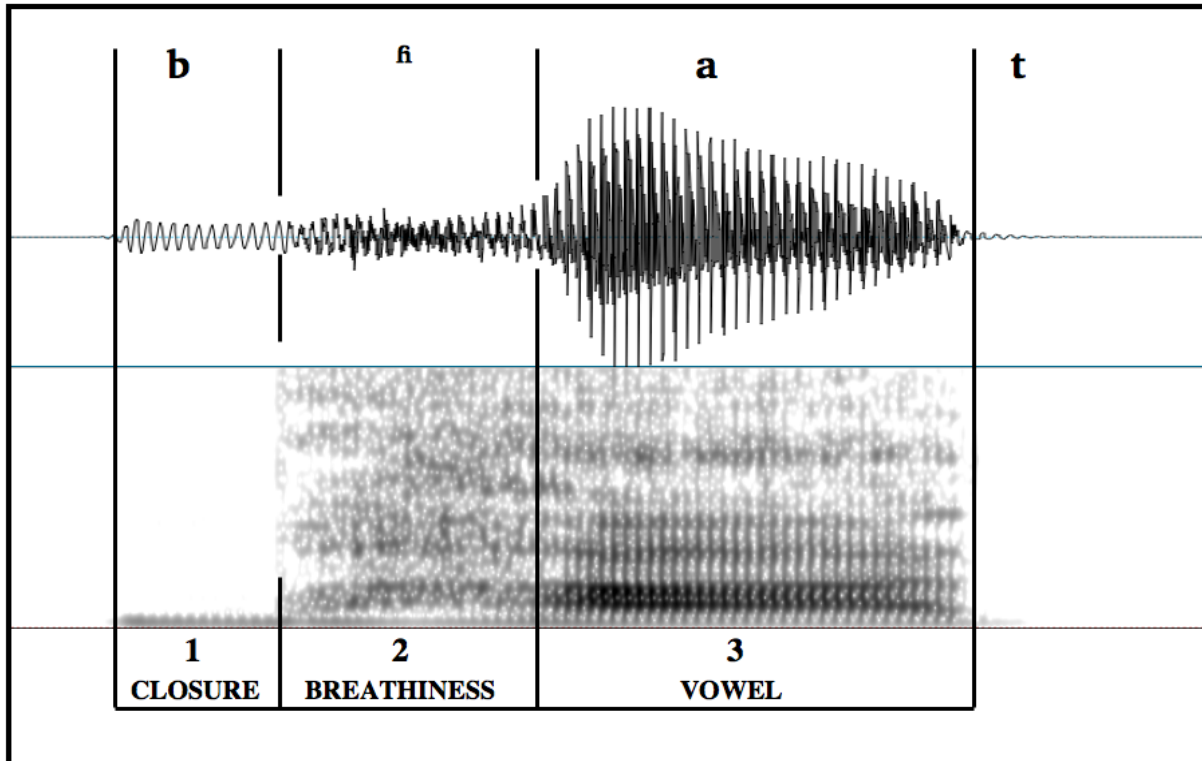
Recognition of breathy-voiced phonemes in Indic languages is uncontroversial, and dates back to ancient Sanskrit grammarians (Davis 1994, Deshpande 1991, Lisker & Abramson 1964, Mikuteit & Reetz 2007). Debate about how the difference between stop types is realized

acoustically, however, is ongoing, with recent work supporting an analysis wherein multiple acoustic cues are utilized to differentiate between the four categories. In his analysis of Hindi stops, for example, Dutta (2007) finds that durational and spectral measures jointly contribute to the distinction. Multiple acoustic parameters emerge as relevant in analysis of Bengali stops and affricates, as well, with durational measures related to the preceding vowel, the stop closure, and the post-release interval proving significant in establishing various distinctions—the preceding vowel is longer before voiced than before voiceless sounds, for instance, while the post release interval serves to distinguish plain from aspirated/breathy voiced sounds (Mikuteit and Reetz 2007).

These and other studies are addressed in more detail in the following sections, which outline previous findings related to the durational and spectral properties of different phonation types. The focus is primarily on the distinctions between modal and breathy voice. Before continuing to a discussion of previous findings, however, note that because the waveforms and spectrograms of breathy voiced sounds are somewhat unique it is useful to have a visual in mind as we proceed. For this reason, the diagrams in Figure 2 and Figure 3 are presented. The first presents a waveform and spectrogram of a breathy voiced obstruent; the second presents waveforms from each of the four stop-types commonly found in Indic languages. Detailed explanation of each figure follows in the body of the text.



Figure 2 Illustration of breathy voiced obstruent: waveform and spectrogram of the word [b<sup>h</sup>at], 'rice'



The illustration above has been separated into three components.

1. The first component represents the consonantal constriction of the voiced segment [b<sup>h</sup>]—the [b] portion, that is—and is characterized by a smoothly periodic waveform.
2. The second component represents the breathy portion—the [h]. This interval begins with the release of the consonantal constriction, visible both because an element of jaggedness is introduced in the periodicity of the waveform and because of the release burst in the spectrogram. This interval represents breath intermingled with the following vowel. During the breathy interval, the waveform is characterized by dampened amplitude, and the formant structure in the spectrogram is quite pale.
3. The third component begins with the cessation of the breathiness from interval 2, and represents the remainder of the vowel [a]—meaning, the portion of [a] that is not intermingled with breath, as was the case in the second interval. The onset of full vowel quality is associated with increased amplitude (as compared with the breathy portion) which is visible in the waveform, and a darkened formant structure.

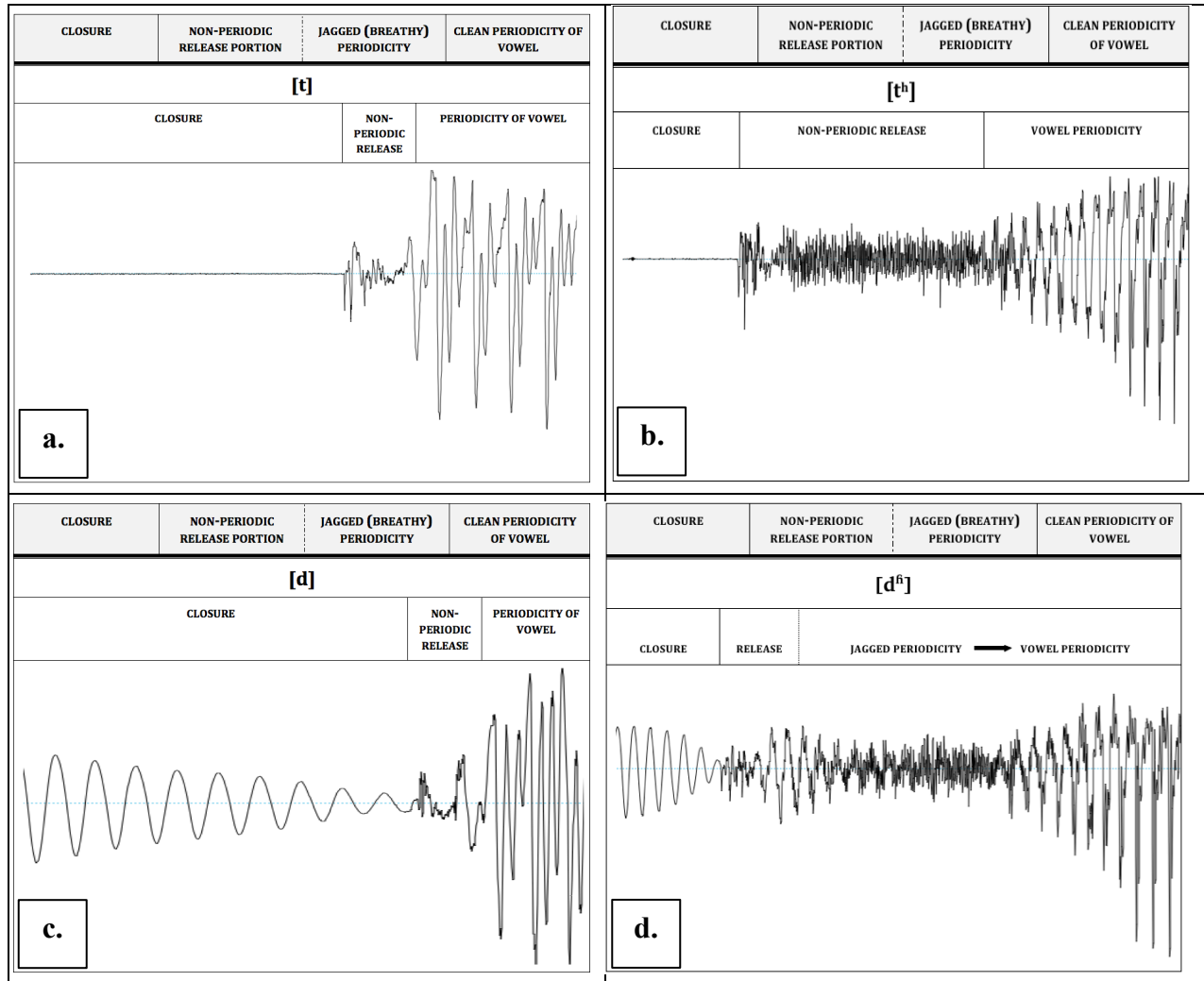
This image illustrates that there are several challenges inherent in dividing breathy voiced sounds into components. When a voiceless sound is aspirated, the aspiration appears as random or aperiodic noise in the waveform, and a following vowel is characterized by an onset of periodicity in the waveform. This distinction makes determining the point at which the aspiration ends and the vowel begins relatively clear. Breathily voiced sounds feature voicing throughout, however, and the consistent vocal fold vibration means that there is often no cessation of periodicity in the waveform.<sup>7</sup> Furthermore, the transition from the dampened amplitude of the breathy interval into the full amplitude of the vowel is gradual. This makes it infinitely more complicated to visually pinpoint the end of the breathiness and the beginning of the vowel, for the post-release interval is not composed of discrete, linear components—breath alone, followed by vowel alone—but by a mixture of the two. Consider the detailed images of waveforms from the four stop-types in Figure 3.

These images illustrate that oral stops may contain four possible intervals, as is indicated at the top of each illustration: the stop closure, a non-periodic release, the jagged, low-amplitude periodicity associated with breathiness, and the clear periodicity of the subsequent vowel. The stop closure contains the clean periodicity of stop voicing for voiced stops. Beginning at the release burst is the interval which is of primary interest to us; this may be composed of a nonperiodic portion and/or a jaggedly periodic portion. The division between these two portions may or may not be clear; thus they are divided by a dotted rather than a solid line. Finally, the cessation of breathiness in the vowel yields clean periodicity.

---

<sup>7</sup> Some breathily voiced tokens do feature a brief interval of aperiodicity immediately following the closure release. This interval is referred to as Mikuteit and Reetz (2007) as the After Closure Time. More discussion of the topic appears in Section 2.6.

**Figure 3 Sample waveforms of four stop-types in Marathi**



Both the plain voiceless stop illustrated in (a.) and the plain voiced stop in (c.) demonstrate clear non-periodic release portions, and an obvious transition into the periodicity of the vowel. Neither breathiness nor aspiration are present to induce the jaggedness associated with buzzy or aspirated noise. The voiceless aspirated stop illustrated in (b.) is characterized by a clear release burst and an interval of random, aperiodic noise, which yields to the periodicity associated with the onset of vocal fold vibration for the subsequent vowel.<sup>8</sup>

<sup>8</sup> Observe that there is some jaggedness in the periodicity immediately following the cessation of aspiration: this is what Mikuteit and Reetz (2007) refer to as Superimposed Aspiration (SA).

The breathy voiced stop is illustrated in (d.), and this image is characterized by the difficulties previously mentioned. While there is clear stop closure voicing and an obvious release burst, no clear divisions are obvious in the interval of periodicity between the closure release and the rise in amplitude observed midway into the vowel. This token of [d<sup>h</sup>] shows a less jagged periodic interval—here identified as the release portion—followed by a low-amplitude interval identified as jagged periodicity. Furthermore, the jagged periodicity first transitions into something slightly less periodic but with no greater amplitude and then becomes more periodic before exhibiting the jump in amplitude which is familiar from the other three waveforms. The jagged periodicity of the breathy interval yields at some point to the cleaner, higher-amplitude periodicity of the pure vowel—meaning vowel after the cessation of breath—but where exactly is the point of transition? The question of where the breathiness ends is simply complicated, and it is not new; previous researchers have struggled with this issue. It is important to keep this challenge in mind as we proceed, for it directly affects the protocols used for measuring duration. The topic is addressed again shortly, and the methodology used in the present study is clearly explained.

One additional point should be discussed before moving on to previous findings. As mentioned, phonation type research dealing with breathy voice often focuses on the distinction between modal and breathy vowels. This research will be referenced in the coming sections, despite the fact that our focus in Marathi is on breathy voiced consonants. This is because the acoustic correlates of breathy voice may well be realized not only via consonant-internal cues, but also in post-consonantal cues. In other words, the duration (or spectral properties, in the case of sonorants) of the consonantal constriction itself may differ based on phonation type, and the subsequent vowel may carry relevant acoustic cues as well—as should seem obvious, given the

way that the breathiness perseverates into the vowel in Figure 3 (d.). In addition, some of the methodology that has been employed in the investigation of breathy voiced vowels will be applied to breathy voiced sonorants in the present work, as well—for example, spectral measures such as H1-H2 are taken both sonorant-internally and on subsequent vowels. For these reasons, literature related to breathy voice in both vowels and consonants is reviewed in the coming sections.

## **2.4 Duration**

Consonant phonation type may affect the temporal properties of at least three portions of the speech signal: consonant duration; the duration of aspiration (for voiceless sounds) and breathiness (for voiced sounds); and vowel duration. These are addressed separately in the following sections.

### **2.4.1 Consonant Duration**

Several durational differences emerge when comparing breathy and modal phonation. First, let us consider the closure duration of oral stops. In an analysis of the four types of oral stops in Hindi, Dutta (2007b) recorded five native speakers producing the 16 oral stops of Hindi (four stops at each of four places of articulation: bilabial, dental, retroflex, and velar). All consonants appeared word-initially before the vowel [a].<sup>9</sup> Based on the results of this study, Dutta reports that plain stops have longer closure durations than their aspirated and breathy counterparts—no overall means are reported, but these results aligns with the results reported by previous researchers. The average closure duration for intervocalic Hindi stops reported by

---

<sup>9</sup> Most tokens were real words, but several nonce words were also included to complete the paradigm.

Benguerel and Bhatia (1980) and the averages for intervocalic Bengali stops reported by Mikuteit and Reetz (2007) appear in Table 3.

**Table 3 Average Closure Duration for Intervocalic Bengali and Hindi Stops**

	Hindi (Benguerel & Bhatia 1980)	Bengali (Mikuteit & Reetz 2007)
<b>PLAIN VOICELESS</b>	154 ms	135 ms
<b>ASPIRATED VOICELESS</b>	140 ms	116 ms
<b>PLAIN VOICED</b>	130 ms	106 ms
<b>BREATHY VOICED</b>	112 ms	96 ms

*from Mikuteit and Reetz (2007: 258)*

As shown in the data above, voiced stops have shorter closure durations than voiceless stops in Hindi and in Bengali, and breathy voiced stops have the shortest closure duration of all in both languages.

While there is an expected correlation between place of articulation and closure duration—with labials expected to show the longest closure duration, followed by dentals and then by velars (Maddieson 1997)—Mikuteit and Reetz (2007) do not find this pattern to obtain to closure duration in Bengali stops. Instead, they report that labial ( $M = 119$  ms), dental ( $M = 122$  ms), and retroflex ( $M = 125$  ms) stops pattern together, with velars ( $M = 110$  ms) showing significantly shorter closure durations than the other three places of articulation. Palato-alveolar affricates ( $M = 91$  ms) show the shortest mean closure durations of all. Dutta (2007b) reports that there is no effect of place of articulation on closure duration in Hindi stop consonants, but notes that there is a marginal effect of phrase position. While tokens contained word-initial consonants, these were inserted into carrier sentences such that they appeared phrase-initially and phrase-medially, and Dutta reports that phrase-initial obstruents have slightly longer closure durations than phrase-medial consonants.

Taking all of his findings together, Dutta claims that, overall, closure duration is an important cue to phonation type. A longer closure duration (and in the case of voiced stops, a longer duration of voicing during closure) is a cue that the sound is plain.

This is opposite the pattern observed in the segment duration of modal and breathy sonorants. In an analysis of sonorant phonation types in Sumi, a Tibeto-Burman language, Harris (2009) reports that breathy nasals and laterals are longer than modal nasals and laterals. This study focused on the production of one female native speaker of Sumi, and should be interpreted accordingly; the finding is valuable, however, as it is to my knowledge the only existing data related to consonant constriction duration and breathy voiced sonorants. A similar pattern has been reported for breathy and modal vowels, as well; Gordon and Ladefoged (2003) report that breathy vowels are significantly longer than modal vowels in Kedang (as reported by Samely 1991) and in Jalapa Mazatec (Silverman *et al.* 1995). One thing that should be mentioned, however, is the data related to geminacy and sonorants reported by Kawahara (2005). Recall that, as mentioned previously, geminate sonorants are more rare than geminate obstruents crosslinguistically and that in a perception experiment, Kawahara found that the more sonorous a sound, the more difficult segment duration is for listeners to perceive. Thus while there is some evidence that breathy sonorants and vowels are longer than modal sonorants and vowels, there is also reason to wonder how salient this temporal difference may be for listeners.

Having reviewed differences related to consonantal constriction durations, we now turn to measures that have been used to capture the duration of aspiration and the breathy interval.

#### **2.4.2 Post-release measures: VOT, NOT, ACT, and the PVI**

Moving beyond the consonantal constriction itself and focusing on the post-release

interval, one durational measure often referenced in discussion of oral stops is voice onset time (VOT). VOT is a measure of the temporal interval between the release of the stop closure and the onset of vocal fold vibration. In their classic paper, Lisker and Abramson (1964) demonstrate both that VOT may be used to reliably distinguish between two and three categories of oral stops and that languages differ with regards to the specific way in which this parameter is employed. If we consider the closure release to be the zero point on the temporal scale, stops which feature pre-release voicing are considered to have a negative VOT, while those which show some lag between closure release and the onset of vocal fold vibration have a positive VOT value.

Lisker and Abramson's 1964 findings—replicated by many others in the intervening years—reveal that, in the two-way English distinction, voiced stops in absolute initial position are characterized by a short positive VOT for some speakers and a negative VOT for others while voiceless stops in the same position show a long positive VOT value (p. 394). The distinction between voiced and voiceless oral stops in Dutch is realized differently, with a negative VOT for voiced stops and a short positive VOT for voiceless stops (p. 392). VOT measurements can be used to acoustically distinguish stop types in languages which contrast voiced, plain voiceless, and aspirated voiceless stops as well. In Thai, for example, voiced stops have a long negative VOT, plain voiceless stops a short positive VOT, and aspirated voiceless stops a long positive VOT (p. 396).

The usefulness of VOT for differentiating between all stop types falls short in a language with a four-way contrast, however, because both the plain voiced stops (such as /d/) and breathy voiced stops (such as /d<sup>h</sup>/) include vocal fold vibration during the stop closure. This means that both stop types have long negative VOTs. Average VOT data for Hindi and Marathi velar stops



appear in Table 4.<sup>10</sup>

**Table 4 Average VOT of velar stops in Hindi and Marathi (in ms)**

	/k/	/k <sup>h</sup> /	/g/	/g <sup>h</sup> /
<b>Hindi</b>	18	92	-63	-75
<b>Marathi</b>	24	87	-116	-89

*data from Lisker and Abramson 1964: 398<sup>11</sup>*

At first glance, mean VOT values for the plain and breathy voiced stops appear to differ. The problem noted by Lisker and Abramson, however, is the significant overlap in their VOT ranges; Hindi /g/ values range from -95 to -30 ms, for example, while /g<sup>h</sup>/ values range from -160 to -40 ms. Virtually the entire /g/ range, then, falls inside the /g<sup>h</sup>/ range. This extensive overlap renders VOT uninformative in distinguishing the plain voiced from the breathy voiced stops.

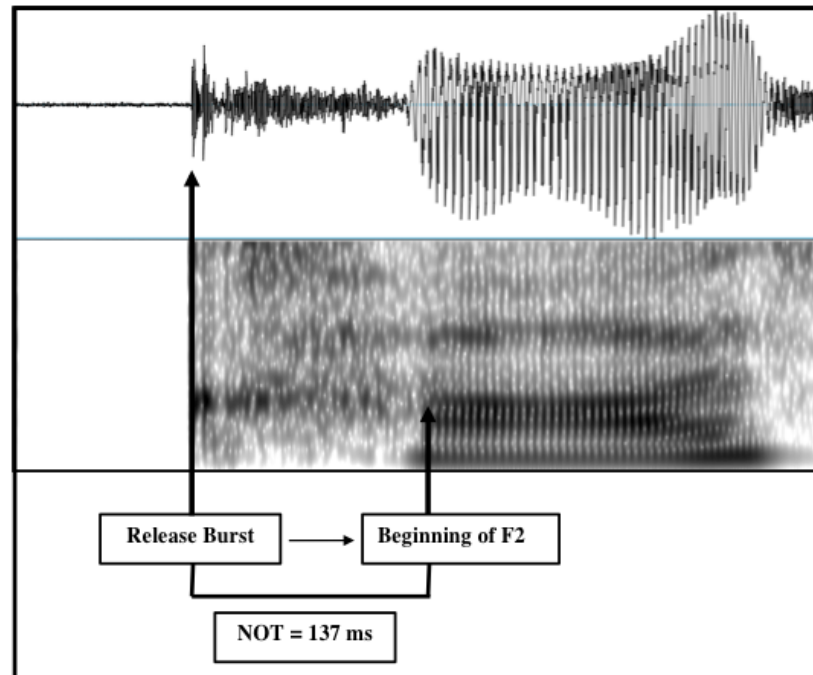
Several alternative durational measures have been proposed, including “Noise Offset Time” (Davis 1994). Noise Offset Time (or NOT) is measured from the burst associated with the release of the stop closure to the beginning of a clear second formant in the following vowel, as illustrated in Figure 4. This image represents the initial portion of the Marathi word [k<sup>h</sup>as], ‘*special*,’ as produced by a female native speaker of Marathi.

---

<sup>10</sup> Although only the velar data are reported here, Lisker & Abramson report VOT values for the labial, dental, and retroflex stops as well. The pattern in these other places of articulation mirrors the pattern found in the velars, and so the velars are used as representative.

<sup>11</sup> Lisker & Abramson recorded a total of seventeen speakers producing samples from eleven distinct languages; the number of speakers recorded for each language is not reported, however.

Figure 4 Illustration of NOT beginning and endpoints in [k<sup>h</sup>as], 'special'



Davis reports average NOT values for the velar stops of Hindi, and these appear in Table 5. Note that Davis did not report data for the other places of articulation.

Table 5 Average NOT of Hindi velar stops (in ms)

	/k/	/k <sup>h</sup> /	/g/	/g <sup>h</sup> /
<b>NOT</b>	26	74	12	34

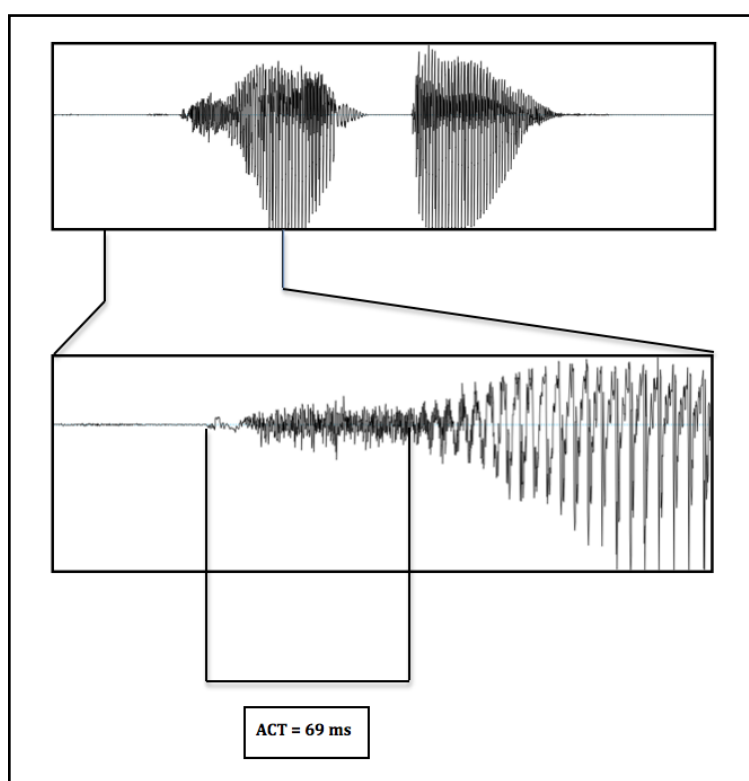
*data from Davis 1994: 184*

Pairwise comparisons between each set of phonemes (i.e. /k/ and /k<sup>h</sup>/, /k/ and /g/, et cetera) confirm the trend observed numerically: each NOT average differs significantly from the others.

Some criticism has been leveled against the NOT, however, in part because the visibility of F2 may vary based on factors such as the intensity of the speech signal itself and on the specific settings of the program being used to generate and view a spectrogram of the sound (Mikuteit & Reetz 2007: 253). The subjectivity involved in pinpointing F2 is obviated in After

Closure Time (ACT), another alternative durational measurement proposed by Mikuteit & Reetz (2007). In proposing the ACT, the authors accurately observe that the post-release interval can be divided into two distinct portions. The first of these, the ACT, consists of the aperiodic buzzy noise associated with the closure release and is found in all four stop types. An ACT measurement is illustrated in Figure 5. Note that for a voiceless aspirated stop, ACT corresponds with VOT.

**Figure 5 Measurement of ACT in [p<sup>h</sup>ɑtɑ], ‘fork (in the road)’**



Mikuteit and Reetz refer to the second portion as Superimposed Aspiration (SA). The SA retains some jaggedness, but this jaggedness is laid over the clear periodicity associated with the vocal fold vibration of the following vowel. This is the portion referred to as jagged periodicity in the illustration in Figure 3. It may be likened to a breathy vowel, and appears in aspirated voiceless and breathy voiced stops. The portion of the waveform demarcated as breathiness in Figure 2,

then, is comprised of both the ACT and the SA.

Consider that the ACT is measured from the closure release to the beginning of clear periodicity; as such, we may think of it as a measure that disregards closure voicing and captures only positive VOT values. Mikuteit & Reetz’s average ACT values (reported as least square means) for the oral stops at four places of articulation in East Bengali appear in Table 6.

**Table 6 Average ACT values of East Bengali stops (least square means, in ms)**

	BILABIAL	DENTAL	RETROFLEX	VELAR
<b>Voiceless Plain</b>	14	15	11	34
<b>Aspirated Voiceless</b>	--	57	48	73
<b>Voiced Plain</b>	8	13	7	23
<b>Breathy Voiced</b>	31	35	36	44

*(data from Mikuteit & Reetz 2007: 184)*

The above data seem to pattern with the NOT data presented in Figure 4; if we consider the velars, the numerical values differ but the trend remains the same. For both the ACT in East Bengali velars and the NOT in Hindi velars, plain voiced stops have the shortest values, followed by the plain voiceless stops; breathy voiced stops have the second to longest values, and aspirated voiceless stops have the longest values (i.e. /g/ < /k/ < /g<sup>h</sup>/ < /k<sup>h</sup>/). It is important to note, however, that statistically significant differences between means for each pair of phonemes are not achieved in all four places of articulation. Davis considered only the velars, and pairwise comparisons within the velars were significant. Mikuteit and Reetz report the same finding for ACT values in East Bengali velars, but the pattern does not hold across labials, dentals, and retroflexes.

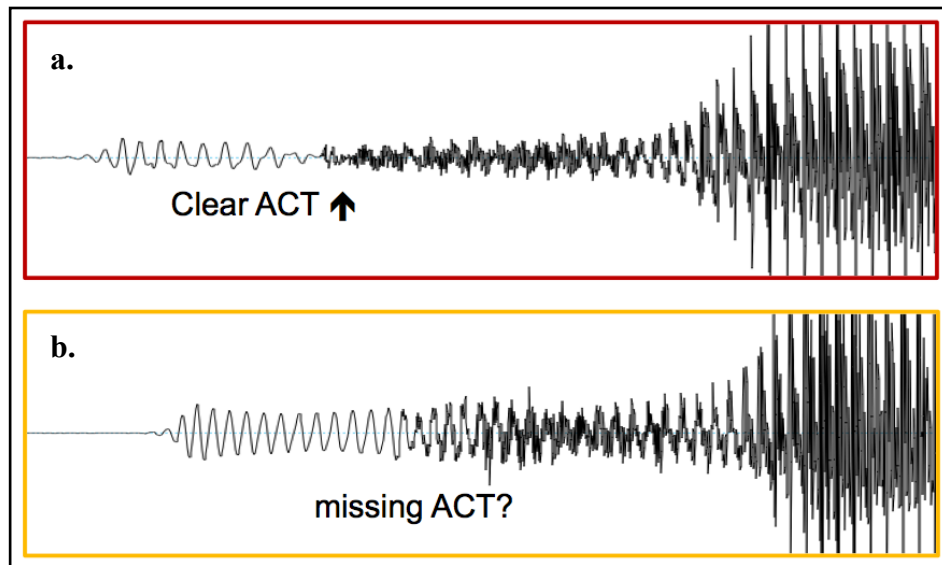
In short, ACT values reliably distinguish the plain from the aspirated/breathy voiced phonemes, but—outside the velar category—they are not sufficient on their own to reflect a four-way contrast. ACT fails to distinguish the plain voiced from the plain voiceless stops. This is not a problem, given that we are not looking for a single acoustic parameter which can differentiate

the stop types; in terms of differentiation, the plain voiced and voiceless stops are clearly distinguished based on the presence (or absence) of vocal fold vibration.

In terms of a comparison, ACT and NOT measurements are similar in that both look at the post-release/pre-vocalic section of the segment. They differ, however, in that the NOT is composed of both the random noise associated with the release burst and a portion of the pre-vocalic glottal buzz which may or may not involve some periodicity, while the ACT is composed of only the post-release random noise, and ends when periodicity becomes visible in the waveform. Periodicity accompanied by “glottal buzz” (à la Lisker & Abramson) in the breathy-voiced stops belongs to the superimposed aspiration portion of the interval, with the periodicity presumed to belong to the following vowel. Mikuteit and Reetz posit that the NOT includes both what they measure as the ACT and some portion of what they designate superimposed aspiration, and when they replicate NOT measurements with their own data they find that this is indeed the case. NOT generally corresponds with ACT measurements for plain stops, and includes some milliseconds of SA for the aspirated and breathy voiced stops.

In the end, both measurements have advantages and both have drawbacks. As noted, the advantage of NOT is that it makes use of a landmark which is unfailingly present in the spectrogram: namely, F2. The drawback, however, is that pinpointing the exact onset of a clear F2 may be open to interpretation. The advantage of measuring the ACT is that, procedurally, it relies on landmarks found in the waveform and is presumably objective rather than subjective. The problem with ACT measurements, however, is that the interval referred to as the ACT and characterized by random noise and a lack of periodicity in the waveform is very often absent in the Marathi data. This is illustrated below.

Figure 6 Clear ACT, missing ACT: two waveforms of the [b<sup>h</sup>] in [b<sup>h</sup>at], ‘rice’



In 6a., we see a clear—if brief—cessation in periodicity, which we can confidently segment as the ACT. In 6b., however, there is no such interval; rather, the smooth periodicity of the waveform representing the prevoicing of the stop gradually becomes slightly jagged, and this is followed by a section of jagged periodicity which eventually transitions into the regular periodicity of the vowel.

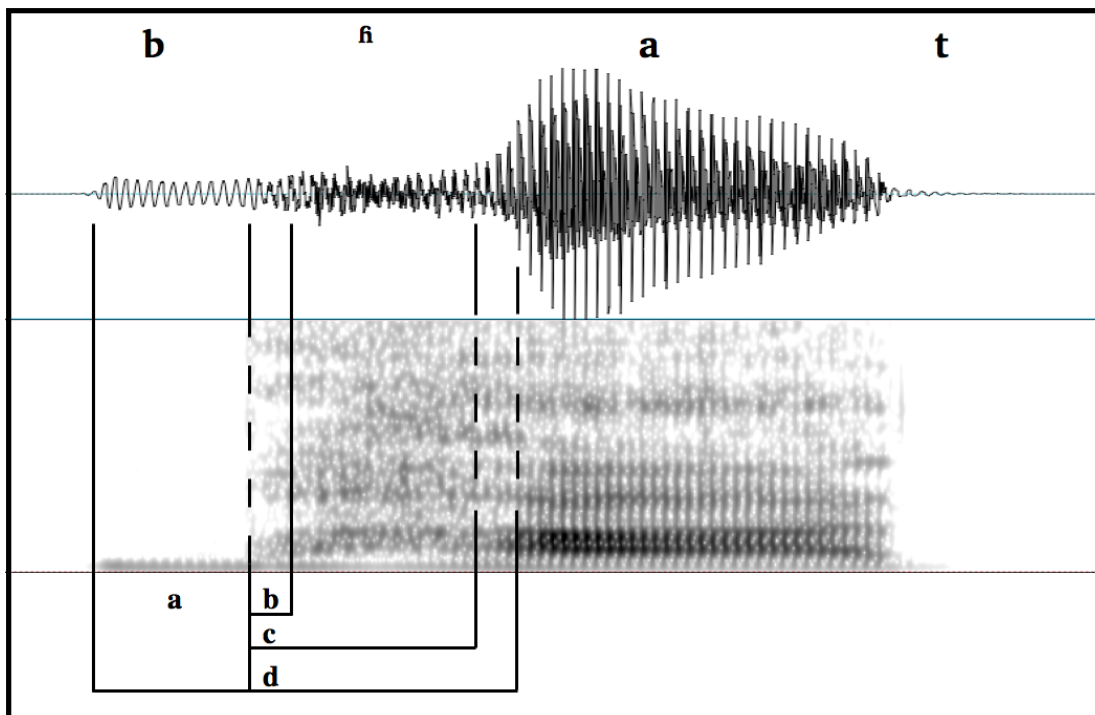
This jagged periodicity is what Mikuteit and Reetz referred to as superimposed aspiration, and there is no difficulty in describing what we see: the jagged periodicity results from the breathiness of the release extending into the following vowel. Dutta (2007) reports that, in Hindi stops, the breathiness extends through approximately the first 30% of the subsequent vowel. The difficulty comes in attempting to divide this interval into the two portions defined by Mikuteit and Reetz, for in the tokens which show no clear ACT there is no abatement in voicing and no cessation in periodicity once the consonant has begun. The initial stop is voiced, the subsequent vowel is voiced, and there is voicing throughout the closure release. Thus while the observation

that the post-release interval of aspirated/breathy voiced stops is composed of two portions is clearly accurate, division of the interval into the portions defined as ACT and SA is not clear-cut.

One additional comment should be made. ACT measurements are relatively long for voiceless aspirated sounds, and much shorter for the other three sounds. Average NOT durations are fairly long for aspirated voiceless and breathy-voiced stops, while plain stops of both types exhibit shorter NOT durations. This grouping more closely mirrors the perceptual experience listeners have when hearing these types of stops; the breathy-voiced part of breathy-voiced stops is perceptually salient, and is clearly distinguishable. While perceptual data cannot help us untangle the acoustic puzzle, we can look to it for clues as to whether we are moving in the correct direction.

As a final illustration of the complications inherent in this task, consider the intervals demarcated in the spectrogram and waveform found in Figure 12.

*Figure 7* Waveform and spectrogram of [b<sup>h</sup>at], 'rice'



Demarcated in 7a is the stop closure. As shown in 7b, which begins with the release burst, there is no measureable ACT in this token because there is no aperiodicity; rather, the wave moves from the smooth periodicity of the stop closure directly into the jagged periodicity of the breathy portion of the sound. 7c, which measures the NOT, ends at the onset of a visible F2 in the spectrogram; as previously noted, however, the start of a visible F2 is mildly subjective at best and wildly subjective at worst. Finally, the interval labeled 7d begins with the release burst and ends with the beginning of a consistent rise in amplitude in the waveform.

At issue here is the fact that the acoustic realization of the breathy voiced stops varies from token to token; in Figure 6 we see waveforms of two breathy-voiced tokens which differ from one another, and this trend of variation holds from speaker to speaker and within a single speaker's tokens. It renders application of ACT as a durational measurement difficult if not impossible, for many tokens simply fail to exhibit a clear ACT. The challenge, then, seems to be to determine which portions of the acoustic signal remain consistent and measurable across all stop types and tokens.

What, then, remains consistent across all tokens? The elements that I would like to point to include the consonant release, which is generally visible in the waveform and which, in the rare instances when it is not, can be identified in the spectrogram. Also consistent is the jump in the amplitude which corresponds with the cessation of the breathy interval. This jump in amplitude is visible in the waveform, and often coincides with the darkening of vowel formants in the spectrogram.

I propose utilizing these consistent hallmarks to measure what I refer to as the Pre-Vocalic Interval (PVI). PVI measurements for aspirated and breathy voiced stops include not only the closure release duration, but also that portion of the vowel which is heavily flavored by



the breathy release—this is because the breathiness perseverates into some portion of the vowel, causing the aforementioned jump in amplitude to occur sometime after vowel onset. (If Marathi exhibits a pattern similar to the Hindi pattern reported in Dutta (2007), breathiness may persist for approximately 30% of the subsequent vowel.) The PVI begins at the release of the consonantal constriction, and ends at the upwards-going zero crossing of the period after which no subsequent periods show decreased amplitude.<sup>12</sup> In other words, the PVI consists of both post-release intervals identified by Mikuteit and Reetz (2007)—the After Closure Time and the Superimposed Aspiration.

Future studies may reveal that the variability observed in the Marathi data is not present in all languages; given Mikuteit and Reetz (2007), for instance, we can presume that the East Bengali data yielded reliable ACTs and less variability. Even in a language with less variability, however, the landmarks utilized in measuring PVI will be clearly visible, rendering it applicable across the board: it can be reliably measured in languages with a great deal of variability, as well as in languages with more consistency. This leads back to the point mentioned previously: by relying on characteristics of the waveform which are consistently present, the PVI avoids the subjectivity inherent in looking for landmarks in the spectrogram and ensures replicability. Future development of a script or algorithm that can use millisecond-by-millisecond intensity readings to automate identification of the PVI end-point will serve to render this measurement even more replicable.

In fact, there is at least some empirical evidence to support the intuition that the waveform yields more reliable measurements. Francis *et al.* (2003) analyzed a number of acoustic measurements which are used fairly commonly to assess voicing onset after stop

---

<sup>12</sup> Development of a script which will make use of the millisecond-by-millisecond intensity readings that can be generated by Voicesauce (Shue *et al.* 2011) is currently underway in order to automate the process of selecting the PVI endpoint. At present, however, this point is picked by hand in PRAAT (Boersma & Weenink 2010).

consonants in syllable onset position; these included measures made from the waveform as well as from the spectrogram. Their findings indicate that those measures which rely on the waveform boasted the greatest accuracy, and were also the least variable, when compared with the true onset of voicing as determined via an Lx waveform generated from a laryngographic recording. The finding that a waveform-based measure most closely captured the onset of voicing bolsters the claim made here when proposing the utility of the PVI: namely, that the waveform provides more reliable landmarks. Thus PVI is valuable both because it is made relying primarily on the waveform—and too, because it utilizes landmarks which are reliably present.

An additional advantage of the PVI is that it can be measured on breathy sonorants in addition to all obstruents. The disadvantage is that it cannot be directly compared with the measures discussed in the previous literature (VOT, NOT, ACT). That said, the expected pattern is in line with previous findings, with one exception. The temporal interval between the closure of the stop to the modal voicing of the subsequent vowel increases from plain voiced sounds to plain voiceless sounds to aspirated voiceless sounds to breathy voiced sounds. Breathily voiced sounds have shorter NOT and ACT values than aspirated voiceless sounds; PVI diverges from this pattern because it includes a greater stretch of the jagged periodicity inherent in breathy phonation.

### **2.4.3 Vowel Duration**

We may also consider the effect of consonant phonation type on the duration of the vowel following the consonants of interest. With regards to temporal measures in this dissertation, this means specifically that portion of the vowel that follows the VOT, ACT, or NOT intervals. Dutta (2007b) reports that, for Hindi stops, vowels following plain stops are

shorter than those following aspirated and breathy voiced stops. Vowels after plain stops average between 160 and 200 ms, while those following aspirated and breathy voiced stops average between 200 and 250 ms. There is a very good reason why we may expect different results in the present study, however, and that is because part of what would be included in vowel duration measurements if we were utilizing VOT, ACT, or NOT is here included in the PVI instead. Thus while the breathy interval as reflected by mean PVI values is expected to be longer than ACT or NOT values reported in previous work, vowel durations are expected to be shorter than those found in previous studies.

To review the durational measures presented above with a focus on breathy and modal phonation, then, breathy voiced stops tend to exhibit shorter closure durations than their modal counterparts. Breathily voiced sonorants and vowels, on the other hand, are reportedly longer than modal sonorants and vowels. The aspirated and breathy interval associated with aspirated and breathily voiced sounds is, obviously, longer than the lag-time associated with plain sounds, and this is reflected in the fact that breathily voiced sounds have longer NOT, ACT, and PVI values than their modal counterparts. Finally, at least one previous study (Dutta 2007b) found vowels after breathily voiced sounds to be longer than those after plain sounds.

## **2.5 Fundamental Frequency**

Crosslinguistically, there are some indications that breathy phonation is associated with a lowered fundamental frequency (Gordon and Ladefoged 2001). This has been found in Hindi, for instance, where F<sub>0</sub> is lower after voiced than after voiceless stops, and even lower after breathily voiced stops than after their plain voiced counterpart (Dutta 2007b, Ohala 1979, Pandit 1957, Schieffer 1986). In fact, Ohala (1979) notes that breathy phonation may lower mean F<sub>0</sub> rates by

as much as 20 Hz. This lowering may arise from the laryngeal configuration necessary for production of breathy voice, which entails less tension in the vocal folds and thus a slower basic rate of vibration (Laver 1980, Ní Chaisade and Gobl 1995, Blankenship 1998, Dutta 2007).

In a 2005 dissertation regarding glottal manner features, however, Pennington warns that “it is not yet clear how universal the effect is” (p. 24). This is an important point. While it is true that the above-mentioned studies have found breathy voice to be associated with lowered F0 values, data from additional languages must be collected before we can make concrete statements about the universality of breathy phonation and lowered F0 values.

One thing that is clear, however, is that in languages where the lowering effect of breathy voice on F0 values has been found, the phenomenon may persevere far beyond the onset of the vowel. Breathily voiced consonants in Hindi may trigger lowered F0 values for at least 30% of the subsequent vowel (Dutta 2007b). This is consistent with the notion that breathiness itself perseverates into some portion of the vowel. In Jalapa Mazatec, which features a phonemic breathy/modal contrast for vowels, the breathiness yields to modal voicing about 43% of the way into the duration of the vowel (Silverman *et al.* 1995). Note that while Silverman *et al.* do not report F0 values for Jalapa Mazatec vowels, a more recent study reports that F0 is not lowered by breathy phonation in this language (Garellek & Keating, 2010).

Several comments should be made. First, results related to F0 differences as a function of phonation type are often restricted to vowels. Previous research into the effect of phonation type on the consonant-internal F0 value of sonorants is, to my knowledge, nonexistent. Recall, however, that F0 measurements for breathy and modal phonation in Marathi can be taken on nasals, laterals, approximants, and rhotics, in addition to the vowels following not only these

sonorants but obstruents as well. As such, the consonant-internal F0 values collected herein constitute a contribution to the existing crosslinguistic findings.

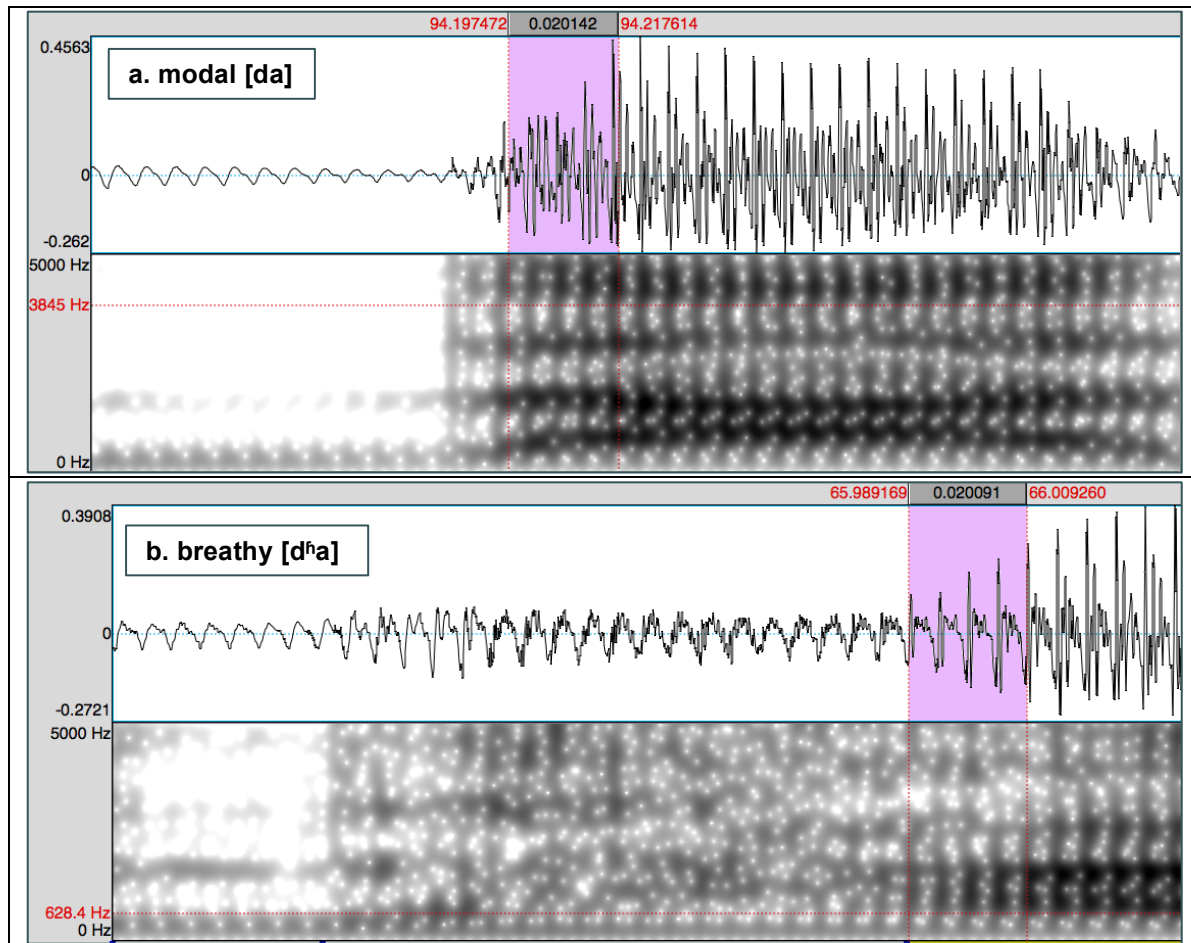
## **2.6 Spectral Measures**

The articulatory characteristics of breathy phonation have a predictable effect on the breathy voice spectrum, which is typically characterized by a strong fundamental component and a rapid drop-off in higher-frequency energy. As such, breathy voiced sounds tend to exhibit a sharp downward spectral tilt (Pennington 2005). This is often assessed by looking at the strength of the fundamental component of the spectrum—or the first harmonic—and comparing it with higher-frequency components. Some of the spectral measures which have been used to capture this include H1-H2, H1-A1, H1-A2, and H1-A3. Each of these measures is discussed in more detail in the following sections, but it is worthwhile to note briefly here that H1 refers to the amplitude of the first harmonic, H2 to the amplitude of the second harmonic, A1 to the amplitude of the harmonic which is dominant in the range of the first formant, A2 to the amplitude of the harmonic which is dominant in the range of the second formant, and A3 to the amplitude of the harmonic which is dominant in the range of the third formant.

In saying that breathy voice is characterized by a sharp downward spectral tilt, we mean that energy in the spectrum drops off rather rapidly. In fact, we can see that there is less high-frequency energy present in breathy voiced sounds than in modal sounds if we look at the images presented in Figure 8. (The associated spectra are presented in Figure 9.) Recall that amplitude is reflected in the spectrogram through darkness—the darker the bands of energy shown in the spectrogram, the greater the amplitude of that component. The spectrogram for the modal example in 8a. retains a fair amount of energy into the higher frequencies, as evidenced by the

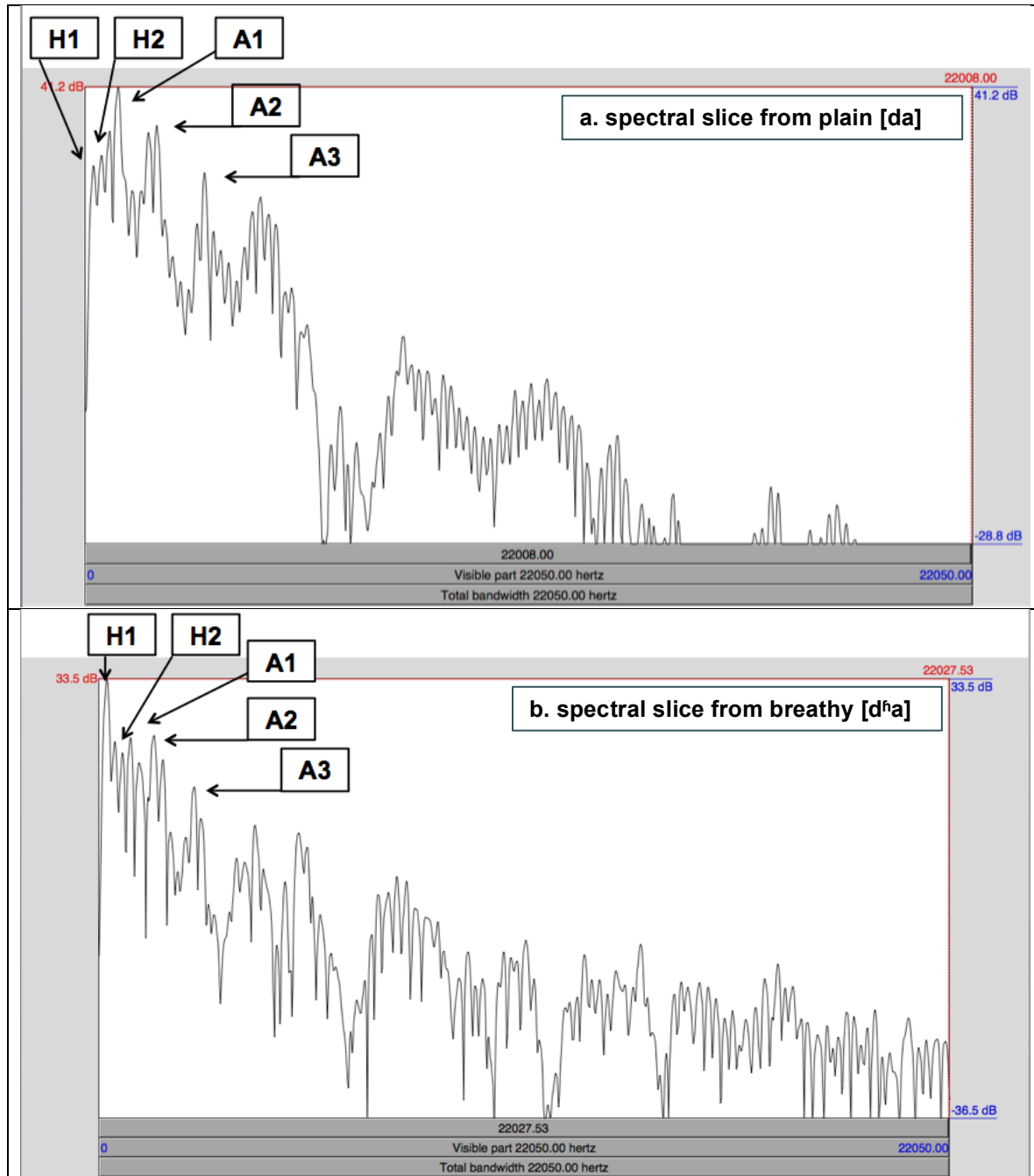
fact that the formant bands in the spectrogram remain dark all the way to the 5000 Hz mark, and this will be reflected in a spectral slice taken from the highlighted portion. In 8b., the portion that has been highlighted is at the tail end of the breathy interval. Here, although the lower formants are darker, the upper frequency range is very light, indicating much less energy in the higher frequencies. Thus, in the spectrum generated from a breathy voice source, we expect a great deal of energy to be associated with the first harmonic, and less energy to be associated with the upper harmonics.

**Figure 8 Waveform and spectrograms of modal and breathy voiced stops**



Spectra generated from the spectral slices highlighted in 8a. and 8b. appear in Figure 9. Note that in 9b., the amplitude of H1 is greater than that of H2, A1, A2, and A3. In the modal voice spectrum shown in 9a., however, where the fundamental component is weaker and the energy does not drop off as steeply, H1 is actually smaller than H2, A1, and A2. If we took an H1-A3 measurement from both these spectra, we would yield positive values for both; the value for 9b., however, would be greater. This is the expected pattern when looking at the spectra of modal and breathy voice sources. With a large H1 and smaller H2, A1, A2, and A3 values than modal voice, breathy voice should yield large positive H1-H2, H1-A1, H1-A2, and H1-A3 values which exceed those values associated with modal phonation.

Figure 9 Illustration of H1, H2, A1, A2, and A3 on spectra generated from the modal and breathy voice sources shown in Figure 8



When discussing the spectral characteristics of breathy phonation, one study that will be referred to repeatedly in the following sections is Esposito (2006), which provides acoustic measurements of modal and breathy vowels in Chong, Fuzhou, Green Hmong, White Hmong,



Mon, Santa Ana del Valle Zapotec, Tlacolula Zapotec, Tamang, and !Xóǝ. An outline of the study is provided here.

In order to determine how a listener's linguistic background affects perception of phonation type contrasts in unknown languages, speakers of Gujarati (which contains phonemic breathy phonation), Mexican Spanish (which contains no breathy phonation), and English (which contains allophonic breathiness) listened to stimuli from the ten languages and dialects listed above and performed a free-sort task. In this task, listeners sort the stimuli according to their own criteria—each participant makes their own determination about what “sounds” breathy, and so forth.

To create stimuli for the free sort task, four tokens were excised from audio recordings of normal speech in each of the ten languages and dialects—for each, the tokens consisted of two breathy and two plain tokens of [a] produced by a male native speaker. Vowel duration was normalized to 250 ms, and F0 was normalized to a slightly falling contour of 115 to 110 Hz. Multiple acoustic measures were then taken for each stimulus: the ones replicated in the present study, and therefore of primary interest to us, include H1-H2, H1-A1, H1-A2, H1-A3, and Cepstral Peak Prominence (CPP). Taking these measures ensured that Esposito could determine how they contributed to (or correlated with) listeners' perceptions. Now that the study has been outlined, relevant results will be discussed in the following sections.

Again, note that while the segments of interest in our investigation of Marathi are consonants rather than vowels, vowel data are of interest because the acoustic cues associated with breathy voiced consonants may well perseverate into subsequent vowels. The present work will diverge from that conducted on vowels, however, in that the spectral measures discussed

below will be conducted on the consonantal constriction of sonorants in addition to vowels following obstruents and sonorants.

### 2.6.1 H1-H2

H1 refers to the amplitude of the first harmonic, and H2 to the amplitude of the second harmonic. Subtraction of the amplitude of H2 from the amplitude of H1 is considered to reflect Open Quotient (OQ), meaning the proportion of one cycle of vocal fold vibration for which the VF are open (i.e., the time the folds are apart divided by the duration of the entire cycle) (Pennington 2005). In a study comparing breathy and modal vowels in Hmong, Huffman (1987) reports that breathy vowels are characterized by a stronger H1 relative to the strength of H2, and that they have a higher open quotient than modal vowels as well (0.8 as opposed to 0.6). Regarding the correlation reported by Huffman, Klatt and Klatt (1990) state that “there is a simple cause and effect relationship between [the two]—an increased open quotient results in a stronger fundamental component of the source spectrum, all else being equal” (p. 824). The correlation between the two has also been proven in an experimental setting by Holmberg *et al.* (1995).

In general, H1-H2 values are greater for breathy voice than for modal voice—by about 7 dB, in the Hmong data reported by Huffman (1987). In addition to Hmong, previous studies have also found that H1-H2 successfully distinguishes breathy from modal vowels in Green Hmong (Andruski and Ratliff 2000), Khmer (Wayland and Jongman 2003), and Gujarati (Esposito 2006, Khan 2012). It also effectively differentiates the three-way distinction between breathy, modal, and creaky vowels in Mazatec (Blankenship 2002).

Furthermore, H1-H2 distinguishes breathy from modal voice for female speakers of Santa Ana del Valle Zapotec (Esposito 2006), and for both males and females for seven of the remaining nine languages analyzed therein. In all of these cases, the H1-H2 values—taken over a 30 ms window from a point either 50 ms into or 50 ms before the end of the vowel, depending on where the phonation type contrast was realized—are greater for breathy than for modal sounds, reflecting the fact that H1 is dominant and that energy in the spectrum drops off rather steeply in breathy voiced sounds.

If we turn to phonation type contrasts among consonants, consonant phonation type triggers reliable differences in H1-H2 values in subsequent vowels in Hindi stops, with breathy voiced stops triggering increased H1-H2 values (Dutta 2007b). While this result held true across all speakers analyzed, Dutta further reports that there was inter-speaker variability in terms of how far into the vowel the difference persisted. The measures in this study were taken from a 30 ms window placed at 10%, 30%, 50%, 70%, and 90% of the vowel. Of the five speakers whose data was analyzed, H1-H2 values after breathy voiced stops were significantly greater for 30% of the vowel for one speaker, 50% for one speaker, and 70% for one speaker. For the remaining two speakers, differences were significant at all five measurement points—meaning that, for those two speakers, differences persisted through at least 90% of the following vowel. These data are interesting, for they underscore that breathy phonation does indeed persevere into subsequent vowels. The specifics may vary in terms of how far breathiness perseverates, but it is true across the board that consonant phonation type has a significant effect on the spectral characteristics of the subsequent vowel.

Moving from obstruents into data related to sonorants, Harris (2009) reports that H1-H2 values do not distinguish breathy from modal nasals and laterals in Sumi. In these analyses,

spectral slices were taken at the midpoint of each modal segment (i.e., at the midpoint of [m]) and at both the midpoint of the sonorant portion and the point where the breathiness began in a breathy segment (i.e., at the midpoint of the [m] portion of an [m<sup>h</sup>] token, and at the boundary of the [m] and the [h] portions). Harris reports that both the nasals and laterals show a high degree of spectral similarity up to at least 1000 Hz, meaning that the lower-frequency measures like H1-H2 are not useful in distinguishing plain from breathy phonation. If Marathi patterns like Sumi, then, there is some question as to whether we will find sonorant-internal differences in this measure in the Marathi data. Furthermore, if it is consonant phonation type which triggers reliable differences in subsequent vowels—as clearly indicated by the Hindi data—then a lack of differentiation sonorant-internally may mean that H1-H2 differences in vowels after sonorants will not be evident either. This remains an open question.

A final comment that should be made here has to do with the fact that there is an increasing awareness that H1-H2 might not be a reliable measure of breathiness because the lower harmonics are vulnerable to being affected by the energy associated with F1 (Simpson 2012). This concern can be minimized by using only low vowels like [a], which have a relatively high F1, and in fact many people do restrict their studies to the low vowel context (Dutta 2007b and Esposito 2006, for instance). This has its own drawbacks, however, as it is very restrictive. Furthermore, phonation type contrasts are not limited solely to the low-vowel context, and so additional parameters that are less vulnerable to such influences have also been utilized in previous work. Another possible fix involves looking at corrected H1-H2 measures, for which algorithms that correct for the influence of formant frequencies and bandwidths are used. This is the choice made in the present work—the spectral measures reported herein are corrected values, meaning that the program used to take spectral measures implemented a correction algorithm

when taking measurements. This is discussed more thoroughly in the subsequent chapter, which deals with methodology.

### **2.6.2 H1-A1, H1-A2, H1-A3**

As previously mentioned, A1 refers to the amplitude of the harmonic that is dominant in the range of the first formant, A2 to the amplitude of the harmonic that is dominant in the range of the second formant, and A3 to the amplitude of the harmonic that is dominant in the range of the third formant. Typically, these measures yield smaller values for modal than for breathy phonation; as explained above, this is because in the breathy voice source spectrum the fundamental component is strong and the amplitude of the higher components drops off rather dramatically. An illustration of these spectral patterns was provided in Figure 9.

As with H1-H2, these spectral measures may be affected by vowel quality because of the potential influence of formant frequencies and bandwidths on the amplitude of the nearby harmonics. Blankenship (1997) reports the results of spectral measures on breathy and modal phonation in several languages, for example, one of which is Mpi. Though not by design, the Mpi data comes from words which contained the high vowel [i], with the result that the amplitude of the first harmonic is consistently boosted by F1. As with H1-H2, this potential issue is controlled for in the present work by reporting corrected spectral measures. The fact that these measures may be impacted by vowel quality is something that should be kept in mind as we move forward, however.

### 2.6.2.1 H1-A1

The usefulness of these three measures—H1-A1, A2, and A3—shows a fair degree of crosslinguistic variation. In a recent study, H1-A1 proved useful in distinguishing breathy from modal phonation in Gujarati (Khan 2012). This study was motivated by several factors: among them is the fact that while previous studies of Gujarati have reported that an increase in the amplitude of H1 relative to other low-frequency components of the spectrum such as H2 and A1 successfully distinguishes between plain and breathy vowels in the language—Fischer-Jorgenson (1967), for instance, as well as Bickley (1982)—these studies did not control for the potential influence of formant frequency and bandwidth interference. As mentioned, however, present methods for taking such measures utilize algorithms that automatically correct for formant interference. One of the ramifications of this is that if formant interference can be corrected for, it is then possible to move beyond the low-vowel context that was so often favored in previous research.

Accordingly, Khan (2012) recorded ten native speakers of Gujarati producing breathy and modal low and mid vowels (/a, e, o/ and their breathy counterparts). Tokens were real words. Acoustic measurements were taken for every millisecond of each vowel. Vowels were divided into nine equal portions, and measurements for each portion were averaged—thereby yielding a mean value for nine equidistant points, which Khan refers to as time points. This not only allows analysis of whether and how harmonic amplitudes differ based on vowel phonation type, but also makes it possible to gain a sense of how things change over time. With regards to H1-A1, Khan reports that breathy vowels show significantly greater values at all nine time points. The difference is significant across the entire duration of the vowel, in other words.

This is the expected pattern: breathy phonation is expected to trigger greater H1-A1 values than modal phonation, and that is the result most commonly seen. According to previous findings, modal vowels are distinguishable from breathy vowels based on H1-A1 in Khmer (Wayland & Jongman 2003) and in Hmong, Fuzhou, Mon, Santa Ana del Valle Zapotec, Tlacolula Zapotec, and !Xóõ (Esposito 2006). Esposito (2006), however, demonstrates that although H1-A1 values are greater for breathy than for modal phonation in most of the languages assessed, the pattern is reversed for Chong and for Mon. This serves to underscore the variability of these measures.

Recall that in the Sumi data reported by Harris (2009), the spectra of plain and breathy nasals and laterals were highly similar up to at least 1000 Hz. As such, if Marathi sonorants pattern with Sumi sonorants, H1-A1 values are not expected to differ sonorant-internally, and the question of whether sonorant phonation type will trigger H1-A1 differences in subsequent vowels remains open.

Data related Hindi obstruents is also informative, and suggests that obstruents may well trigger significant H1-A1 differences in subsequent vowels. Recall that Dutta (2007b) assessed the effect of Hindi stops on subsequent vowels in tokens produced by five native speakers. Measurements were at 10%, 30%, 50%, 70%, and 90% of the vowel, and H1-A1 values after breathy voiced stops were significantly greater at all five time points for two of the five speakers, and at four of the five time points for the remaining three speakers. The effect of consonant phonation type on H1-A1 values in subsequent vowels in Marathi, then, may well be significant.

### 2.6.2.2 H1-A2

As with H1-A1, the expected pattern when looking at H1-A2 values associated with plain and breathy phonation is to find greater values associated with breathy than with plain phonation. This has in fact been found in Mazatec (Blankenship 2002) and in Hmong, Fuzhou, Mon, Santa Ana del Valle Zapotec, Tlacolula Zapotec, and Tamang (Esposito 2006). In these languages, modal vowels are distinguishable from breathy vowels based on H1-A2. Dutta (2007) found a similar effect in Hindi, with breathy voiced stops triggering increased H1-A2 values in subsequent vowels as compared with modal voiced stops. Dutta further reports that the effect of breathy phonation on H1-A2 persisted up to the 50% measurement point for one of the five speakers whose data was analyzed, and up to the 70% measurement point for the remaining four speakers.

With regards to plain and breathy vowels in Gujarati, Khan (2012) reports that H1-A2 values associated with breathy phonation are significantly greater for all time points measured. He further reports that there is no interaction between H1-A2 and vowel quality—the effect was significant in both the low- and the mid-vowel contexts. Thus it seems likely that consonant phonation type will trigger differences in subsequent vowels in the Marathi data, though the question of how far into the vowel the differences will persist remains open.

If we consider the consonant-internal measures reported by Harris (2009), we may guess that because the spectral profile of plain and breathy sonorants began to diverge above 1000 Hz, differences in H1-A2 values would have been found. This measure was not taken and so we cannot say for sure, but it seems a likely extension of the observation that the spectral profile diverged above 1000 Hz based on sonorant phonation type. That said, Harris did report H1-A3 values, and these did not differ significantly based on phonation type. This seems puzzling, and



diverges from the findings reported by Traill and Jackson (1987). In an investigation of plain and breathy nasals in Tsonga, a Bantu language, ten female and five male native speakers were recorded producing a set of real Tsonga words that contained plain and breathy nasals before the vowel [a]. Measures of spectral tilt were calculated by hand based on spectra generated for the nasal consonants and for the first 40 ms of the subsequent vowel. These measures were not H1-A2 or H1-A3, but are somewhat similar to them: the first was H1 minus the amplitude of the most intense harmonic around 1400 Hz, and the second was H1 minus the amplitude of the most intense harmonic above 2000 Hz. While there was some inter-speaker variation, the measure which is closest to H1-A2—i.e., that of H1 minus the amplitude of the harmonic nearest 1400 Hz—was consistently reliable in distinguishing between plain and breathy nasals. This was true for both male and female speakers. The result obtained only in the nasal-internal measurements, however; the measurements taken on the first 40 ms of the subsequent vowel did not show the same reliable differentiation.

Given these results, it is perhaps unclear what we should expect with regards to the Marathi data. The Tsonga findings indicate that sonorant-internal differences may be significant, but the Sumi findings contradict that. The results from Hindi indicate that obstruent phonation type triggers differences in subsequent vowels (Dutta 2007b), but the Tsonga data does not show the same thing for vowels after sonorants (Traill and Jackson 1987). Overall, however, what does seem clear is that where consonant phonation type does trigger differences in H1-A2 values, breathy phonation should be associated with larger positive values than modal phonation.

### 2.6.2.3 H1-A3

Previous research regarding H1-A3 falls very much in line with the results reported above for H1-A1 and H1-A2, meaning that there is some amount of variability. Some previous studies have found this parameter to successfully distinguish between plain and breathy phonation, while others have not. Furthermore, while values associated with breathy phonation are expected to be greater than those associated with modal phonation, this is not always the case.

In terms of languages that employ phonemic breathy voice in vowels, reliable differences in the H1-A3 values associated with modal and breathy vowels have been found in Hmong, Green Hmong, White Hmong, Mon, Tlacolula Zapotec, Tamang, !Xóǀ, and for male speakers of Santa Ana del Valle Zapotec (Esposito 2006). One surprising wrinkle here, however, is that Esposito further reports that while the values for these spectral measures are generally greater for breathy than for modal phonation, differences go in both directions for H1-A3, with modal values greater for five languages and breathy values greater for four. H1-A3 also distinguishes plain from breathy vowels in Gujarati, with differences significant in time points 1 and 3 through 9 in the data reported by Khan (2012).

As with the other spectral measures he reported, Dutta (2007) finds that there is a great deal of inter- and intra-speaker variability for H1-A3. Still, mean H1-A3 values in subsequent vowels did differ based on consonant phonation type, with values significantly higher after breathy voiced stops for between 50 and 70% of the subsequent vowel. We may thus expect consonant phonation type to trigger reliable H1-A3 differences in subsequent vowels in Marathi as well.

Turning now to sonorant-internal results, recall that—as mentioned in the previous section—differences in H1-A3 values in Sumi sonorants did not attain statistical significance.

That said, the pattern in Sumi appeared to be in the expected direction, with breathy sonorants showing numerically greater H1-A3 values than modal sonorants (Harris 2009). Furthermore, in the Tsonga nasal data, Traill and Jackson (1987) used discriminant analysis to determine how well each acoustic parameter assessed contributed to correct phonation type identification of the tokens analyzed, and found that the measure that is most similar to H1-A3—which was the amplitude of the first harmonic minus the amplitude of the strongest harmonic component above 2000 Hz—was marginally useful for the tokens produced by male speakers. It did not aid in classification for the tokens produced by female speakers. The parameters that worked best for the male speech were H1-H2 and the measure closest to H1-A2; for females, the parameters that contributed most to accurate identification were the measure closest to H1-A2 and a measure of the harmonics-to-noise ratio.

In short, then—as with H1-A2—it is unclear what to expect with regards to H1-A3. While we may expect consonant phonation type in to trigger differences in vowels following obstruents, whether or not differences will emerge sonorant-internally or in vowels after sonorants remains to be seen.

## **2.7 Cepstral Peak Prominence**

Cepstral peak prominence is a measure of harmonic amplitude that takes overall amplitude into account. What this means is as follows: the overall amplitude of the voice source results not only from the amplitude of the harmonic components but also from the amplitude contributed by whatever background noise there is in the signal. By calculating the CPP, we subtract the background noise contribution and get an accurate representation of the harmonic

structure of the voice signal. In an article frequently cited with regards to CPP, Hillenbrand *et al.* (1994) state that:

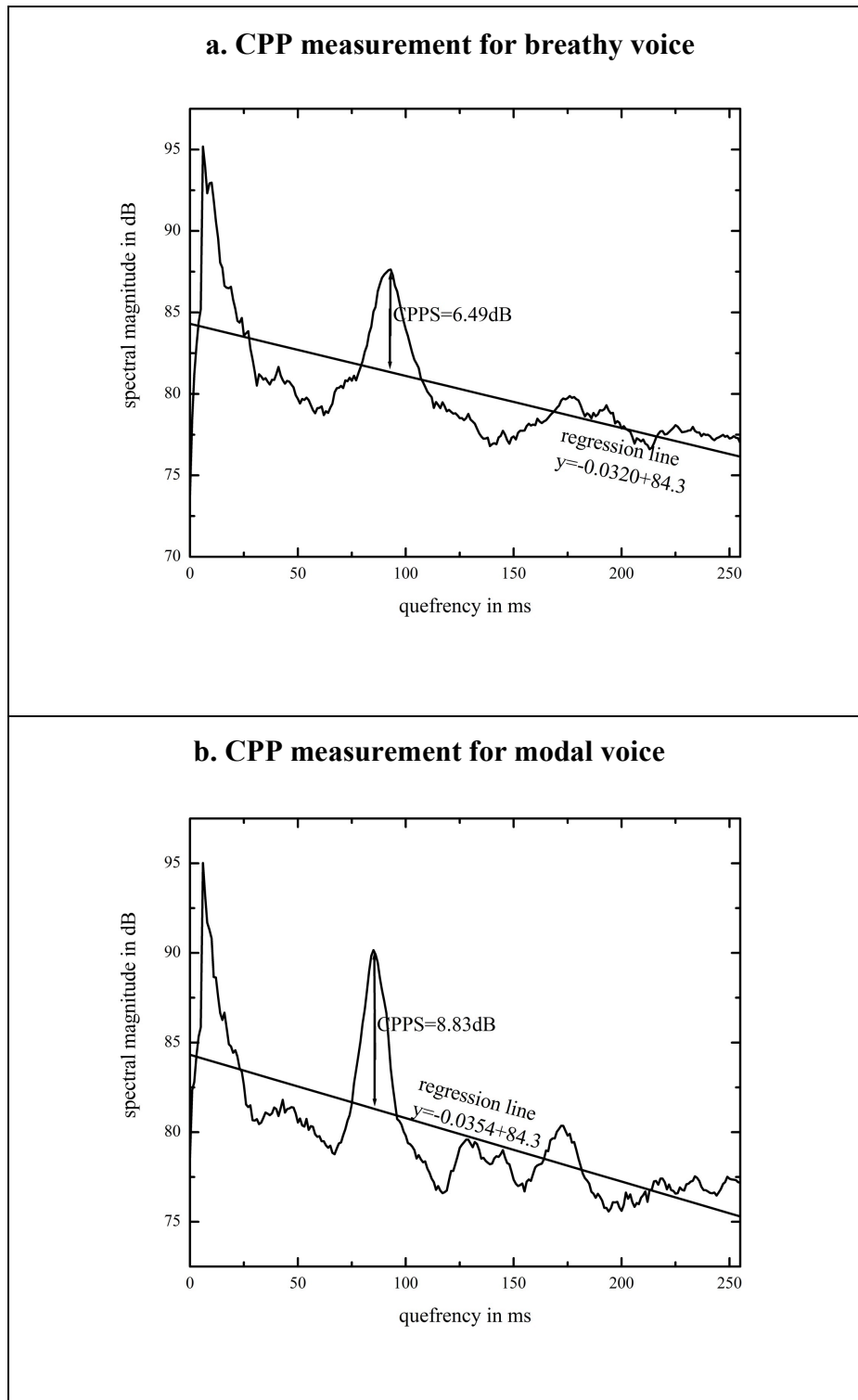
The idea behind the CPP measure is that a highly periodic signal should show a well defined harmonic structure and, consequently, a more prominent cepstral peak than a less periodic signal. What is needed is a measure of the *prominence* of the peak rather than its absolute amplitude. This is because the amplitude of the cepstral peak is affected not only by the degree of periodicity but also by overall energy and the window size of the cepstrum analysis (p. 772).

CPP is measured from the cepstrum. Pennington (2005) describes generation of the cepstrum as follows:

The cepstrum is obtained by calculating the Fourier transform of the time waveform, taking the logarithm of the magnitude of the transform, then performing an inverse Fourier transform on the logarithmic function...as the degree of periodicity of the voice source grows...[the cepstrum] displays increasingly higher peaks (p. 79).

Note that in a cepstrum, the frequency energy from the spectrum is converted to quefrequency (1/frequency), which is measured in milliseconds (Samlan & Story 2011). In the cepstrum, the harmonic structure of the sound is rendered very clearly, and it is possible to separate the harmonic components of the signal from the noise in the background. This is illustrated in Figure 10, which shows that CPP values are calculated by drawing a regression line that separates the harmonic peaks from the background noise. The distance from the regression line to the tip of the cepstral peak is then measured.

Figure 10 Illustration of CPP measures from modal and breathy voiced tokens



Images courtesy of Erik Thomas

In the production of breathy phonation, the vocal folds may never close completely, meaning that there is at least some amount of continuous airflow. The functional result of this vocal fold configuration is increased aspiration noise as compared with that which is observed in modal voicing (Balasubramaniam *et al.* 2011). There is more noise in the signal, in other words, and this increased noise is often visible in the waveform as more jagged periodicity and in the cepstrum as weaker cepstral peaks. Strong peaks in the cepstrum, on the other hand, are an indicator of less noise in the signal, a clear harmonic structure, and a very periodic sound (Keating & Esposito 2007). Thus we expect breathy voiced sounds to have lower CPP values than sounds produced with modal voice, as illustrated in Figure 10a. and b., because breathiness is associated with more noise and reduced harmonic energy.

Note that while the use of CPP to capture linguistically meaningful differences in phonation type is relatively new, it has been used somewhat extensively in research related to the perception of non-linguistic breathiness (Hillenbrand *et al.* 1994) as well as in the realm of vocal pathology, where it is reported to be a reliable measure of pathological breathiness (Heman-Ackah *et al.* 2003, Watts and Awan, 2011). It is also being utilized more and more commonly in computer science, engineering, and text-to-speech technology applications (Patil *et al.* 2006). It is useful in these contexts for a number of reasons, one of which is that CPP is not affected by F0 in the same way that some of the lower-frequency spectral measures may be (Murphy 2006). This means that while some spectral measures such as H1-H2 may be more accurately measured for male than for female voices—due to the fact that mean fundamental frequency tends to be lower for males, thereby ensuring greater distance between H1 and the formants and therefore less interference from the formants—the same is not true for CPP. Furthermore, while some perturbation-based measures related to, for example, jitter and shimmer are sensitive to F0, and

can only be reliably attained when accurate estimation of the fundamental period is possible (Maryn *et al.* 2009), CPP does not rely on accurate F0 estimation. As such, CPP provides a more stable measure of voice quality than a number of other measures reported in the literature (Moers *et al.* 2012).

As noted, modal phonation is generally characterized by clear periodicity and a weak noise component in the signal, and thus with higher CPP values. Blankenship (2002) reports that CPP is especially useful for distinguishing modal from breathy phonation in Mazatec vowels. This is supported by Esposito (2006), who finds that breathy phonation correlates with significantly lower CPP values than modal phonation in all ten of the languages analyzed acoustically therein. In the very recent study on Gujarati, Khan (2012) reports that while CPP values are lower in breathy than in modal vowels, as expected, the difference is marginal overall and is significant in the middle portion of the vowel—at time points 4-6 (out of nine total time points).

In sum, then, the total number of studies using CPP to assess phonemic voice quality differences is small as of yet, but is steadily growing as more and more researchers turn to CPP. Results thus far are promising with regards to the utility of this measure, which seems to provide a reliable measure of voice quality across different vowel qualities and F0 ranges.

## **2.8 Broad review**

The results presented in the above sections are summarized in Table 7. Marathi does involve a four-way obstruent contrast, and this four-way contrast will be investigated, but the contrast which is more unique is the two-way contrast between breathy and modal phonation in Marathi's sonorants. Thus the comparison which is of the greatest interest to us is that between

breathy and modal phonation; accordingly, these are the results which are summarized below. The three parameters discussed above where modal phonation is associated with greater values are presented first. The remaining parameters, for which breathy phonation is associated with greater values, follow.

**Table 7 Summary of Previous Findings**

<b>CLOSURE DURATION (OBSTRUENTS)</b>	Modal > Breathy	Dutta 2007b, Mikuteit & Reetz 2007
<b>FUNDAMENTAL FREQUENCY</b>	Modal > Breathy	Dutta 2007b, Ohala 1979, Pandit 1957, Schiefer 1986
<b>CPP</b>	Modal > Breathy	Blankenship 1997, Esposito 2006, Khan 2012
<b>SEGMENT DURATION (SONORANTS)</b>	Breathy > Modal	Harris 2009
<b>POST-RELEASE INTERVAL</b>	Breathy > Modal	Davis 1994 (NOT), Mikuteit & Reetz 2007 (ACT)
<b>H1-H2</b>	Breathy > Modal	Andruski & Ratliff 2000, Blankenship 1997, Dutta 2007b, Esposito 2006, Huffman 1987, Khan 2012, Wayland & Jongman 2003,
<b>H1-A1</b>	Breathy > Modal	Dutta 2007b, Esposito 2006, Khan 2012, Wayland & Jongman 2003
<b>H1-A2</b>	Breathy > Modal	Blankenship 2002, Dutta 2007b, Esposito 2006, Khan 2012, Traill & Jackson 1087
<b>H1-A3</b>	Breathy > Modal	Dutta 2007b, Esposito 2006, Khan 2012

Broadly, then, modal phonation tends to be characterized by longer closure duration, higher fundamental frequency, and greater CPP values. Breathy phonation, meanwhile, exhibits longer segment durations and post-release intervals, and greater H1-H2, H1-A1, H1-A2, and H1-A3 values.

These are the generalizations that can be drawn from the existing literature, and not all prove equally useful. Some, such as H1-H2, repeatedly prove significant in distinguishing breathy from modal phonation, yet are accompanied by questions as to their reliability. Others, such as H1-A3, prove significant less often. Furthermore, data for breathy voiced vowels



outweighs that available for consonants, and consonant data relates primarily to breathy voiced obstruents. Even with these limitations in mind, however—and despite the fact that very few studies have investigated breathy phonation in sonorants—we can take some broad lessons from the literature with regards to the acoustic characteristics which are likely to arise as a result of breathy phonation, such as increased spectral values and decreased CPP values.

It is with these thoughts in mind that we move into a discussion of the current project. With its extensive use of phonemic breathy voice across multiple manners of articulation, Marathi allows for a tremendous enrichment of our understanding of the acoustic correlates of breathy voiced consonants, and may help to resolve questions about the utility of the above-mentioned measures.

## **2.9 An overview of Marathi**

As noted, Marathi is an Indo-Aryan language spoken by upwards of 70 million people primarily in the state of Maharashtra. Despite being such a widely-spoken language, however, very little work has been conducted with Marathi phonetics or phonology over all, and acoustic analysis of the language is almost nonexistent. Marathi scholarship on Marathi linguistics certainly exists—notably, an unpublished 1958 dissertation on Marathi phonology and morphology by Ashok Kelkar. In addition, two descriptive grammars of Marathi have been published in the last decades, one by Rajeshwari Pandharipande in 1997, and one by Ramesh Dhongde and Kashi Wali in 2009. In both of these volumes, however, the combined phonetics/phonology sections are limited in length and detail, particularly as compared with the sections devoted to syntax and morphology (on the order of 30 pages out of more than 250 in Dhongde and Wali and about 40 pages out of more than 600 in Pandharipande).

This is particularly unfortunate given Marathi's rich phonetic inventory, shown in Figure 11 (Masica 1991, Dhongde and Wali 2009). Commentary about the chart appears below.

Figure 11 Marathi consonant chart

		Place of Articulation													
		Labial		Non-retroflex Apical				Retroflex		Alveo-Palatal		Velar		Glottal	
				Dental		Alveolar									
Manner of Articulation	Stop	<b>p</b> <b>p<sup>h</sup></b>	<b>b</b> <b>b<sup>h</sup></b>	<b>t</b> <b>t<sup>h</sup></b>	<b>d</b> <b>d<sup>h</sup></b>			<b>ʈ</b> <b>ʈ<sup>h</sup></b>	<b>ɖ</b> <b>ɖ<sup>h</sup></b>			<b>k</b> <b>k<sup>h</sup></b>	<b>g</b> <b>g<sup>h</sup></b>		
	Nasal		<b>m</b> <b>m<sup>h</sup></b>		<b>n</b> <b>n<sup>h</sup></b>				<b>ɳ</b> <b>ɳ<sup>h</sup></b>				<b>(ŋ)</b>		
	Affricate					<b>ts</b> <b>(ts<sup>h</sup>)</b>	<b>dz</b> <b>(dz<sup>h</sup>)</b>			<b>tʃ</b> <b>tʃ<sup>h</sup></b>	<b>dʒ</b> <b>dʒ<sup>h</sup></b>				
	Fricative					<b>s</b>	<b>(z<sup>h</sup>)</b>	<b>(ʂ)</b>		<b>ʃ</b>					<b>h</b>
	Approximant		<b>v</b> <b>v<sup>h</sup></b>		<b>l</b> <b>l<sup>h</sup></b>				<b>ɭ</b>		<b>j</b>				
	Rhotic										<b>r</b> <b>r<sup>h</sup></b>				

Although the voiceless aspirated labial stop [p<sup>h</sup>] is listed above, this sound is undergoing a change to [f] both in standard Hindi and in much of the Marathi-speaking world. My anecdotal observation before recording the data presented in this study was that male speakers tend not to produce [p<sup>h</sup>], but produce [f] instead. Female speakers are also trending that way, but three of the four female speakers who have served as my primary language consultants still produce the occasional [p<sup>h</sup>]. Agreement on the status of the labial approximants is lacking. They are categorized as [w] by Dhongde and Wali (2009), as three distinct phonemes ([f, v, w]) by Kavadi & Southworth (1965) and Pandharipande (1997), as two phonemes ([v, w]) by Jha (1977), and as [v] by Masica (1993). They are characterized here as approximants, and are treated as

sonorants in the following analyses.

The retroflex fricative appears in this inventory, but is considered to be a ‘marginal phoneme’ by Dhongde and Wali. The retroflex fricative appears only in Sanskrit words, and is pronounced as an alveo-palatal by almost all speakers in almost all instances. The same is true for [ts<sup>h</sup>]; this sound does occur in Marathi, but very rarely, only word-medially, and only in words of Sanskrit origin (Jha 1977). The breathy voiced alveolar affricate and fricative, [dz<sup>h</sup>] and [z<sup>h</sup>], are reported to be in free variation (Southworth 2000). For the consultants with whom I work, [dz] and [z] vary freely as well.<sup>13</sup> The functional result of this variation is that Marathi speakers do in fact produce both plain and breathy voiced alveolar fricatives. The breathy voiced fricative is certainly a sound of interest, but is outside the scope of the current investigation.

The rhotic consonants included above are marked as being alveo-palatal, but there is debate in the literature as to how to characterize these sounds. In the current study they are sometimes produced with a more trill-like articulation, sometimes with a more flap-like articulation, and sometimes with a more approximant-like articulation. This is true for both the modal and the breathy voiced versions, for all speakers and in all contexts. At present, then, I consider the different articulations to be in free variation.

Two final points are that the velar nasal is also considered a “marginal phoneme” by Dhongde and Wali (2009), as it arises almost exclusively due to place assimilation. Also, the glottal fricative is included above but is not coded as being definitively voiceless or voiced. This is because it is described as voiceless by some (Pandharipande 1997, Dhongde and Wali 2009) and voiced as others (Masica 1991, Southworth 2000). In fact, Jha (1977) mentioned the glottal

---

<sup>13</sup> In fact, variation between the alveolar and alveo-palatal fricatives and affricates is rampant. While the place of articulation distinction is widely considered to be phonemic for these sounds, there is variation along this parameter as well, in part because the alveolars are frequently palatalized before front vowels. One of my participants also de-fronted the alveo-palatals in all vowel contexts. This is beyond the scope of the present investigation, but deserves investigation in future work.

fricative, but pointedly avoids making any commentary about whether or not it is voiced.

The above discussion is intended to provide a brief overview of the Marathi phonetic inventory. The major point to keep in mind here, however, is that because Marathi employs phonemic breathy voice across so many manners and places of articulation, it provides rich territory for making the comparisons which are of critical interest in the present work: those based on both phonation type and obstruency. We turn now to an overview of the present study.

### **3 Overview of the Present Study – Methods and Procedures**

As noted in the previous sections, current knowledge about the acoustic correlates of breathy phonation has been gained primarily through research focused on vowels and obstruents. Very little research has focused on phonation type distinctions in sonorants, and so our understanding of the acoustic parameters associated with breathy voiced sonorants is far from complete. Therefore, the present study serves to augment our understanding of breathy phonation in general and in sonorants in particular by providing instrumental acoustic analysis of breathy voiced obstruents and sonorants in Marathi. Male and female native speakers of Marathi were recorded producing real words that contained plain and breathy voiced consonants in different word positions and vowel contexts.

Breathy voiced sonorants are more rare crosslinguistically than breathy voiced obstruents: the results presented herein reveal that phonation type contrasts are also cued more poorly by sonorants than by obstruents. I argue that the acoustic patterns can help us to understand the crosslinguistic typology. Phonation type contrasts are less auditorily distinct in sonorants than in obstruents, and as such they make for poorer phonemic contrasts.

#### **3.1 Participants**

Participants recorded in this study included ten native Marathi speakers (five female, five male). They ranged in age from 29 to 64 years old; males had a mean age of 45.4 years, and females a mean age of 37.2 years. All participants are multilingual, speaking Hindi and English in addition to Marathi, and were living in and around Lawrence, Kansas or Kansas City, Missouri at the time of the recording. Additional demographic data can be found in Appendix A.

### 3.2 Materials

The labial, dental, and alveo-palatal consonant series were chosen as representative, because together they include the full set of phonation type contrasts of interest and allow for comparison across different manners and places of articulation.<sup>14</sup> The relevant sounds are highlighted in Figure 12.

Figure 12 Marathi consonants recorded in the present study

		Place of Articulation												
		Labial		Non-retroflex Apical				Retroflex		Alveo-Palatal		Velar		Glottal
				Dental		Alveolar						k	g	
Manner of Articulation	Stop	p	b	t	d			t̪	ɖ			k	g	
		p <sup>h</sup>	b <sup>h</sup>	t <sup>h</sup>	d <sup>h</sup>			t̪ <sup>h</sup>	ɖ <sup>h</sup>			k <sup>h</sup>	g <sup>h</sup>	
	Nasal		m		n				ɳ				(ŋ)	
			m <sup>h</sup>		n <sup>h</sup>				ɳ <sup>h</sup>					
	Affricate						ts	dz			tʃ	dʒ		
							(ts <sup>h</sup> )	(dz <sup>h</sup> )			tʃ <sup>h</sup>	dʒ <sup>h</sup>		
Fricative						s	(z <sup>h</sup> )	(ʃ)		ʃ			h	
Approximant		v		l					ɭ		j			
		v <sup>h</sup>		l <sup>h</sup>										
Rhotic											r			
											r <sup>h</sup>			

A description of the 88 target words recorded in the present study is provided here, and a full list can be found in Appendix B. The list includes the labial, dental, and alveo-palatal consonants of Marathi in two word positions (word-initial and word-medial) and in two vowel contexts (before [a] and before [e]). Different vowel contexts were utilized because they may trigger articulatory, acoustic, and perceptual differences—as noted previously, a vowel like [i]

<sup>14</sup> Only consonant pairs that highlight the modal/breathy contrast are recorded; although [j] and [ʃ] are alveo-palatal they not recorded, as they do not have breathy counterparts.

with a low F1 may influence the amplitude of the lower harmonics, thereby interfering with measures like H1-H2 (Blankenship 1997). The appeal of [a] is that it boasts a high F1, and is therefore unlikely to boost the lower harmonics. This study is designed to enhance our understanding of the acoustic correlates of breathy phonation overall, however—not just in a particular vowel context—and so it is important to take multiple vowel contexts into account. As the aim of the present study is to develop a comprehensive understanding of the acoustic correlates of phonation type contrasts in Marathi sonorants and obstruents, we also deemed it important to investigate the sounds in both word-initial and word-medial contexts.

Very little work has investigated stress patterns in Marathi (Dhongde & Wali 2009); nonetheless, the tokens selected for this experiment were chosen such that the vowel of interest generally carried the stress according to the tentative pattern regarding stress set forth by Pandharipande (1997) and referenced in Dhongde & Wali (2009). In disyllabic words like those included in this study that contain the vowel [a], [a] carries the stress. If the word contains two [a]s, the first [a] carries the stress. If [e] appears in the final syllable, and the vowel in the first syllable is anything other than [a], then the stress is on the second syllable. Thus for all of the tokens in the wordlist that contain [e] in the second syllable, that [e] carried the stress provided the first vowel was not [a]. Finally, if [ʌ] is one of the two vowels in a disyllabic word and the other vowel is [i, e, o, u], then the other vowel—which, for the purposes of this experiment was always [e]—always carries stress. In short, then, the generalizations outlined above can be summarized as follows:

- an initial [a] always carries stress
- a final [e] carries stress if it is not preceded by [a], and
- an initial [e] carries stress if it is followed by [ʌ].<sup>15</sup>

---

<sup>15</sup> Thus the tokens in the list that diverged from this pattern were mainly those that contained [e] in the initial syllable; as often as possible, the word chosen contained a [ʌ] in the second syllable, but this was not always possible given the constraints of the Marathi lexicon.

Appearance of some of the sounds under investigation is limited in certain vowel contexts, but a near-complete wordlist was compiled for both [a] and [e]. Five of the 88 target words included in this study were nonce words, created in collaboration with a native language consultant to fill the gaps. Note that these gaps are accidental—the nonce words created do not violate any phonotactic constraints in Marathi, and did not give the participants pause.

A total of 2640 tokens were recorded (10 speakers X 22 consonants X 2 word positions X 2 vowel contexts X 3 repetitions). Some of those tokens were discarded, however, for the following reasons. Notably, [p<sup>h</sup>] is almost never produced by male speakers, who produce [f] instead. As mentioned, this is consistent with an ongoing sound-change in Marathi. Female speakers produced [p<sup>h</sup>] irregularly, but consistently. Each female speaker produced [p<sup>h</sup>] at least once, and produced [f] for the remainder of the repetitions. Since [f] is not of interest to us in this study, [f] data were discarded. Free alternation occurred between [dʒ], [dz], and [z], and between [dʒ<sup>h</sup>], [dz<sup>h</sup>], and [z<sup>h</sup>]. Fricative productions were excluded from the analysis, and affricate productions were grouped.

As might be expected given the debate about them in the literature, the rhotic sounds [r] and [r<sup>h</sup>] also showed a high degree of variability. They were sometimes produced with a more trill-like articulation, sometimes with a more flap-like articulation, and sometimes with a more approximant-like articulation. This was true for all speakers and in all contexts, but there was a reliable acoustic cue to be found within the variable productions. As might be expected, all were accompanied by a very low F3 which rose steeply as it transitioned into the vowel. If the boundary between consonant and vowel was ever unclear in the case of the rhotics, the steep rise in F3 was used to demarcate the two during data segmentation.



### 3.3 Recording and token presentation

Speakers were recorded onto a flash disc using an Electro-Voice N/D 767 cardioid microphone and a Marantz Portable Solid State Recorder (PMD 671) in the anechoic chamber at the University of Kansas, Lawrence. Half of the participants were recorded at a sampling rate of 22.05 kHz. Midway through data collection the sampling rate was changed to 44.1 kHz, but the difference in sampling rates caused no problem for data analysis because of the analysis program utilized, which will be described momentarily.

Tokens were recorded in the carrier sentence [sita word bʌgte], which means “Sita saw word”. The carrier sentence, which was presented in Devanagari script, was designed so that the initial consonant of the target word would be intervocalic (following the vowel in “Sita”). Originally, the intention was to measure consonant closure duration on consonants in both initial and medial position. Speakers consistently paused before producing the target word, however, and so initial consonant duration was measured only for the voiced sounds. Duration for the voiceless sounds was measured only word-medially.

To render the carrier sentence more meaningful, and to introduce the nonce words, the speaker was provided with a context scenario. They were asked to imagine that a girl named Sita had read a story. While reading, she of course saw a great number of words written on the page, and that is what the sentences represent—an accounting of the individual words she saw on the page while reading. Speakers were also told that the story included some terms they might not be familiar with, and were given a list of those words along with their definitions. They were able to read through the list as many times as they wanted to, in order to become comfortable with the nonce words. Speakers typically read the list only once, although several speakers read the list twice. Carrier sentences were presented using PowerPoint, with one sentence per slide. Speakers

controlled the rate of presentation themselves, by clicking a button to move from one slide to the next at their own pace.

### 3.4 Data analysis

The digitized recordings were first imported into PRAAT, freeware which allows simultaneous examination of the waveform and (wide-band) spectrogram of each token (Boersma and Weenink 2010). Segmentation and annotation was conducted in PRAAT, using text grids. The audio files and text grids were then run through VoiceSauce (Shue *et al.* 2011), a free application implemented via Matlab® (The Mathworks, 2011). VoiceSauce is used for acoustic analysis, and can be used to take a wide range of acoustic measurements. For this study, it was used to generate F0, CPP, and corrected H1-H2, H1-A2, and H1-A3 measures, which are explained momentarily. Durational measures were also extracted from the VoiceSauce outputs using Matlab.

VoiceSauce identifies paired wav and text grid files in a specified folder, and measures every millisecond of every labeled portion of the text grid. Using the Straight algorithm developed by Kawahara *et al.* (1998), F0 is measured in 1 ms intervals. This measurement is critical, as VoiceSauce uses it as the default in locating and measuring the amplitude of the harmonics, on which the subsequent measures depend.

The formant and harmonic amplitudes—H1, H2, A1, A2, and A3—are computed in VoiceSauce across a four-pitch-period window. A pitch-synchronous window is used to avoid the variability that may result from a fixed window. These amplitude measures may be affected by vowel formants and thereby rendered inaccurate, and so in addition to taking the straight, uncorrected measurements, VoiceSauce also uses an algorithm developed by Iseli & Alwan (2004) to correct for the effect of formants. The resulting corrected measures are those denoted

by a “c”—as in H1-H2c or H1-A1c. Note that while the algorithm used to perform corrections minimizes the effect of formants, the correction is not perfect. Thus, while measures on high vowels—with low F1s, which would cause trouble in uncorrected measures—are rendered more accurate, there is still room for error in high vowel contexts.

Cepstral peak prominence is calculated using the algorithm from Hillenbrand *et al.* (1994). Again, a variable window is used, this one equal to five pitch periods. The VoiceSauce manual explains the process as follows:

A variable window length equal to 5 pitch periods was used for the calculations. After multiplying the data with a Hamming window, the data is then transformed into the real cepstral domain. The CPP is found by performing a maximum search around the quefreny of the pitch period. This peak is normalized to the linear regression line which is calculated between 1 ms and the maximum quefreny. (*taken from the VoiceSauce manual, available online at <http://www.ee.ucla.edu/~spapl/voicesauce/documentation/parameters.html#cpp>*)

The output generated by VoiceSauce, then, includes measurements for every millisecond of every interval annotated in the text grid.

The output generated by VoiceSauce was then pulled into R, a freeware package used for statistical computing (R Core Team, 2012). Within R, data for each segment was divided into five equal portions representing 0-20% of the segment, 21-40% of the segment, 41-60% of the segment, 61-80% of the segment, and 81-100% of the segment. The data within each of these intervals was averaged, yielding five measurement time points for each segment. These are referred to throughout as Windows 1-5; Window 1 represents the average of the first 20% of the segment, Window 2 the average of the second 20%, Window 3 the average of the third 20% (i.e., 41-60%), Window 4 the average of the fourth 20%, and Window 5 the average of the final 20% of the segment. This allows presentation and analysis of time-normalized data.

### 3.5 Intervals measured

The duration of the consonantal constriction, the PVI, and the subsequent vowel were measured. With regards to consonant closure/constriction duration, sonorants and voiced obstruents were measured in both word-initial and word-medial position. Voiceless obstruents were measured only in word-medial position. PVI and vowel duration were always measured in both word positions.

F0 and CPP were taken for the consonantal portion of sonorants and voiced obstruents, and the remaining spectral measures—H1-H2c, H1-A1c, H1-A2c, H1-A3c—were taken during the consonantal constriction of sonorants. All measures were taken on the vowel following each consonant of interest. While PVI and vowel duration are considered separate intervals in the durational results reported, they are collapsed in the F0 and spectral results (following Dutta 2007) and are referred to simply as the vowel. In other words, while the PVI is important in that it provides a measurement of how long the breathy interval is, it is important to remember that this interval is still a part of the vowel—it is breath mixed with vowel, and as such, is considered part of the vowel for the non-temporal acoustic analyses.

### 3.6 Predictions

As has been noted in the preceding sections, little prior research on breathy sonorants exists; given the previous findings related to breathy phonation in obstruents and vowels, however, the following predictions can be made.

- I. **Consonant Duration:** plain obstruents are expected to have longer closure durations than aspirated/breathy obstruents. In common with Bengali (Mikuteit and Reetz 2007) and Hindi (Benguerel and Bhatia, 1980), alveo-palatal affricates are expected to exhibit the shortest closure duration, while the labial and dental obstruents are expected to pattern

together. Breathy sonorants are expected to have longer consonantal constrictions than plain sonorants.

- II. **Pre-Vocalic Interval:** Breathy voiced sounds are expected to exhibit the longest PVI, with aspirated voiceless sounds, plain voiceless sounds, and voiced sounds following in descending order.
- III. **Vowel Duration:** Given the findings related to vowels after breathy voiced stops in Hindi (Dutta 2007b), vowels after aspirated and breathy voiced sounds may be expected to be longer than those following plain sounds. As was mentioned in the discussion of the PVI in Chapter 2, however, that part of the vowel which is heavily mixed with breath is included in the PVI in this study, and so overall vowel duration is not expected to show an effect of consonant phonation type.
- IV. **F0:** Breathy voiced consonants are expected to be associated with lowered F0 in subsequent vowels. F0 differences are further expected to be apparent in consonants themselves, with breathy voiced obstruents and sonorants showing a lower F0 than their plain counterparts.
- V. **H1-H2c, H1-A1c, H1-A2c, H1-A3c:** Breathy phonation is consistently associated with a sharply decreasing spectral tilt, resulting from a dramatic drop-off of energy in the higher frequency range. Due to this sharp drop-off, the spectral measures which reflect this—H1-H2c, H1-A1c, H1-A2c, H1-A3c—are expected to be higher for breathy than for modal sounds, although there is one caveat that should be mentioned.
  - a. There is some evidence that female voices are more prone to breathiness overall than male voices, and that the lack of complete vocal fold closure which characterizes breathy phonation may be present in female speech much of the

time, regardless of phonation type (Södersten and Lindestad 1992). While values associated with breathy voiced consonants are still predicted to be lower than those associated with modal consonants, if there is less of a difference in the breathiness of breathy vs. modal sounds in female speech than in male speech, then it is possible that the lower spectral measures—in particular H1-H2c—may be less informative with regards to distinguishing phonation type contrasts for female speakers than for male speakers.

- VI. **CPP:** Breathily phonation yields more noise in the signal, more jagged periodicity, and less prominent harmonic peaks, while modal sounds show a more regular harmonic structure and more prominent cepstral peaks. As such, CPP values are expected to be lower when associated with breathy than with modal phonation.

### 3.7 Statistical analyses

Results for duration, fundamental frequency, the spectral measures H1-H2c, H1-A1c, H1-A2c, H1-A3c, and CPP are reported in the following sections. Several explanatory comments regarding statistical methodology should be made at the outset, however.

Statistical analyses were conducted using R, a free, open-source platform used for statistical computing (R Core Team, 2012). Analyses comprised two parts. First, an omnibus repeated measures analysis of variance (Anova) was conducted for each of the seven measures presented in the following sections—duration, F0, H1-H2c, H1-A1c, H1-A2c, H1-A3c, and CPP. In the omnibus Anova, the measure under investigation was the dependent variable and the independent variables included Phonation type (two levels: plain and breathy), Obstruency (two levels: obstruent and sonorant), Voice (two levels: voiceless and voiced), Place of articulation

(three levels: labial, dental, alveo-palatal), Word position (two levels: initial and medial), Vowel context (two levels: [a] context and [e] context), Gender<sup>16</sup> (two levels: male and female), and Window (five levels: Windows 1, 2, 3, 4, and 5). Recall that Window 1 represents a mean value averaged over the first 20% of the segment, Window 2 the mean value averaged over the second 20% of the segment, and so forth. The omnibus equation accounts for the fact that voice and obstruency are not crossed; obstruents contrast for voice, but sonorants do not.

The significance level for all omnibus Anovas, which were used to gather information about significant interactions, was set at 0.05. A series of post hoc comparisons was conducted following the omnibus Anova, and so the conservative Bonferroni correction was utilized to limit the risk of Type I errors. As this study includes investigation of seven measures, the Bonferroni correction dictates that the significance level for follow-up analyses must be  $0.05/7$ , or 0.007 (Baayen 2008). The level for marginal significance in post hoc tests was set at 0.014 (twice the significance level).

We move now into the first results section, which deals with durational measures.

---

<sup>16</sup> Again, it is important to point out that it would be more accurate to use the term Sex instead of Gender, for what I refer to here are the biological differences arising from sexual dimorphism. I emphatically do not mean to point to any sort of sociolinguistic, societally-constructed gender distinction. That said, the term Gender is used much more commonly than the term Sex in linguistic research to refer to physiological sex-based differences, and so I use the term Gender in order to conform with the general disciplinary trend.

#### 4 Duration Results

The temporal measures reported herein are threefold, and include: consonant duration for sonorants, and closure duration for voiceless medial and voiced initial and medial obstruents; the post-release breathy interval, as captured by the PVI; and subsequent vowel duration.

Recall that previous findings support a number of hypotheses about the way phonation type may affect temporal measures. Breathily voiced obstruents are associated with a shorter closure duration than plain voiced obstruents in Bengali (Mikuteit & Reetz 2007) and in Hindi (Benguerel & Bhatia 1980; Dutta 2007). Pre-vocalic interval measures, which are reported here, cannot be directly compared to the measures used by previous researchers, but general predictions can still be made: whether utilizing the NOT (Davis 1994), the ACT (Mikuteit & Reetz, 2007), or the PVI, as reported herein, the prediction is that breathy and aspirated sounds should be associated with larger positive values than modal sounds, regardless of voicing. In other words, aspirated voiceless sounds are expected to have a longer PVI than plain voiceless sounds, and breathily voiced sounds are expected to have a longer PVI than plain voiced sounds.

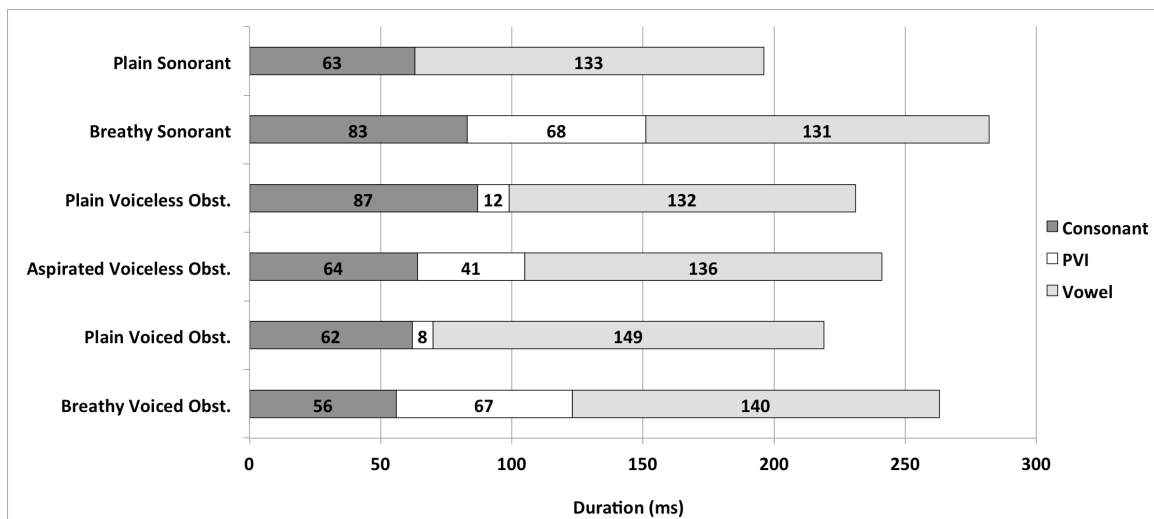
Recall that the PVI includes an interval of breathiness overlaid on some portion of the vowel. In other words, PVI values tend to be greater on the whole than ACT or NOT values, because the PVI includes that portion of the vowel which is heavily mixed with breath. This is deemed valuable in the current study, because the PVI is designed to garner an accurate measurement of the duration of the breathy interval. It is not the claim of this researcher, however, that the PVI is breath/aspiration alone; rather, it is a mixture of breath and vowel. It is also important to note that the functional result of this measurement technique is that some part of the signal which was apportioned to the vowel in previous studies is here apportioned to the PVI. This means that although Dutta 2007b, for instance, reports that breathily voiced obstruents



in Hindi are associated with longer subsequent vowels, the same may not be true here because part of what Dutta measured as vowel duration is here included in the PVI instead.

Each of the temporal measures—consonant, PVI, and vowel duration—is presented in more detail momentarily, but Figure 13 provides an overview of full syllables. This figure is valuable in that it allows us to envision the way the three portions come together to form the big picture. Data are for word-medial syllables collapsed across gender, place of articulation, and vowel context.

**Figure 13 Full syllable duration (consonant, PVI, vowel) by phonation type, obstruency, and voicing**



In looking at Figure 13, we can make several general observations. Of note is the fact that phonation type has an overall effect on syllable duration, although this is perhaps true only numerically for the voiceless obstruents. As we will see, this is confirmed statistically for the sonorants and voiced obstruents. For the voiced consonants, syllables that contain breathy consonants are longer overall than those that contain plain consonants. This durational difference is driven in large part by the PVI, and—in the case of sonorants—by the sonorant constriction duration as well. Let us turn now to a more complete discussion of consonant duration.

#### 4.1 Consonant Duration

Recall that the wordlist was designed to elicit consonants in both word-initial and word-medial positions, and tokens were embedded in a carrier sentence. Speakers consistently paused before saying the target word, however, and so closure duration for voiceless consonants could only be measured in word-medial position. Thus voiced consonants—meaning sonorants and voiced obstruents—can be compared across word positions, but analyses that include the plain and aspirated voiceless obstruents are conducted on word-medial data only.

The consonant duration omnibus Anova revealed no gender effects, and so male and female data are collapsed. The factors we are most interested in are phonation type and obstruency, and these are discussed below. In addition, however, the omnibus revealed some overall patterns which are not surprising but are of interest because—to my knowledge—this is the first such analysis of Marathi acoustic data. To that end, and in the interest of thoroughness, they are included here.

Voice significantly affects the duration of obstruents in medial position ( $F(1,7) = 66.66$ ,  $p < 0.0001$ ): voiced obstruents ( $M = 59$  ms) have a shorter closure duration than voiceless obstruents ( $M = 76$  ms). Obstruency also affects the duration of voiced consonants ( $F(1,7) = 8.332$ ,  $p = 0.018$ ): voiced obstruents ( $M = 59$  ms) are shorter than sonorants ( $M = 73$  ms). Vowel context has a significant effect on consonant duration, as well ( $F(1,7) = 22.38$ ,  $p = 0.001$ ), with consonants before [a] ( $M = 80$  ms) showing a slightly but significantly shorter mean duration than consonants before [e] ( $M = 86$  ms).

The results also indicate that the duration of voiced consonants varies significantly according to word position ( $F(1,7) = 46.54$ ,  $p < 0.0001$ ). Initial voiced consonants ( $M = 94$  ms)

are longer on average than medial voiced consonants ( $M = 72$  ms). Furthermore, there is a Phonation type by Word position by Obstruency interaction ( $F(1,7) = 9.12, p = 0.015$ ) which can be explained as follows. Voiced obstruent durations do not differ significantly by phonation type in either word-initial or word-medial position, but they do show a shorter mean closure duration (CD) in medial ( $M = 59$  ms) than in initial position ( $M = 96$  ms).

That consonant phonation type does not trigger significant differences in obstruent closure durations is somewhat surprising, given the fact that previous research has reported shorter CDs for breathy than for plain voiced obstruents. Benguerel and Bhatia (1980) report mean CDs for plain voiced obstruents (130 ms) and breathy voiced obstruents (112 ms) in Hindi; Mikuteit & Reetz (2007) report the values for plain voiced obstruents (106 ms) and breathy voiced obstruents (96 ms) in Bengali. In both instances, the difference was significant—the current data reveals that the pattern does not hold in Marathi, however. Although the differences between average plain and breathy voiced values are somewhat similar—a difference of 6 ms in the current Marathi data, 10 ms in Bengali, and 18 ms in Hindi—CD differences for voiced obstruents in Marathi do not differ statistically by phonation type.

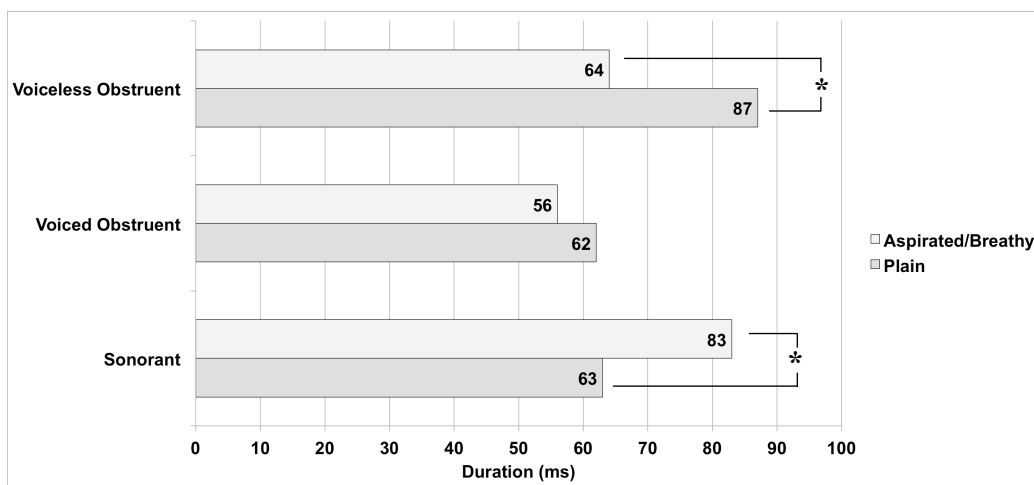
One additional point that should be made is that in both the Hindi and Bengali data, mean CDs were longer than those reported here. This difference is presumably explained in part by the fact that Bengali, Hindi, and Marathi all contain both singleton and geminate obstruents; the Bengali data reported above, for instance, included closure duration measurements for both singletons (mean CD = 66 ms) and geminates (mean CD = 161 ms). In contrast, the Marathi data analyzed in the present work include singletons only.

Plain sonorants pattern with voiced obstruents in that they are shorter in medial position ( $M = 63$  ms) than in initial position ( $M = 90$  ms). Breathly sonorants break from this pattern,

however. The difference between their mean duration in medial position ( $M = 83$  ms) does not differ significantly from their mean duration in initial position ( $M = 93$  ms). As a result, plain and breathy sonorant durations do not differ significantly in initial position, but in medial position breathy sonorants are significantly longer than plain sonorants ( $F(1,9) = 21.89, p = 0.001$ ). Recall that the mean difference between plain and breathy voiced sonorants in medial position is 20 ms, and that segment duration is harder for listeners to perceive for sonorants than for obstruents (Kawahara 2005). This means that although phonation type does trigger statistically significant durational differences in word-medial sonorants, the extent to which those durational differences serve as a useful cue for listeners remains an open question and should be investigated in future work.

Interestingly, no overall effect of phonation type emerges. As illustrated in Figure 14, this is potentially due to the fact that the sonorants and voiced and voiceless obstruents pattern differently.

**Figure 14** Consonant duration by obstruency, voice, and phonation type



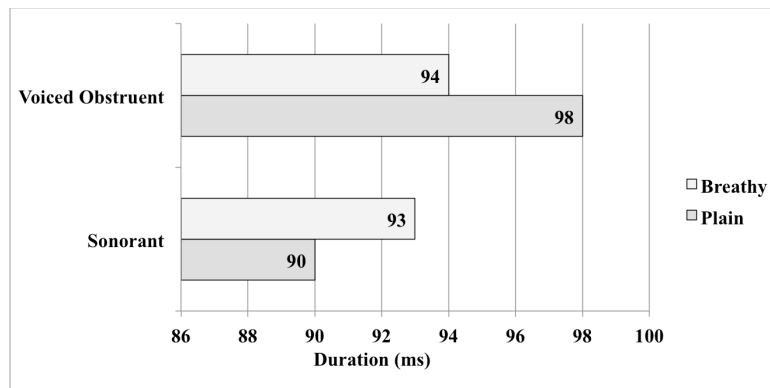
Recall that voiceless obstruent data is available only for word-medial position, and as such the data above reflects mean duration in word-medial position. As mentioned, voiced obstruent

duration is unaffected by phonation type (in either position), whereas medial sonorants are longer when breathy than when plain. The reverse is true for voiceless obstruents, however: plain voiceless obstruents are longer than their breathy counterpart ( $F(1,9) = 47.71, p < 0.0001$ ).

Let us consider this difference in the way that phonation type affects consonant duration in voiceless obstruents and sonorants. In Figure 13, it is clear that aspirated voiceless obstruents have a shorter mean closure duration (64 ms) and a relatively long period of aspiration (41 ms) as compared with their plain counterpart (CD = 87 ms, PVI = 12 ms). Vowel duration does not differ based on consonant phonation type, and so the closure duration differences do not ultimately yield a large difference in mean syllable duration. The story is very different for sonorants, however; breathy sonorants have a significantly longer mean constriction duration (83 ms vs. 63 ms for plain sonorants), in addition to a breathy interval that averages 68 ms. Although vowel duration is once again unaffected by phonation type, the result of the increased constriction duration and the PVI is that syllables containing breathy voiced sonorants are quite a bit longer on average (287 ms), than those containing plain sonorants (199 ms). Duration appears to cue breathy phonation in sonorants in a way that it does not in obstruents. The reasons for this remain to be investigated; as we move into the following sections and report spectral results, however, we may ask whether these temporal cues come into play to compensate for reduced spectral cues.

As has been stated, neither sonorants nor voiced obstruents show durational differences in word-initial position. For the sake of interest, this is illustrated in Figure 15.

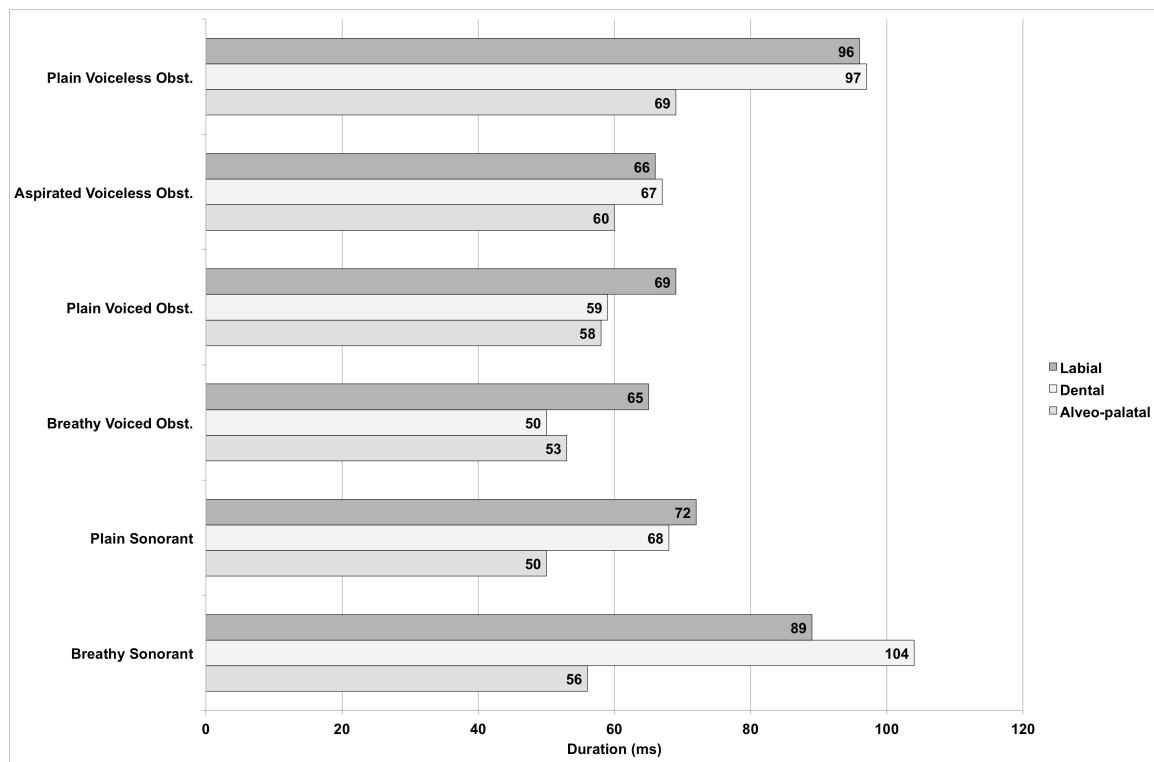
**Figure 15 Duration of sonorants and voiced obstruents in word-initial position**



Finally, Place of articulation has a marginal effect on consonant duration ( $F(1,7) = 4.53$ ,  $p = 0.062$ ), and there is a significant Phonation type by Obstruency by Place of articulation interaction ( $F(1,7) = 18.33$ ,  $p = 0.002$ ). Mean duration of medial consonants broken down by obstruency, voice, place of articulation, and phonation type appears in Figure 16. In looking at this figure, note that the pattern described previously obtains in large part to the data when broken down by place of articulation. The aspirated voiceless obstruents are shorter than their plain counterparts. Voiced obstruents show some overall differences based on place of articulation—labials show the greatest mean consonant duration, while dentals and alveo-palatals are shorter. Alveo-palatals as a group tend towards the shortest mean durations, and in this regard they align with the pattern reported in Bengali (Mikuteit and Reetz 2007) and Hindi (Benguerel and Bhatia, 1980). In these studies, the affricates of Bengali and Hindi had shorter mean durations than the labials and dentals. As such, the short mean CDs reported for Marathi are unsurprising.

What is surprising, however, is that labials and dentals patterned together in Bengali and Hindi, and were longer than the affricates. This is the case for the voiceless obstruents shown below, but not for the voiced obstruents. Rather, the labials are longer, and the dentals and alveo-palatals are shorter and do not differ from one another statistically.

**Figure 16** Medial consonant durations by obstruency, voice, phonation type, and place of articulation



Furthermore, while the labial and dental sonorants are longer when breathy than when plain, the alveo-palatal sonorants are not. These data reflect only two sounds—[r] and [r<sup>h</sup>]— both of which are highly variable. They are sometimes produced as a tap, sometimes as a trill, and sometimes as an approximant, and this is true for all speakers in all contexts. The story told by these data, however, is that when we average across all the variability, the rhotic is not longer when breathy than when plain.

It is important to note that consonant duration is not necessarily straightforward. Here, it is measured as the closure/constriction duration only. Closure release is apportioned to the PVI, which also contains the breathy interval signified by low-amplitude periodicity. Yet the release portion could also be considered part of the consonant duration, and the interval containing

intermingled breath and vowel could be considered part of the vowel. I mention this only to point out that while these temporal measures carry important information, the boundaries imposed by a researcher when dividing them into consonant duration, PVI (or ACT, NOT, or VOT), and vowel duration are discrete boundaries laid on top of a continuous signal. The signal could be divided differently, which is why it is important to establish a clear and replicable methodology.

These points aside, while the duration of consonantal constriction is only one piece of the temporal puzzle, we can already make some broad generalizations. This measure does not cue breathiness in voiced obstruents, but it does distinguish plain from aspirated voiceless consonants in medial position, with plain consonants being longer on average than their aspirated counterparts. In addition, it distinguishes plain from breathy sonorants in certain contexts. Breathy labial and dental sonorants are longer on average than their plain counterparts, but only in medial position. If there are durational cues that distinguish plain from breathy voiced obstruents—or plain from breathy alveo-palatal sonorants—they are not found segment-internally. Rather, they may appear in the PVI or in the subsequent vowel. We turn now to investigation of the pre-vocalic interval.

## **4.2 Pre-Vocalic Interval**

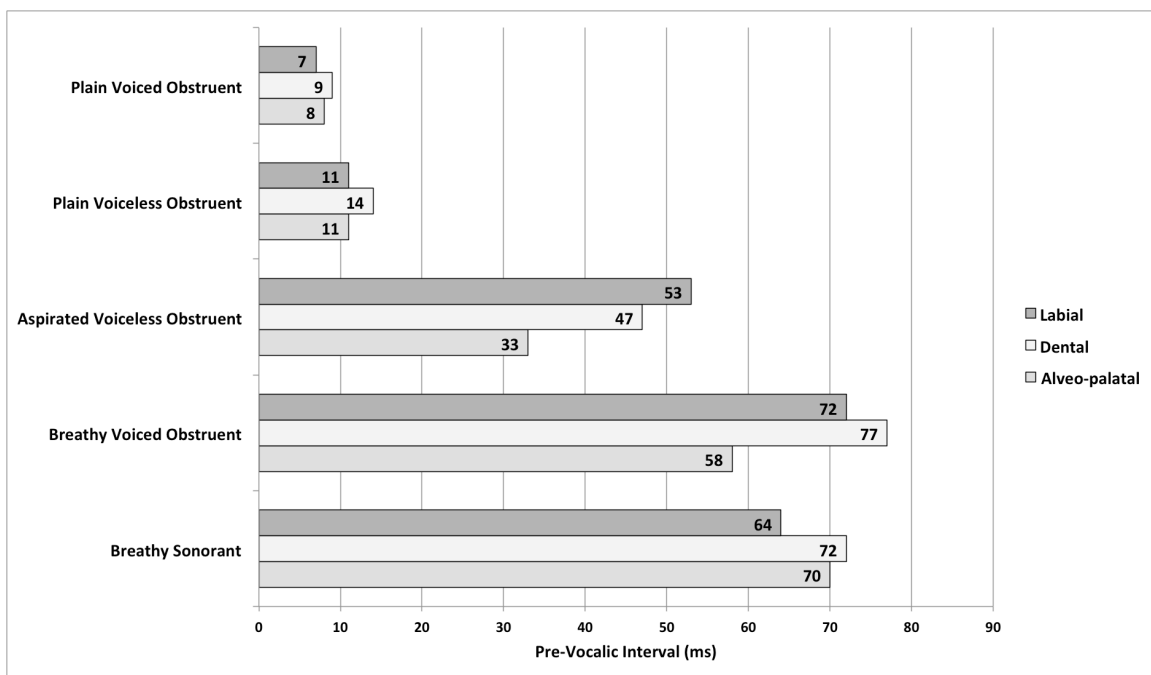
The pre-vocalic interval was developed for use with obstruents, but because breathy sonorants show a clearly defined breathy interval like that of the breathy voiced obstruents it can be measured for both sonorants and obstruents. Breathiness in the breathy sonorants and voiced obstruents is visible in the waveform as jagged periodicity and dampened amplitude; in the spectrogram, it is rendered visible because it has the effect of causing the formants to be very faint. The only way in which sonorants diverge from obstruents in this regard is that the plain



sonorants do not contain any measurable PVI. Whereas plain obstruents often show a short but measurable PVI—Mikuteit and Retz’s After Closure Time (2007)—plain sonorants transition directly from the consonantal constriction into the vowel. Thus PVI durations are reported for all of the obstruents and for the breathy sonorants, but plain sonorants have a PVI equal to zero.

Once again, it is useful to gain an overview of PVI at the outset, and to this end mean PVI values by Obstruency, Voice, Phonation type, and Place of articulation appear in Figure 17. There is no effect of Gender or Word position on PVI durations, and so the data are collapsed across gender and word position.

**Figure 17 PVI durations by obstruency, voice, breath, and place of articulation**



Unsurprisingly, there are main effects of Phonation type ( $F(1,7) = 233.6, p < 0.0001$ ) and Voice ( $F(1,7) = 77.75, p < 0.0001$ ), and these can be elaborated as follows. As predicted, breathy sounds have longer mean PVIs than plain sounds regardless of voice or obstruency. Plain voiceless sounds have shorter PVIs than aspirated voiceless sounds, and plain voiced obstruents have shorter PVIs than breathy voiced obstruents or sonorants. Furthermore, the mean PVI of

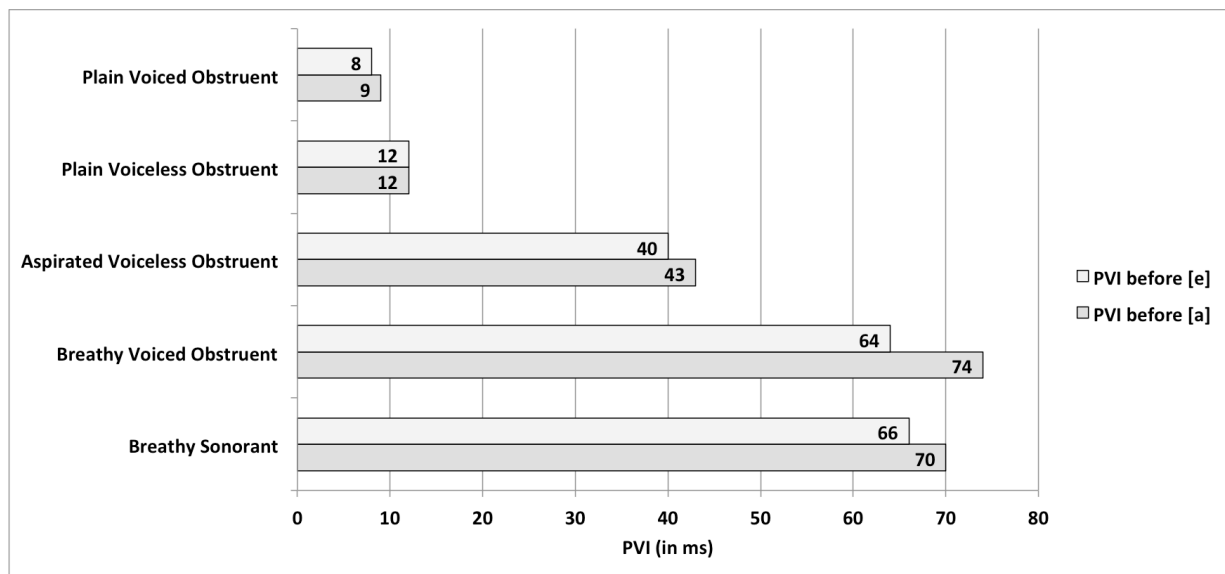
breathy voiced sounds is longer than that of aspirated voiceless sounds. The reverse is true for plain sounds, however. A post-hoc analysis reveals that PVI duration in the plain consonants is significantly affected by voicing ( $F(1,9) = 21.15, p = 0.001$ ). This was not predicted, as it is an effect that did not emerge in a pilot study (Berkson 2012). Nonetheless, the mean PVI of plain voiced obstruents ( $M = 8$  ms) is shorter than that of plain voiceless obstruents ( $M = 12$  ms). At a difference of only 4 ms, however, it is difficult to claim that this makes for a strong cue, regardless of statistical significance. PVI does not differ significantly across breathy voiced obstruents and breathy sonorants.

Despite the fact that there is an overall effect of Place of articulation ( $F(1,7) = 12.36, p = 0.007$ ), place does not trigger significant differences in mean PVI values in the plain obstruents (both voiced and voiceless), nor in the breathy sonorants. The breathy voiced and aspirated voiceless alveo-palatal obstruents, however, have shorter mean PVIs than their labial and dental counterparts. This is rather surprising. Note that for these voiceless sounds, PVI does not differ from VOT, and for VOT we would expect the alveo-palatals to have the greatest VOT, followed by dentals and then labials. Even with the alternative measure of ACT, this is precisely the pattern reported for Bengali (Mikuteit & Reetz 2007): the affricates (palato-alveolar in the case of Bengali) showed significantly greater ACT values than the labials, dentals, retroflexes, or velars. Recall that the aspirated and breathy voiced alveo-palatal obstruents have shorter mean closure durations than their labial or dental counterparts; now we can see that they have shorter mean PVIs, as well. This is not terribly surprising, given that the alveo-palatals are affricates while the labials and dentals are oral stops, but it is interesting nonetheless: the consistent story

here is that the alveo-palatal obstruents diverge somewhat from the pattern of their labial and dental counterparts.<sup>17</sup>

A vowel context effect also emerges, as shown in Figure 18. While the numerical difference between the PVI in [a] and [e] vowel contexts for each consonant type (i.e. “plain voiced obstruent”) is minimal, the pattern is consistent—longer PVI durations in [a] than in [e] vowel contexts—and the test is significant ( $F(1,9) = 53.33, p < 0.0001$ ).

**Figure 18 PVI durations by obstruency, voice, breath, and vowel context**



Despite these differences, it is important to point out that the PVI remains a valuable measurement of lag-time in both vowel contexts—the significance does not disappear in the [e] context.

To review the findings presented thus far, then, note that PVI works to cue a great many differences. It distinguishes between plain and breathy sonorants and voiced obstruents, and

<sup>17</sup> There is a catch here: Mikuteit and Reetz include the friction part of the affricate in their ACT measures, whereas the post-release friction prior to aspiration/breathiness is not included in any of the measures reported in this study—namely, closure duration, PVI, and vowel duration. If future work seeks to delineate the full temporal profile of these affricates, or syllables containing these affricates, a measure of the friction portion of them—or inclusion of the friction in one of the existing measures—will need to be incorporated.

between plain and aspirated voiceless obstruents. It also distinguishes between plain voiced and voiceless obstruents.

Let us turn now to the vowel results. While Dutta (2007b) reported differences in the duration of subsequent vowels based on consonant phonation type, it is very possible that we will not find a similar pattern because of the different measurement techniques utilized herein: some portion of the signal which Dutta presumably included as vowel duration has, in the present study, been apportioned to the PVI instead. Thus, we expect vowel durations to be shorter overall than those reported in previous studies, and the effect of phonation type on vowel duration may be absent as well. This is important to keep in mind as we move forward.

### **4.3 Vowel Duration**

Although we may expect a main effect of Phonation type not to emerge based on the fact that much of what others measured as vowel is here measured as part of the PVI, the effect of Phonation type and the interaction of Phonation type with other factors remain of key interest. As with the temporal measures related to consonant duration, however, several additional overall effects emerge, and these are reported here both in the interest of being thorough and because prior acoustic analysis of Marathi is scant.

The difference between the mean duration of [a] ( $M = 133$  ms) and that of [e] ( $M = 127$  ms) attains significance, with  $F(1,9) = 12.73$  and  $p = 0.006$ . Other factors also trigger significant differences in mean vowel duration, but the numerical differences are so small that it is difficult to interpret them as being meaningful in any concrete way. For example, there is a significant effect of Place of articulation on mean vowel duration, with ( $F(1,7) = 13.08, p < 0.001$ ). Vowels

after bilabial sounds ( $M = 134$  ms) are longer than vowels after alveo-palatal sounds ( $M = 129$  ms), which are in turn longer than those after dental sounds ( $M = 127$  ms).

If we turn to Phonation type, and the way it interacts with other factors, we find that—as predicted at the end of the previous section—vowels after aspirated/breathy consonants ( $M = 126$  ms) do not differ from those that follow plain consonants ( $M = 134$  ms) ( $F(1,7) = 4.512, p = 0.062$ ). This diverges from the findings reported by Dutta 2007 for Hindi, wherein vowels after breathy voiced obstruents were longer than those following plain obstruents, but the explanation for this is that the bulk of that increased duration is apportioned to the PVI in the present study. Thus, phonation type distinctions are still cued temporally; the only difference lies in whether we point to the vowel or to the PVI as carrying that temporal cue.

The omnibus Anova also revealed an effect of Gender ( $F(1,7) = 5.882, p < 0.04$ ). Vowels produced by female speakers are longer on average ( $M = 133$  ms) than those produced by male speakers ( $M = 127$  ms); again, however, this difference is so small that, regardless of its statistical significance, it is not clear that it is informative. It is perhaps worth noting that gender has not triggered significant durational differences elsewhere—neither in the consonant results nor in the PVI results—and it does not interact with other factors in the rest of the vowel data that will be presented here. One message to take from this is that gender does not trigger large-scale durational differences overall. The broad pattern remains consistent between male and female speakers.

Two higher order interactions also emerge in the vowel duration data. These include a Phonation type by Obstruency by Vowel context interaction ( $F(1,3) = 78.64, p < 0.001$ ), and a Phonation type by Vowel context by Place of articulation interaction ( $F(1,3) = 14.026, p =$

0.006). There is no effect of Voice in the obstruent data, and mean durations separated according to vowel context, obstruency, and phonation type appear in Table 8.

**Table 8 Mean vowel durations by vowel context, obstruency, and breath**

	<b>[a]</b>		<b>[e]</b>	
	plain	breathy	plain	breathy
<b>Labial Obstruents</b>	150	119	132	141
<b>Dental Obstruents</b>	151	128	149	130
<b>Alveo-Palatal Obstruents</b>	151	118	113	133
<b>Labial Sonorants</b>	145	116	153	131
<b>Dental Sonorants</b>	131	130	116	120
<b>Alveo-Palatal Sonorants</b>	137	145	106	131

Duration of both [a] and [e] differ significantly based on phonation type in almost all contexts, but the difference is often in the unexpected direction—meaning, had our data looked like that of Dutta 2007, we would expect breathy phonation to trigger longer vowels. That it does not has already been explained—this is a function of the measurement techniques used herein, where the PVI includes a portion of the vocalic interval. The differences above that are of note are the [e] values after breathy alveo-palatal obstruents and sonorants. These are two instances where vowels are significantly longer after breathy than after plain consonants. Again, the extent to which this durational difference is useful for listeners remains to be investigated in future perception studies, but one possible explanation is that the cues we will discover in later chapters—spectral cues, or those related to F0 or CPP—may be weaker after breathy alveo-palatals than after breathy labials or dentals. If this proves to be the case, then the increased duration illustrated in Table 8 could be interpreted as evidence of compensation.

No additional significant interactions emerge. In many ways, there is not a great deal to report regarding vowel duration precisely because the differences found in previous work are, in my opinion, due to the fact that breathy voiced sounds are associated with significant breathy intervals. Mikuteit and Reetz (2007) refer to this interval as superimposed aspiration—it is breathiness, or aspiration, intermingled with the vowel, and extends into the vowel for as much as 60 or more milliseconds. If this breathy interval is considered to be part of the vowel, then we expect to find significant differences in vowel duration. If it is instead measured separately, then the differences will be attributed not to the vowel but to the breathy interval itself, herein referred to as the PVI, and this is exactly what we see in the present study.

#### **4.4 A brief recap of duration results**

While the larger significance of these findings is best understood as part of the whole—meaning after the F0, spectral, and CPP results have been reported—several points may be made now. First, we may begin to wonder whether we will find sonorant-internal cues to breathiness, given their increased duration. In other words, perhaps increased segment duration is useful precisely because the breathy voiced sonorants carry additional segment-internal cues to phonation type. Second, the alveo-palatal consonants differ to some extent from the others. While the labials and dentals often pattern together, the alveo-palatals diverge: they do not show many consonant duration differences, and breathy alveo-palatals trigger increased duration in subsequent [e] vowels, whereas the labials and dentals do not. This may simply be a result of the specific consonants involved—the alveo-palatal obstruents are affricates rather than stops, and the sonorant is a rhotic that is not realized consistently. It may also be an indication, however,

that phonation type is cued differently in the alveo-palatals than in the other consonants. This is a possibility that remains to be investigated as we get to additional results.

Phonation type distinctions appear to strongly influence consonant and PVI durations. While voiced obstruents do not differ based on phonation type, sonorants and voiceless obstruents do, and they diverge in the expected directions—with plain voiceless obstruents showing longer mean closure durations than aspirated voiceless obstruents, and breathy sonorants showing longer mean constriction durations than plain sonorants. Nonmodal phonation—meaning aspiration and breathy voice—is associated with increased PVI values, and breathy voice is associated with longer PVI values than aspirated voiceless sounds. Vowel durations do not show a consistent lengthening effect triggered by breathy phonation, and this diverges from previous findings related to Hindi (Dutta 2007). The reasons for this divergence have been mentioned several times—namely, part of what was measured as vowel duration in the Hindi data was instead considered to be part of the breathy interval in the current data, and was therefore apportioned to the PVI.

Duration will be discussed again at the end of the results section, when it can be fitted into a discussion that takes into account all of the measures reported herein. We turn now to the fundamental frequency results.



## 5 Fundamental Frequency Results

There is some crosslinguistic evidence that phonation type differences trigger differences in fundamental frequency. Of particular interest in the present study are the findings reported for Hindi (Dutta 2007, Ohala 1979, Schiefer 1986). Given the findings of these previous studies, fundamental frequency is expected to be lower after voiced obstruents than after voiceless obstruents, and potentially even lower after breathy voiced than after plain voiced obstruents. The present study does not look at vowels exclusively, however: it looks at the fundamental frequency of voiced consonants as well, meaning the consonant-internal fundamental frequency of sonorants and voiced obstruents. As such, we are able to investigate whether the pattern found in vowels emerges in consonants as well. If it does, we expect breathy sonorants and voiced obstruents to have lower mean fundamental frequency values than their plain counterparts. The idea here is that if breathiness has a consistent lowering effect on fundamental frequency—as seems to be the case, given the consistency with which such an effect is reported for Indic languages—then that effect may be expected to emerge both consonant-internally and in subsequent vowels.

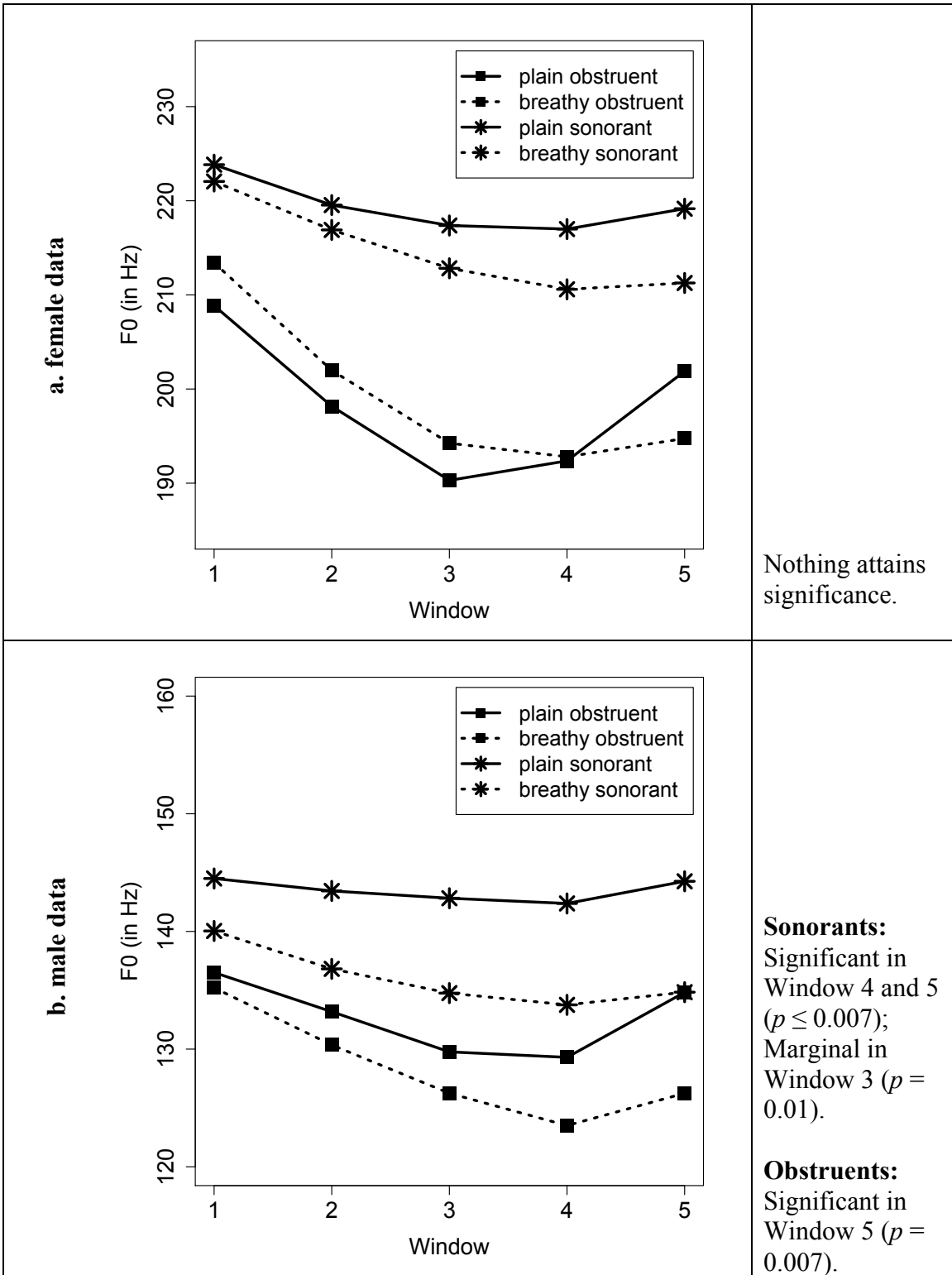
Male and female data are presented separately throughout this section, and we will see that F0 patterns differently for the two genders. Overall, male data often show the expected patterns, with breathy phonation having a fairly consistent lowering effect on fundamental frequency. For the female speakers, however, breathy phonation is only inconsistently associated with lowered fundamental frequency.

## 5.1 Fundamental frequency in voiced consonants

An omnibus repeated measures analysis of variance was conducted to determine whether Phonation type triggers significant differences in the F0 values of voiced consonants, and to determine whether there were interactions between Phonation type and the remaining independent variables of Place of articulation, Vowel context, Word position, Obstruency, and Window. As mentioned above, Gender was not included in the omnibus: rather, male and female data were tested separately.

The results indicate that there is a main effect of phonation type on mean F0 values for voiced segments for males, with ( $F(1,4) = 26.44, p = 0.007$ ). Their plain voiced consonants ( $M = 138$  Hz) have a higher F0 than their breathy voiced consonants ( $M = 132$  Hz). The same is not true for women: their plain voiced consonants ( $M = 212$  Hz) are slightly higher than their breathy voiced counterparts ( $M = 208$  Hz) numerically, but this difference does not attain statistical significance. Although this is not one of the comparisons of note, it is also the case that both genders show a main effect of obstruency—this is visible in Figure 19, where the mean value of F0 across the time course of plain and breathy sonorants and voiced obstruents is illustrated. Men, with  $F(1,4) = 125.9$  and  $p = 0.0004$ , have a mean sonorant F0 of 140 Hz and a mean voiced obstruent F0 of 131 Hz. Women, with  $F(1,4) = 173.7$  and  $p = 0.0002$ , have a mean sonorant F0 of 217 Hz and a mean voiced obstruent F0 of 199 Hz. Of more interest to us is the fact that both genders show a Phonation type by Obstruency by Window interaction (females –  $F(1,4) = 74.98, p = 0.013$ ; males –  $F(1,4) = 32.81, p = 0.005$ ).

Figure 19 Time course of F0 (in Hz) in voiced consonants



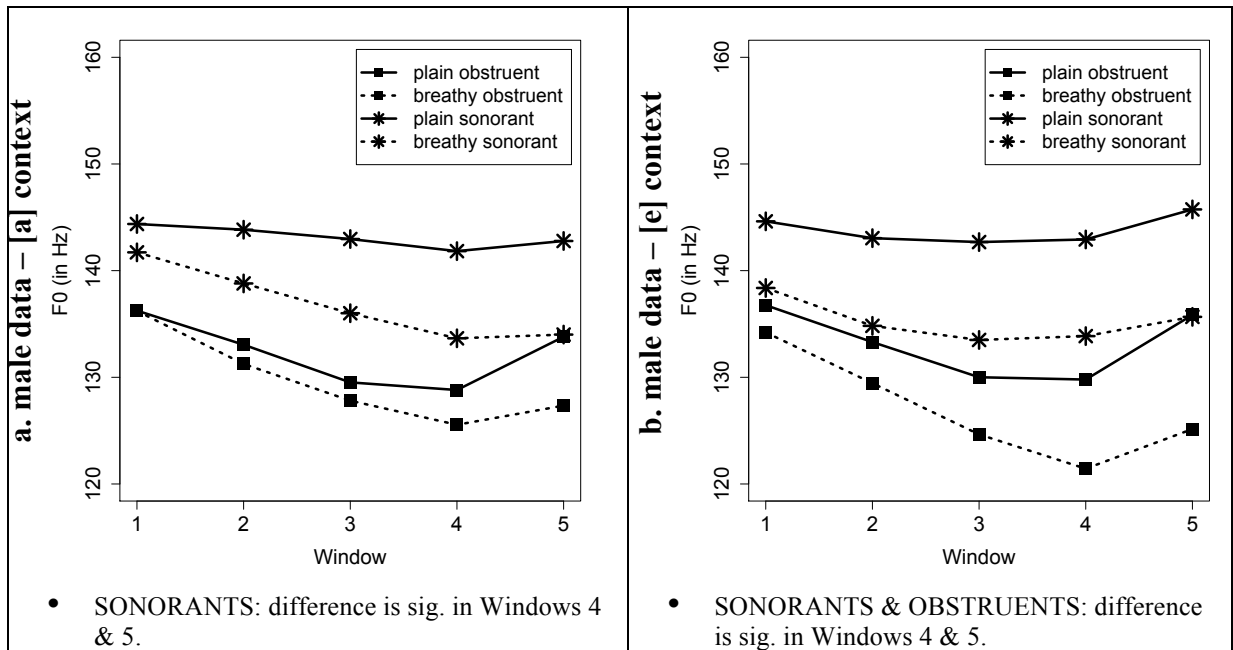
Post hoc analyses do not help explain this interaction for females, as they fail to reveal significant F0 differences between plain and breathy obstruents or sonorants.<sup>18</sup> For male speakers the difference between plain and breathy sonorants is marginal in window 3 and attains significance in windows 4 – 5, with  $p \leq 0.005$ .

Additional interactions include a Phonation type by Vowel context by Window interaction ( $F(1,4) = 87.45, p = 0.01$ ) and a Phonation type by Word position by Window interaction for males ( $F(1,4) = 104.8, p < 0.001$ ). In Figure 20, we see male mean F0 values across voiced consonants separated by phonation type, obstruency, and vowel context. Note that for the sonorants, F0 in Windows 4 and 5—meaning the final 40% of the consonants—differs significantly based on phonation type in both vowel contexts, with  $p \leq 0.007$ . What changes from the [a] to the [e] vowel context are the obstruents: while no significant differences between plain and breathy obstruent values emerge in the [a] context, in the [e] context they differ significantly in Windows 4 ( $F(1,4) = 44.09, p = 0.003$ ) and 5 ( $F(1,4) = 33.73, p = 0.004$ ). This is interesting: recall that the higher vowel context is the one where there is a greater likelihood of interference from vowel formants in spectral measures like H1-H2. Here, we see F0 cueing breathiness in the very context where spectral measures may prove to be a weaker cue.

---

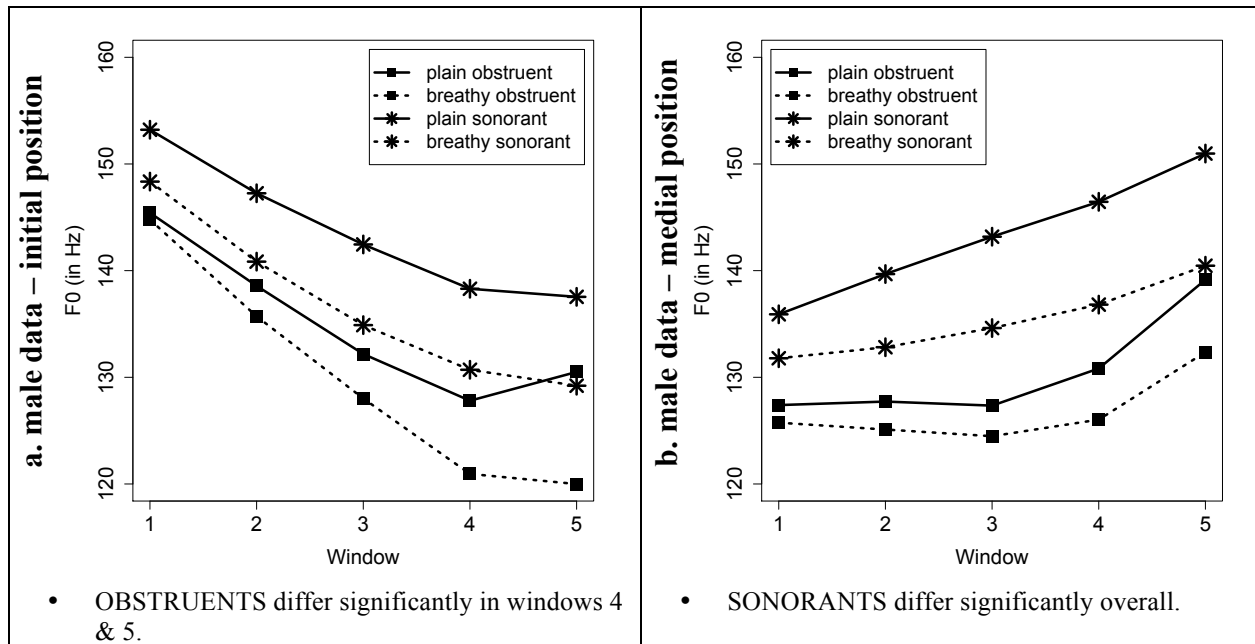
<sup>18</sup> This is because R does not automatically correct for multiple comparisons, and so sometimes significant interactions emerge in the omnibus Anova but none of the post hoc tests meet the significance level established for the current research ( $p \leq 0.007$ ).

Figure 20 Time course of F0 across voiced consonants by vowel context – male data



We turn now to the interaction involving word position. In Figure 21, we see mean male F0 values plotted across voiced consonants separated by phonation type, obstruency, and word position. Here, breathy obstruents have lower mean F0 values than plain obstruents in word-initial position in the final 40%—Windows 4 and 5—of the consonant. In addition, breathy sonorants have lower F0 overall when they occur in word-medial position.

Figure 21 Time course of F0 across voiced consonants by word position



If we can make broad observations about these patterns, they are as follows. For the females, F0 is not informative. While several interactions attain significance in the omnibus Anova, post hoc tests reveal no significant differences for female data based on Phonation type.

The parameter of F0 is more informative for the male speakers, however. For them, breathy phonation has a more consistent lowering effect on sonorants than on obstruents, although breathy obstruents do show lowered F0 values in Windows 4 and 5 in the [e] context and in word-initial position. Taken together, the conclusion we must draw is that F0 in consonants may be a cue for breathiness in male speech, but does not consistently distinguish plain from breathy consonants in female speech.

The next question we consider is whether the same holds true in vowels. Much of the previous research on breathy voiced sounds in Indic languages, and in Hindi in particular, has looked not at the fundamental frequency of consonants themselves but at the effect of a consonant's phonation type on the subsequent vowel (Dutta 2007, Ohala 1973). The expectation

based on these previous findings is that breathy phonation has a lowering effect on F0 values in the following vowel. This is explored in the next section.

## 5.2 Fundamental frequency in vowels

Before proceeding with the results, recall that the breathy interval measured as the PVI is collapsed into the vowel for analysis of the F0, spectral, and cepstral results. Thus while there were three subsections of data presented in the duration chapter—consonant, PVI, and vowel—there are only two subsections in this and the following chapters. For each, results related to the consonant are presented first, while those related to the vowel are presented next. It is important to remember this as we proceed: here, and in the coming chapters, “vowel” refers to the interval following the release of the consonantal constriction. The breathy interval is considered to be a part of the vowel.

As with the consonants, an omnibus repeated measures analysis of variance was conducted to determine whether Phonation type triggers significant differences in mean F0 in vowels, and to determine whether there were interactions between Phonation type and the remaining independent variables of Place of articulation, Vowel context, Word position, Obstruency, and Window. Again, male and female data were tested separately.

That said, the two genders show similar results in many ways: *F* and *p* values for significant interactions are presented in Table 9.<sup>19</sup> Before moving into the post hoc analyses of these interactions, note that the omnibus Anova reveals that there is no overall phonation type effect for either gender. Both show a phonation type by window interaction, however. This is not

---

<sup>19</sup> In the instances where interactions are so complex that it is not clear how to disentangle them, I do not attempt to do so.

unexpected, and is interpreted as reflecting the fact that the effect of consonant phonation type on F0 does not perseverate throughout the entire vowel, but is restricted to the initial portion.

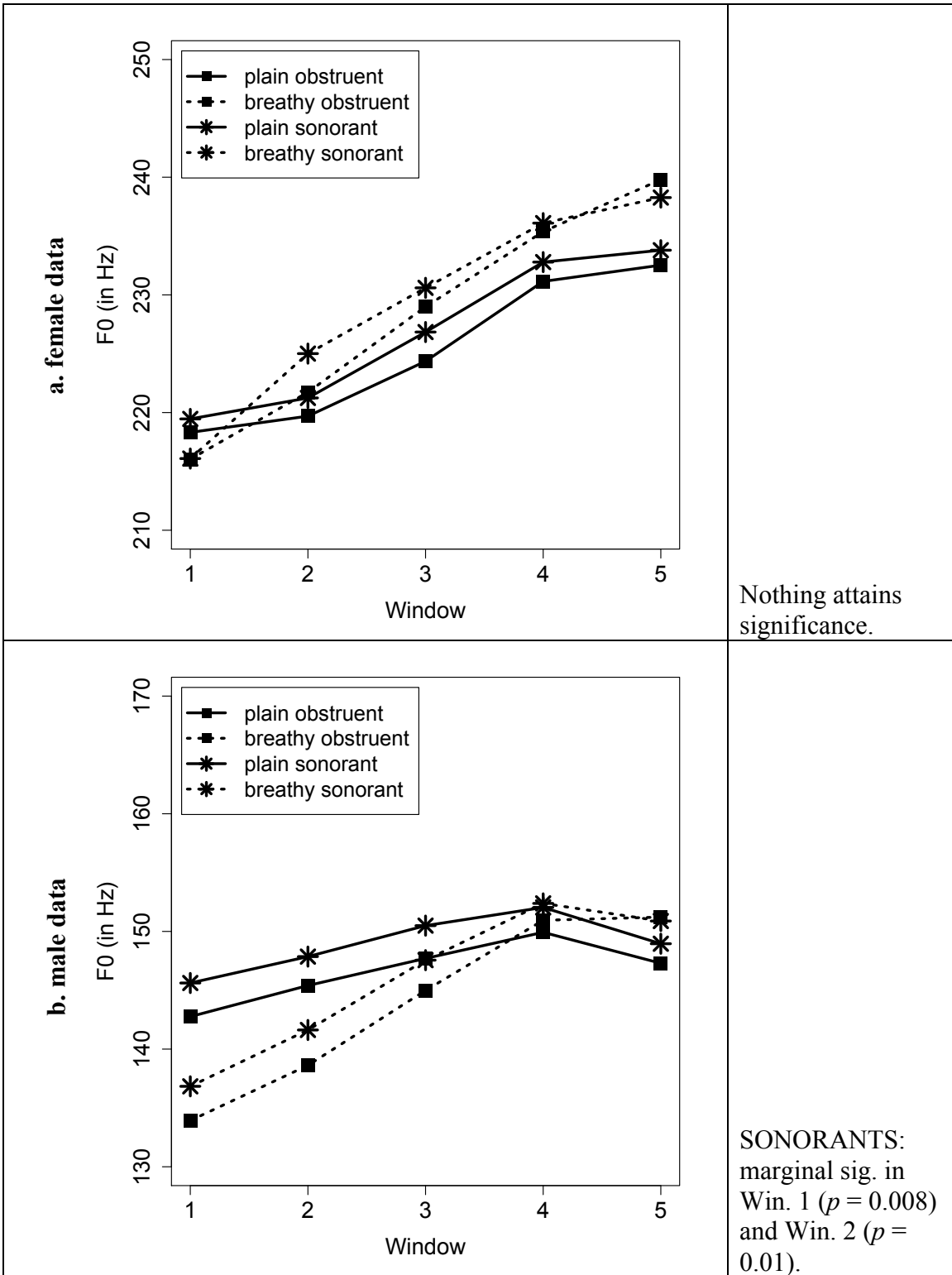
**Table 9 F0 of [a] and [e] by obstruency and place of articulation (omnibus)**

	Female		Male	
	<i>F</i> (1,3)	<i>p</i>	<i>F</i> (1,3)	<i>p</i>
<b>Phonation type x window</b>	21.34	0.02*	70.41	0.004**
<b>Phonation type x voice x window (obstruents only)</b>	38.27	0.03*	54.28	0.005**
<b>Phonation type x vowel context</b>	61.83	0.004**	64.59	0.004**
<b>Phonation type x word position</b>	114.2	0.002**	158.59	0.001**
<b>Phonation type x obstruency x window</b>	--	--	18.58	0.02*
<b>Phonation type x vowel context x window</b>	38.9	0.008**	--	--
<b>Phonation type x obstruency x vowel context</b>	291.41	<0.001**	11.15	.04*
<b>Phonation type x obstruency x vowel context x window</b>	--	--	14.4	.03*

Data in the following line graphs are separated by obstruency and plotted across all 5 Windows, because of the Phonation type by Obstruency by Window interaction. Figure 22 presents the overall time course of F0 across vowels after sonorants and voiced obstruents. Female data appear in (a.), and male data in (b.). Discussion follows.



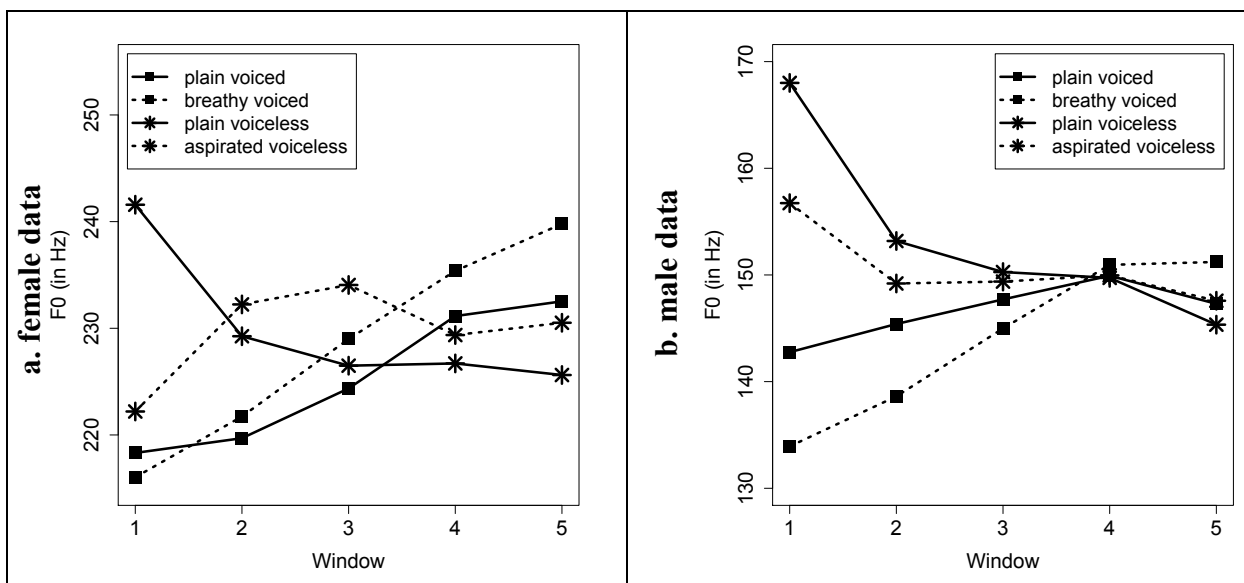
Figure 22 Time course of F0 across vowels after voiced consonants by obstruency and phonation type



As was the case when looking at F0 within consonants, post hoc tests reveal that significant differences triggered by phonation type do not arise for female speakers. Marginal significance is attained for vowels after sonorants in Windows 1 and 2 for male speakers, with F0 values in vowels following breathy sonorants marginally lower than those following plain sonorants.

Recall that there is also a phonation type by voice by window interaction. This is shown in Figure 23. Once again, breathy phonation does not trigger F0 differences in vowels after voiced obstruents for women; however, the difference between vowels following plain and those following aspirated voiceless obstruents attains significance in Window 3 ( $F(1,4) = 67.18, p = 0.001$ ). This is the only significant result, and yet it is not very informative: this point of significance arises not at the beginning of the vowel, closest to the consonants that presumably trigger the difference, but in the middle of the vowel. Furthermore, our interest lies heavily in the direction of breathy phonation, which triggers no significant differences for female speakers. For the male speakers, post hoc tests do not reveal significant differences in vowel F0 values based on the phonation type of the preceding consonant.

**Figure 23 Time course of F0 across vowels after obstruents by voice (post hoc)**

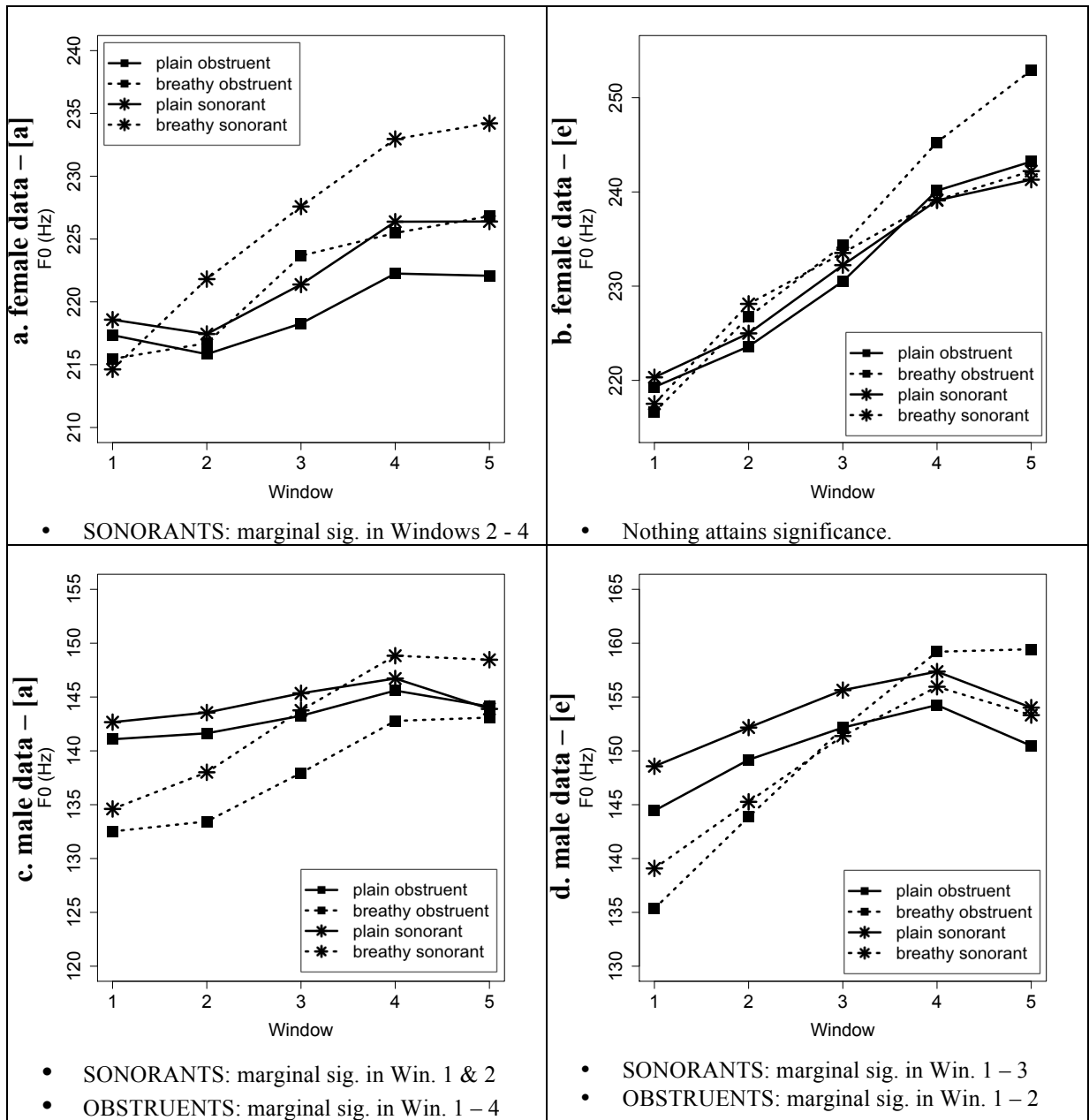


In the figure above, and in Figure 19, breathy voiced obstruents in the female data are associated with mean F0 values that are numerically higher than those associated with modal phonation. This difference does not attain significance, but it is puzzling in that it is not in the expected direction. Sonorants, meanwhile, show very little in the way of significant F0 differences based on phonation type, but they do pattern in the expected direction. Breathily sonorants are associated with F0 values that are numerically lower than plain sonorants. The expected pattern disappears completely in the females' vowels; none of the differences between plain or breathy sonorants or voiced obstruents attain significance, but the numerical difference is in the unexpected direction across the board.

We should be cautious about interpreting this because it does not attain significance. We will see that it is consistent throughout vowel contexts and word positions, however, and attains marginal significance in some contexts. It is also the case that an unexpected pattern in the fundamental frequency has the potential to affect most of the spectral measures that will be reported subsequently—namely H1-H2c, H1-A1c, H1-A2c, and H1-A3c. These values depend on measures of the first harmonic—the F0—and so it is at least worth drawing the reader's attention to the unusual F0 pattern found for female speakers, which makes it look as though breathy voice is having the same effect as aspiration. Furthermore, because males and females pattern differently, it raises the very real question of whether breathy voice is being produced differently by the two genders. This is something that bears additional investigation: the same question will be raised multiple times by the results reported throughout the next chapters, and so the proposal that breathy phonation is cued differently by male and female speakers will be returned to in greater detail in the discussion section.

Post hoc analyses were conducted to investigate the Phonation type by Vowel context by Obstruency interaction, which was significant for both genders in the omnibus Anova, and the Phonation type by Vowel context by Obstruency by Window interaction, which was significant for men. Data are plotted in Figure 24. Differences here are marginal rather than significant.

Figure 24 Time course of F0 across [a] and [e] (post hoc)

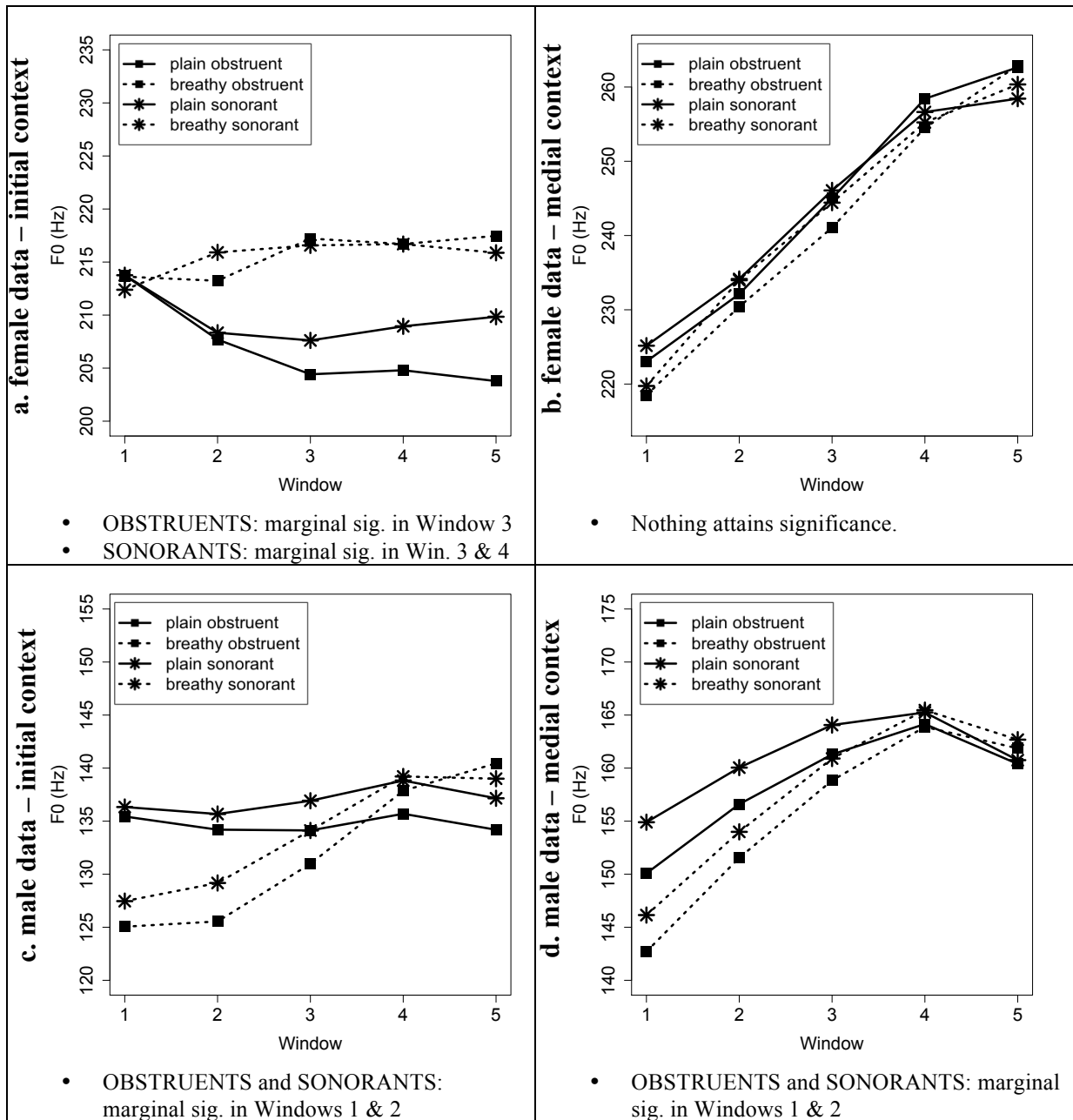


Differences in F0 following plain and breathy segments near significance for [a] after sonorants in windows 2-4 for female speakers. For males, F0 differences consistently approach significance in Windows 1 and 2 – in the first 40% of the vowel, in other words. This is true for both obstruents and sonorants, in both [a] and [e]. While other pockets of marginal significance

do emerge, as annotated above, the take-home message seems to be that F0 in vowels does not reliably cue breathy phonation.

Both genders show a Phonation type by Word position interaction, and Word position data are plotted in Figure 25. Once again, male F0 values attain near-significant differences based on phonation type in the early windows of both initial and medial word positions. Female F0 values differ based on phonation type in window 3 following voiced obstruents and sonorants in initial position, and in window 4 following sonorants. (Vowels after sonorants, in other words, are different for 40% of the vowel, while those following obstruents are different for only 20% of the vowel.) Again, it is difficult to know what to say about these results: why is the phonation type of the preceding consonant triggering a difference so late in the vowel, and why do the values diverge opposite the expected pattern? This is strange, and is left unexplained at present.

Figure 25 Time course of F0 across vowels by word position (post hoc)



No significant differences are found after consonants in medial position, but those data are included here because they illustrate a point of interest. Tokens were embedded in carrier sentences for several reasons, one of which was to aid in eliciting natural pronunciation as much as possible and to avoid getting a list effect. As you can see when you look at the time course of

F0 after consonants in initial and medial positions—in other words, at F0 across first and second syllables in disyllabic words—F0 across syllable one is relatively flat; F0 across syllable two rises. Speakers did indeed fall into a sort of sing-song rhythm at times, which included a rising pitch on the second syllable of target words, and this is clearly reflected in the graphs above.

No additional results related to mean fundamental frequency in vowels is reported. While the omnibus Anova for males and females returned a number of significant interactions, post hoc tests do not greatly aid us in disentangling those effects.

### **5.3 F0 recap**

Perhaps the most noteworthy finding related to F0 is that while other researchers have found consonant phonation type in Indic languages to trigger reliable differences in F0, the same is not found in the Marathi data reported here. While F0 works better for male than for female speakers in terms of cueing breathy phonation, results tend to be marginal rather than significant.

For the male speakers, consonant-internal F0 differences do cue phonation type differences. Breathily obstruents have lower mean F0 values than plain voiced obstruents in the final 20% of the consonant, and breathily sonorants have lower mean F0 values than plain sonorants in the final 40% of the consonant. Furthermore, breathily sonorants are lower overall than plain sonorants in word medial position. Breathily sonorants also trigger marginally lower mean F0 values in the first 40% of the subsequent vowels, but again this is for male speakers only. On the whole, then, F0 simply is not highly informative as a cue for breathy phonation. This diverges from findings related to Hindi (Ohala 1979, Schiefer 1986, Dutta 2007). It is perhaps not entirely surprising, however. Pennington (2005) makes a point of noting that the association between breathy phonation and a lowered fundamental frequency, although



suggested by previous findings, has not yet been established as a universal phenomenon. To that end, the current findings suggest that the association is not universal—although found fairly consistently for Hindi, the correlation between breathy phonation and lowered fundamental frequency is spotty at best in the Marathi data presented herein, and is more reliably present for male than for female speakers.

Thus far, then, we have seen that breathy phonation is cued by temporal parameters such as consonant constriction duration for sonorants and Pre-Vocalic Interval measurements for both sonorants and obstruents, while F0 does not provide such a strong cue. We turn now to the lower frequency spectral measures of H1-H2c and H1-A1c, and the higher frequency spectral measures of H1-A2c and H1-A3c, to determine whether these prove more informative than fundamental frequency.

## **6 Spectral Measure Results**

The following sections present the results for the spectral measures of H1-H2c, H1-A1c, H1-A2c, H1-A3c, and the cepstral measure of CPP. Recall that the general expectation for the spectral measures is that values associated with breathy phonation will be larger than those associated with modal phonation. For CPP, on the other hand, modal phonation is expected to trigger increased values as compared with breathy phonation. We begin with H1-H2c.

### **6.1 H1-H2c Results**

H1-H2 is a measure of the amplitude of the second harmonic subtracted from the amplitude of the first harmonic. During the production of breathy phonation, the fundamental frequency is dominant in the spectrum, and so H1 values are large and positive. Breathly phonation is also characterized by a sharper decline in the amplitude of higher frequency energy than modal voicing—for instance, Fant (1979) reports that breathy phonation is associated with a widened F1 bandwidth and decreased F1 amplitude. Breathly phonation is thus expected to show a steep spectral slope: high-amplitude H1 values and lower H2 values, resulting in large positive H1-H2 measures. This is the prediction to bear in mind throughout the next section, then: breathly voiced sounds are expected to be associated with greater H1-H2 values than modal sounds.

Recall that, as mentioned in Chapter 2, spectral measures such as H1-H2 are vulnerable to influence from formant frequencies and bandwidths. For this reason, the measures reported herein are automatically corrected for the possible influence of formants by VoiceSauce, and are thus referred to as, for example, H1-H2c rather than H1-H2.

There are a number of questions of interest to address in this portion of the results. First, will the Marathi data follow the expected pattern, with H1-H2c values being higher for breathly

than for modal sounds, particularly given the fact that we saw some unexpected patterns in the F0 results for females? Whereas the prediction was that breathy voice would have a lowering effect on fundamental frequency, this was not in fact the case for the female speakers. In addition, will sonorant-internal differences in H1-H2c values distinguish plain from breathy sonorants, given that they did not cue phonation type differences in Sumi nasals and laterals (Harris 2009)? Finally, obstruent phonation type has been shown to affect H1-H2c values in subsequent vowels in Indic languages such as Hindi (Dutta 2007); it is expected that the same will be found in Marathi. Marathi differs from Hindi in that it contains breathy voiced sonorants in addition to obstruents, however, whereas Hindi employs breathy phonation only in obstruents. As such, Marathi allows us to investigate whether sonorants and obstruents affect vowels similarly, or whether there are differences in the way they pattern. The comparison between sonorants and obstruents is a key element of the present study. The expectation is that making such a comparison will provide insight into the fact that phonemic breathy phonation is more commonly employed in obstruents than in sonorants crosslinguistically.

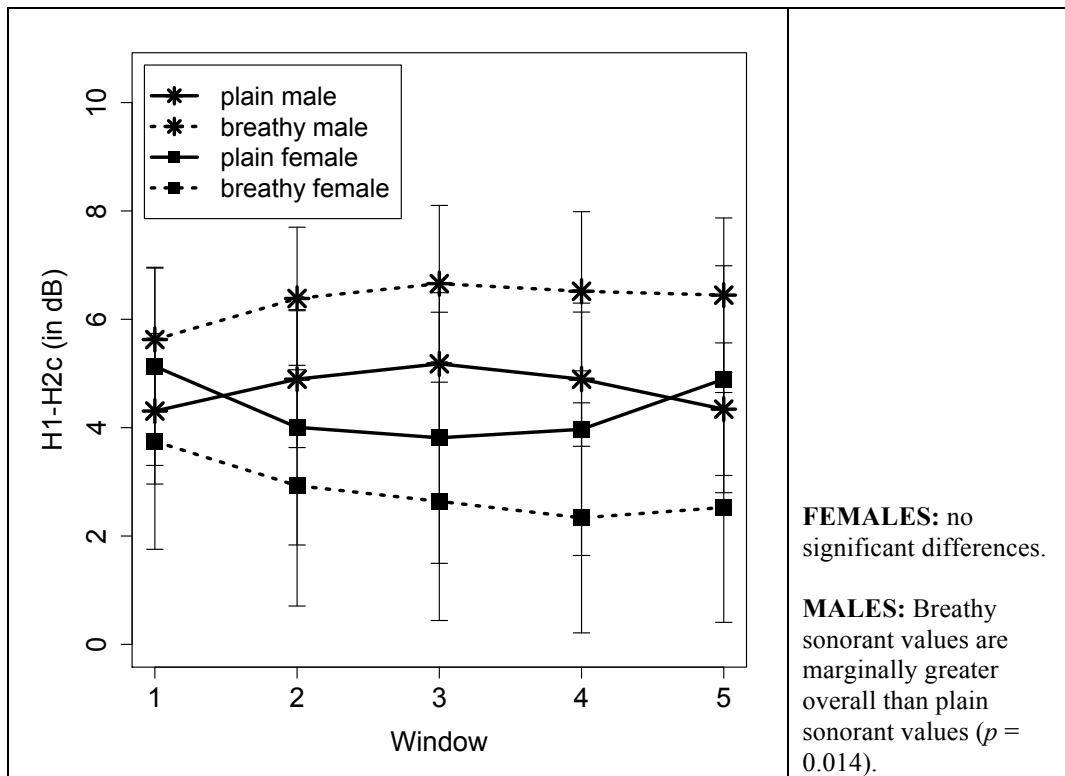
We begin with results related to H1-H2c values in consonants, and then move into those related to vowels.

### **6.1.1 H1-H2c – consonant results**

In the case of H1-H2c, when we discuss consonant results they pertain only to the sonorants. This is true for most of the spectral measures: with the exception of CPP, spectral measures were obtained for the consonantal constriction of the sonorant segments but not across the closure duration of obstruents.

A plot of H1-H2c values over the time course of sonorant segments appears in Figure 26. Several observations can be made about this figure. First, it looks as though male and female data pattern differently, and we will see that this observation is statistically supported. The male data looks as we might expect, with breathy sonorants yielding numerically higher values than plain sonorants. This persists across the entire duration of the segment. Results for the female speakers almost seem to mirror this pattern, however; values for breathy sonorants are numerically lower than those for plain sonorants. Again, this persists throughout the entire segment. We also see a large degree of overlap in the error bars. For the female data, error bars on the breathy plot regularly extend beyond the means of the plain plot, and vice versa, whereas the same is not true for the male data. This suggests that we will find a significant effect of phonation type on H1-H2c values for the male speakers, but not for the female speakers.

**Figure 26 Time course of H1-H2c across sonorants by phonation type and gender**

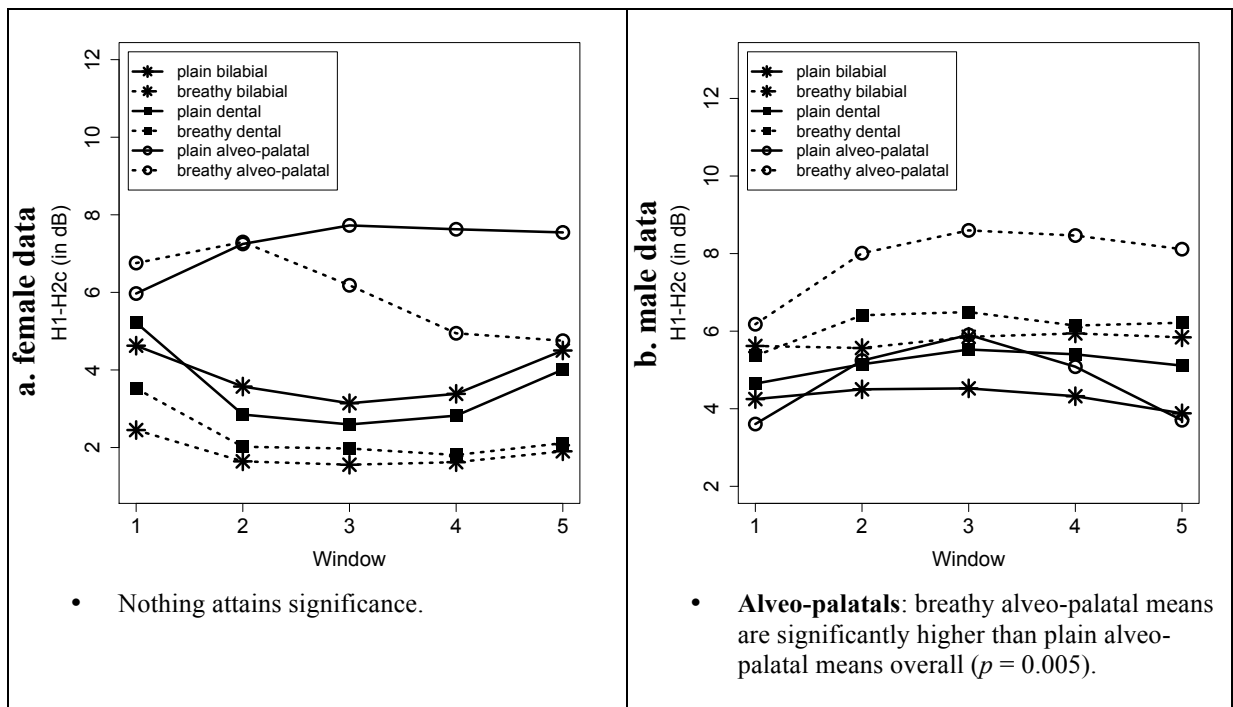


Note that if we interpret H1-H2c as a measure that reflects Open Quotient (OQ)—and the correlation between the two is widely accepted, having been established in an experimental setting by Holmberg *et al.* (1995)—then the above values make it look as though the females in this study have a lower mean OQ than the males. This would be surprising if true; in a study of 25 male and 20 female speakers of American English, Holmberg *et al.* (1989) found that the mean OQ for male speakers (0.609) in normal speech is lower than that of female speakers (0.735) (p. 518). Mean H1-H2c values for the plain sonorants produced by male and female speakers plotted above do not differ significantly, however, and so perhaps the data is not as unexpected as it may appear at first glance. That said, breathy sonorants produced by females are unquestionably baffling in many regards, not least because they pattern opposite the expected direction. This will be explored in greater detail as we move forward.

As with the other measures that have been reported, an omnibus repeated measures analysis of variance was conducted to determine whether Phonation type triggers significant differences in the H1-H2c values of sonorant segments, as well as whether there were interactions between Phonation type and the remaining independent variables of Gender, Place of articulation, Vowel context, Word position, and Window. The results indicate that while there is no overall effect of Phonation type, there is a Phonation type by Gender interaction ( $F(1,7) = 8.32, p = 0.02$ ). This is not surprising, given what we see in Figure 26. The effect of Phonation type is marginally significant overall for males ( $F(1,4) = 17.73, p = 0.014$ ), but not for females. In other words, for males the difference between plain and breathy sonorant H1-H2c values approaches significance and is in the expected direction. For females, no significant differences arise.

The omnibus also reveals a Phonation type by Place of articulation interaction ( $F(1,7) = 7.49, p = 0.03$ ), shown in Figure 27. In (a.), we can see that the female data follows the overall female pattern for each of the three places of articulation; in other words, breathy sonorants consistently show H1-H2c values that are numerically lower than the plain H1-H2c values, contrary to my predictions. Impressionistically, the bilabial and dental sonorants seem to pattern together, while the alveo-palatals have higher H1-H2c values overall, an effect which is significant ( $F(1,4) = 35.32, p = 0.004$ ). The alveo-palatals also seem to show greater differentiation based on phonation type, but they do not diverge from the overall pattern: no differences based on phonation type attain statistical significance on the female data.

**Figure 27 Time course of H1-H2c across sonorants by breath, place, and gender**



In the male data, plotted in (b.), breathy sonorants appear to follow the expected trend, with H1-H2c values numerically higher than their plain counterparts. The bilabials and dentals appear to pattern similarly, while the alveo-palatals once again appear to show greater

differentiation based on phonation type. Post hoc analyses reveal that this is indeed the case. There is no significant effect of Phonation type on H1-H2c values in the bilabials or dentals, but it is significant in the alveo-palatals ( $F(1,4) = 30.44, p = 0.005$ ). Recall that Phonation type triggered significant durational differences in bilabial and dental sonorants, but not in the alveo-palatals. Now, we see that sonorant-internal H1-H2c differences cue breathiness for the alveo-palatal sonorants, which lack the durational cue, but not in the bilabials and dentals, where breathiness is cued temporally via increased word-medial duration. Finally, here it looks like the dentals might be slightly higher overall than the bilabials, but the post hoc tests reveal that this is not the case. Bilabial and dental sonorants are not statistically different from one another in the male data. No window by window analysis was conducted, as Window was not involved in the interaction revealed by the omnibus.

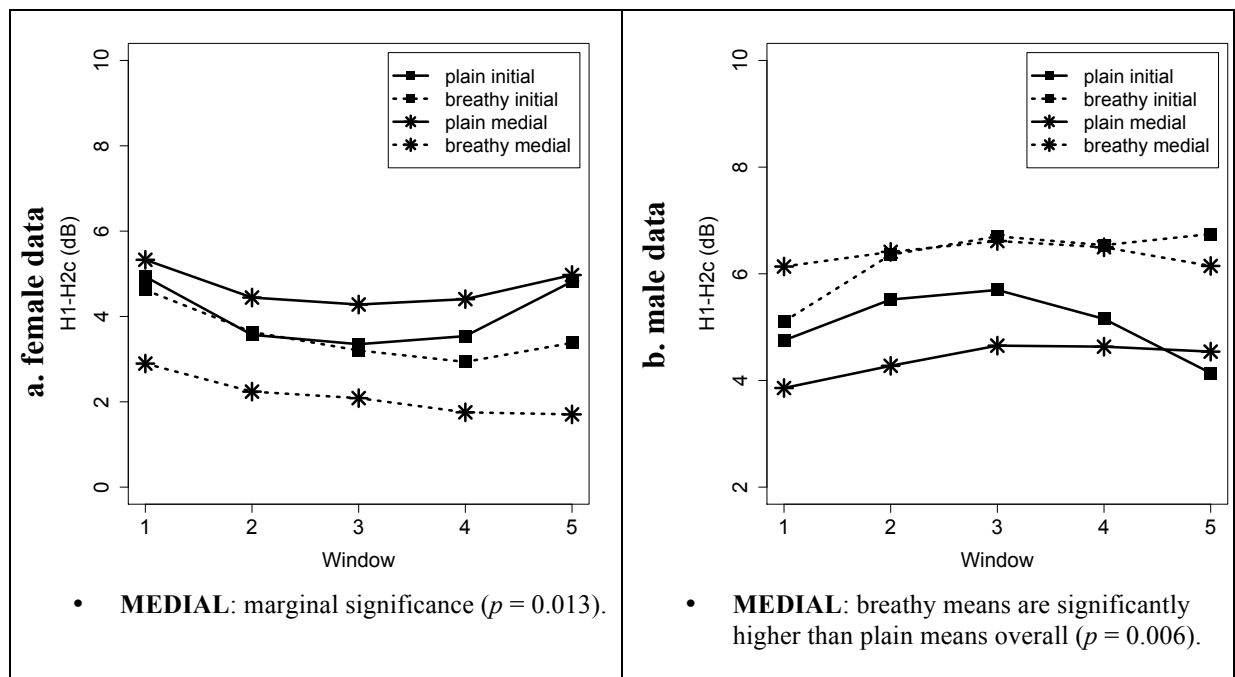
The final interaction that we will consider is a marginal Phonation type by Word position by Gender interaction ( $F(1,7) = 3.81, p = 0.09$ ). These data are presented in Figure 28. Once again, female data are shown in (a.) and male data in (b.). As with previous graphs, data are plotted across the time course of the segment for the purposes of visualization but there is no interaction between Word position and Window and so the post hoc analysis is for overall significance only.

The female results in (a.) show that in both word positions, breathy voiced sonorants have lower H1-H2c values—numerically—than plain sonorants, which is the unexpected direction. Sonorants in word-medial position show greater differentiation than those in word-initial position, however. Post hoc analyses reveal that differences in H1-H2c values based on Phonation type do not attain significance in word-initial position in the female data. In word-

medial position, however, the story is different: word-medially, mean H1-H2c values are marginally lower overall for breathy than for plain sonorants ( $F(1,4) = 17.92, p = 0.013$ ).

For the males, mean H1-H2c values are in the expected direction. Breathily sonorants have greater mean values than plain sonorants, and the difference is significant overall in word-medial position ( $F(1,4) = 29.43, p = 0.006$ ). This means that males pattern like females in that it is word-medial position where H1-H2c means are more differentiated based on phonation type. They pattern differently, however, in that the means diverge in the expected direction.

**Figure 28** Time course of H1-H2c across sonorants by phonation type and word position



There is a theme that begins to emerge here. As was the case with fundamental frequency, H1-H2c is not a good cue for breathiness for female speakers. Numerically, the female data pattern opposite what we would predict, but in most contexts the difference does not attain even marginal significance. Male data pattern in the expected direction: there is an overall effect of Phonation type, with breathy voiced sonorants yielding greater mean H1-H2c values than plain

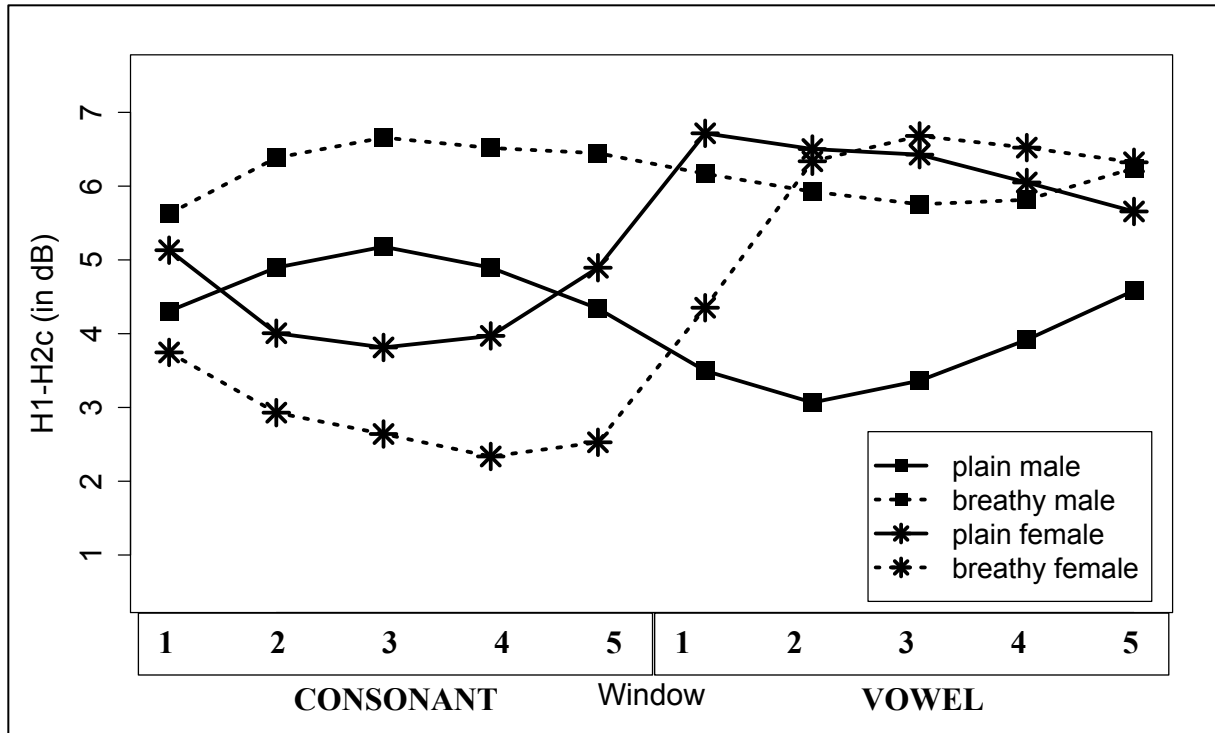


sonorants. Post hoc tests show that this overall significance is likely driven by specific contexts, however (i.e. word-medial position, or alveo-palatal sonorants).

The extent to which segment-internal spectral information in sonorants can distinguish plain from breathy phonation has never before been studied in detail, and thus remains an open question. Given the present results, however, we can now say that sonorant-internal H1-H2c values are not a strong cue for breathiness: breathy sonorants may differ from plain sonorants to some degree with regards to this measure, but the difference is not present in all contexts and is not sustained across both genders.

This finding is new; as it has not been investigated before, we cannot compare the present findings with previous research. Previous research has, however, investigated the effect of consonant phonation type on H1-H2 differences in subsequent vowels. To that end, we now turn to H1-H2c results for vowels. Before doing so, however, a plot of H1-H2c across full sonorant-vowel syllables is provided. While analysis of vowel data appears in the next section, this plot provides a picture of how the sonorant and vowel data fit together, allowing us to envision how things change over the time course of the syllable. It is a good image to have in mind as we proceed.

Figure 29 H1-H2c across syllables by gender (sonorant data only)



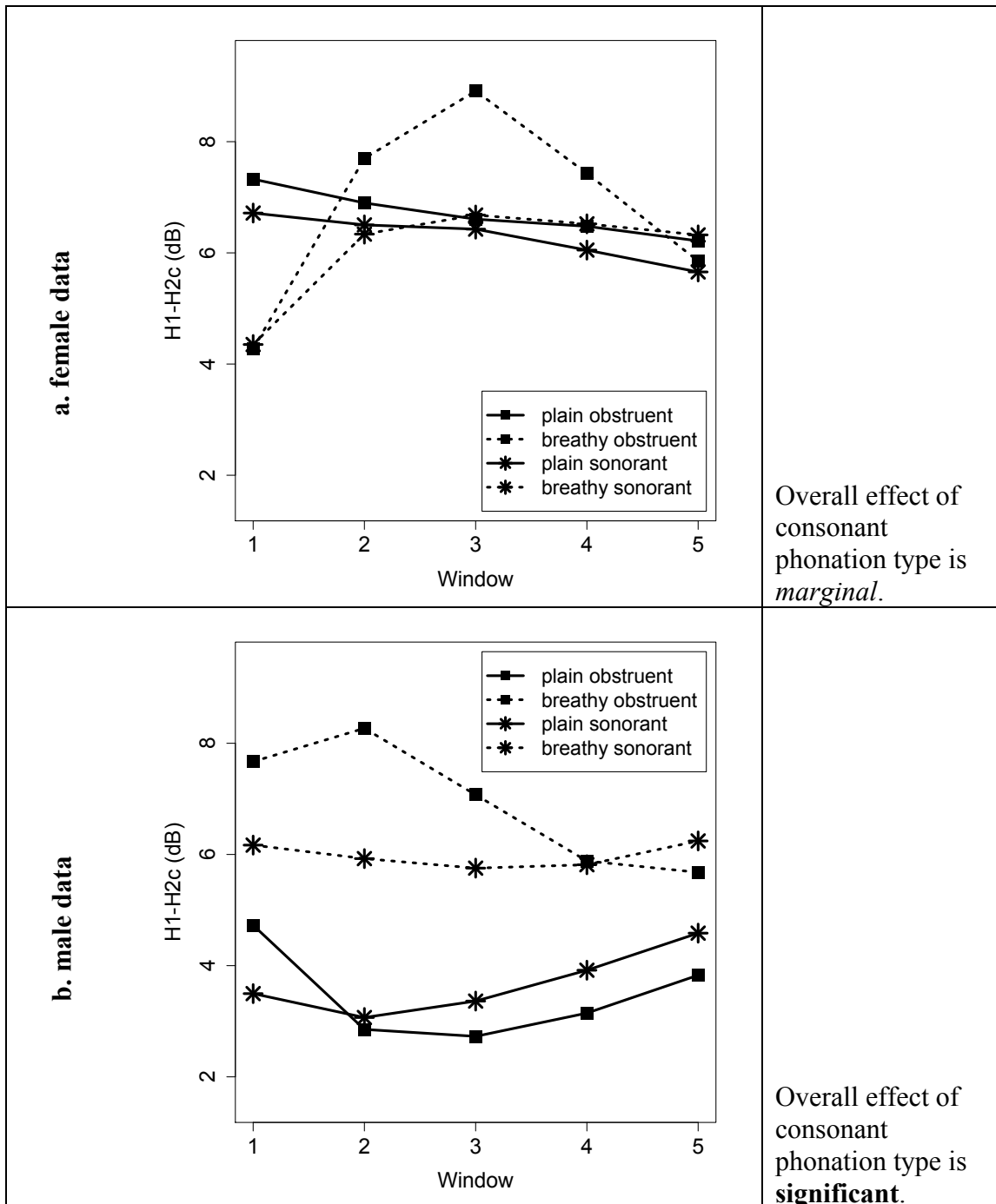
### 6.1.2 H1-H2c – vowel results<sup>20</sup>

As we have seen, segment-internal H1-H2c differences based on phonation type are inconsistent for sonorants—they are present more often for male than for female speakers, and even within male speakers reliable differences emerge only in specific contexts. When we observe the effect of consonant phonation type on subsequent vowels, however, we see a very different picture. The time course of H1-H2c across vowels following voiced segments—that is, sonorants and voiced obstruents—appears in Figure 30. Female data are plotted in (a.), and male data in (b.). Impressionistically, these data look more in line with what was predicted: vowels following breathy consonants have higher H1-H2c values, while those following plain

<sup>20</sup> Recall that the intervals reported as the PVI and as the vowel in the duration section are concatenated in the F0, spectral, and CPP results sections.

consonants have lower values. This appears to be true for both men and women. Furthermore, for both genders, the differentiation appears to be greater for vowels after obstruents than for vowels after sonorants in at least some windows.

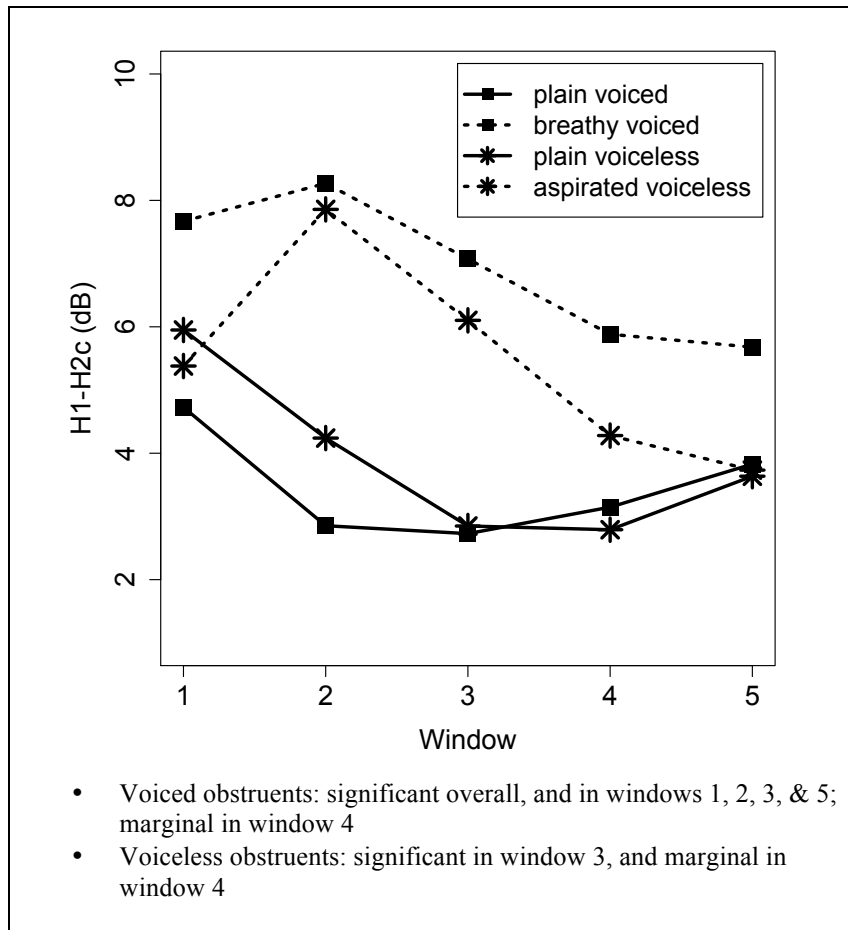
**Figure 30 H1-H2c across vowels after voiced segments by phonation type and obstruency**



An omnibus repeated measures analysis of variance was conducted to determine whether consonant phonation type triggers significant differences in the H1-H2c values of following vowels, as well as whether there were interactions between Phonation type and the remaining independent variables of Gender, Place of articulation, Vowel context, Word position, and Window. The omnibus revealed both a main effect of Phonation type ( $F(1,6) = 6.59, p = 0.04$ ) and a Phonation type by Gender interaction ( $F(1,6) = 6.59, p = 0.04$ ). A post hoc analysis conducted to investigate this interaction reveals that while there is an overall effect of Phonation type for the male speakers ( $F(1,4) = 81.24, p < 0.001$ ), the overall effect is marginal for the female speakers ( $F(1,4) = 33.98, p = 0.01$ ).

Before moving into a more thorough discussion of the results with regards to vowels following voiced segments, note that there is a Phonation type by Voice by Window interaction for vowels following obstruents ( $F(1,6) = 15.08, p = 0.01$ ). There was no overall effect of Phonation type or of Voice in the female data, nor was there an effect that attained significance in any specific window. This was not true for the male data, however, which appears in Figure 31. Vowels following breathy voiced obstruents have significantly higher H1-H2c values overall than those following plain voiced obstruents ( $F(1,4) = 135, p < 0.001$ ); furthermore, the window-by-window analysis reveals that the difference is significant with  $p \leq 0.007$  in all windows except window 4, where it is marginal ( $F(1,4) = 17.63, p = 0.014$ ). The difference between H1-H2c values following plain and aspirated voiceless consonants, on the other hand, attains significance in window 3 ( $F(1,4) = 34.5, p = 0.004$ ), and is marginal in window 4 ( $F(1,4) = 22, p < 0.009$ ), but these are not enough to drive overall significance. With regards to the cue of H1-H2c, then, voiced obstruents have an advantage over voiceless obstruents in terms of the effect of Phonation type on H1-H2c values in subsequent vowels.

Figure 31 H1-H2c across vowels after obstruents by voice – male data



Having assessed the effect of Voice as an independent variable, the remainder of the analysis focuses on vowels following sonorants and voiced obstruents.

The first finding addressed is the Phonation type by Obstruency by Window interaction ( $F(1,6) = 31.36, p = 0.001$ ) revealed by the omnibus Anova. Recall that, in Figure 30, it looks as though there might be greater differentiation based on consonant phonation type in vowels after voiced obstruents than in vowels after sonorants. This is supported by the statistics, but only for the male speakers. Post-sonorant differences do not attain significance in any window of the female data; post-obstruent differences are significant in Window 1 ( $F(1,4) = 101.6, p < 0.001$ ). As with the sonorant-internal results, however, the values here diverge from the expected

direction, with breathy voiced obstruents triggering lower mean H1-H2c values in the subsequent vowel than plain voiced obstruents.

In the male data, H1-H2c values after plain and breathy voiced consonants differ significantly overall for both sonorants ( $F(1,4) = 63.99, p = 0.001$ ) and voiced obstruents ( $F(1,4) = 227.9, p = 0.0001$ ), though the values appear to be more greatly differentiated after obstruents. Post hoc tests included a window-by-window analysis, in order to attempt to disentangle the Window interaction found in the omnibus Anova. Mean H1-H2c values in vowels after voiced obstruents differ significantly in three windows, with a marginal difference in one of the remaining window, while the sonorants differ significantly in two windows and are marginal in two windows. Test results appear in Table 10.

**Table 10** ANOVA results for H1-H2c of vowels following voiced segments by obstruency and window

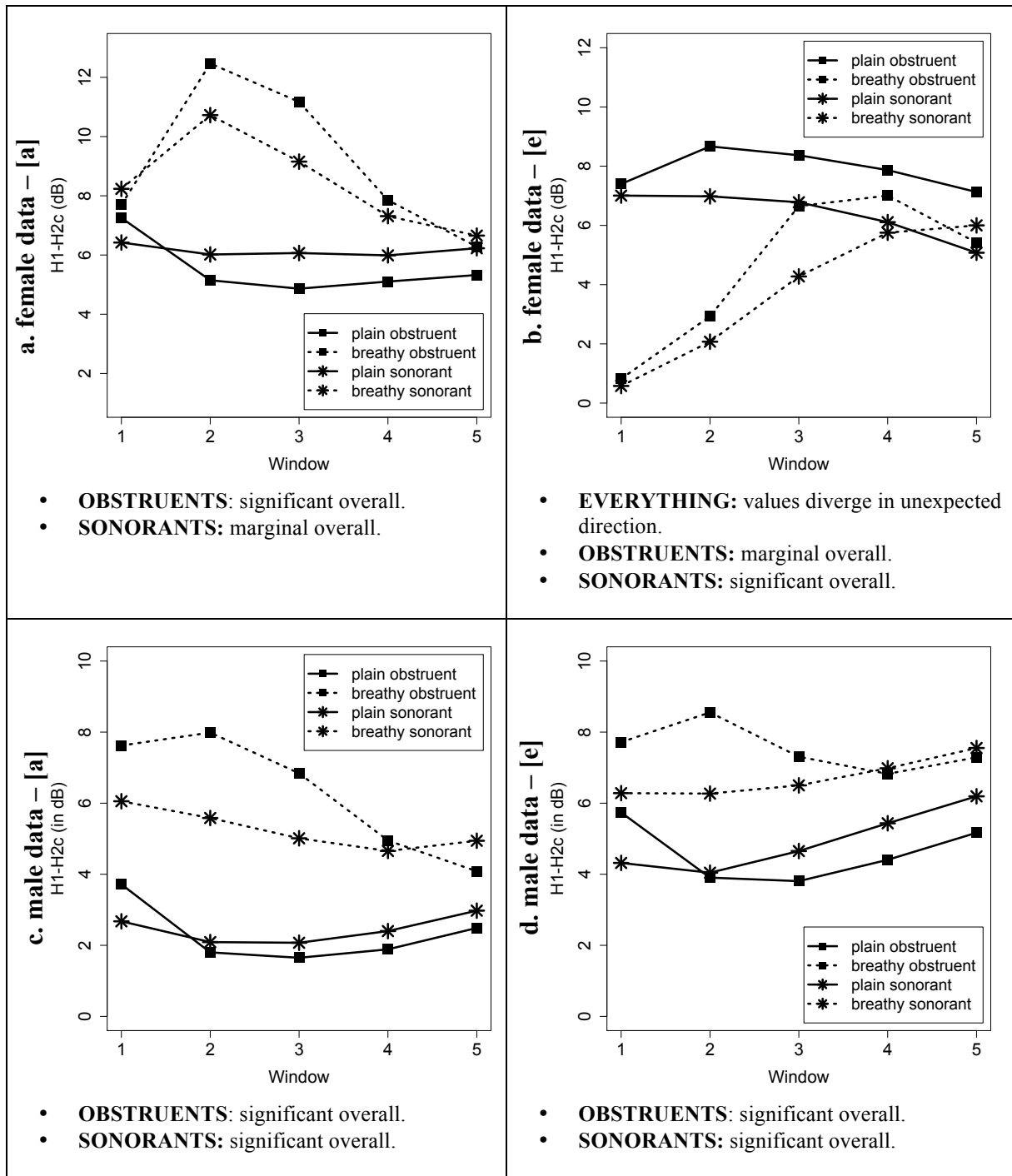
	Voiced Obstruents		Sonorants	
	$F(1,4)$	$p$	$F(1,4)$	$p$
<b>Window 1</b>	8.38	0.04	10.1	0.03
<b>Window 2</b>	57.46	0.002*	24.44	0.008
<b>Window 3</b>	79.94	0.0009*	30.31	0.005*
<b>Window 4</b>	18.14	0.0131	17.92	0.013
<b>Window 5</b>	41.14	0.003*	28.78	0.006*

Note that H1-H2c values in vowels after both sonorants and obstruents differ significantly based on phonation type overall: in other words, post-sonorant differences in H1-H2c do not attain significance in as many windows, but the differences are nonetheless great enough to trigger overall significance.

When designing this experiment, we considered it critical to include multiple vowel contexts. This was in part motivated by the fact that vowel formants can influence the harmonic

amplitudes that yield spectral measures such as H1-H2. The measures reported herein are corrected for this influence, but even with the correction the influence cannot be eliminated completely. Thus, in order to develop a complete picture of the spectral correlates of breathy phonation, it is important to investigate multiple vowel contexts. Indeed, the H1-H2c omnibus Anova reveals a Phonation type by Vowel context by Gender interaction ( $F(1,6) = 14.85, p = 0.008$ ), suggesting that this measure might not be consistent across distinct vowel contexts. When we observe the female data presented in Figure 32, it seems clear that H1-H2c values differ markedly from one vowel context to the next. Whereas we see the expected pattern in the [a] data in (a.)—with H1-H2c values higher after breathy sounds than after plain sounds—we see the reverse in (b). The data in (b.) are for the vowel [e], and these data are very surprising. Whether following a sonorant or a voiced obstruent, [e] after breathy consonants has lower mean H1-H2c values than [e] after plain voiced consonants, at least initially. In the male data, shown in (c.) and (d.), differences based on phonation type are significant overall—in [a] and in [e], after both sonorants and voiced obstruents.

Figure 32 H1-H2c across [a] and [e] by phonation type and obstruency



These data, in conjunction with the results reported in the F0 chapter, indicate that females are producing breathy phonation quite differently than males, such that the expected F0 and low-



frequency measures diverge from the expected pattern. This is an unexpected finding, and will need to be addressed in the coming chapters. To my knowledge, this is the first study to reveal systematic, sex-specific strategies for cueing breathy phonation in sonorants. The assumption is that these differences are due to physiological differences in the larynx and vocal tract that arise from sexual dimorphism. More commentary appears in the Discussion section, after the remainder of the results have been presented.

Post hoc analyses reveal that the H1-H2c differences in [a] are significant after voiced obstruents and marginal after sonorants. Differences in [e] are significant after sonorants. While the differences in [a] are in the expected direction, those in [e] are once again opposite the predicted pattern. The relevant test results appear in Table 11, along with results for the male data, which is affected by neither vowel context nor obstruency. For the males, H1-H2c values in [a] and [e] differ significantly based on phonation type—with breathy values higher than plain values—for both sonorants and obstruents.

**Table 11** ANOVA results for H1-H2c of [a] and [e] by obstruency, gender, and window

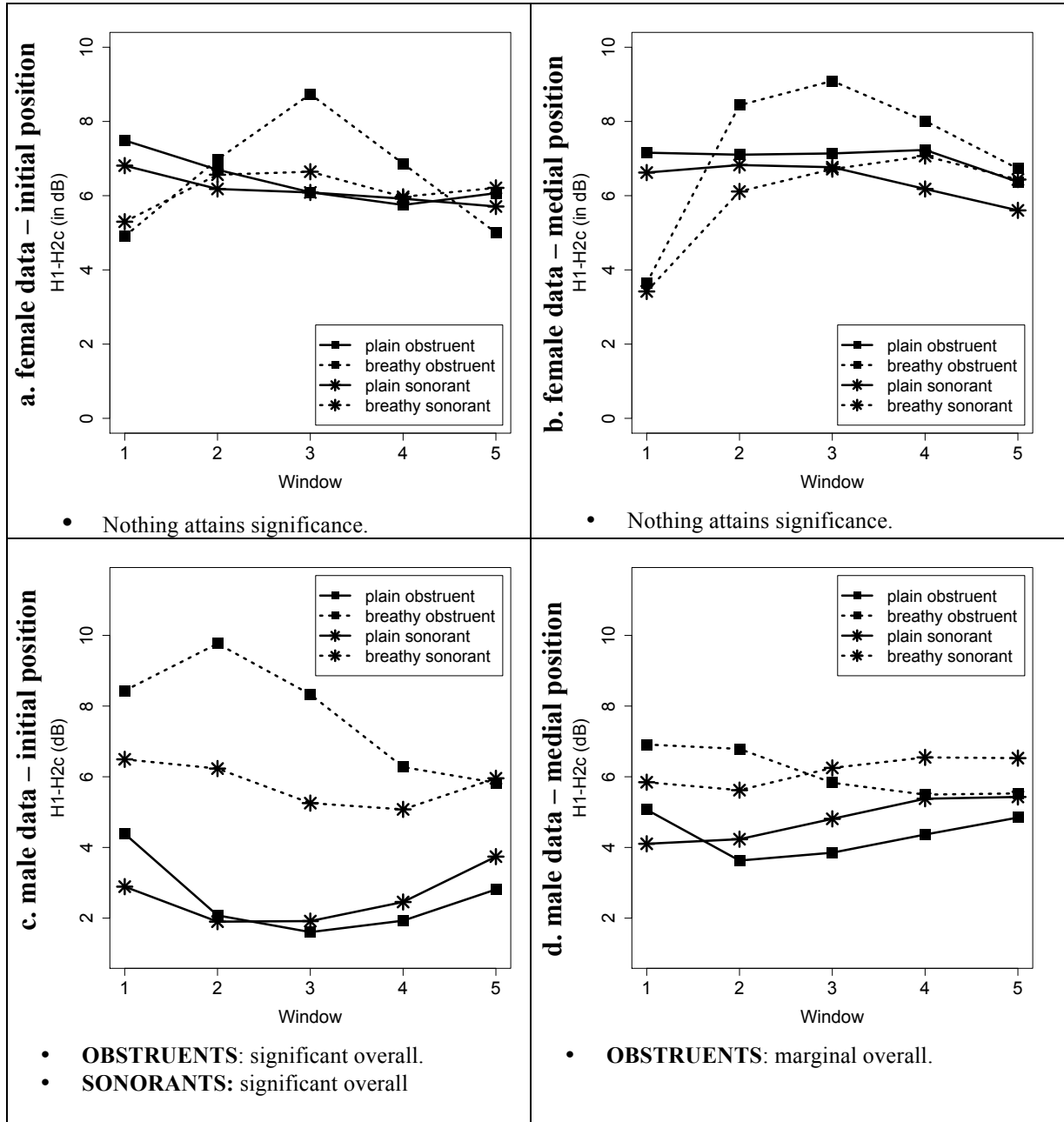
	Female Data		Male	
	<i>F</i> (1,4)	<i>p</i>	<i>F</i> (1,4)	<i>p</i>
<b>[a] after voiced obstruents</b>	25.61	0.007*	147.4	<0.001*
<b>[e] after voiced obstruents</b>	21.41	0.01	124	<0.001*
<b>[a] after sonorants</b>	8.928	0.04	48.02	0.002*
<b>[e] after sonorants</b>	25.47	0.007*	44.71	0.002*

The overall results presented in Table 10 and the vowel-specific results presented in Table 11 highlight the relevance of looking at both an overall measure and a window-by-window measure. In natural language, after all, we have access to the overall spectral composition of a

speech sound and to the way sounds change over time. A significant difference in H1-H2c at a particular time point in the vowel may be highly relevant in terms of distinguishing one sound from another; and too, lesser differences that add up to a significant overall effect may also be informative.

The final interaction to present has to do with word position, and this is a marginal Phonation type by Obstruency by Word position by Gender interaction ( $F(1,6) = 5.54, p = 0.06$ ). H1-H2c has more consistently reflected differences in phonation type for the male than for the female speakers, and we see the same here. When divided according to word position, nothing attains significance in the female data; for the males, however, the phonation type of the preceding consonant triggers significantly different H1-H2c values overall—values are higher after breathy consonants than after plain consonants—for both voiced obstruents ( $F(1,4) = 397.6, p < 0.0001$ ) and sonorants ( $F(1,4) = 164.3, p < 0.0001$ ) in initial position. Differences after medial consonants are marginal for voiced obstruents ( $F(1,4) = 19.15, p = 0.012$ ), however, and do not attain significance for sonorants ( $F(1,4) = 8.9, p = 0.04$ ). These data are plotted in Figure 33. Window does not enter into the above interaction, and so a window-by-window analysis was not conducted.

Figure 33 H1-H2c across vowels after voiced segments by phonation type, obstruency, and word position



Taken together, the H1-H2c results highlight the importance of investigating multiple contexts. H1-H2c differences may well be informative, but the value of these differences varies along several parameters. This measure is more informative for [a] than for [e]. It is more

informative in word initial position than in word medial position. And, it is more informative for male than for female speakers. This mirrors the findings related to Tsonga nasals reported in Traill and Jackson (1987): in those data, H1-H2 proved an important variable in delineating plain from breathy nasals for male speakers, but not for female speakers. While data from additional languages must be assessed before making a claim about universal tendencies, the Tsonga findings considered in tandem with the Marathi data presented herein encourage us to hypothesize that H1-H2 may cue phonemic breathy phonation more consistently for male than for female speakers.

### **6.1.3 H1-H2c recap**

Breathy voiced sonorants are more rare than breathy voiced obstruents typologically, and as such previous investigation into the correlates of breathy phonation in sonorants is scant. Whether sonorants cue breathy phonation with segment-internal spectral differences remains an open question, yet we can now say something definitive at least with regards to the measure of H1-H2c. The consonant results reported in this section tell us that phonation type does not trigger strong H1-H2c differences segment-internally. We now see that, beyond this, obstruents trigger H1-H2c differences in more windows of subsequent vowels than sonorants do. In terms of spectral differences based on phonation type, then, obstruents may have a leg up on sonorants; the segment-external cues in the vowel are more extensive, and even though the sonorants could potentially contain segment-internal cues, in reality those cues are not highly significant.

However, we may also consider that while the spectral correlates of phonation type are perhaps not as strong in sonorants or in vowels following sonorants, the segment-internal and segment-external cues may be additive. This will be addressed more thoroughly in the discussion, but it is an interesting conundrum and is something to keep in mind as we proceed. After all, H1-

H2c is only one of the spectral measures investigated in the present study. We turn now to the next of these, H1-A1c.

## 6.2 H1-A1c

H1-A1 is a measure of the amplitude of the peak harmonic in the range of the first formant subtracted from the amplitude of the first harmonic. As such, this measure is vulnerable to influence from the first formant itself, and so a corrected measure is reported herein. This measure—referred to as H1-A1c—is generated by VoiceSauce, which utilizes an algorithm to correct for the effect of formant frequencies and bandwidths.

While H1-H2 is correlated with open quotient, H1-A1—along with H1-A2 and H1-A3—is correlated with the speed with which the vocal folds close during the glottal cycle. As compared with modal or creaky phonation, breathy voice is associated with less abrupt closure during the glottal cycle (Esposito 2006, Gordon & Ladefoged 2001, Huffman 1987), as well as with the presence of more noise in the signal and less energy in the higher frequencies (Pennington 2005, Stevens 1998). In looking at the spectra of vowels produced in modal and breathy phonatory states, Stevens (1998) reports that the spectrum representing breathy voice is similar to that of modal phonation in terms of the amplitude of the fundamental frequency, but that high-frequency<sup>21</sup> amplitudes in breathy phonation are lower by as much as 15 dB (p. 89-90). The expectation herein, then, is that breathy voice will be characterized by spectra dominated by the fundamental frequency and showing lower high-frequency amplitudes. Accordingly, H1-A1, H1-A2, and H1-A3 measures are expected to yield larger positive values for breathy than for modal phonation.

In previous research, H1-A1 has been shown to successfully distinguish plain from breathy vowels in Jalapa Mazatec (Blankenship 1997) and in Gujarati (Fischer-Jørgensen 1967). This was true for both [a] and [e] in Gujarati—the same vowels assessed in the present study.

---

<sup>21</sup> “High frequency” here refers to energy above 2000 Hz.

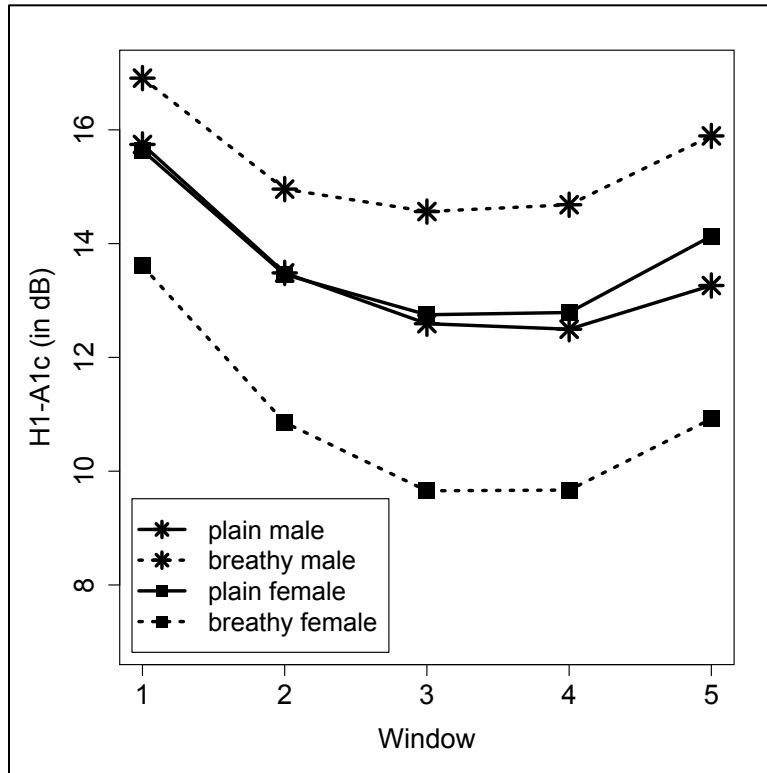
Before moving to the vowel results, however, let us begin with the results related to measures of H1-A1 in sonorant segments. Harris (2009) found very little difference in the spectral makeup of plain and breathy nasals and laterals in Sumi, particularly below 1000 Hz. If the same holds in Marathi, it is possible that H1-A1 will not reveal much differentiation in the sonorants themselves. Thus while phonation type is expected to trigger differences in H1-A1c values, we may find that these differences are greater in subsequent vowels than in the segments themselves.

### **6.2.1 H1-A1c – consonant results**

An omnibus repeated measures analysis of variance was conducted to determine whether consonant phonation type triggers significant differences in sonorant-internal H1-A1c values, as well as whether there were interactions between Phonation type and the remaining independent variables of Gender, Place of articulation, Vowel context, Word position, and Window. The results indicate that there is no overall effect of phonation type in the H1-A1c values of sonorant segments: values within plain and breathy sonorants do not differ overall.

There are several interactions of note, however, including a Phonation type by Gender interaction ( $F(1,7) = 7.3, p = 0.03$ ). As we can see in Figure 34, males and females differ in terms of the direction in which plain and breathy sonorants diverge. The pattern mirrors that which we saw for sonorant-internal H1-H2c results: males show greater H1-A1c values for breathy than for plain sonorants, and this is the direction in which values are expected to diverge. As with H1-H2c, however, the females pattern opposite what we predicted, with plain sonorants showing greater values overall.

Figure 34 H1-A1c values across sonorants by gender

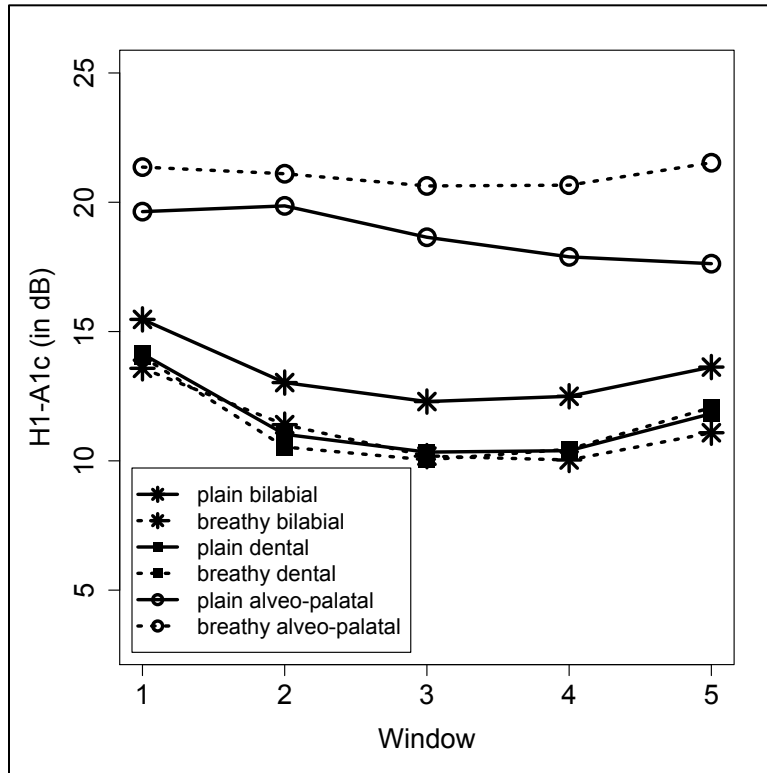


Post hoc analyses do not aid us in disentangling this interaction any further; neither male nor female values attain significance overall, and as there is no Phonation type by Gender by Window interaction we did not conduct a window-by-window analysis. The Phonation type by Gender interaction, then, is interpreted as simply reflecting that males pattern differently than females overall with regards to phonation type differences and H1-A1c.

The omnibus Anova also revealed a significant Phonation type by Place of articulation interaction ( $F(1,7) = 73.94, p < 0.0001$ ). The relevant data are shown in Figure 35. In looking at this figure, we can see that the alveo-palatals appear to diverge from the bilabial and dental sonorants in two ways. First, their H1-A1c values are higher overall; second, the breathy alveo-palatal values are higher than the plain alveo-palatal values, whereas the reverse pattern is seen in the other places of articulation.



Figure 35 H1-A1c of sonorants by phonation type and place of articulation



Again, however, post hoc tests do not reveal anything further. The difference in values for breathy and plain sonorants attains significance for none of the three places of articulation. The omnibus Anova revealed neither a Gender nor a Window interaction, so no additional analyses were conducted. The interaction, then, is interpreted as reflecting the fact that the alveo-palatals pattern differently than the bilabials and dentals. This is perhaps not so interesting in and of itself, for it does not necessarily aid us in understanding how H1-A1c differences within segments help distinguish plain from breathy sonorants. However, think back to the consonant duration and H1-H2c results: while the medial bilabial and dental sonorants showed marked durational differences based on breath, the alveo-palatals did not. Breathly alveo-palatal sonorants did show significant H1-H2c differences for male speakers, however, whereas the bilabials and dentals did not. The H1-A1c results show the same pattern numerically—it simply fails to attain statistical

significance. The message we can take from this is that while the alveo-palatals lack a durational cue that the other sonorants have, they differ with regards to the spectral pattern.

That said, it is important to note that no overall effect of Phonation type is revealed. While phonation type does interact with other factors, post hoc analyses by Gender and by Place of articulation reveal no overall significance in the difference between plain and breathy sonorants. While sonorants could hypothetically hold acoustic cues related to breath, H1-A1c seems relatively uninformative.

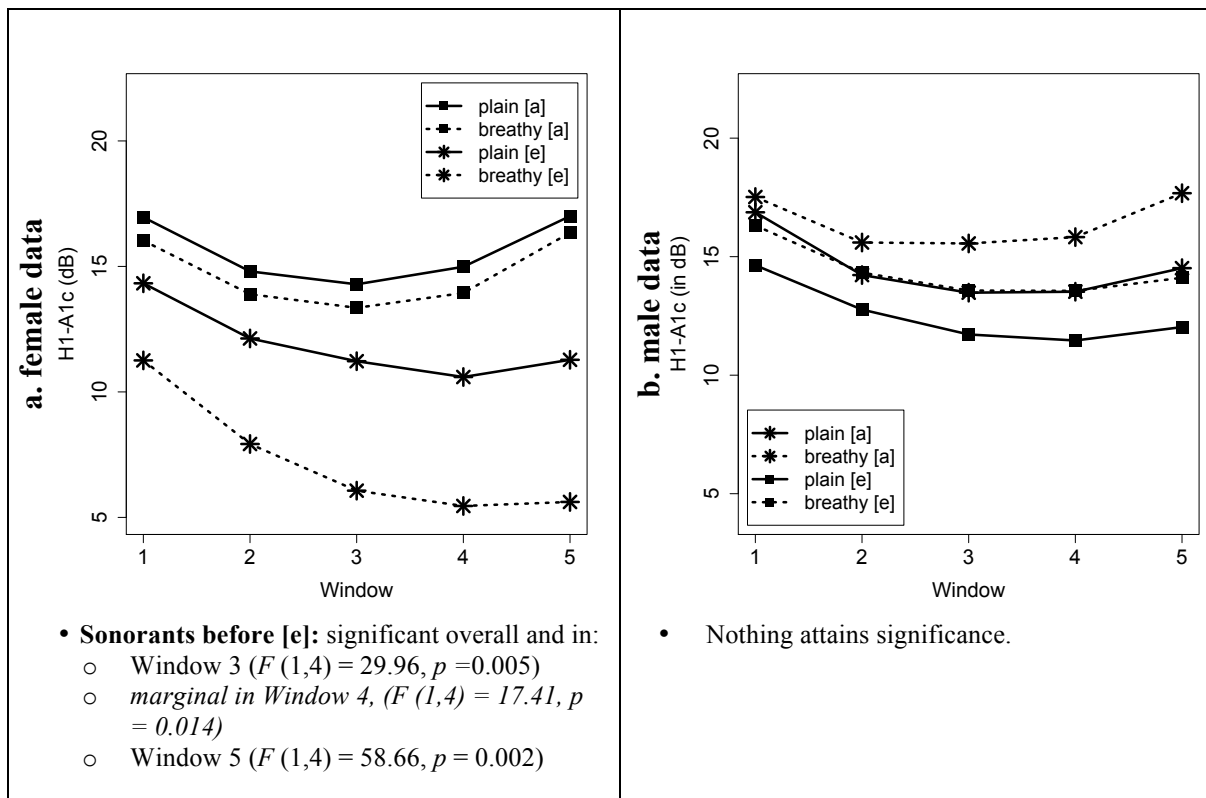
There is one final interaction to consider, however. Recall that data were collected in both the [a] and the [e] vowel contexts, in order to investigate how stable various acoustic cues were across distinct vowel contexts. The low vowel context is ideal when looking at spectral values, as the high first formant is unlikely to boost harmonic amplitudes and spectral measures are therefore less likely to be affected by formant structure. The [e] context has not been investigated quite as often, in part because, with a lower first formant, formant interference is more of a concern. In fact, what we see when we look at the H1-A1c data across the two vowel contexts suggests the validity of this concern. The omnibus Anova reveals a Phonation type by Vowel context by Gender by Window interaction ( $F(1,7) = 7.52, p = 0.03$ ).

Male and female data for the distinct vowel contexts appears in Figure 36. What is notable about these data is the female pattern in the [e] vowel context. Indeed, while the rest of the data look more or less like what we have seen before, H1-A1c values for females' breathy sonorants diverge widely from the plain values. Furthermore, the difference goes opposite the predicted direction.

The interaction, then, seems to arise because female data pattern very differently than male data, particularly in the [e] context. In fact, this is the only significant result revealed by the

post hoc tests. No significant differences emerge for the males, or for the females in the [a] context. For the females, however, H1-A1c values in plain sonorants are significantly different than those in breathy sonorants before [e] ( $F(1,4) = 167.4, p < 0.001$ ). The window-by-window analysis also revealed significant differences in Windows 3 and 5 and a marginal difference in Window 4. ( $F$  and  $p$  values appear below the graph.)

**Figure 36** H1-A1c values across sonorants by breath, vowel context, and gender



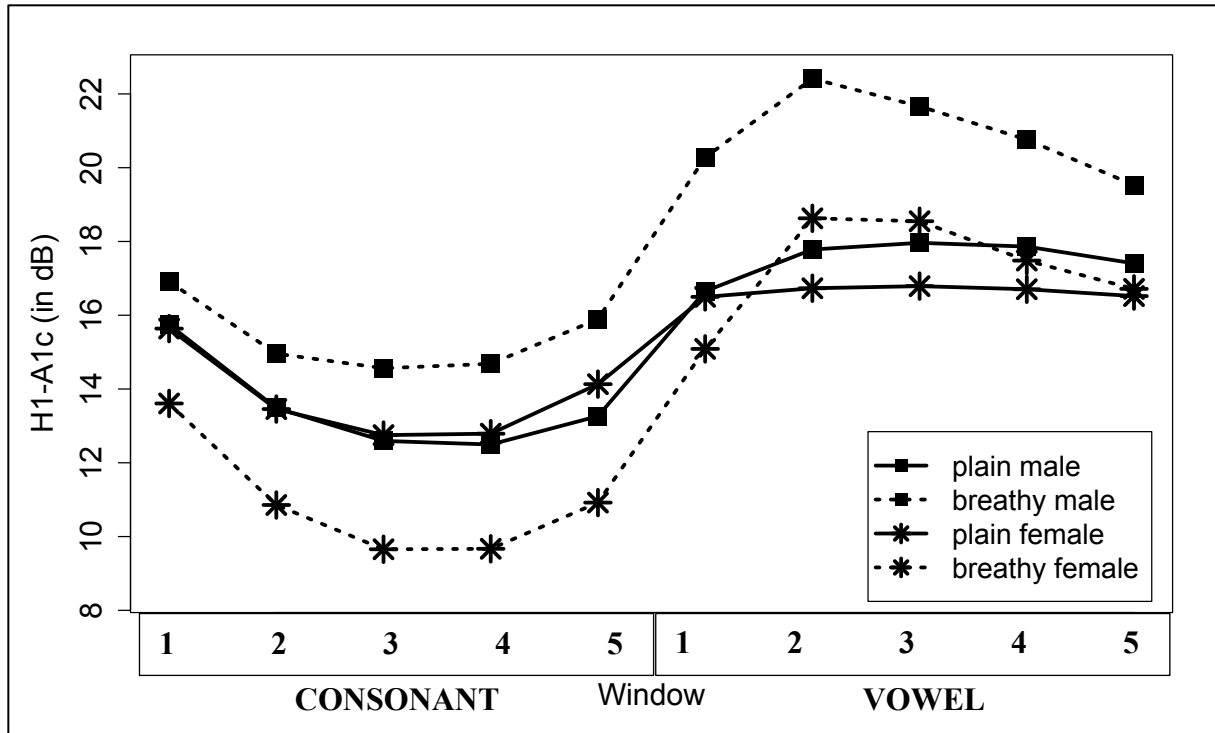
I believe that there are several ways to consider this. In the first interpretation, we can highlight the fact that H1-A1c does in fact reliably distinguish plain from breathy sonorant segments, albeit in a very limited way. For females, in the [e] context, this acoustic parameter distinguishes the two. However, the divergence is against the predicted direction. Breathly sounds

are expected to have higher H1-A1 values, but this is not in fact the case here, presumably because interference from F1 boosted the amplitude of the harmonic nearest F1.

We can also look at this from the other direction, however, and recognize that much of the time, H1-A1c differences based on breathiness simply do not emerge as significant. Instead we see interactions with gender, place, and vowel context, indicating that phonation type does in fact have some kind of effect on segment-internal H1-A1c values, but that the effect simply is not very strong. Thus while it is true that sonorants have the capacity to carry segment-internal spectral cues, they do not seem to do so in a reliable manner with regards to the lower-frequency spectral measures of H1-H2c and H1-A1c.

Before moving into the vowel data, a plot of H1-A1c across full sonorant-vowel syllables is provided. This plot allows us to see how the sonorant and vowel data fits together, and lets us envision how things change over the time course of the syllable. It is a good image to have in mind as we proceed.

Figure 37 H1-A1c across syllables by gender (sonorant data only)



We turn now to the question of whether consonant phonation type triggers reliable H1-A1c differences in subsequent vowels.

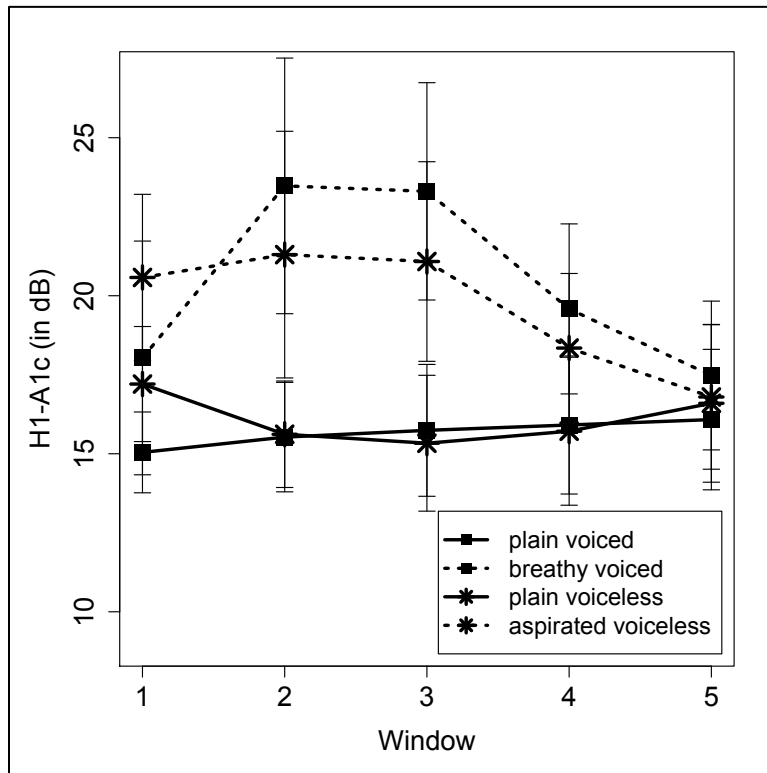
### 6.2.2 H1-A1c – vowel results<sup>22</sup>

One of the primary comparisons of interest in the present work is that of sonorants versus obstruents. As in the previous sections, this section compares vowels after sonorants with vowels after obstruents. Before presenting those data, however, let us consider the H1-A1c values in vowels after voiced and voiceless obstruents, presented in Figure 38. As we can see, it appears that voiced and voiceless obstruents trigger similar patterns in subsequent vowels: those following plain consonants have lower H1-A1c values overall than those following breathy

<sup>22</sup> Recall that the intervals reported as the PVI and as the vowel in the duration section are concatenated in the F0, spectral, and CPP results sections.

consonants. An omnibus repeated measures analysis of variance was conducted to determine whether consonant phonation type triggers significant differences in the H1-A1c values of following vowels; also of key interest here was the question of whether there were interactions between Phonation type and Voice within the post-obstruent data, and with the remaining independent variables of Gender, Place of articulation, Vowel context, Word position, and Window. The omnibus Anova revealed a Phonation type by Voice interaction ( $F(1,4) = 15.39, p = 0.017$ ), and so post hoc analyses were conducted. Results indicated that the effect of phonation type on vowel H1-A1c values is significant overall after both voiced obstruents ( $F(1,9) = 41.99, p = 0.0001$ ) and voiceless obstruents ( $F(1,9) = 28.47, p = 0.0005$ ), however. Since the interaction of Phonation type and Voice is not our primary concern, further disentanglement of this effect is left for future work..

**Figure 38 H1-A1c in vowels after obstruents by phonation type and voice**

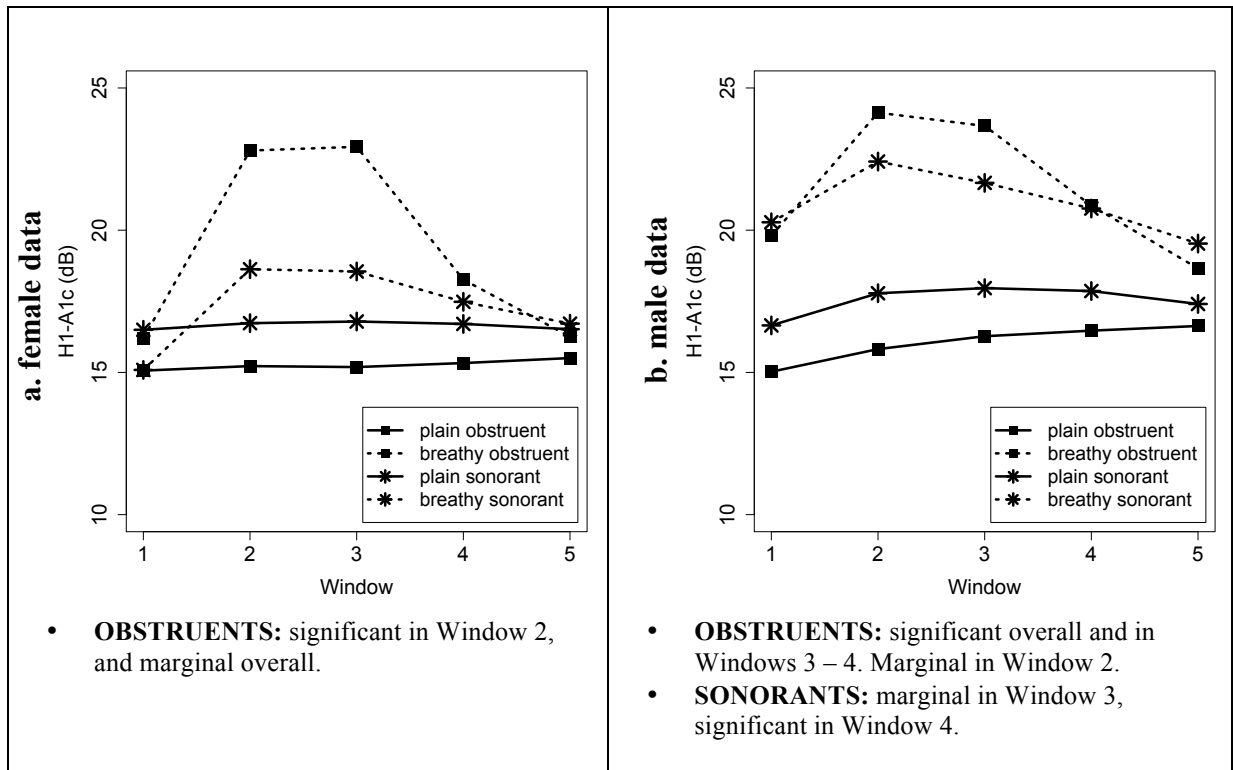


While the rest of the chapter will focus on vowels following sonorants and voiced obstruents, the take away message from Figure 38 is that breathy phonation in obstruents triggers significant differences in H1-A1c values—in the expected direction—in subsequent vowels. We turn now to a comparison of vowels after sonorants and voiced obstruents.

The H1-H2c results showed us that although sonorant segments could carry spectral information, segment-internal H1-H2c differences did not strongly distinguish plain from breathy sonorants. In addition, obstruents triggered greater differentiation in subsequent vowels than sonorants did. When we consider mean H1-A1c values plotted over the time course of vowels following sonorants and voiced obstruents, shown in Figure 39, we may anticipate that the same is true of H1-A1c as well. These data are separated by gender, because we will see that—again, as was the case with H1-H2c values—gender interacts with a number of other factors.

The data in Figure 39 indicate that, as with H1-H2c, H1-A1c values in vowels may differ more widely based on consonant phonation type after obstruents than after sonorants. This looks to be true for both male and female speakers, a point which was investigated statistically because the omnibus Anova revealed a Phonation type by Obstruency by Gender interaction ( $F(1,6) = 46.36, p = 0.0005$ ). Post hoc tests indicated a significant overall effect of consonant phonation type for males for vowels after obstruents ( $F(1,4) = 29.25, p = 0.006$ ) but not after sonorants ( $F(1,4) = 16.66, p = 0.015$ ). For females, consonant phonation type did not trigger significant overall differences in vowels after sonorants ( $F(1,4) = 1.35, p = 0.31$ ); the effect was marginal after obstruents, however ( $F(1,4) = 19.37, p = 0.011$ ).

Figure 39 H1-A1c in vowels after voiced consonants by obstruency and breath



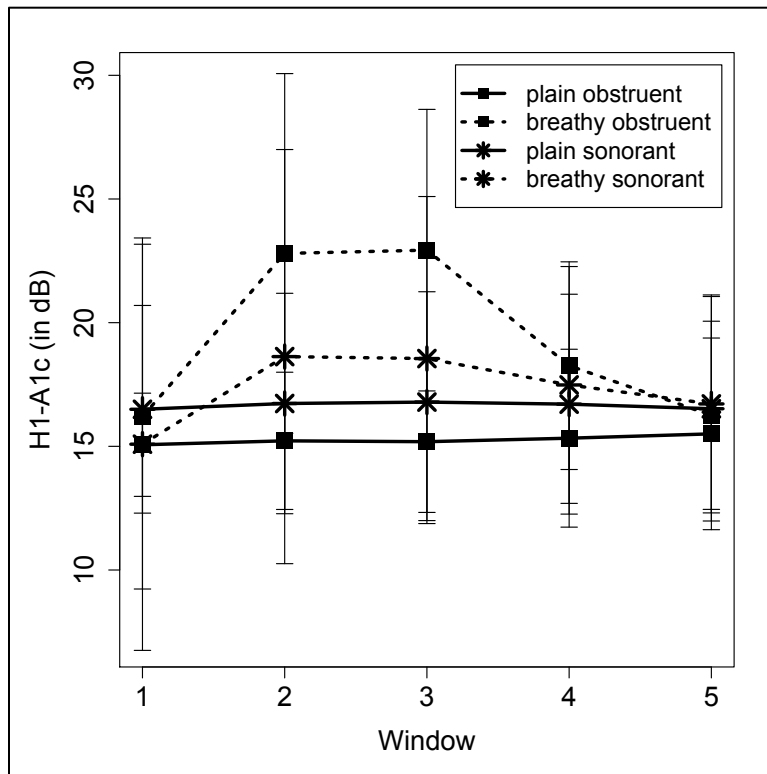
The omnibus Anova also revealed a Phonation type by Obstruency by Window interaction ( $F(1,6) = 8.59, p = 0.03$ ), which is discussed below.

A series of post hoc test was conducted to investigate the Window interaction. For females, values after obstruents differ significantly in Window 2 ( $F(1,4) = 32.48, p = 0.004$ ), but not enough to drive overall significance. For males, H1-A1c differences in vowels after obstruents are marginal in Window 2 ( $p = 0.011$ ) and significant in Windows 3 and 4 ( $p < 0.007$ ). Differences after sonorants are marginal in Window 3 ( $p = 0.01$ ) and significant in Windows 3 and 4, with  $p < 0.007$ . The distinction here, though, is that while those windows of significance after obstruents are enough to drive overall significance, the same is not true for sonorants. Again, then, the data suggests that obstruents cue phonation type in subsequent vowels more robustly than sonorants do.



There are two elements of these facts that bear discussion. First, in simply looking at the plots, it is not obvious impressionistically that the female post-obstruent values do not differ significantly. The magnitude of the differences seem in line with what is seen for males, where significance is attained. This raises the question of why, and it turns out that there is a tremendous amount of variation in the female data. This is illustrated in Figure 40, where error bars have been added to the plot from above.

**Figure 40** H1-A1c in vowels after voiced consonants by phonation type and obstruency – female data, with error bars



These error bars obscure much of the rest of the graph, and so they were omitted from the original plot; they are included here, however, because they provide insight into the fact that so few points of significance arise despite the fact that post-obstruent values seem to diverge based on consonant phonation type in a manner akin to the pattern shown in the male data.

The second point worth discussing is that male values attain significant differences based on phonation type not in the initial portion of the vowel, but in windows 3 and 4. This is unexpected: if consonant phonation type triggers differences in subsequent vowels, we might expect those differences to be realized in the early portion of the vowel, closest to the consonant. In fact, large standard deviations could serve as one possible explanation for both of these points—lack of significance for females despite apparently large differences in mean values, and significance for males in later windows of the vowel rather than immediately adjacent to the preceding consonant. The first two windows of the vowel are where the breathiness—the PVI—is located, and is therefore contributing to the overall means. It could be the case, then, that the early windows are subject to more variability as a result. To provide a sense of the variation under discussion, standard deviations are presented in Table 12.

**Table 12** Standard deviation of H1-A1c values (in dB) in vowels by gender, obstruency, phonation type, and window

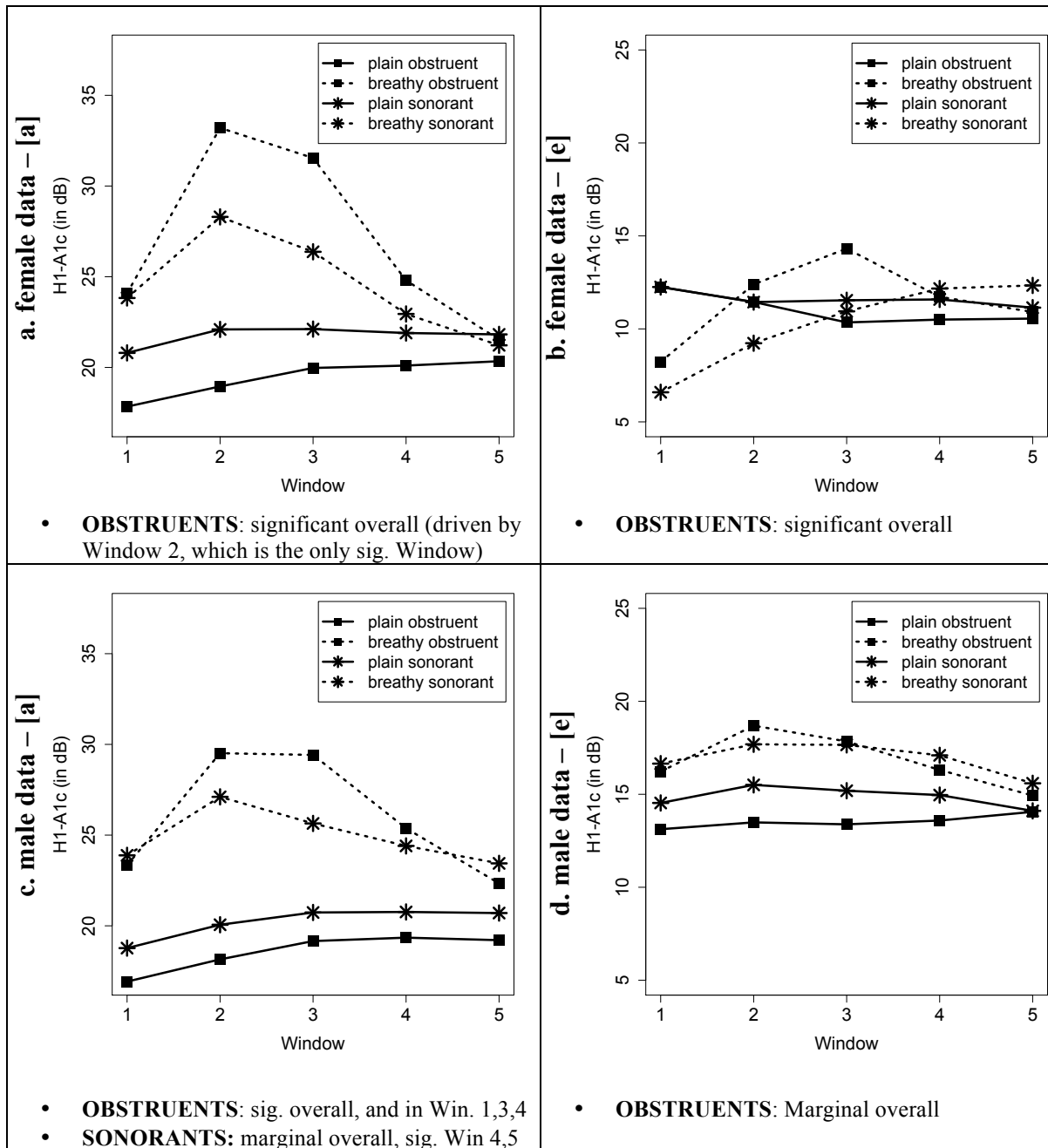
	Female Data					Male Data				
	WINDOW					WINDOW				
	1	2	3	4	5	1	2	3	4	5
<b>Breathy obstruent</b>	13.9	14.5	11.4	8.4	7.6	7.0	9.2	9.2	7.5	6.3
<b>Plain obstruent</b>	4.2	5.5	6.6	7.2	7.7	3.5	4.8	5.9	5.8	5.4
<b>Breathy sonorant</b>	16.7	16.7	13.1	9.6	8.8	8.1	9.1	7.7	6.9	6.4
<b>Plain sonorant</b>	8.4	8.9	8.9	8.9	9.1	4.2	5.3	5.6	5.6	6.0

Here, note that breathy sonorants and breathy voiced obstruents in the female data exhibit large standard deviations in the first three windows—these being the same windows wherein mean values seem to diverge greatly based on consonant phonation type yet do not attain statistical significance. The story is the same in the male data, with the exception of Window 1 for the

breathy obstruents. Though the absolute numbers are not as great as those in the female columns, standard deviations are greater in the earlier windows, where the PVI is found. For the time being, then, it remains plausible that we see significant differences emerging in the latter portions of the vowel for male speakers because those windows are subject to less variation.

The omnibus Anova also revealed a Phonation type by Obstruency by Vowel context by Gender by Window interaction ( $F(1,6) = 6.84, p = 0.04$ ), and this interaction seems fairly straightforward when looking at the pattern in [a] and comparing it with the pattern in [e], which can be seen in Figure 41. Female data appear in (a.) and (b.), and male data in (c.) and (d.). Differences in H1-A1c values based on phonation type appear to be greater in [a] than in [e], and values seem to diverge more after obstruents than after sonorants.

Figure 41 H1-A1c in [a] and [e] after voiced consonants by phonation type, obstruency, and gender



The observations made above are partially supported by the statistics: [a] shows a significant effect of Phonation type overall after voiced obstruents for both males and females, with  $p < 0.007$ . For males, differences in [e] after voiced obstruents and in [a] after sonorants are

marginal overall, with  $p < 0.014$ . Window interacts with these variables as well, and so a window-by-window analysis was conducted.

For the female speakers, differences in H1-A1c values in [a] were significantly higher after breathy than plain obstruents in Window 2, with  $F(1,4) = 100.5$  and  $p = 0.0005$ . Nothing else, in any other window of either vowel, attained significance. For the male speakers, H1-A1c values in [a] were significantly higher after breathy sounds than after their plain counterparts in Windows 1, 3, and 4 for obstruents, and in Windows 4 and 5 for sonorants, with  $p < 0.007$ . H1-A1c values in [e] were significantly higher after breathy sounds than after their plain counterparts in Windows 3 and 4 for obstruents and Window 5 for sonorants, with  $p < 0.007$ .

The female pattern is perhaps in line with what we would expect—i.e., we see H1-A1c as an acoustic correlate that gives information about consonant phonation type behaving more reliably in the low vowel context,<sup>23</sup> and the difference manifests in the initial portion of the vowel, meaning the portion closest to the consonant that is triggering the differences. The lack of significant differences after sonorants is perhaps not unexpected, either—we are beginning to see a clear pattern of obstruents triggering greater differences in vowels and sonorants triggering differences that are marginal or not significant.

The male data again requires some thought, however. We see significant differences based on phonation type for both vowel contexts after both sonorants and obstruents, but differences become significant in the latter portions of the vowel. It is odd that spectral differences appear not in the initial portions of the vowel, closest to the consonant presumably triggering the changes, but at a point more temporally removed from the consonant. Again, however, one possible explanation for finding significant differences later in the vowel is that the

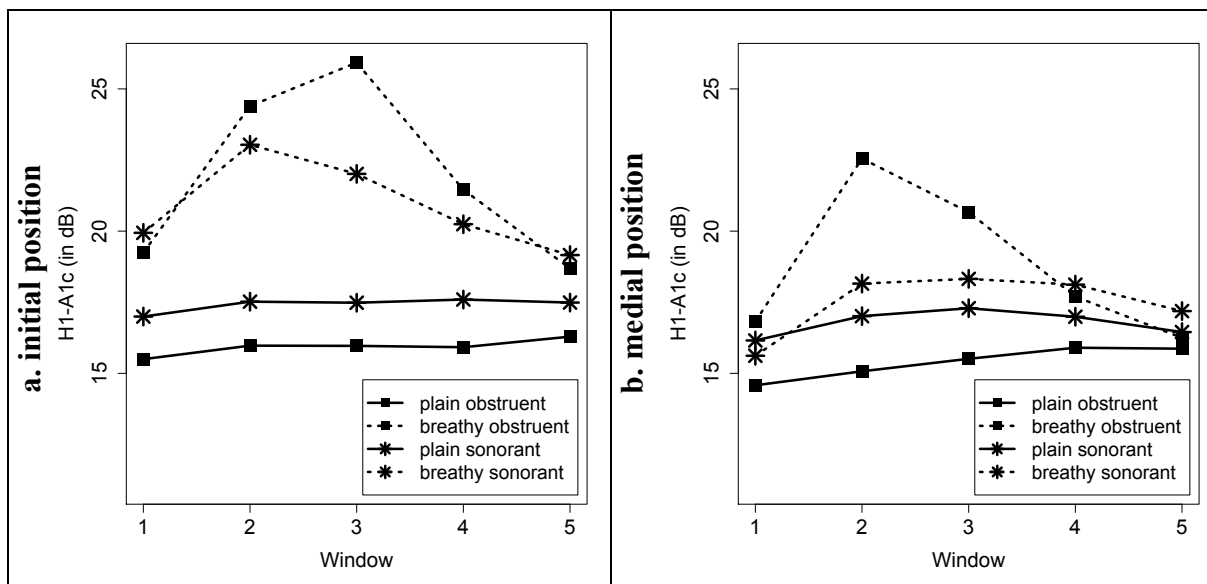
---

<sup>23</sup> In addition to this, recall that differences in sonorants before [e] did differ significantly, but that the difference was not in the expected direction.

breathy interval is highly variable—thus, even though the mean H1-A1c values diverge more numerically in the early windows, the variance is higher as well. This seems like a valid possibility, given the standard deviations noted in Table 12.

Finally, let us consider word position. The omnibus Anova reveals a Phonation type by Obstruency by Word position by Window interaction ( $F(1,6) = 6.36, p = 0.05$ ). Plots for the relevant H1-A1c values in vowels following voiced obstruents and sonorants appear in Figure 42. Note that Gender did not enter into the interaction, and the data are not separated by Gender.

**Figure 42 H1- A1c after voiced consonants by breath, obstruency, and word position**



Vowels following consonants in word initial position appear in (a.), and those after medial consonants appear in (b.). Impressionistically, differences in the vowel triggered by consonant phonation type seem to diverge more sharply after consonants in initial position than in medial position, particularly with regard to post-sonorant values. Values appear to be much higher after breathy sonorants in initial position than after breathy sonorants in medial position, and this is indeed what is shown by the post hoc tests. While H1-A1c values in vowels following breathy obstruents are significantly higher overall than those following plain obstruents in both

word positions, with  $p < 0.007$ , differences based on phonation type after sonorants are significant after initial consonants ( $F(1,9) = 18.76, p = 0.002$ ), but not after medial consonants ( $F(1,9) = 2.08, p = 0.18$ ).

As Window interacted with these factors as well, a window-by-window analysis was conducted. This revealed that values after obstruents differ based on phonation type in all windows in word initial position, and in Windows 1 – 3 in word medial position ( $p < 0.007$ ). After sonorants in initial position, differences based on phonation type are significant in Windows 1 – 4. After sonorants in medial position, however, differences are not significant in any window.

Once more, then, we see obstruents triggering greater and more sustained differences than sonorants. In particular, sonorants are more sensitive to word position than obstruents: while obstruents trigger differences in H1-A1c values in subsequent vowels in both word-initial and word-medial positions, sonorants trigger differences only when they occur word-initially. The difference is lost after sonorants in word-medial position. One point worth making briefly here, however, is the fact that languages which contain breathy voiced sonorants do not seem to feature positional restrictions that align with the acoustic picture we are developing: in other words, while word-medial position may seem a weaker position based on the acoustic results reported thus far, breathy voiced sonorants do in fact surface word-medially. This is true in Marathi—consider [n<sup>h</sup>avi] ‘barber’ and [pun<sup>h</sup>a] ‘again’, for instance. It is also true in several Tibeto-Burman languages which feature breathy voiced sonorants both word-initially and word-medially. In Sumi we find words such as [n<sup>h</sup>a] ‘close (dish)’ and [an<sup>h</sup>a] ‘mucus’ (Harris 2009: 82), and in Camling we find words like [l<sup>h</sup>amma] ‘catch’, [lamma] ‘look for’, and [lamma<sup>h</sup>a] ‘dump’ (Ebert 2003: 506). The idea, then, is that while these contrasts may be better cued

initially, the phonological component of the grammar overcomes, or ignores, the phonetic variability. How?

One thing we may turn our attention to are the durational results presented previously. Sonorants in initial position do not differ based on phonation type, but word-medial sonorants are significantly longer than word-medial plain sonorants. Like the alveo-palatals, then—which lack a durational cue that the bilabials and dentals have but have spectral cues that the bilabials and dentals lack—this appears to be another example of a trading relationship. Word-initially, sonorants cue breathiness not temporally but via spectral cues encoded in the subsequent vowel. Word-medially, on the other hand, temporal cues are present and the spectral cues are absent. Stepping back from the specific data, the picture is rather elegant. Phonation type cues are present both word-initially and word-medially, but the cues trade off from one context to the next. Where one cue is weak or absent, another is present. The phonetic picture, then, is characterized by a tremendous amount of variability, but there are a vast number of acoustic parameters which can contribute to making a phonemic contrast.

### **6.2.3 H1-A1c recap**

The H1-A1c data not only falls in line with but actually strengthens the pattern that is beginning to emerge. While sonorants could potentially contain segment-internal spectral differences based on phonation type, these differences are minimal at best. Furthermore, vowels are more affected by consonant phonation type after obstruents than after sonorants. Another way to conceive of this is to say that obstruent phonation type triggers greater acoustic differentiation than sonorant phonation type. Furthermore, while H1-A1c differences in subsequent vowels are affected by factors such as speaker gender, vowel quality, and word position, we can say the following: like H1-H2c, H1-A1c is a more reliable correlate of



breathiness for male speakers than for female speakers, and it is more reliably affected by obstruents than by sonorants.

It is true that H1-A1c is reported to reliably distinguish plain from breathy phonation in vowels in a few languages—namely, in Jalapa Mazatec (Blankenship 1997) and in Gujarati (Fischer-Jørgensen 1967). Dutta (2007b) also reports that H1-A1c differences in vowels based on the phonation type of the preceding consonant in Hindi sometimes persevere for 50, 70, or even 90% of the vowel. Dutta also reports a great deal of variation between speakers, however, and H1-A1c is not reliable for all five of the speakers whose data was analyzed. Thus it is perhaps not surprising that we find differences based on phonation type manifesting in vowel H1-A1c measures, but that these differences are subject to variability based on the factors outlined above.

One factor to keep in mind when interpreting these results is that H1-A1c—like H1-H2c—is still a relatively low-frequency spectral measure, in that we are looking at the amplitude of the fundamental frequency and the amplitude of the harmonic nearest to F1, which varies from [a] to [e] but is certainly well below 1000 Hz in both cases. Recall that Harris (2009) found a great deal of similarity in the spectral profile of plain and breathy Sumi nasals and laterals below 1000 Hz. For these reasons, the fact that we found segment-internal H1-A1c differences to be minimal at best is perhaps to be expected. As we move into H1-A2c and H1-A3c, however, we will be moving above that 1000 Hz mark in the spectrum. As such, we may therefore expect these higher-frequency spectral measures to be more informative.

### 6.3 H1-A2c

H1-A2 is a measure of the amplitude of the harmonic exciting the second formant subtracted from the amplitude of the first harmonic. Several prior studies have shown this measure to effectively distinguish breathy phonation from modal voice, particularly in vowels. In Jalapa Mazatec, for instance, vowels show a three-way phonation type contrast—between plain, creaky, and breathy—and Blankenship (1997) reports that H1-A2 reliably distinguished breathy vowels from the other phonation types in 12 out of 14 samples. Esposito (2006) reports the results of a discriminant analysis of modal and breathy vowels from Mazatec. In this study, 48 tokens (16 from each of three male speakers) were normalized so that each had a duration of 250 ms and a normalized F0 contour falling from 115 to 110 Hz. Eight acoustic parameters of the manipulated tokens were measured, after which discriminant analysis revealed that H1-A2 differences alone accounted for 53% of the variance. This seems to indicate that H1-A2 is a very strong cue for breathiness.<sup>24</sup> In a similar analysis for tokens from ten<sup>25</sup> different languages and dialects, however—2 modal and 2 breathy vowel tokens from each language, for a total of 40 tokens—Esposito reports that H1-A2 accounted for only 10% of the variance, while CPP accounted for 46% of the variance and H1-H2 for 27% of the variance.

Given the fact that prior studies reveal crosslinguistic variation in the degree to which H1-A2 is informative in cueing phonation type differences, then, the question of whether H1-A2 will prove to be a strong cue for breathy phonation in Marathi is wide open. One indication that

---

<sup>24</sup> Note that these tokens were then used in a perception task. Participants were speakers of Gujarati (which has a phonemic breathy/modal vowel distinction) and Spanish and American English (which do not). Surprisingly, none of the groups of speakers used H1-A2 as a cue for breathiness in the perception task. This highlights an important fact: that something reliably distinguishes a phonation type contrast—or any contrast—acoustically does not mean that speakers rely heavily on that cue when it comes to perception. Critically, however, the more we know about the acoustic differences that are present in the signal, the better we can refine our hypotheses about the cues speakers are using to distinguish contrasts and the more knowledgeably we can design experiments to test these hypotheses.

<sup>25</sup> Chong, Fuzhou, Green Hmong, White Hmong, Mon, Santa Ana del Valle Zapotec, San Lucas Quiavini Zapotec, Tlacolula de Matamoros Zapotec, Tamang, and !Xóǀ.

it may prove informative, however, comes from the analysis of Sumi nasals and laterals presented in Harris (2009). As has been mentioned previously, one cautionary note to bear in mind is that this study involved data from a single female speaker of Sumi. It is still valuable, however, particularly given the relative dearth of acoustic information related to phonation type contrasts in sonorants. Harris found that phonation type triggered differences in the spectral make-up of sonorants above 1000 Hz, while below 1000 Hz, breathy and modal consonants were highly similar. H1-A2 was not one of the measures Harris reported, but the parallel can nonetheless be drawn: we are now dealing with a higher-frequency measure, and so it may be the case that H1-A2 will reflect segment-internal phonation-type differences in Marathi sonorants more reliably than the lower-frequency measures of H1-H2c and H1-A1c. If the Marathi data herein patterns like Harris's Sumi data, we may also expect that this measure will prove useful in cueing phonation type for the female speakers in addition to the male speakers.

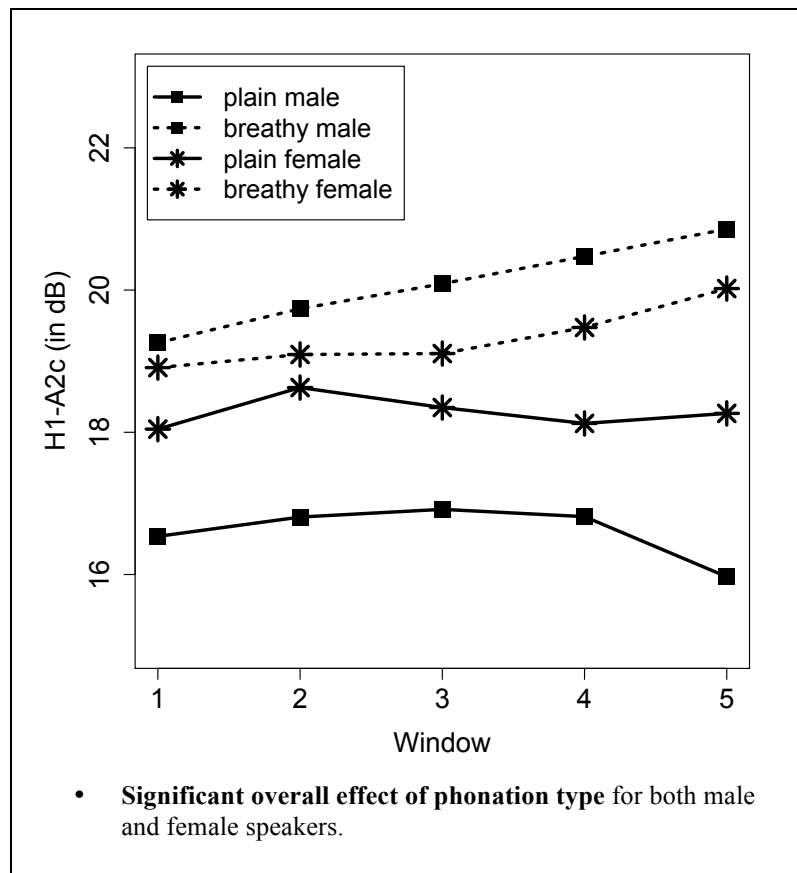
Recall that breathy phonation is associated with a steeper drop-off in high frequency energy than modal phonation, while the amplitude of the first harmonic is expected to be as strong or stronger than in modal phonation. As such, the prediction is that breathy voiced sounds will be associated with greater positive H1-A2c values than modal voice. We turn now to the consonant results to investigate whether the expected differences are present sonorant-internally, and to determine whether female and male speakers pattern more similar in this measure than they did in H1-H2c and H1-A1c.

### **6.3.1 H1-A2c – consonant results**

In the previous section, we learned that sonorants do not show an overall effect of phonation type with regards to segment-internal H1-A1c values. Furthermore, females patterned

opposite the expected direction, with breathy voice triggering lowered sonorant-internal H1-A1c values. This is not the case for H1-A2c. The data in Figure 43 are plotted by phonation type and by gender. These look very different than the H1-A1c results: both males and females pattern in the expected direction. As predicted, plain sonorants have lower H1-A2c values than breathy sonorants.

**Figure 43 H1-A2c across sonorants by consonant phonation type and gender**



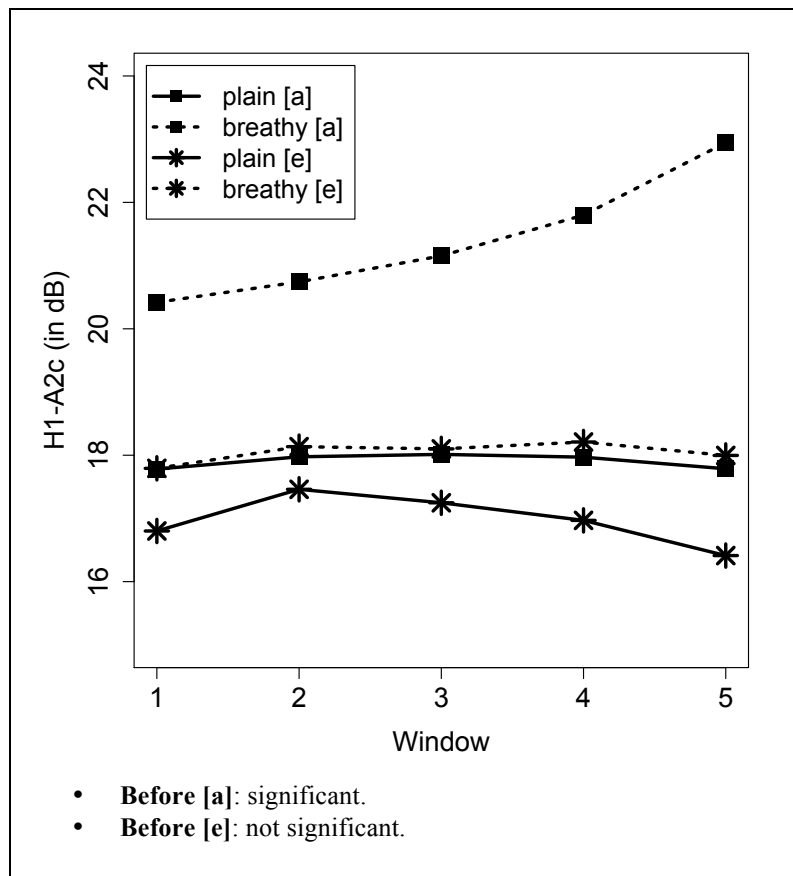
A repeated measures Anova was conducted to evaluate the effect of the independent variables (Phonation Type, Place of Articulation, Word Position, Vowel Context, Window, and Gender) on H1-A2c values in sonorants. Although impressionistically the data above may make it look as if males showed greater differentiation than females, the Anova revealed that this is not

in fact the case. Gender did not interact with any of the other variables. There was, however, an overall phonation type effect ( $F(1,7) = 11.44, p = 0.012$ ). The observation made when looking at the above plot is supported, then: breathy voiced sonorants have significantly greater segment-internal mean H1-A2c values than plain sonorants. This is quite interesting. The acoustic parameters reviewed in the preceding chapters often revealed Gender differences, with phonation type cued differently in male and female speech. H1-A2c, however—a higher frequency measure—shows no such gender effect.

Recall that plain Sumi nasals and laterals did not differ from breathy Sumi nasals and laterals below 1000 Hz (Harris 2009). This is not exactly what is shown in the present results: H1-H2c and H1-A1c, which are the lower-frequency measures and are dependent on spectral components below 1000 Hz, are often reliable acoustic cues for phonation type segment-internally for male speakers, but not for female speakers. Harris's results are for a single female speaker, however, and so it is not surprising that the female data herein align with the Sumi data. Furthermore, since males pattern with females with regards to this measure the indication is that while phonation type may be cued by the lower frequency spectral measures for male but not for female speakers, the higher frequency spectral measures are useful for both genders.

Gender is not the only variable for which we find no interaction emerging with regards to segment-internal H1-A2c values. There is no interaction with Word position, either. There is, however, a Phonation type by Vowel context interaction ( $F(1,7) = 6.38, p = 0.04$ ). The relevant data are plotted in Figure 44. Impressionistically, it looks as though the same thing we have seen elsewhere is true here as well: sonorants before [a] show greater differentiation than sonorants before [e].

Figure 44 H1-A2c across sonorants by consonant phonation type and vowel context



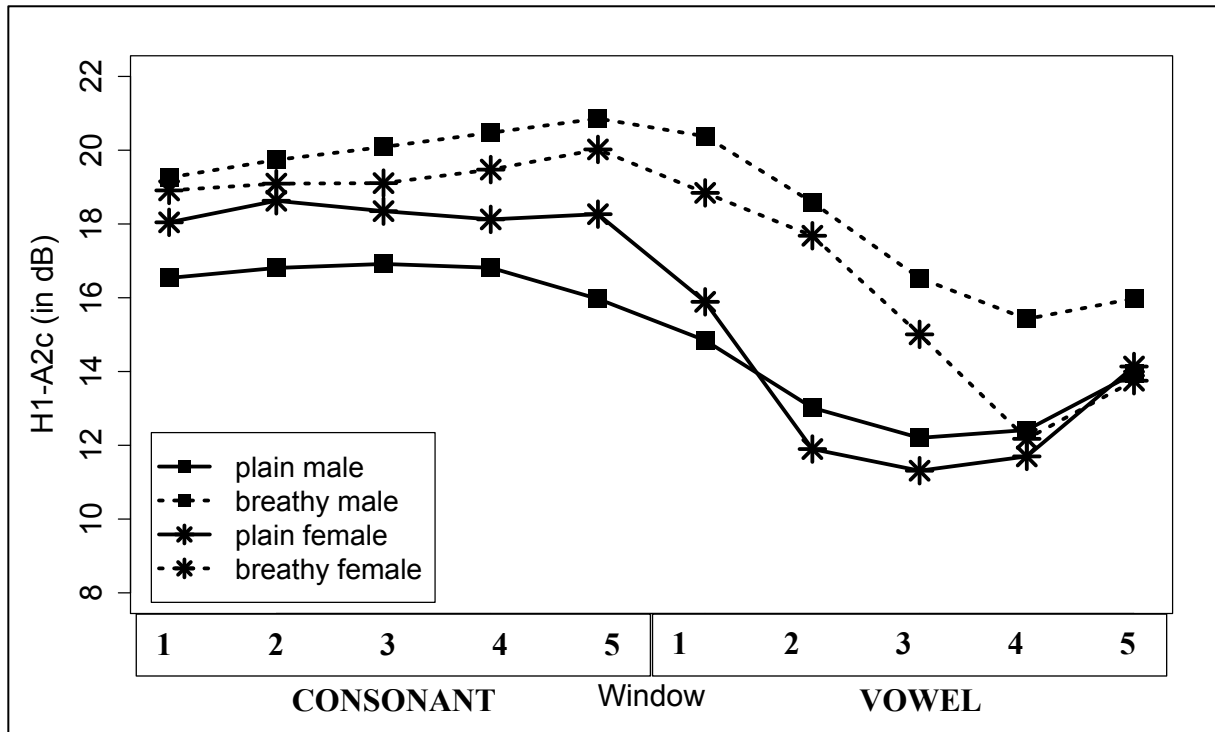
This observation is supported by the post hoc tests, which reveal that the difference between plain and breathy sonorants is significant overall before [a] ( $F(1,7) = 17.06, p = 0.003$ ). This is not present in the [e] context, however, where the effect of phonation type is not significant ( $F(1,7) = 1.45, p = 0.258$ ).

Recall that [a] is ideal for looking at spectral measures because of its formant structure, and that the acoustic parameters investigated in the present study have been assessed in some previous studies only in the [a] context (Blankenship 1997, Dutta 2007b). The findings herein indicate, however, that acoustic cues which are reliably present in low vowel contexts diminish or disappear in high vowel contexts—and this is not just vowel-internally, but also in preceding sonorants. That is not to say that the cues are not important, for of course they are present in low

vowel contexts; rather, it is important to note that additional or alternate cues must cue breathiness in high vowel contexts.

The omnibus Anova reveals no other interactions affecting H1-A2c values segment-internally, so we turn now to the vowel results. Again, before doing so a plot of H1-A2c values across full sonorant-vowel syllables is provided. This figure illustrates how the sonorant and vowel data fit together, and allows us to envision how things change over the time course of the syllable.

**Figure 45 H1-A2c across syllables by gender (sonorant data only)**



### 6.3.2 H1-A2c – vowel results<sup>26</sup>

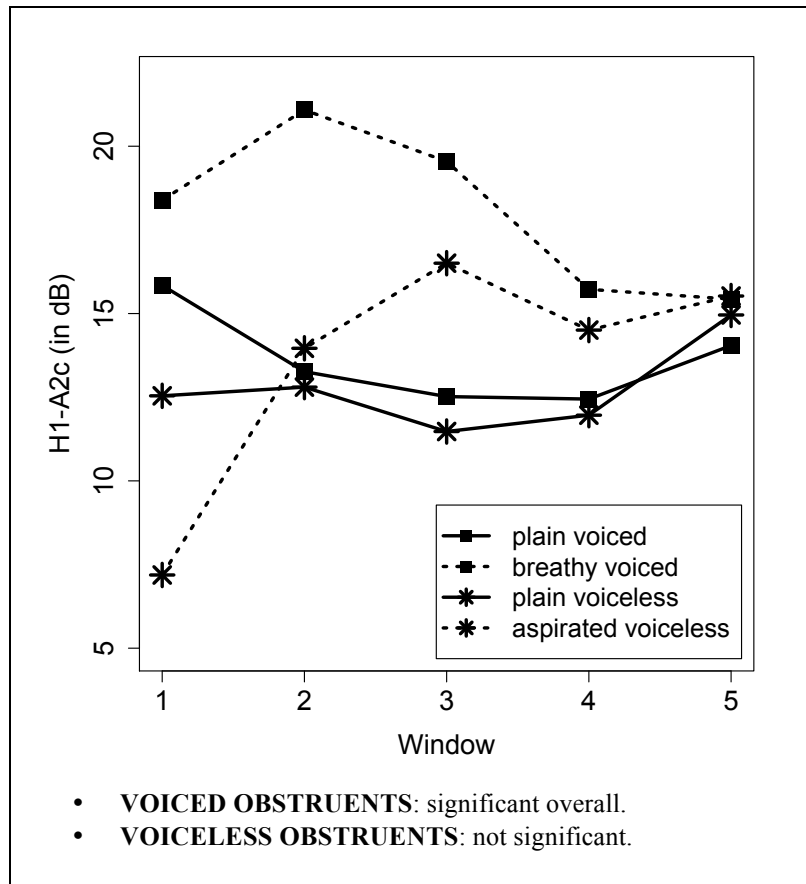
As with H1-A1c, consonant phonation type triggers a number of significant H1-A2c differences in subsequent vowels. First we consider vowels after obstruents, which appear in Figure 46. An omnibus Anova was conducted to investigate the effects of the Phonation type, Voice, Place of articulation, Vowel context, Word position, Window, and Gender on H1-A2c values in subsequent vowels. The test revealed a significant Phonation type by Voice interaction for vowels following obstruents ( $F(1,4) = 21.07, p = 0.01$ ). Based on this result, post hoc analyses were conducted and revealed that vowels after voiced obstruents differ significantly based on the phonation type of the preceding consonant ( $F(1,9) = 19.7, p < 0.001$ ). Vowels following breathy voiced obstruents have higher H1-A2c values overall than their plain counterparts. The same is not true for voiceless sounds—H1-A2c values in vowels following voiceless obstruents are not as greatly differentiated, and the difference does not attain overall significance.

---

<sup>26</sup> Recall that the intervals reported as the PVI and as the vowel in the duration section are concatenated in the F0, spectral, and CPP results sections.



Figure 46 H1-A2c across vowels after obstruents by consonant phonation type and voice



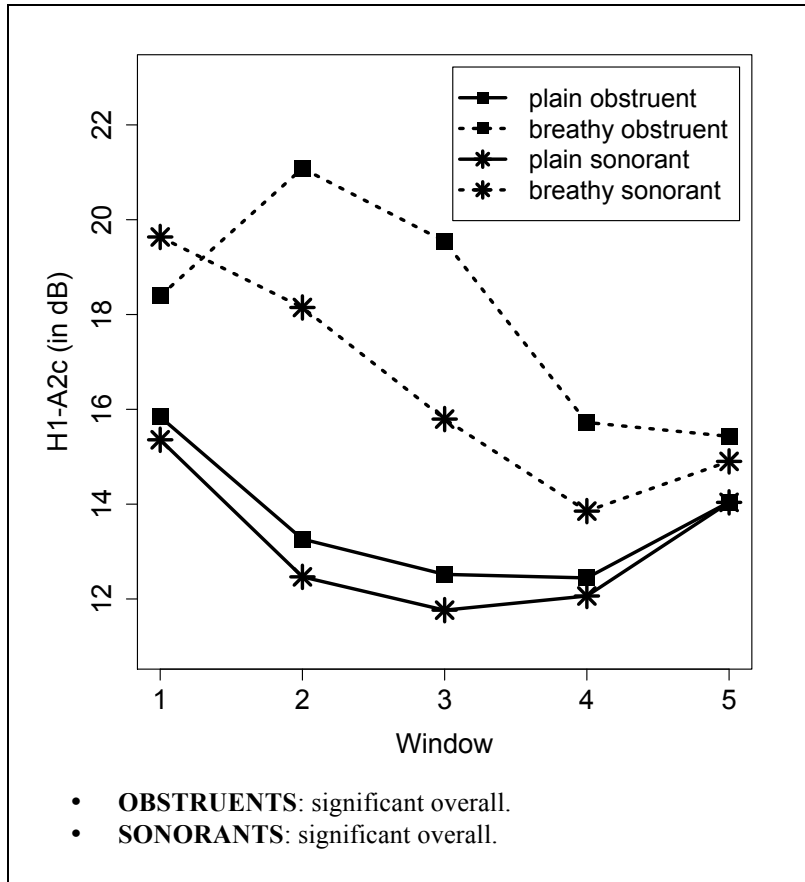
There is a Phonation type by Voice by Window interaction ( $F(1,4) = 361.92, p < 0.0001$ ), which potentially arises because differences based on phonation type reach significance in all windows of the voiced obstruents but not the voiceless obstruents. As this is not of key interest in the present study, however, it is not investigated any further.

Having established that the phonation type of voiced obstruents triggers significant differences in vowels, we turn now to a comparison of vowels following sonorants and voiced obstruents. We have seen that voicing and phonation type interact; we now ask whether obstruency and phonation type interact. The relevant data appear in Figure 47.

Impressionistically, it looks like the trend we have come to expect is once again in evidence:

breathy phonation triggers increased H1-A2c values, and obstruents show greater differentiation than sonorants.

Figure 47 H1-A2c across vowels after voiced segments by consonant phonation type and obstruency



The omnibus Anova confirms that there is an overall effect of Phonation type ( $F(1,6) = 13.75, p = 0.01$ ) as well as a Phonation type by Obstruency interaction ( $F(1,6) = 25.57, p = 0.002$ ).

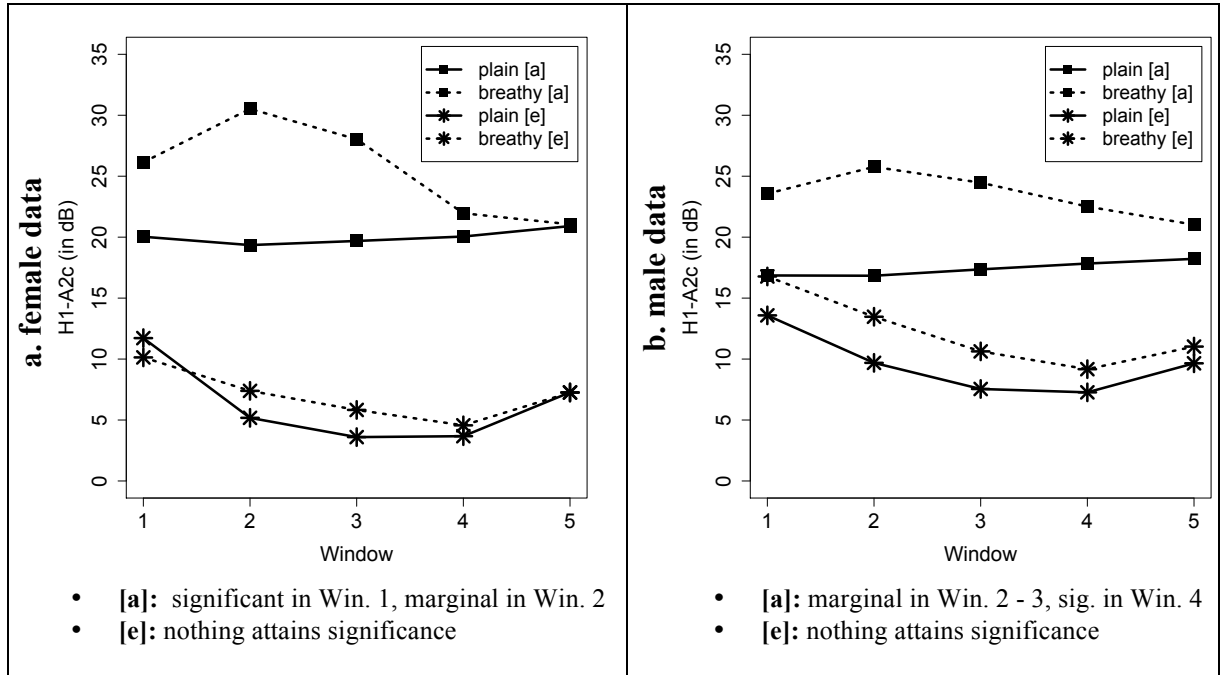
Interpreting the Phonation type by Obstruency interaction is not straightforward, however, because vowels following both sonorants and obstruents attain significant overall differences in H1-A2c values based on phonation type. Furthermore, no Phonation type by Window interaction arises, so a window-by-window analysis was not conducted.

One possibility is to interpret this interaction as reflecting the fact that while both sonorants and obstruents trigger differences in subsequent vowels based on phonation type, obstruents again trigger greater differentiation than sonorants. Another interesting point here is that there is no Phonation type by Obstruency by Gender interaction—meaning that, overall, gender does not trigger different patterning in vowel H1-A2c values. In other words, H1-A2c patterns similarly across both genders, in contrast with the lower-frequency measures of H1-H2c and H1-A1c.

Gender is involved in two higher order interactions, however—namely, those involving Vowel context and Word position. Let us address these one at a time. The omnibus Anova revealed a Phonation type by Vowel context by Window by Gender interaction ( $F(1,6) = 40.48$ ,  $p < 0.001$ ). Data are shown in Figure 48. In observing these data, it appears that H1-A2c values are higher overall in [a] than they are in [e]. This observation holds across genders. Furthermore, it appears as though consonant phonation type triggers greater differentiation in [a] than it does in [e], again regardless of gender.

Post hoc analyses were conducted in order to determine whether this interaction could be untangled, and in general the results are in line with the observations mentioned above. H1-A2c values in [a] are significantly greater than those in [e], and this is true for male speakers ( $F(1,4) = 85.55$ ,  $p < 0.001$ ) and for female speakers ( $F(1,4) = 153.4$ ,  $p < 0.001$ ).

**Figure 48 H1-A2c across [a] and [e] following voiced segments by consonant phonation type and gender**



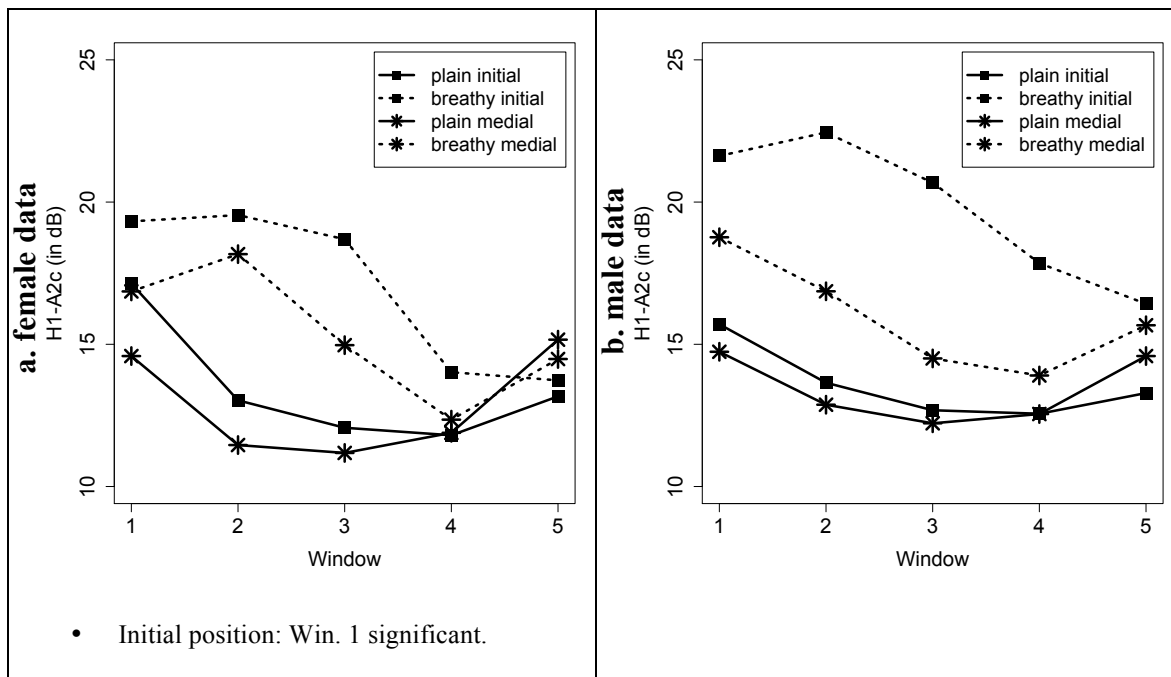
With regards to differences in H1-A2c values triggered by consonant phonation type, significance is attained only for [a] and only in Window 1 ( $F(1,4) = 26.21, p = 0.007$ ) for the female speakers, with marginal significance in Window 2 ( $F(1,4) = 20.1, p = 0.011$ ). Results for the male speakers again show a somewhat surprising pattern: no differences are significant in [e]. In [a], however, differences based on consonant phonation type are marginal in Windows 2 ( $F(1,4) = 20.24, p = 0.011$ ) and 3 ( $F(1,4) = 25.3, p > 0.007$ ), and significant in Window 4 ( $F(1,4) = 26.69, p < 0.007$ ).

While it is true that relatively few of these tests meet the alpha level established for post hoc analyses, the attempt to untangle a four-way interaction (Phonation type by Vowel context by Window by Gender) should not distract us from the fact that H1-A2c values in vowels are significantly different overall based on consonant phonation type after both sonorants and voiced obstruents. That is the key lesson here. The interaction indicates that the results vary both by

gender and by vowel context, and so perhaps the lesson that we can take from this again deals with reliability. Post-consonantal H1-A2c differences are used to cue breathiness, but use of this cue differs. It is more reliable in [a] than in [e], for both male and female speakers. Furthermore, the differences are more pervasive—meaning that they attain marginality or significance in more windows—for the male speakers.

A similarly complex picture emerges with regards to Word position. The omnibus Anova revealed a Phonation type by Word position by Window by Gender interaction ( $F(1,6) = 6.113$ ,  $p = 0.048$ ), and these data appear in Figure 49.

**Figure 49** H1-A2c values across vowels following voiced segments by consonant phonation type, word position, and gender



This interaction tells us that H1-A2c values within the vowel may be more or less differentiated based on phonation type depending on the word position of the preceding consonant.

Impressionistically, vowels after word-initial consonants appear to show greater differentiation.

The vowel context differences that arose were twofold—[a] showed greater values overall, and greater differentiation. Here, vowels after initial consonants do not appear to show greater values overall—rather, it looks like H1-A2c values are simply boosted in vowels after breathy voiced consonants in word initial position. This turns out to be the case: there is no significant effect of word position overall, for either gender. Furthermore, when post hoc analyses are conducted on subsets of the data, it is again the case that very few points of significance emerge. Here the difference between H1-A2c values after plain versus breathy consonants is significant after initial consonants in Window 1 for female speakers ( $F(1,4) = 25.96, p = 0.007$ ); nothing else emerges as significant.

Taking a step back from specific comparisons, it is important to note that vowel context and word position interact with the effects of phonation type and gender in ways that are less than straightforward, and it is not clear what we should make of these four-way interactions. The lesson I would like to pull from them is as follows. First, the patterns are in line with what we have come to expect—when looking at spectral measures, the higher vowel context of [e] is subject to more variability and is less likely to show significant differentiation in terms of the acoustic measures reported herein. In addition, word initial position is acoustically privileged, meaning that spectral differences are often more prevalent in word initial consonants and in vowels following word initial consonants.

These phonetic facts, however, need to be considered in tandem with the phonological reality, which is that languages do not care about word position and vowel context in the same way. The relevant phonemic contrasts can be found across various word positions and vowel contexts, indicating that the phonology is able to ignore all of the phonetic variability being brought to light in this dissertation. How does the phonological component of the grammar

accomplish this? It is a highly interesting question, and needs to be addressed in more detail in future work.

One point which will be treated more fully in the Discussion in Chapter 8 has to do with retention of breathy voiced obstruents and loss of breathy voiced sonorants in the phonetic inventories of languages in Nepal. Newar, a Tibeto-Burman language spoken in Nepal, has multiple dialects, all of which have retained breathy voiced obstruents but most of which have lost breathy voiced sonorants (Genetti 2005). A similar pattern is seen in the Kiranti languages Eastern Nepal. While many of the 32 extant Kiranti languages retain breathy voiced obstruents, only one—Camling—has retained breathy voiced sonorants (Ebert 2003). These could be indicative of a larger pattern; breathy voiced contrasts on sonorants are not limited to word initial or [a] contexts in languages where they exist, but the Marathi data tells us that they are subject to great variability in terms of how well they are cued acoustically, and the Newar and Kiranti patterns indicate that these contrasts may be lost from phonetic inventories more commonly than similar contrasts in obstruents. If so, this can be interpreted with reference to Dispersion Theory (Flemming 1995)—phonemic contrasts are good, because they allow for the creation of contrasting lexical items. If the contrasts are poorly cued acoustically, however, they may fail to satisfy the requirement for auditory distinctiveness and the contrast may eventually be lost. This possibility will be returned to in Chapter 8.

Moving our attention back to the specific parameter of H1-A2c, it is perhaps surprising to find so few points of significance when conducting post hoc analyses on [a] and [e] data, or on word initial and word medial data. Recall, however, that obstruency entered into neither the vowel context nor the word position interactions. Vowels after obstruents showed greater differentiation than those after sonorants. One explanation, then, could be that obstruency does

not enter into the interactions under discussion at present because vowels after obstruents and sonorants pattern similarly. When merging obstruent and sonorant data, however—which is not only justified but necessary here because obstruency does not enter into the interaction—we both increase standard deviations and lose access to the extent to which breathy voiced obstruents trigger greater H1-A2c values in vowels. We therefore lose overall significance in the difference between vowels after plain and breathy voiced segments. In exchange for this loss, however, we are able to see the ways in which vowel context and word position affect the overall pattern.

### **6.3.3 H1-A2c recap**

Again stepping back from the specific data in order to gain a broader view of the H1-A2c results, it is important to recall several facts. Segment-internal H1-A2c differences related to phonation type emerged here, and there was no gender interaction, meaning that this acoustic parameter cues breathiness segment-internally in sonorants for both male and female speakers. There was a vowel context interaction, however: the difference is significant before [a] and not before [e], meaning that once again we get better consonant-internal spectral cues in the [a] context than in the [e] context. Moving to consonant-external cues, we saw that consonant phonation type triggered overall differences in H1-A2c values in vowels after both sonorants and voiced obstruents, but that there was potentially greater differentiation after obstruents. Thus far, these findings align both with our predictions—that breathy voice will be associated with greater H1-A2c values than modal phonation—and with previous findings from, for example, Harris (2009), who reported that Sumi nasals and laterals began to show spectral differentiation based on phonation type only above 1000 Hz.



It is important to note here that the absence of reliable spectral cues for the higher vowel context in the low-frequency measures investigated herein does not mean that breathiness is not cued in the higher vowel context. Clearly, it is: while it is true that breathy voiced sonorants in particular are more limited within the Marathi lexicon as a whole and are more commonly found in the context of [a] or [ʌ] than in high vowel contexts, the phonation-type contrast is not limited to low-vowel contrasts. Breathily voiced sounds do appear in high vowel contexts in Marathi—for example, [b<sup>h</sup>iti] ‘*fear*’, [d<sup>h</sup>ir] ‘*courage*’, [d<sup>h</sup>il] ‘*slack*’, and [tum<sup>h</sup>i] ‘*you all*’. The same is true in other languages with breathily voiced sonorants—in Sumi, for instance, which contains words such as [n<sup>h</sup>i] ‘*marry/betroth*’, [m<sup>h</sup>i] ‘*clouded sky*’, [al<sup>h</sup>i] ‘*transaction*’, and [al<sup>h</sup>u] ‘*flea*’ (Harris 2009: 81) It has to be the case, then, that additional cues are present when the lower-frequency cues are absent—even if, as posited in the above discussion of the loss of breathily voiced sonorants in Newar and Kiranti languages, phonation type contrasts truly are cued more poorly by sonorants than obstruents. In the next sections, we will see whether the cues investigated in this study which have yet to be reported—H1-A3c and Cepstral Peak Prominence—prove more informative.

H1-A3c findings are presented in the next section. Like H1-A2c, H1-A3c is again a higher-frequency measure. If the trend established in this section continues, then, we expect the findings related to H1-A3c to align more closely with those reported for H1-A2c than those reported for H1-H2c or H1-A1c.

#### 6.4 H1-A3c

H1-A3 is a measure of the amplitude of the harmonic exciting the third formant subtracted from the amplitude of the first harmonic, and is correlated with the speed with which the vocal folds close during the glottal cycle. Abrupt closure of the vocal folds triggers greater excitation of high frequency energy, yielding larger A3 values and, in turn, smaller H1-A3 values. When closure is slower, and less abrupt, high frequency energy is not excited in the same way. Accordingly, A3 values are lower and H1-A3 values greater.

Previous findings regarding H1-A3 are mixed. While H1-A2 was relatively accurate in distinguishing plain from breathy vowels in Jalapa Mazatec, the same was not true for H1-A3, which successfully distinguished only seven out of 14 samples (Blankenship 1997). H1-A2 was a better measure than H1-A3 of breathiness in Gujarati, as well—meaning it more reliably distinguished plain from breathy vowels (Fischer-Jørgensen 1967). Furthermore, in the discriminant analysis mentioned in the last section—that conducted by Esposito (2006) on two modal and two breathy vowel tokens from each of ten<sup>27</sup> different languages and dialects—H1-A3 accounted for just 8% of the variance.<sup>28</sup> H1-A2 accounted for slightly more (10%).

Indications from Sumi are give reason to think that the story may be different for sonorants, however. Harris (2009) found that plain and breathy nasals in Sumi showed differences in amplitude in the region of the third formant, in the expected direction; although H1-A3 was not directly measured in that study, Harris's findings—that the amplitude of F3 was greater for plain consonants than for breathy consonants— indicate the presence of amplitudinal differences in the region of F3, where the A3 measurement is made. As such, it is reasonable to posit that H1-A3 measurements would have yielded greater positive values for breathy than for

---

<sup>27</sup> Chong, Fuzhou, Green Hmong, White Hmong, Mon, Santa Ana del Valle Zapotec, San Lucas Quiavini Zapotec, Tlacolula de Matamoros Zapotec, Tamang, and !Xóǿ.

<sup>28</sup> Recall that in that analysis, CPP accounted for 46% of the variance and H1-H2 for 27%

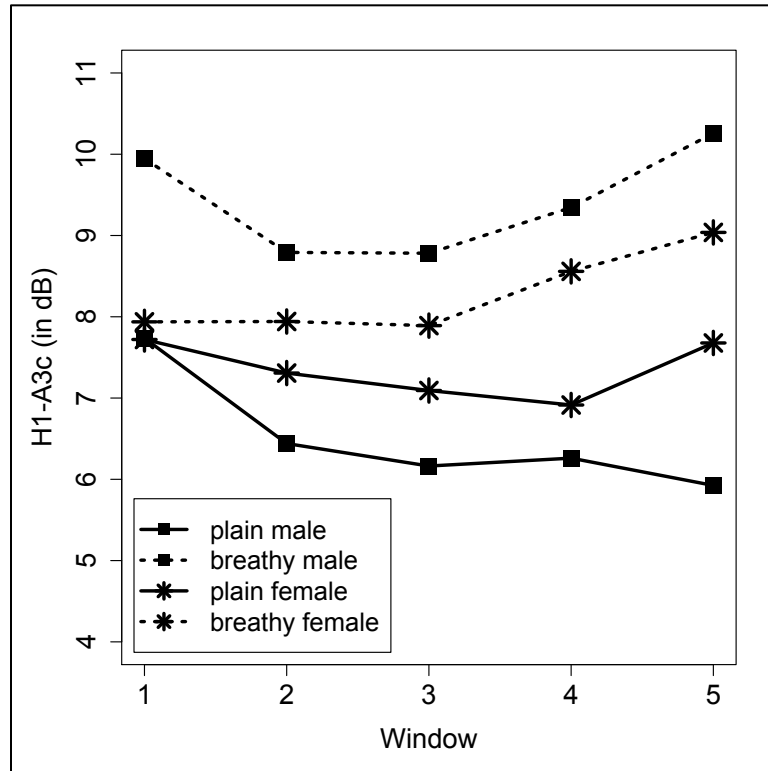
plain nasals in Sumi, if such measurements had been taken. If this pattern holds in Marathi, then we expect the measure of H1-A3 to yield consonant-internal cues—in the form of greater H1-A3 values for breathy than for plain sonorants—and this is what we explore first.

#### **6.4.1 H1-A3c – consonant results**

Overall H1-A3c values plotted across the time course of sonorant segments appear in Figure 50. Impressionistically, male and female data pattern similarly, in that plain sonorants show lower H1-A3c values overall than breathy sonorants. A repeated measures Anova was conducted to evaluate the effect of the independent variables (Phonation Type, Place of Articulation, Word Position, Vowel Context, Window, and Gender) on H1-A3c values in sonorants.

The results of this test reveal that the above observation is statistically supported. There is an overall effect of Phonation type ( $F(1,7) = 6.59, p = 0.037$ ), meaning that breathy phonation does trigger greater segment-internal H1-A3c values. Although it appears as though male data may show greater differentiation based on Phonation type than female data, this observation is not supported statistically. The omnibus reveals no Phonation type by Gender interaction.

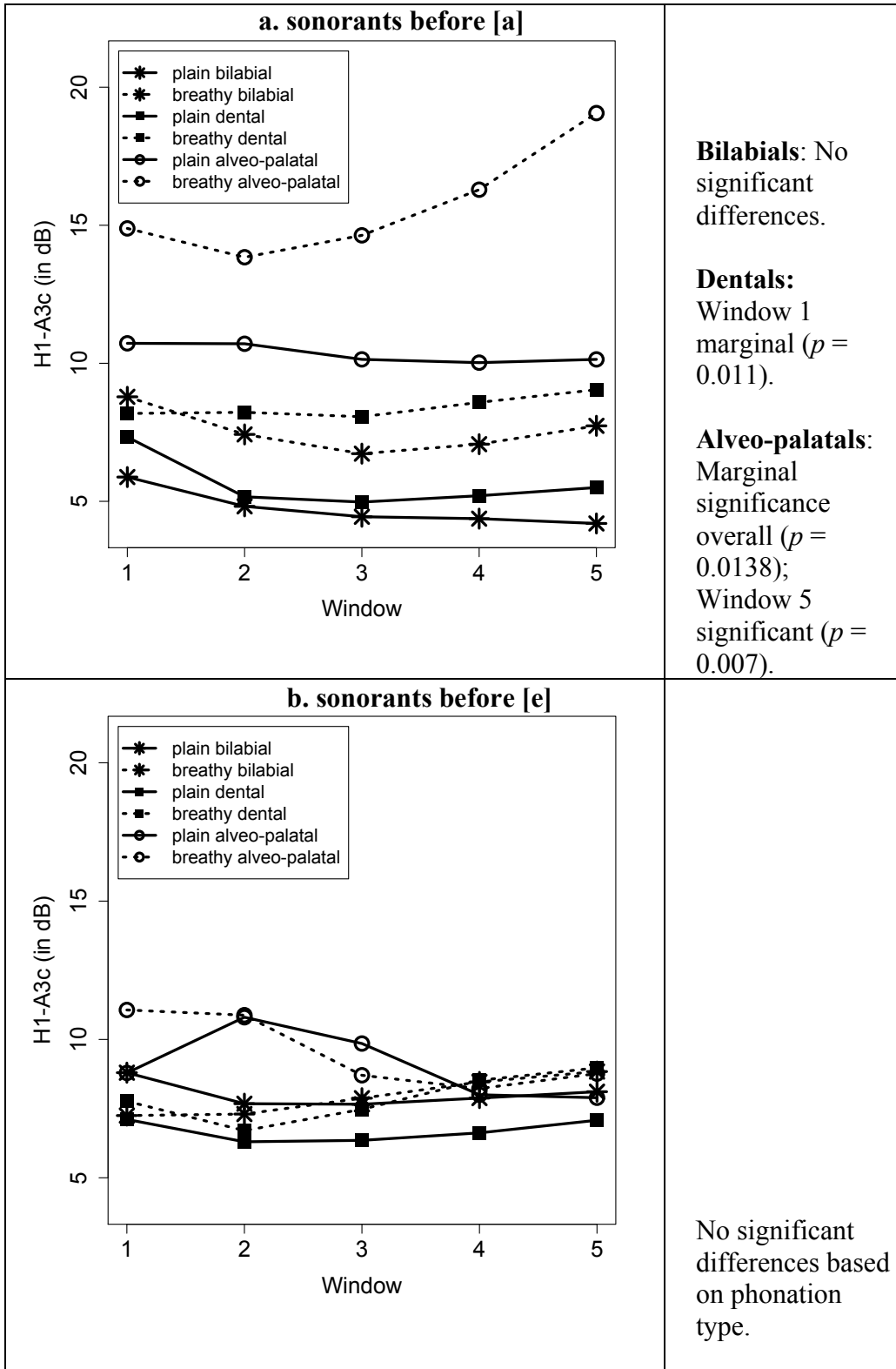
Figure 50 H1-A3c across sonorants by phonation type and gender



In fact, beyond the overall effect of phonation type on H1-A3c values, the omnibus Anova reveals very few significant interactions. This is interesting in and of itself, for it indicates that the effect of consonant phonation type on H1-A3c values is relatively stable.

One effect that does emerge, however, is a Phonation Type by Place by Vowel context by Window interaction ( $F(1,7) = 14.73, p = 0.006$ ). The pattern underlying this interaction is presented in Figure 51, and what we see is not unfamiliar: once again, sonorants before [a]—shown in (a.)—pattern in the expected directions. Breathy voiced sonorants are associated with greater H1-A3c values than their plain counterparts. This pattern disappears in sonorants that occur before [e], however, as shown in (b.).

Figure 51 H1-A3c values across sonorants by phonation type, place of articulation, and vowel context



In (a.), bilabials and dentals appear to pattern together, while alveo-palatal sonorants have greater H1-A3c values overall as well as greater differentiation between plain and breathy sonorants. Post hoc analyses do not greatly clarify the interaction, however. They do reveal that the overall difference in H1-A3c values based on Phonation type in the alveo-palatal sonorants approaches significance ( $F(1,9) = 9.30, p = 0.0138$ ), but little else emerges. We have seen bilabial and dental sonorants pattern together before, and they do the same here in that differences between plain and breathy sonorants do not attain significance for the bilabials or dentals. In a window-by-window analysis, the only point of significance is between plain and breathy alveo-palatals in Window 5 ( $F(1,9) = 11.88, p = 0.007$ ). Furthermore, no H1-A3c differences based on phonation type are significant when sonorants appear before [e].

The overall message regarding H1-A3c cues within sonorants seems to be this. There is an overall effect of phonation type, with breathy sonorants yielding greater H1-A3c values overall than their plain counterparts. This is true for both male and female speakers. The interaction involving vowel context is hard to interpret, because when the data are separated according to vowel context very few comparisons reach the alpha level established for post hoc analyses. Nonetheless, Figure 51 indicates that the significant effect of phonation type overall may be driven largely by sonorants in the [a] vowel context.

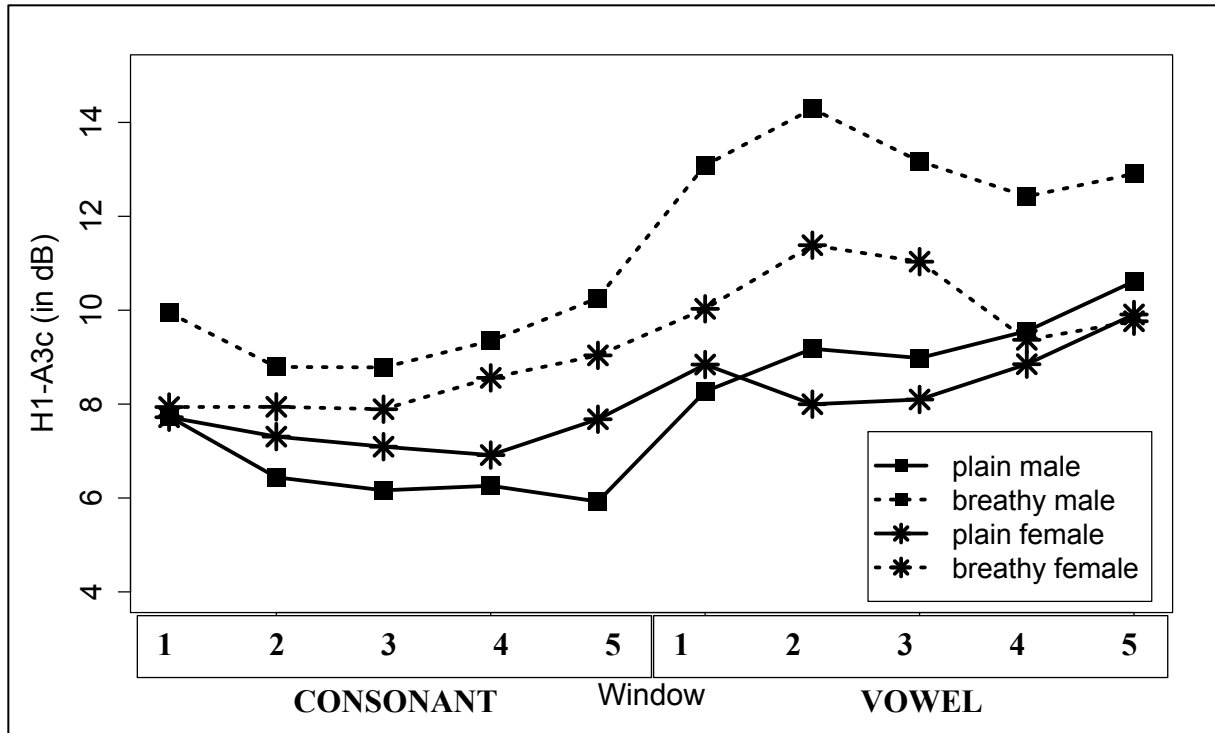
This is interesting because when we consider A3 measurements, we are presumably considering portions of the signal that are well above the 1000 Hz mark where Harris (2009) reports distinct spectral differences in the plain and breathy nasals of Sumi. While vowel quality—and specifically, the height of F1—may render the lower frequency measures less effective in the higher vowel context of [e], it is not necessarily expected that the same would be true in the higher frequency measure of H1-A3c. These results indicate, however, that this is in

fact the case. The lower vowel context of [a] still has a benefit over the higher context of [e] with regards to these spectral measures.

Again, it is important to note here that breathy voiced sounds do occur before high vowels. Such sequences are not illegal, and the fact that the spectral cues are weaker or absent in the high vowel context only tells us that breathy phonation in those contexts is cued by something else. Recall that [e] is significantly longer after breathy alveo-palatals than after their plain counterparts, for instance. It is worth noting, however, that—impressionistically—breathy sonorants in Marathi lexical items appear far more commonly before [a] than before any other vowel, with the possible exception of [ə]. I would argue that while these sounds pick up other cues in higher vowel contexts, the fact that breathy sonorants lack the spectral cues under discussion in this and the preceding sections renders them less distinct from their plain counterparts, and this is a contributing factor to the low incidence of such sequences in the Marathi lexicon as a whole. At least with regards to sonorants, breathy phonation simply is not cued as well in the higher vowel contexts.

Before moving ahead into the vowel data, a plot of H1-A3c across full sonorant-vowel syllables is provided. Again, the value in looking at a plot like this is that it allows us to see how the sonorant and vowel data fit together, thereby enabling us to envision how things change over the time course of the syllable.

Figure 52 H1-A3c across syllables by gender (sonorant data only)



We turn now to the results related to H1-A3c values within vowels.

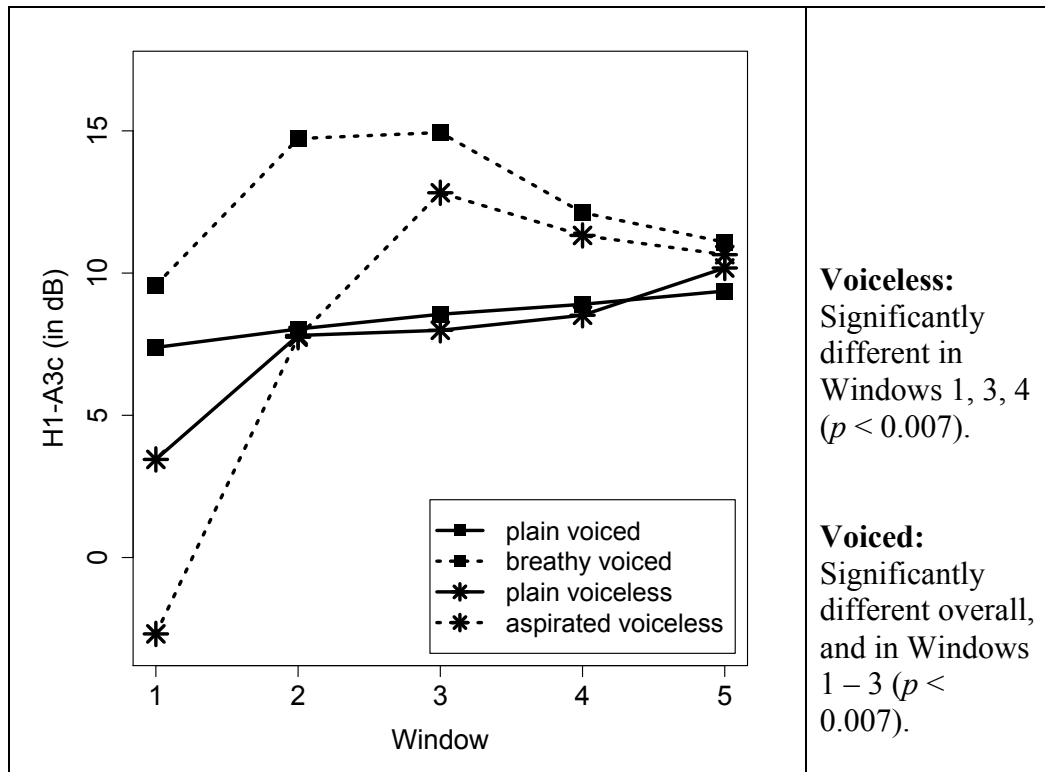
#### 6.4.2 H1-A3c – vowel results<sup>29</sup>

Before moving into a comparison of vowels after voiced segments, H1-A3c values after obstruents are shown in Figure 53. Note that vowels after aspirated voiceless obstruents begin with very low H1-A3c values. In fact, it is not until Window 3—in the middle portion of the vowel—that the voiced and voiceless obstruents begin to pattern similarly, with lower values appearing in vowels after plain obstruents.

<sup>29</sup> Recall that the intervals reported as the PVI and as the vowel in the duration section are concatenated in the F0, spectral, and CPP results sections.



Figure 53 H1-A3c across vowels after obstruents by consonant phonation type and voice



To investigate the effects of Consonant Phonation Type, Voice, Place of Articulation, Vowel Context, Word Position, Window, and Gender on H1-A3c values in subsequent vowels, an omnibus Anova was conducted. As we might expect, both a Phonation type by Voice interaction ( $F(1,4) = 22.44, p = 0.009$ ) and a Phonation type by Voice by Window interaction ( $F(1,4) = 96.38, p < 0.001$ ) were revealed. Post hoc analyses—the results of which appear in Table 13—reveal that consonant phonation type triggers significant differences in vowel H1-A3c values overall and in Windows 1-3 for vowels after voiced obstruents. No significant overall difference arises for voiceless obstruents, however, though Windows 1, 3, and 4 attain significance. This is presumably due to the fact that differences in Window 1 are in the opposite direction of those in Windows 3 and 4.

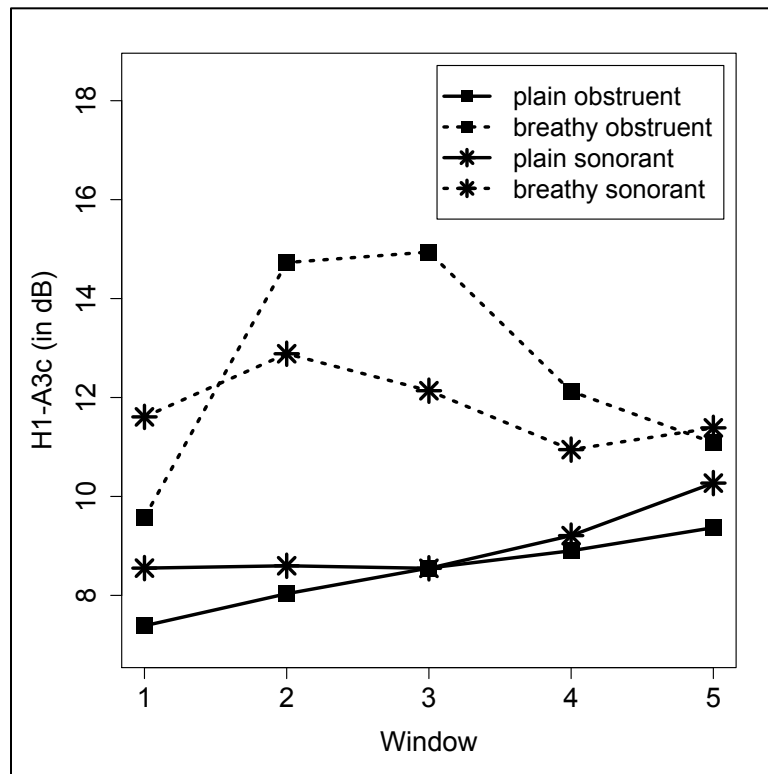
**Table 13** Post hoc analyses of H1-A3c differences triggered by phonation in vowels following voiced and voiceless obstruents

	Voiced Obstruents		Voiceless Obstruents	
	<i>F</i> (1,9)	<i>p</i>	<i>F</i> (1,9)	<i>p</i>
<b>Window 1</b>	15	0.004*	12.5	0.006*
<b>Window 2</b>	22.32	0.001*	4.53	0.062
<b>Window 3</b>	15.44	0.003*	14.23	0.004*
<b>Window 4</b>	8.36	0.018	14.61	0.004*
<b>Window 5</b>	6.81	0.028	3.16	0.109
<b>Overall</b>	17.52	0.002*	1.57	0.242

Higher order interactions related to obstruent voicing are not investigated here, as they are not the focus of the present study. The message for us to take from the above data is that obstruent voicing does trigger differences in the time course of H1-A3c across subsequent vowels. We are primarily interested in breathy voiced phonation, however, and in the ways in which breathy voice is instantiated differently in and after obstruents and sonorants, so we move now to a comparison of vowels after voiced obstruents and sonorants. An overall plot of the relevant H1-A3c values appears in Figure 54.

Two things may be noted about the data in this plot. First, it appears as though consonant phonation type affects H1-A3c values in vowels in the expected direction—vowels after plain consonants are associated with lower values overall, while those following breathy voiced consonants are associated with greater values overall. Second, as we have seen previously, obstruents appear to trigger greater differentiation than sonorants. To investigate whether this was in fact the case, and to determine whether there were interactions between Phonation type and the remaining independent variables (Obstruency, Place of articulation, Vowel context, Word position, Window, and Gender), an omnibus repeated measures Anova was conducted.

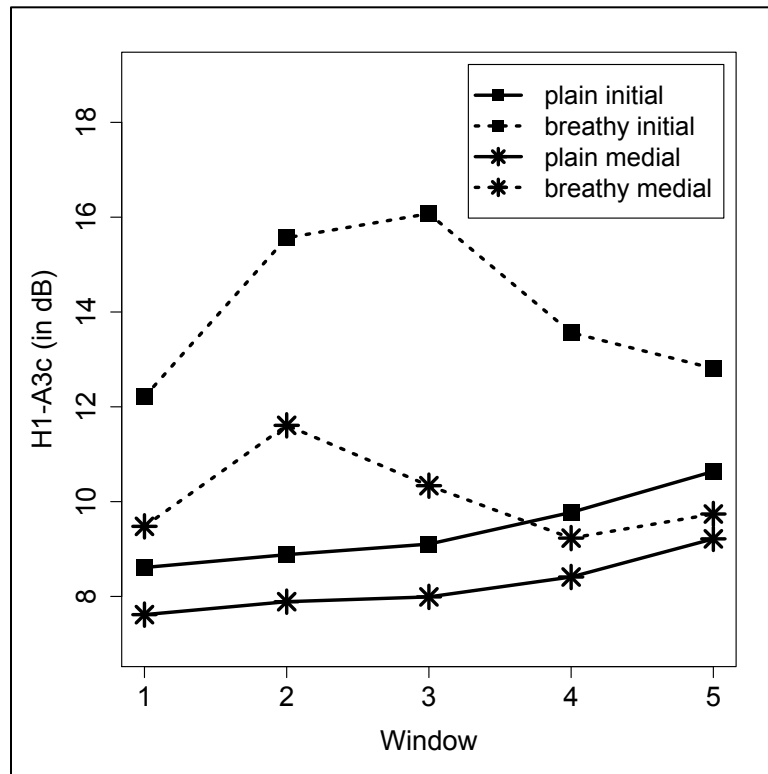
Figure 54 H1-A3c across vowels after voiced segments by consonant phonation type and obstruency



The first observation is supported. There is a main effect of consonant phonation type ( $F(1,6) = 10.85, p = 0.017$ ), meaning that values after breathy voiced consonants are greater than those after plain voiced consonants. Interestingly, there is no Phonation type by Obstruency interaction, nor is there a Phonation type by Gender interaction. H1-A3c differs from some of the measures investigated previously, then, in that here sonorants and obstruents pattern alike—meaning that the differences they trigger in subsequent vowels do not differ. Male and female speakers pattern together, as well.

There is a Phonation type by Word position interaction ( $F(1,6) = 17.31, p = 0.006$ ), plotted in Figure 55. The pattern seen here is not unexpected. Vowels after consonants in initial position show greater differentiation than those following consonants in medial position.

Figure 55 H1-A3c values in vowels after voiced segments by consonant phonation type and word position

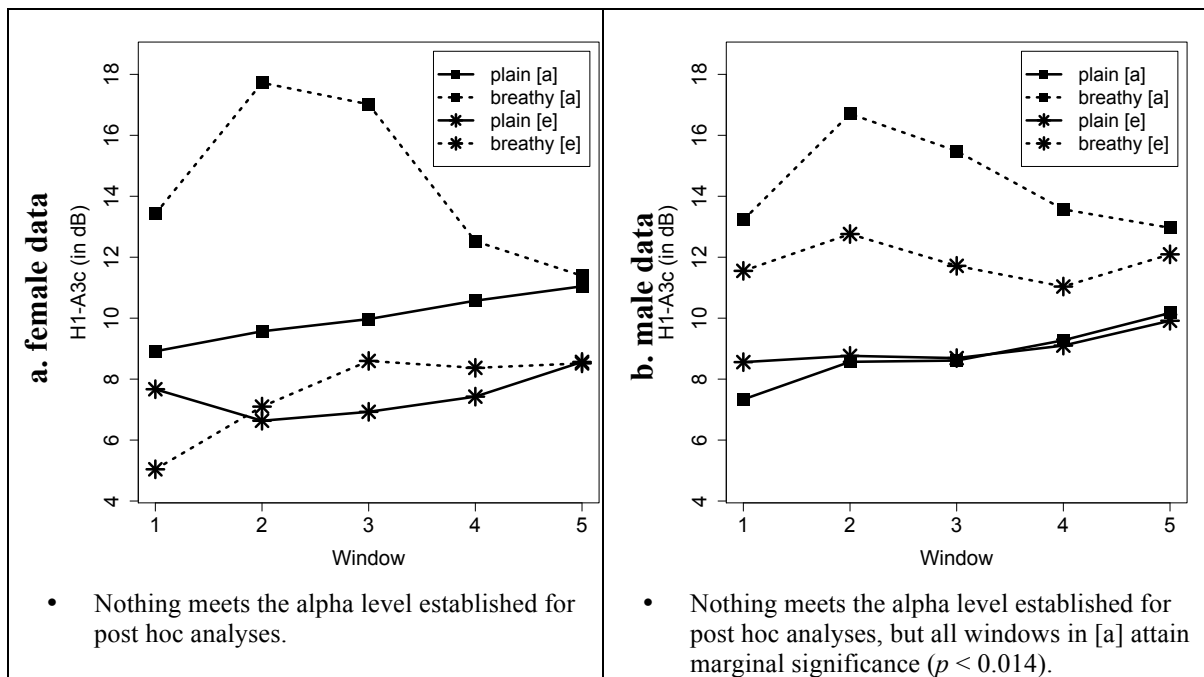


Post hoc analyses reveal that differences after initial consonants attain overall statistical significance ( $F(1,9) = 23.3, p < 0.001$ ), while those after medial consonants do not ( $F(1,9) = 4.68, p = 0.059$ ). Note that Obstruency does not enter into this interaction; the finding holds true for vowels after both obstruents and sonorants. A window-by-window analysis was not conducted as Window does not enter into the interaction, but impressionistically we may observe that vowels after initial consonants show greater differentiation overall, and that the values remain farther apart even in the later windows of the vowel. The idea here is that for consonants in initial position, breathy voice is associated with increased H1-A3c values that persist throughout much or all of the subsequent vowel, but that H1-A3c does not cue breathiness in the

same way after medial consonants. After word-medial consonants, the effect of breathy phonation is reduced and is more localized.

The omnibus Anova also reveals a Phonation type by Vowel context by Window by Gender interaction ( $F(1,6) = 40.48, p < 0.001$ ), and these data are shown in Figure 56. Again, these higher order interactions are complicated to untangle, but in general [a] shows greater differentiation of H1-A3c values based on the phonation type of the preceding consonants. Values after [e] are less differentiated, and—for females—are lower overall.

**Figure 56 H1-A3c in vowels after voiced segments by consonant phonation type, vowel context, and gender**



Post hoc tests were conducted in an attempt to untangle the interaction, and these data appear in Table 14. While no result meets the alpha level established for post hoc tests in this study, values for [a] are marginal for male speakers in all windows, and for females in Window 1, while  $p$  values for [e] never approach significance.

**Table 14** Post hoc analyses of H1-A3c in [a] and [e] by consonant phonation type and gender

	Female				Male			
	[a]		[e]		[a]		[e]	
	<i>F</i> (1,4)	<i>p</i>	<i>F</i> (1,4)	<i>p</i>	<i>F</i> (1,4)	<i>p</i>	<i>F</i> (1,4)	<i>p</i>
<b>Window 1</b>	21.48	0.01	0.1	0.767	19.55	0.012	2.1	0.221
<b>Window 2</b>	15	0.018	0.4	0.563	14.56	0.019	0.68	0.143
<b>Window 3</b>	4.8	0.798	0.79	0.423	22.97	0.009	6.49	0.063
<b>Window 4</b>	1.09	0.355	0.84	0.411	21.49	0.01	7.91	0.048
<b>Window 5</b>	0.441	0.543	0.39	0.568	21.19	0.01	13.58	0.021

The take home message here, then, is similar to what we saw in the previous section when looking at H1-A3c values in sonorants before [a] and [e]. H1-A3c patterns differently in the different vowels, and it patterns differently in sonorants before the different vowels. It is a more reliable cue to breathiness in [a] than in [e] (and before [a] than before [e]). Based on the images in Figure 56, one might guess that this is more true for females than for males, but that observation is not supported by the statistics.

### 6.4.3 H1-A3c recap

Importantly, the consonant results in this section reveal an overall effect of consonant phonation type. Plain and breathy sonorants differ significantly from one another in terms of this acoustic parameter, and the effect is relatively unaffected by the independent variables investigated herein—meaning that few interactions emerge in the omnibus Anova. The genders pattern together, as do the different word contexts. There is a somewhat complicated Phonation type by Place of articulation by Vowel context by Window interaction, indicating that the sonorants pattern differently before [a] than before [e]—an effect that we have seen multiple times in the preceding sections. Even taking these interactions into account, however, the most

important point to cull from the rest of the consonant results is the significant main effect of Phonation type.

What is interesting about the vowel results is that Obstruency—which has interacted with many other factors when analyzing other acoustic results—does not enter into interactions related to H1-A3c. In previous measures, vowels after obstruents showed greater differentiation than those after sonorants; in H1-A3c, however, this is not the case. If we consider this in tandem with the fact that gender interacts with so few other variables, here, we may posit that H1-A3c is valuable in specific ways. While we see phonation type cued differently in many ways in male and female speech, and in obstruents and sonorants, in H1-A3c we find a measure where those differences disappear. In this way, H1-A3c is similar to duration, which was another parameter wherein we did not find gender-based differences.

How all of these parameters fit together to form the big picture remains to be addressed, but we can anticipate the following. Some parameters work very well for males, but not for females, and vice versa. Others work very well for both genders. These parameters are additive, however—meaning, whatever set of parameters cue breathiness for males, we know that duration and H1-A3c must be a part of the set. And while female speakers may utilize a different set of parameters, duration and H1-A3c will again be a part of the set. These issues—and a synthesis of the findings from all measures into a unified picture of the acoustic cues associated with phonation type in Marathi—will be presented after the results of the final acoustic measure are reported. We turn now to a discussion of Cepstral Peak Prominence.

## 6.5 Cepstral Peak Prominence

The measure of cepstral peak prominence reflects a property of the voice source, meaning that it reflects what is happening at the level of the vocal folds. The cepstrum from which this measure is taken is generated by performing a Fourier transform on the voice signal, and then performing an additional Fourier transform on the spectrum; doing so renders a clear image of the harmonic structure of the signal. Breathy voice, which is often produced with incomplete closure of the vocal folds, is characterized by more noise in the signal and a steeper drop-off in harmonic energy: in this phonatory state, then, the harmonics do not stand out as clearly against the noise in the signal, and the peaks in the cepstrum are therefore lower. The modal voice source spectrum, on the other hand, is characterized by a very clear harmonic structure. Accordingly, modal voice should yield greater CPP values, while breathy voice is expected to be associated with decreased CPP values. As mentioned previously, this measure is often interpreted as reflecting periodicity (Keating & Esposito 2007). The idea here is that, in addition to affecting the harmonic structure of the sound, the noise present in the signal in the production of breathy voice also contributes to more jagged periodicity, which is often visible in the waveform.

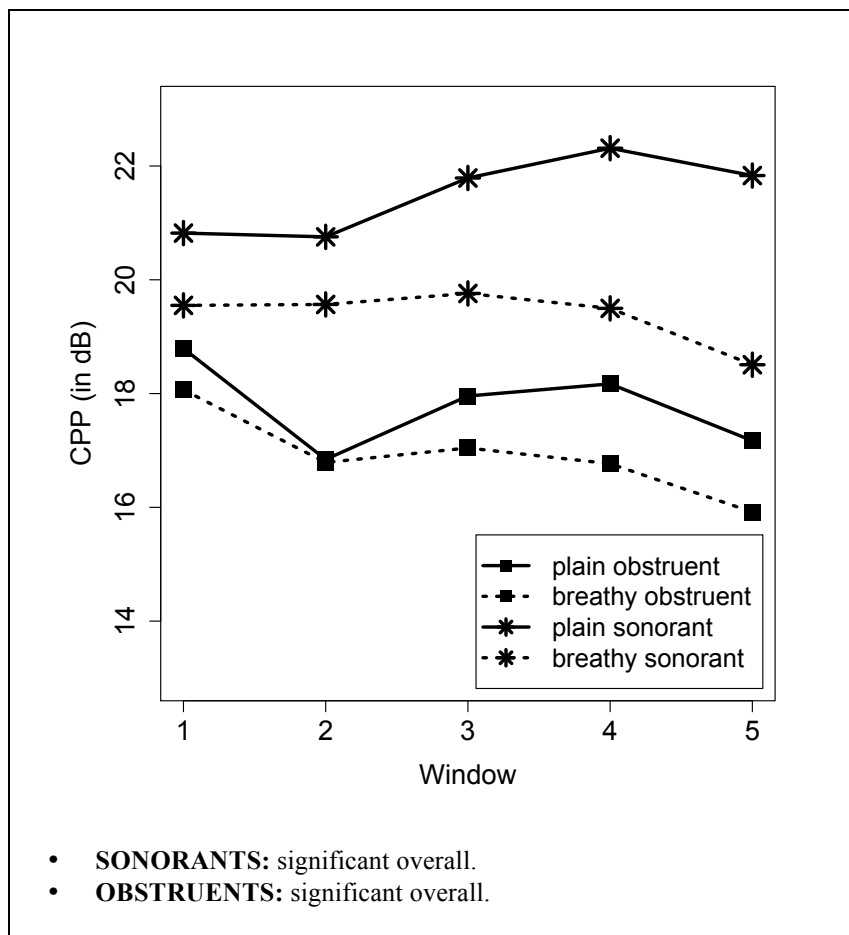
The present research contributes new information on at least two fronts. First, while CPP has gained more prominence in the literature of late, it has thus far been used primarily to assess voice quality in vowels. Sonorants and voiced obstruents are also periodic, however, and so we can investigate whether there are consonant-internal CPP differences in addition to whatever effects we may find triggered by consonants in subsequent vowels. Second, if CPP values in vowels differ based on the phonation type of the preceding consonant, we don't know how far into the vowel the effect perseverates. The following results shed light on both issues.



### 6.5.1 CPP – consonant results

An omnibus repeated measures Anova was conducted to investigate whether Phonation type had an effect on CPP values within voiced consonants, and to determine whether there were interactions between Phonation type and the remaining independent variables (Obstruency, Place of articulation, Vowel context, Word position, Window, and Gender). A main effect of Phonation type ( $F(1,6) = 83.27, p < 0.0001$ ) and a Phonation type by Obstruency interaction ( $F(1,6) = 24.82, p = 0.003$ ) emerged. Data are plotted in Figure 57.

Figure 57 Time course of CPP across voiced segments by phonation type and obstruency

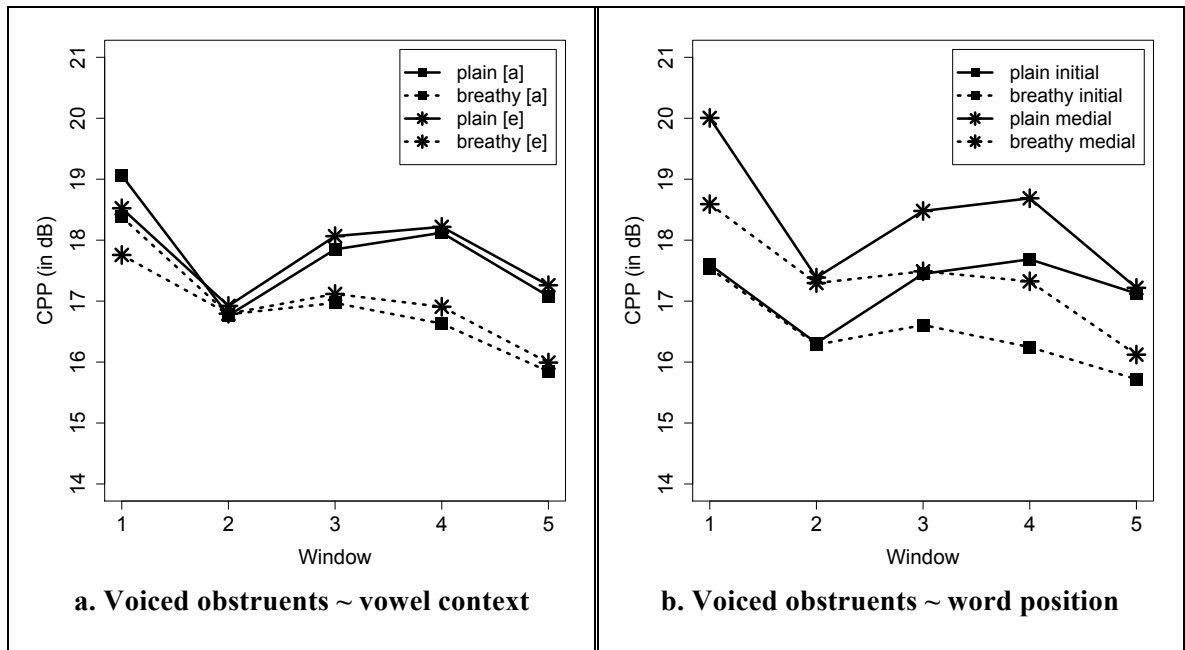


As can be seen in this figure, sonorants appear to show higher mean CPP values overall. This is not surprising. They also appear to show greater differentiation between plain and breathy

segments. Post hoc analyses do not necessarily help us to untangle this interaction, however, because they reveal a significant effect of Phonation type overall for both sonorants ( $F(1,9) = 18.93, p = 0.002$ ) and obstruents ( $F(1,9) = 17.02, p = 0.003$ ). It seems likely that the interaction may arise because the obstruents pattern differently than the sonorants in Window 2; let us consider obstruent data alone for a moment.

To further investigate the pattern found within the voiced obstruent data, an omnibus repeated measures Anova was conducted on voiced obstruent data only. The omnibus revealed that Phonation type interacts neither with Gender nor with Vowel context. These data are plotted in Figure 58a, where it is clear that CPP means in the two vowel contexts pattern together. What did emerge in the omnibus was a Phonation type by Word position interaction ( $F(1,7) = 26.8, p = 0.0006$ ), as seen in Figure 58b. Here, initial segments show slightly lower mean CPP values overall; furthermore, while post hoc tests reveal that the differences between plain and breathy obstruent CPP values is significant overall in word-medial context ( $F(1,9) = 18.32, p = 0.002$ ), differences are not significant in word-initial position ( $F(1,9) = 8.91, p = 0.017$ ). Although a window-by-window analysis was not conducted, impressionistically it looks like mean CPP values for plain and breathy voiced obstruents differ significantly for the final 60% of the segment, regardless of vowel context or word position. The omnibus indicates that no additional factors—including gender—interact with phonation type for voiced obstruents.

**Figure 58 Time course of CPP across voiced obstruents**

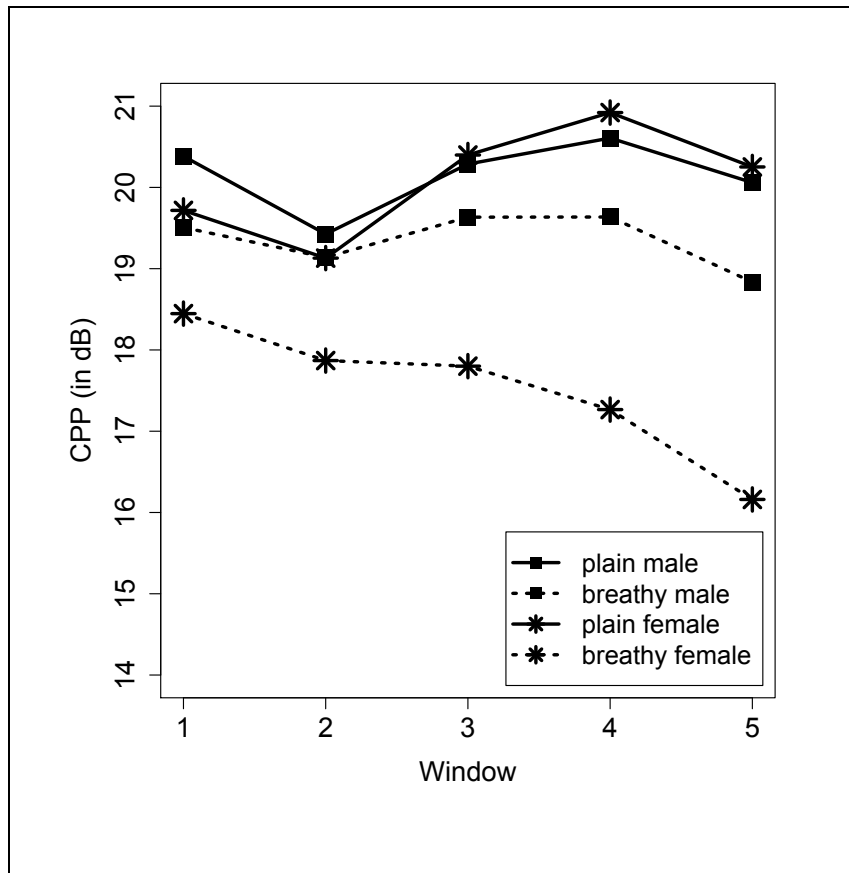


The take home message, then, seems to be that CPP distinguishes plain from breathy voiced obstruents in the final 60% of the segment, and that these differences are enough to drive overall significance in most contexts—with the failure to attain significance in word-initial position standing out as an exception. This is yet another example of the kind of compensation that we have seen before: while word-initial position has been the stronger position with regards to some of the measures presented previously—for H1-A2c within sonorants, for instance, which differed significantly based on phonation type word-initially but not word-medially—here, the reverse is true. Here, word-medial voiced obstruents pick up a cue that word-initial voiced obstruents lack.

Moving away from the voiced obstruent data, we now turn to results related to sonorants, wherein the omnibus revealed a number of interactions. The first one that we will address is a Phonation type by Gender interaction ( $F(1,6) = 37.79, p = 0.0008$ ), shown in Figure 59. Here, we see plain and breathy sonorant means for male and female speakers. It looks at first glance as

though there is greater differentiation between plain and breathy sonorants produced by female speakers than by male speakers, and this observation is borne out in the post hoc analyses. While mean CPP values within sonorants show an overall effect of Phonation type for female speakers ( $F(1,4) = 45.84, p = 0.002$ ), the same is not true for male speakers ( $F(1,4) = 9.36, p = 0.04$ ).

**Figure 59** Time course of CPP across sonorants by phonation type and gender



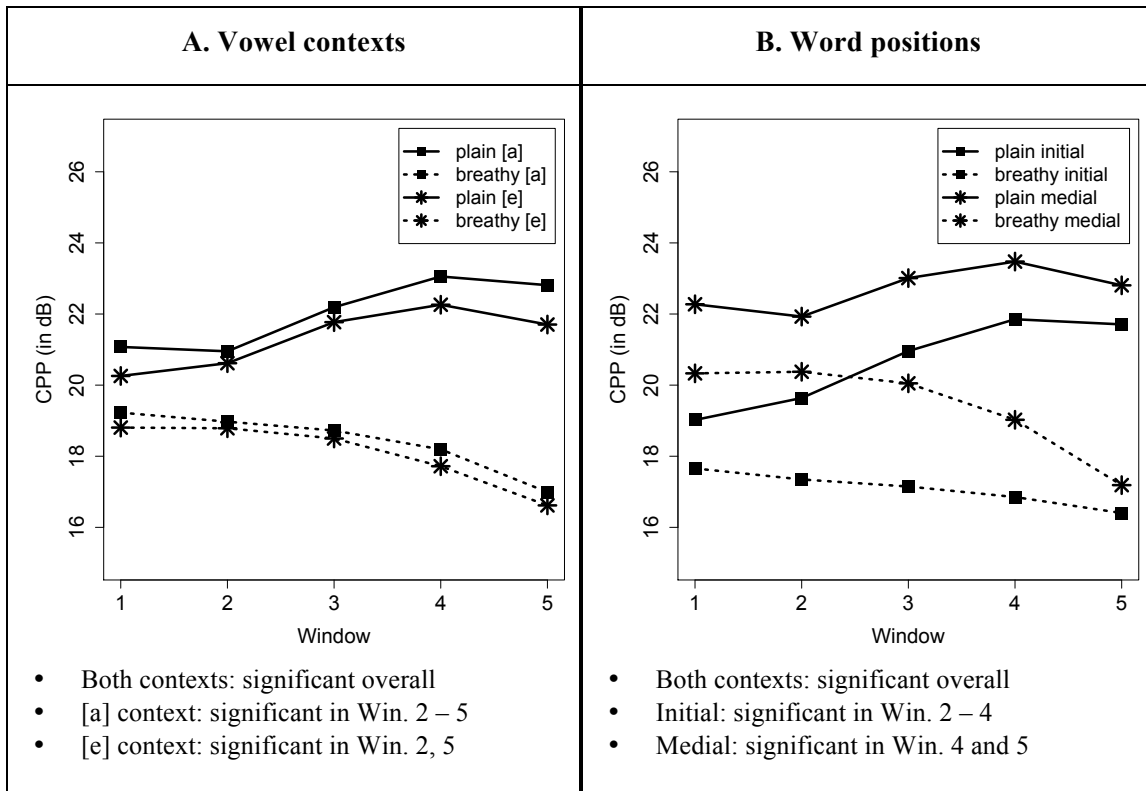
This pattern is of interest in part because it is opposite that which we have seen previously; while the lower frequency measures like H1-H2c and H1-A1c successfully distinguished plain from breathy sonorants for male speakers, significant differences for those measures did not arise sonorant-internally for female speakers. Now, we see a measure that cues breathy phonation sonorant-internally for female but not for male speakers.

The omnibus Anova returned several additional higher order interactions, which are explored in the female sonorant data in the following pages. Note that several of these interactions include Obstruency. I interpret this as arising from the fact that obstruents are unaffected by most factors, while the same is not true for sonorants. One effect that emerges, for instance, is a Phonation type by Vowel context by Obstruency by Window interaction ( $F(1,3) = 7.6, p = 0.03$ ). This interaction arises because while obstruents do not differ by vowel context, sonorants do.<sup>30</sup> There is also a Phonation type by Word position by Obstruency by Window interaction ( $F(1,3) = 6.03, p = 0.05$ ). Relevant data are shown Figure 60, and discussion follows.

---

<sup>30</sup> It is at first puzzling that there is no gender interaction here, given the fact that I just reported an overall effect of Phonation type for female but not for male speakers. The explanation for this, however, is that the default significance level in R is 0.05. Any result  $\leq 0.05$ , then, is coded as significant by R. Put plainly, R thinks that male sonorant CPP values differ significantly based on breath, but the  $p$ -values are above the significance level established for post hoc analyses herein.

Figure 60 CPP across sonorants ~ vowel context and word position (female data)



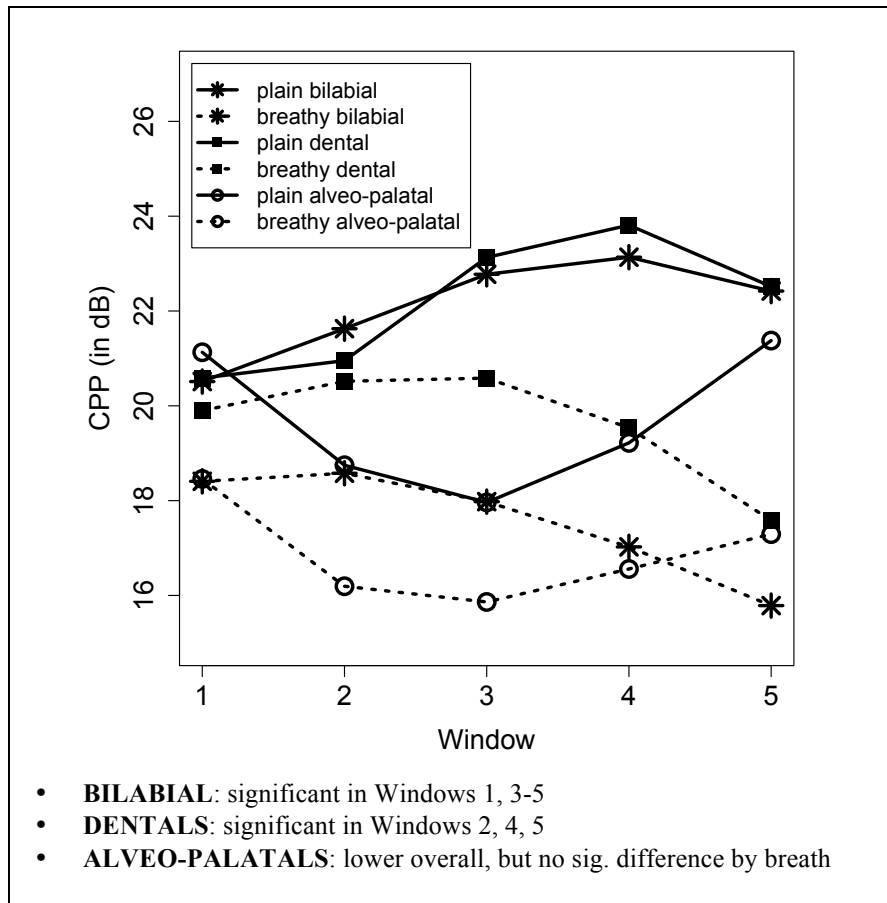
Post hoc analyses were conducted to aid in interpreting these interactions. Focusing on the vowel context interaction first, the post hoc tests reveal that CPP means for sonorants produced by females are significantly different overall in both the [a] context ( $F(1,4) = 46.74, p = 0.002$ ) and the [e] context ( $F(1,4) = 26.28, p = 0.007$ ). When we do a window by window analysis, however, we see that different windows attain significance for the [a] and the [e] contexts. While means are significantly different in Windows 2-5 in the [a] context, the differences attain significance only in Windows 2 and 5 in the [e] context. These differences are enough to drive overall significance, however, and so it is not clear how much importance to attribute to the effect of vowel context. Perhaps the core message is that although the specific windows of significance vary, breathiness in sonorants is cued by CPP differences in both vowel contexts.

With regards to the word position interaction, post hoc tests reveal that plain sonorant mean CPP values are higher overall than their breathy counterparts in both word initial position ( $F(1,4) = 49.78, p = 0.002$ ) and in word medial position ( $F(1,4) = 29.46, p = 0.006$ ). Once again, a window-by-window analysis reveals distinct windows of significance—and these results are listed in Figure 60b. As with the vowel context window, however, perhaps the important lesson to take away from this is that differences in those specific windows are great enough to drive overall significance.

This is not to say that there is no value in looking at the line graphs plotted in Figure 60: gaining more information about precisely which windows yield significant differences gives us more information about how CPP plays out over the time course of the sonorants. We have to be careful, however, not to let it obscure the big picture. Plain and breathy sonorants may have significantly different mean CPP values in only two windows of the word-medial context, but this still drives an overall difference. As listeners—and speakers, of course—have access to the entire segment during speech, so the overall significance is an important piece of the puzzle.

The omnibus also revealed that, like Vowel context and Word position, Place of articulation interacts with Phonation type, Obstruency, and Window ( $F(1,3) = 46.75, p = 0.0005$ ). CPP means for sonorants produced by female speakers appear in Figure 61, and discussion appears below.

Figure 61 CPP across sonorants by phonation type and place of articulation (female data)



The obstruency piece of the interaction arises because obstruent means are unaffected by place. With regards to Place of articulation, the alveo-palatal sonorants have lower CPP means overall; furthermore, the difference between plain and breathy means does not attain significance for the alveo-palatals overall, nor in any individual window. The labial and dental sonorants pattern together in that both show a significant effect of Phonation type overall (labials:  $F(1,4) = 45.52$ ,  $p = 0.003$ ; dentals:  $F(1,4) = 39.38$ ,  $p = 0.003$ ). A window-by-window analysis further reveals that bilabial plain and breathy means differ significantly in all Windows except Window 2, while mean differences for the dental sonorants attain significance in Windows 2, 4, and 5.



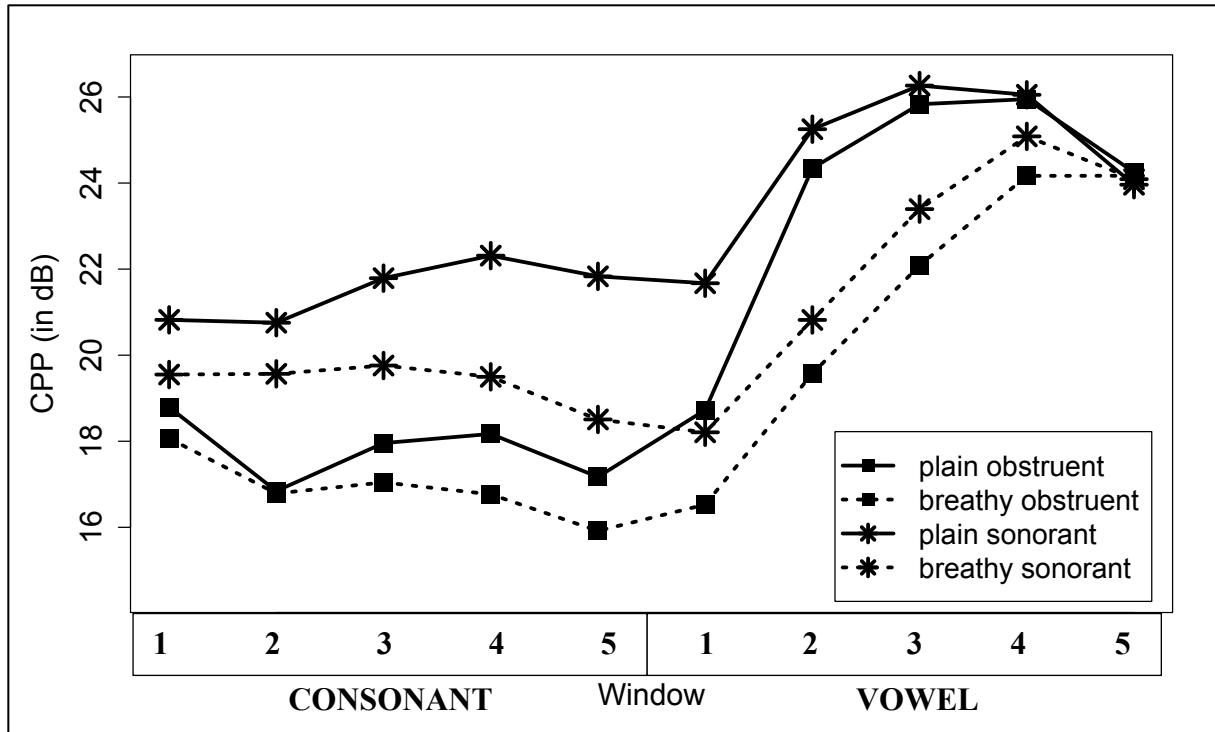
If we step back for a broader view, the picture so far is as follows. Mean CPP values in voiced consonants reflect phonation type: plain and breathy voiced obstruents differ significantly in windows 3 – 5 for both male and female speakers. CPP differences between plain and breathy sonorants, however, differ by gender. Female speakers show very consistent differences in mean CPP values based on phonation type, and the effect holds across vowel contexts and word positions. The one exception is the alveo-palatal sonorants. For the alveo-palatal sonorants—which, recall, are the rhotics [r] and [r<sup>h</sup>]—mean differences based on phonation type do not attain significance. This is perhaps not surprising, as there have been numerous other instances wherein the alveo-palatals diverged from the pattern shared by the labials and dentals. The overall picture, nonetheless, is that CPP cues phonation type differences sonorant-internally for female speakers. No so for the males: sonorant-internal CPP means do not differ for male speakers. Taken together these facts suggest that while CPP is a reliable phonation type cue for obstruents for all speakers, it distinguishes plain from breathy sonorants for females only.

Recall that the reverse was true for some of the parameters examined previously. H1-H2c, for instance, more reliably distinguished plain from breathy sonorants for males than for females—sonorant-internal differences were marginal overall for male speakers, but there was no effect for female speakers. The H1-H2c pattern held true in vowels, as well, acting as a more reliable cue in the male data than in the female data. We may wonder whether the same pattern will emerge here: will CPP in vowels cue phonation type differences more reliably for female than for male speakers?

We move into analysis of CPP values in vowels momentarily, but before doing so a plot of CPP across full syllables is provided in Figure 62. As stated with regards to the full-syllable

plots that have appeared in the previous results sections, the value in looking at a plot like this is that it enables us to envision how things change over the time course of the syllable.

Figure 62 CPP across full syllables by obstruency



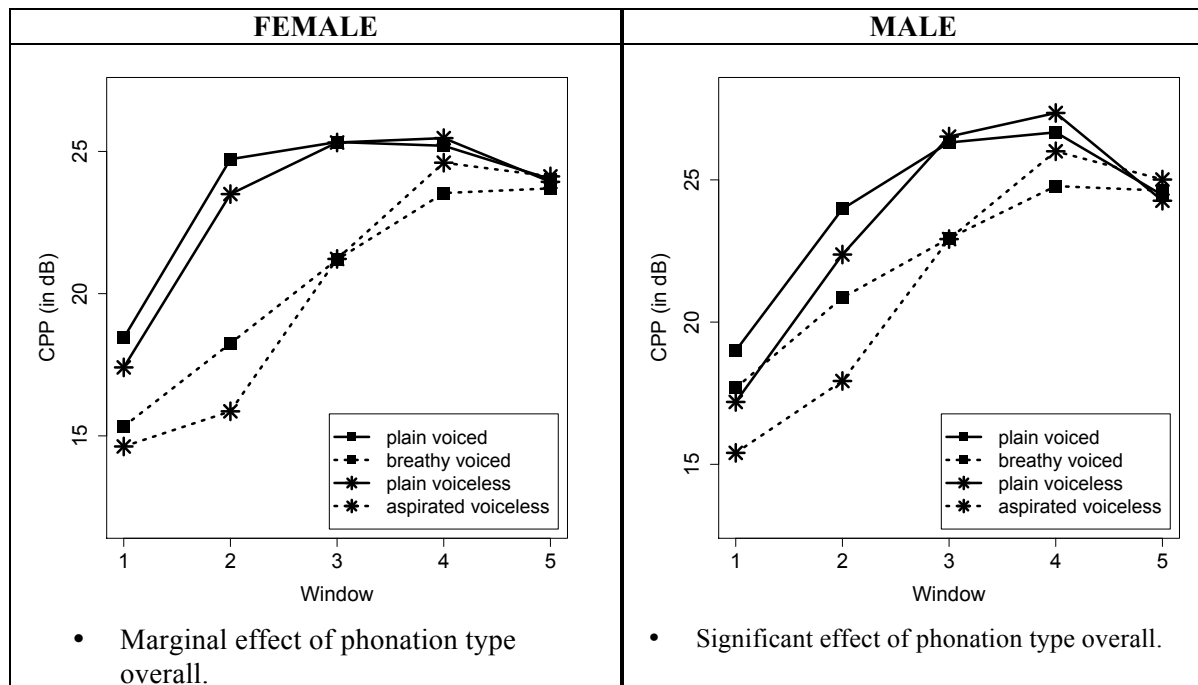
### 6.5.2 CPP – vowel results<sup>31</sup>

An omnibus repeated measures Anova was conducted to investigate whether the phonation type of the preceding consonant affected mean CPP values within vowels, and to determine whether there were interactions between Phonation type and the remaining independent variables of Obstruency, Place of articulation, Vowel context, Word position, Window, and Gender.

<sup>31</sup> Recall that the intervals reported as the PVI and as the vowel in the duration section are concatenated in the F0, spectral, and CPP results sections.

Let us consider vowels after obstruents before turning to vowels after voiced consonants. In the post-obstruent data, there is a significant main effect of Phonation type ( $F(1,6) = 36.54, p = 0.003$ ), as well as a Phonation type by Voice by Gender by interaction ( $F(1,6) = 22.06, p = 0.005$ ). The relevant data are plotted in Figure 63.

Figure 63 CPP in vowels after obstruents by breath, voice, and gender

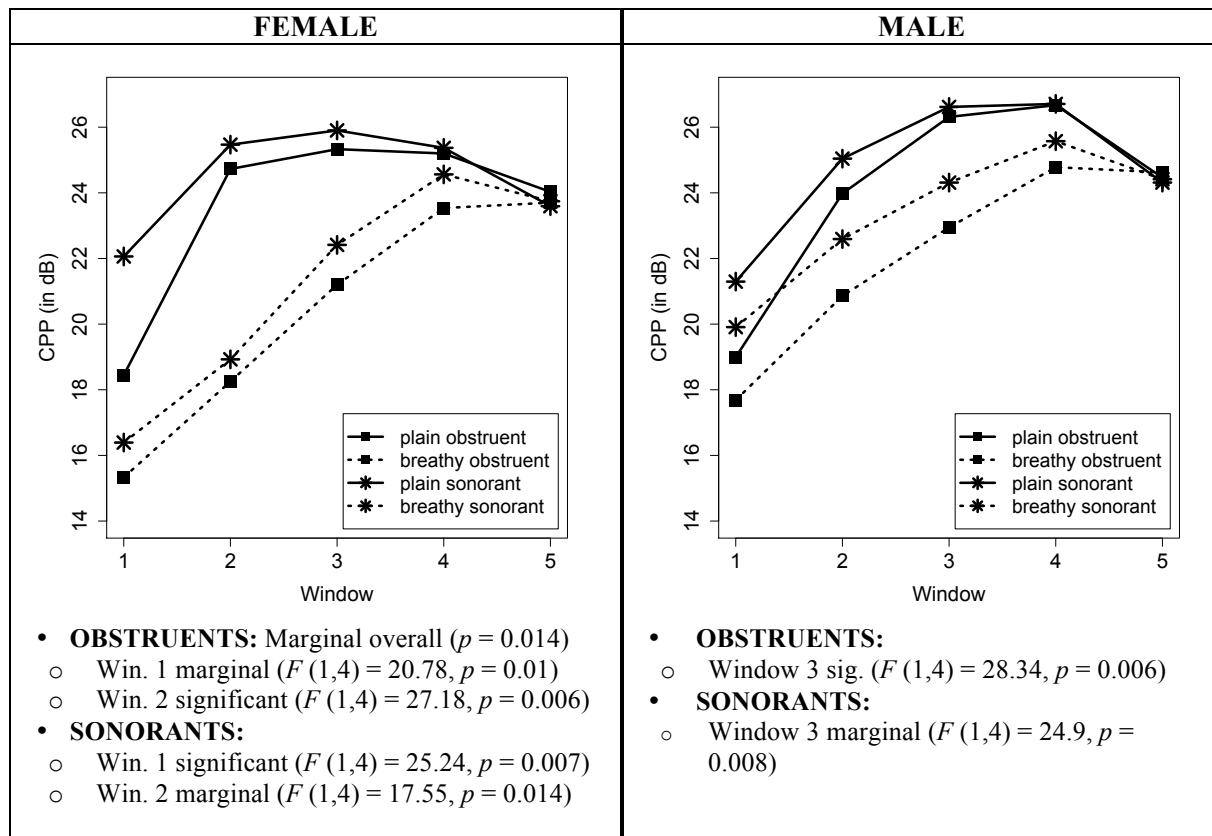


The difference between plain and breathy means approaches overall significance for female speakers ( $F(1,4) = 18.65, p = 0.013$ ) and attains overall significance for male speakers ( $F(1,4) = 33.67, p = 0.004$ ). There is no effect of Voice for either gender. As in the preceding sections, the comparison we are most interested in is the way in which sonorants and obstruents affect subsequent vowels: do they pattern similarly, or are there marked differences? Having ascertained that voiced and voiceless obstruents pattern similarly, then, we will turn now to vowels following voiced consonants (i.e., sonorants and voiced obstruents).

In addition to a main effect of Phonation type on mean CPP values in vowels following voiced consonants ( $F(1,6) = 23.63, p = 0.005$ ), there is also a Phonation type by Obstruency by Gender by Window interaction ( $F(1,6) = 6.073, p = 0.05$ ). These data are presented in Figure 64. Recall that when we looked at mean values for H1-H2c, for example, values after obstruents showed greater differentiation than values after sonorants. The same does not appear to be true with CPP means, at least impressionistically: the differences after sonorants and obstruents appear to be comparable.

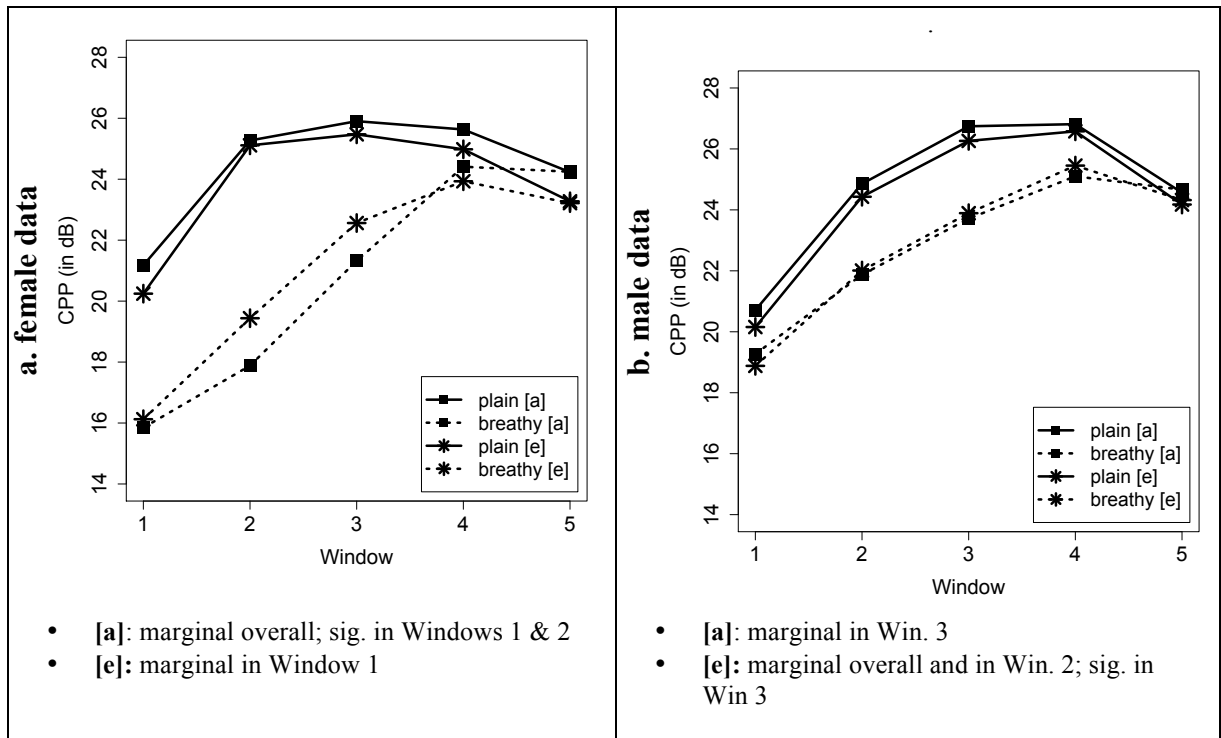
Post hoc tests conducted to untangle the interaction seem to support this observation, although they do not do anything to greatly clarify the story: while the pattern differs for the male and female data, no gender-internal obstruency effect emerges. What that means is that for the females, obstruents and sonorants pattern very similarly: differences after both are marginal or significant in Windows 1 and 2. (All relevant  $f$  and  $p$  values can be found below the graph in Figure 64). For the male speakers, differences in mean CPP values in vowels are significant or marginal in Window 3. Nothing attains overall significance—neither vowels after obstruents nor vowels after sonorants for females, nor for males.

Figure 64 CPP in vowels after voiced consonants by breath, obstruency, and gender



The omnibus Anova also revealed a number of interactions which the post hoc tests help untangle. One of these has to do with Vowel context, which interacts with Phonation type, Window, and Gender ( $F(1,6) = 8.8, p = 0.03$ ). Data are shown in Figure 65, and can be understood as follows. For female speakers, the difference between mean CPP values after plain and breathy voiced consonants is marginally significant overall in [a] ( $F(1,4) = 20.54, p = 0.011$ ), and attains significance in Windows 1 ( $F(1,4) = 33.77, p = 0.004$ ) and 2 ( $F(1,4) = 32.12, p = 0.005$ ). Differences in [e] are not significant overall ( $F(1,4) = 11.78, p = 0.027$ ), but are marginal in Window 1 ( $F(1,4) = 18.49, p = 0.013$ ).

Figure 65 CPP across vowels after voiced consonants ~ vowel contexts



Turning now to the male data, no overall differences emerge as significant, although differences in [e] are marginal overall ( $F(1,4) = 21.45, p = 0.01$ ). In the window-by-window analysis, male means in [e] are marginal in Window 2 ( $F(1,4) = 18.13, p = 0.013$ ) and differ significantly in Window 3 ( $F(1,4) = 43.37, p = 0.003$ ). In fact, differences in [a] are marginal in Window 3 as well ( $F(1,4) = 17.15, p = 0.014$ ). This is in keeping with the trend that we have seen before for male speakers' vowels, wherein differences are significant not in the early windows of the vowel but in later windows.

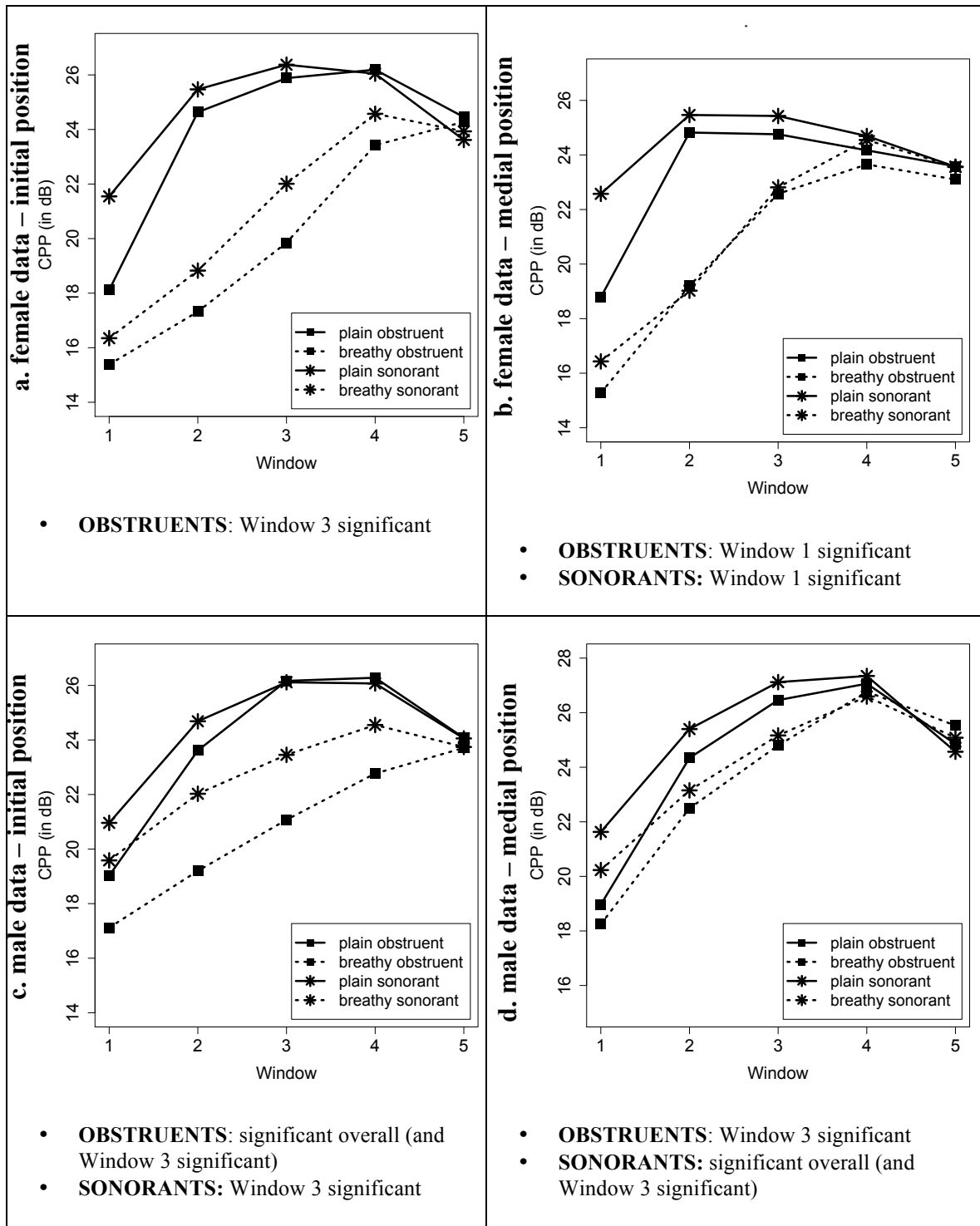
Perhaps the big-picture message to take away from these results is that although CPP is a measure which in theory should not be affected by vowel context, the same pattern we have seen elsewhere emerges here as well. While CPP can act as a cue for phonation type, it does not do so consistently. The vowel context differences that emerge here raise an interesting point: namely, we cannot pin this difference on interference from the vowel formants. How, then, can we

understand it? This questions should be addressed in future work. Overall, however, we once again see a parameter which distinguishes plain from breathy phonation more reliably for one gender than for the other—in this instance, more reliably for the female than for the male speakers. This is true only with regards to sonorant-internal measurements: obstruents show an overall effect of phonation type for both male and female speakers, while sonorants show an effect of phonation type only for the female speakers.

The omnibus Anova also revealed a Phonation type by Word position by Obstruency by Window by Gender interaction ( $F(1,6) = 7.59, p = 0.03$ ). These data are in Figure 66. Female data do not show an overall effect of phonation type, but male data do. For the males, plain consonants trigger higher CPP values in subsequent vowels, but only in certain contexts—after obstruents in word initial position ( $F(1,4) = 26.87, p = 0.007$ ), and after sonorants in word medial position ( $F(1,4) = 46.55, p = 0.002$ ).

The window-by-window analysis returns significant results for both genders. For female speakers, mean CPP values after plain voiced segments are significantly higher than those following breathy voiced sounds in Window 1 in word-medial context. This is true for both sonorants and obstruents. In word initial position, however, obstruent means differ in Window 3, and sonorant means do not attain significant differences overall or in any individual window.

Figure 66 CPP across vowels after voiced consonants ~ word positions



Recall that males showed overall differences after obstruents in word initial position and after sonorants in word medial position. The window by window analysis reveals that, in both



instances, the difference between plain and breathy means is significant only in Window 3. Once again, it is not yet clear why differences should manifest midway into the vowel, so far removed from the consonant triggering the differences. The pattern is consistent for males, however, and it is important to keep in mind that while differences are significant at only this one time point, it is enough to drive an overall significance.

### 6.5.3 CPP recap

Like H1-A3c, CPP diverged from the measures reported in previous sections in that a significant consonant-internal effect of Phonation type was found. Differences were in the expected direction: overall, breathy voiced consonants have lower mean CPP values than plain voiced consonants. As with the measures discussed previously, however, many differences related to gender also emerged. While F0 and the lower-frequency spectral measures of H1-H2c and H1-A1c cued phonation type distinctions sonorant-internally for male speakers but not for female speakers, the reverse is true here. Sonorant-internal differences in mean CPP values are significant overall for female speakers, but are not significant for male speakers. Interestingly, voiced obstruents were very consistent across gender, vowel context, and word positions—differences in voiced obstruents were significant in Windows 3-5 across the board.

Moving into vowels, a main effect of Phonation type was again present, but post hoc analyses revealed something interesting. Overall, differences after voiced obstruents were marginal for female speakers but not for male speakers. Differences after sonorants did not attain significance for either gender. Here again, then, we see more evidence of the gender differences that have come up multiple times. Male and female speakers cue breathy phonation differently, and females utilize CPP to cue breathiness in sonorants more reliably than males. Furthermore,

consonant phonation type triggers differences in subsequent vowels more reliably for females than for males, and voiced obstruents have an advantage over sonorants in that they trigger greater differentiation in subsequent vowels.

Having presented the findings from the last of the measures taken herein, we can now step back from the individual results in order to develop a sense of the big-picture view. The rest of the dissertation proceeds as follows. The next chapter provides a brief overview of the results presented thus far. How do they fit together, and what do they tell us about how and where breathy phonation is cued—by males and by females, by sonorants and by obstruents, in consonants and in vowels, and so forth? We then move into Chapter 8, which provides a discussion of the lessons that can be drawn from these results as well as some of the issues they raise.

Foremost among the topics that must be discussed are the gender effects that have been revealed again and again in the results sections. The implication is that breathy voice is cued very differently in males and females, and this point is addressed in detail. Another critical point of discussion revolves around obstruency. Obstruents and sonorants have shown some marked differences with regards to how breathy phonation is cued: do these differences serve to illuminate the question of why phonemic breathy voice is so typologically rare? As noted previously, I argue that the answer is yes: I argue that the fact that breathy voice is cued more robustly by obstruents than by sonorants and the fact that breathy voiced obstruents are more typologically common than breathy voiced sonorants are connected. These and other points are addressed in Chapter 8. First, however, we move into a review of all the results.

## 7 An overview of all results

In the preceding chapters, results from a number of measures taken on both consonants and vowels have been presented. These results reveal a number of differences in the way that breathy voice is cued: we have seen differences between male and female speakers, as well as differences between sonorants and obstruents, between the low vowel [a] and higher vowel [e] contexts, and between word-initial and word-final contexts. The details provided in each of the preceding sections are valuable, for they contribute to a more comprehensive understanding of the way in which each the acoustic parameters assessed contribute to making phonation type contrasts in Marathi. It is also worthwhile to pull away from the fine-grained detail, however, and consider the big picture. To this end, a review of the major findings is presented in Table 15.

**Table 15 Success of the measures of duration, F0, H1-H2c, H1-A1c, H1-A2c, H1-A3c, and CPP in distinguishing phonation types in sonorants and obstruents produced by male and female speakers.**

		MEASURE															
		CONSONANT						VOWEL									
		CONSONANT DUR.	F0	H1-H2c	H1-A1c	H1-A2c	H1-A3c	CPP	PVI	VOWEL DUR.	F0	H1-H2c	H1-A1c	H1-A2c	H1-A3c	CPP	
FEMALES	SON.	✓ (medial)				✓	✓	✓	✓	✓	✓ ([e] after alveo-palatals)	✠ (initial)	✠		✓	✓	✓ (Window 1)
	OBST.							✓	✓	✓ ([e] after alveo-palatals)	✠ (initial)	✠		✓	✓	✓ (Window 2)	
MALES	SON.	✓ (medial)	✓			✓	✓		✓	✓ ([e] after alveo-palatals)	✠ (Windows 1&2)	✓		✓	✓	✠ (Window 3)	
	OBST.		✓					✓	✓	✓ ([e] after alveo-palatals)	✠ (Windows 1&2)	✓	✓	✓	✓	✓ (Window 3)	

✓ = significant; ✠ = marginal; dark grey boxes indicates that the parameter in question was not measured (i.e. H1-H2c was not measured within obstruents).

We may also set the higher-level interactions aside for the moment and considering only those parameters which showed a main effect of phonation type. In looking at these data—

presented in a slightly different format in Table 16—it becomes clear that while CPP is more reliable for females and F0 and H1-H2c for males, there are three parameters which remain consistent for male and female speakers across vowel contexts and word positions. These are **H1-A2c, H1-A3c, and the Pre-Vocalic Interval.**

**Table 16 Parameters which reflect an overall effect of Phonation type**

<b>WOMEN</b>		
SONORANT	VOWEL	
<ul style="list-style-type: none"> <li>• H1-A2c (<math>m^h &gt; m</math>)</li> <li>• H1-A3c (<math>m^h &gt; m</math>)</li> <li>• CPP (<math>m &gt; m^h</math>)</li> </ul>	<ul style="list-style-type: none"> <li>• PVI (<math>m^h &gt; m</math>)</li> </ul>	<ul style="list-style-type: none"> <li>• H1-A2c (<math>m^h &gt; m</math>)</li> <li>• H1-A3c (<math>m^h &gt; m</math>)</li> </ul>
OBSTRUENT	VOWEL	
<ul style="list-style-type: none"> <li>• CPP (<math>b &gt; b^h</math>)</li> </ul>	<ul style="list-style-type: none"> <li>• PVI (<math>b^h &gt; b</math>)</li> </ul>	<ul style="list-style-type: none"> <li>• H1-A1c (<math>(b^h &gt; b - \text{ONLY in [a]})</math>)</li> <li>• H1-A2c (<math>b^h &gt; b</math>)</li> <li>• H1-A3c (<math>b^h &gt; b</math>)</li> </ul>
<b>MEN</b>		
SONORANT	VOWEL	
<ul style="list-style-type: none"> <li>• F0 (<math>m &gt; m^h</math>)</li> <li>• H1-A2c (<math>m^h &gt; m</math>)</li> <li>• H1-A3c (<math>m^h &gt; m</math>)</li> </ul>	<ul style="list-style-type: none"> <li>• PVI (<math>m^h &gt; m</math>)</li> </ul>	<ul style="list-style-type: none"> <li>• H1-H2c (<math>m^h &gt; m</math>)</li> <li>• H1-A1c (<math>(b^h &gt; b - \text{ONLY after initial consonants})</math>)</li> <li>• H1-A2c (<math>m^h &gt; m</math>)</li> <li>• H1-A3c (<math>m^h &gt; m</math>)</li> </ul>
OBSTRUENT	VOWEL	
<ul style="list-style-type: none"> <li>• F0 (<math>b &gt; b^h</math>)</li> <li>• CPP (<math>b &gt; b^h</math>)</li> </ul>	<ul style="list-style-type: none"> <li>• PVI (<math>b^h &gt; b</math>)</li> </ul>	<ul style="list-style-type: none"> <li>• H1-H2 (<math>b^h &gt; b</math>)</li> <li>• H1-A1c (<math>b^h &gt; b</math>)</li> <li>• H1-A2c (<math>b^h &gt; b</math>)</li> <li>• H1-A3c (<math>b^h &gt; b</math>)</li> </ul>

H1-A2c and H1-A3c values reliably distinguish between plain and breathy phonation for both male and female speakers sonorant-internally as well as in vowels after both sonorants and voiced obstruents. The temporal measure of Pre-Vocalic Interval reliably distinguishes plain from breathy voiced consonants for both genders, as well. Again, this is true for both sonorants and obstruents, across all vowel and word-position contexts. Although not included in Table 16 because it is contextually limited to word-medial position, **sonorant constriction duration** also cues phonation type for both male and female speakers. In medial position, breathy sonorants are significantly longer than plain sonorants overall.

While more extensive treatment of these findings—including a discussion of how they fit with those of previous researchers—can be found in the next chapter, a few brief comments can be made here as well.

First, let us address the consistently useful measures of H1-A2c and H1-A3c. That these prove reliable across genders and vowel contexts is not surprising: of the spectral measures assessed within this dissertation, these are the higher-frequency measures. As such, it is to be expected that they are less affected by the relative closeness of F1 to H1, which has proven problematic in this research both in high vowel contexts and in the lower frequency measures for female speakers. Furthermore, Harris (2009) noted that the spectral characteristics of Sumi sonorants differed based on phonation type above 1000 Hz—a threshold above which A2 and A3 fall. For these reasons, finding H1-A2c and H1-A3c to be relatively stable parameters with which to cue phonation type is unsurprising. Previous research has been divided as to the utility of these measures, however, and so this point will be returned to in the Discussion.

It also becomes clear in looking at Tables 15 and 16 that speakers employ multiple acoustic parameters to cue breathy phonation, and that males and females do so quite differently.

In addition to the temporal and high-frequency measures which are reliable cues for both genders, consonant-internal CPP values reliably distinguish plain from breathy sonorants for female speakers. For the females, CPP values are significantly different in the first 20% of subsequent vowels after sonorants, as well.

Male speakers, meanwhile, do not show the same pattern. For them, consonant-internal CPP differences cue phonation type only for obstruents; in sonorants, the contrast is cued by F0, H1-H2c, and H1-A1c. For males, fundamental frequency cues phonation type consonant-internally for both obstruents and sonorants, with breathy voiced sounds showing lower F0 values than plain sounds. They also show significant differences in vowels in terms of the low-frequency measure of H1-H2c, and this is true after both sonorants and obstruents. Vowels also show significant differences based on consonant phonation type after obstruents in terms of H1-A1c; after sonorants, the difference is only significant word-initially.

In the end, the large-scale picture revealed by the Marathi data is that multiple acoustic parameters contribute to cueing phonation type differences in Marathi sonorants and obstruents. These results are in keeping with what we may expect, for acoustic research increasingly affirms that this is the pattern in human language. Contrasts are cued not by a single parameter, but by a multitude of parameters—or, as stated by Esposito (2006), “[in] linguistics contrasts, redundancy is the rule...not the exception” (p. 45).

## 8 Discussion

This dissertation provides the first in-depth acoustic analysis of phonation types in Marathi obstruents and sonorants. Recordings from five male and five female speakers have been analyzed, and data from a number of acoustic parameters have been presented. Measures assessed include the temporal measures of consonant duration (meaning constriction duration for sonorants, and closure duration for obstruents), a measure of lag-time or the breathy interval here referred to as the PVI, and vowel duration. Fundamental frequency values both within voiced consonants and in subsequent vowels was also measured, as were the spectral measures of H1-H2c, H1-A1c, H1-A2c, and H1-A3c. Finally, the cepstral measure of Cepstral Peak Prominence was also included. This chapter addresses the research questions posed at the beginning of the dissertation in light of the data that has been collected. It also includes discussion of several important issues raised by the data. The rest of the chapter proceeds as follows.

First, in Section 8.1, the present findings are compared and contrasted with previous findings. This is done in order to situate the current work: how does it fit with what we know about the acoustic correlates of phonation type contrasts from other languages? Where does it diverge, and what new insights does it give us?

Section 8.2 focuses on the idea of trading relations. Time and again in the present data, we have seen that in contexts where one cue is weak or absent, other cues are reliably present. Specific examples from the dissertation are highlighted, as is the larger relevance of these findings.

Section 8.3 revolves around the gender effect found so consistently in the Marathi data analyzed herein. In numerous acoustic parameters, male and female speakers patterned differently. This is not shocking: gender differences have been reported frequently in previous

research on voice quality and phonation types. Nonetheless, discussion of this finding is critical. I propose sexual dimorphism as one possible explanation: the basic laryngeal differences between men and women have multiple acoustic ramifications, which may easily underlie many of the differences seen in this data. The suggestion is that phonation type contrasts are cued differently in male and female speech because the reliability of cues differ by gender—and again, here I must stress that while I use the term gender for the sake of conformity, what I really mean is biological sex.

The final portion of the discussion, Section 8.4, addresses one of the major issues posed at the beginning of this dissertation, namely that of the relative crosslinguistic rarity of breathy voiced sonorants. Phonemic breathy phonation is rare in general, and to our knowledge it is employed more often with obstruents than with sonorants. Furthermore, the existing literature on breathy voiced sonorants is scarce. This dissertation aims to shed light on the acoustic correlates of breathy voice in sonorants, and to determine whether and how the cues for breathiness in sonorants pattern with or differ from the cues found in obstruents. The supposition was made that by developing an understanding of the acoustic correlates of phonation type contrasts in sonorants, we may gain insight into why these sounds are so typologically uncommon. By comparing the acoustic correlates of breathiness in obstruents and sonorants, we have discovered that sonorants do not cue phonation type as robustly as obstruents in numerous measures. I propose that this fact underlies the typological rarity of breathy voiced sonorants—that when taking all the acoustic information into account, phonation type contrasts are less auditorily distinct in sonorants and as such make for poorer contrasts in sonorants than in obstruents.



## 8.1 The present results in relation to previous findings

We may wonder how the Marathi results correlate with those reported in previous studies on breathy phonation, and the answer is that they fall in line quite nicely. In fact, a very recent study on phonation type contrasts in Gujarati reports very similar findings (Khan 2012). Ten native speakers of Gujarati (3 male, 7 female) produced 26 stimuli each, which contained either breathy or modal vowels. A host of measures were taken, and the results indicated that the spectral measures of H1-H2, H2-H4, H1-A1, H1-A2, and H1-A3 successfully distinguished breathy from modal phonation—sometimes over the course of an entire vowel, and other times at distinct time points in the vowel.<sup>32</sup> With the exception of H2-H4, which was not measured in this dissertation, these results overlap with those from the present study to a large degree.

Interestingly, Khan also assessed the effect of breathy phonation on F0 values: the plain and breathy versions of [e] diverged in the expected direction, with breathy [e] (denoted in the literature on Gujarati as [ɛ̃]) showing lower values than modal [e], but [a] and [o] showed the opposite pattern. Breathly [a] and [o] returned higher F0 values than their plain counterparts. This is of particular interest to us because females in the present study showed a similar pattern.

Many previous studies have found an increase in the amplitude of H1 relative to that of H2 or of F1 to successfully distinguish breathy from modal phonation, with breathy phonation associated with higher values. Recall that this was the pattern found for male speakers herein, and that males were the only subjects who showed consistently significant differences in H1-H2c based on phonation type. A similar pattern is reported for !Xóõ (Bickley 1982, Ladefoged 1983), Gujarati (Bickley 1982, Fischer-Jorgensen 1967), and Hmong (Huffman 1987) among others.

---

<sup>32</sup> Cepstral peak prominence, three harmonics-to-noise ratio measures, and Closed Quotient also proved reliable cues.

We may also consider H1-A2 and H1-A3, for there is some debate in the literature on the utility of these measures—or better said, less consistency with regards to the results found in different languages. Esposito (2006), for instance, found that H1-A2 distinguished modal vowels from breathy vowels in seven out of the ten languages and dialects assessed in that study, while H1-A3 cued the distinction in nine out of ten languages. Interestingly, however, H1-A3 cued breathy phonation only for male speakers of Santa Ana del Valle Zapotec, and there was an added wrinkle. While H1-A3 values are expected to be greater for breathy than for modal phonation, and Marathi patterns in the expected direction, Esposito reports with modal values of H1-A3 were greater than breathy values for five out of the ten languages assessed, and lower in four.

With regards to Hindi, Dutta (2007b) reports a great deal of both inter- and intra-speaker variability in the measures of H1-A2 and H1-A3. Dutta further reports that the distinction between modal and breathy consonant phonation did not trigger reliable or consistent H1-A3 differences in subsequent vowels. Clearly, the pattern is different here; in Marathi, differences in these acoustic measures are reliable, pattern in the expected direction, and distinguish plain from breathy phonation across the board.

This is perhaps enough of an overview for the present purposes. The idea here is simply to highlight the fact that the findings presented herein are in keeping, overall, with those reported by previous researchers. Also in keeping with current lines of thought in the body of literature devoted to linguistic voice quality is the idea that no one cue will be found that acts as a sole cue to phonation type: rather, multiple cues combine and interact in order to make the relevant contrasts. Khan (2012) notes that “breathy voice in Gujarati is a dynamic, multidimensional feature, surfacing through multiple acoustic cues that are potentially relevant to the listener.” (p.

780). While the present study has not focused as much on tracking the way cues trade off dynamically—meaning that our focus has been on the way cues trade off in different contexts rather than on tracking the way cues shift over the time course of a segment—the same is clearly true in Marathi, and we now move into a discussion of the way in which the various acoustic correlates of breathiness interact with one another.

## **8.2 Trading relations**

Time and again in the Marathi data reported in this dissertation, we have seen evidence of trading relations. Where one cue is weak or absent, another cue is present. One clear example comes from the effects related to word position. As noted in an earlier chapter, phonation type contrasts in sonorants are not positionally restricted in Marathi, or in Sumi or Camling—in at least these three languages which contain breathy voiced sonorants in their phonetic inventories, the contrast is realized word-initially and word-medially. The current study shows, however, that the acoustic cues related to phonation type contrasts in sonorants are variable based on word positions.

Recall that there are significant durational differences based on phonation-type in sonorants, but not universally: breathy sonorants are longer than plain sonorants in word-medial position only. Word-initially, then, sonorants lack this temporal cue for breathiness. It turns out that the reverse pattern applies for the spectral measure of H1-A1c in vowels, however: while voiced obstruents in both word-initial and word-medial position trigger significant H1-A1c differences in subsequent vowels, sonorants trigger differences only when they occur word-initially.

Furthermore, recall that H1-A2c is one of the parameters that has most reliably distinguished between phonation types in Marathi in this dissertation. Although there was a main

effect of phonation type on H1-A2c values in vowels after sonorants, post hoc tests revealed that the effect after medial sonorants was not significant. This indicates that the effect after initial sonorants is great enough to drive overall significance. Here again, then, we have a similar story: word-initial breathy sonorants lack the temporal cue of increased constriction duration, but are cued spectrally through increased H1-A2c values in subsequent vowels. Meanwhile, word-medial breathy sonorants lack the spectral cue, but are cued temporally through increased constriction duration.

It is not the case that breathiness triggers increased constriction duration in all word-medial sonorants, however: while the labial and dental sonorants are longer when breathy, alveo-palatal sonorants show no differences in constriction duration based on phonation type, and this leads us to another trading relation. What the alveo-palatal sonorants lack in terms of a consonantal temporal cue, they make up for in several ways. First, they pick up low-frequency spectral cues that the other sonorants do not have. While the bilabials and dentals do not show sonorant-internal H1-H2c differences for male or female speakers, the alveo-palatals do—for male speakers only. Breathily alveo-palatal sonorants show significantly greater H1-H2c values than their plain counterparts in the male data.

In fact, the alveo-palatals often pattern differently than the other sonorants. While the bilabial and dental sonorants consist of the nasals [m, m<sup>h</sup>, n, n<sup>h</sup>] and the approximants [v, v<sup>h</sup>, l, l<sup>h</sup>], the alveo-palatal sonorants are the rhotics [r] and [r<sup>h</sup>]. Beyond this, as has been mentioned previously, they are produced with great intra-speaker variability.<sup>33</sup> Additional ways in which they diverge from the bilabials and dentals can be seen with regards to CPP. Female speakers

---

<sup>33</sup> Impressionistically, this seems like free variation in the data I have analyzed for this dissertation. The rhotics are sometimes produced with a more trill-like articulation, sometimes with a more tap-like articulation, and sometimes as clear approximants. This happens both word-initially and word-medially, in both vowel contexts, and with both the plain and the breathy rhotic.

consistently cue breathy phonation with consonant-internal CPP differences. This is one of the most reliable cues for female speakers, and yet it is absent in the alveo-palatal sonorants.

Finally, the alveo-palatals as a group also trigger an interesting effect with regards to vowel duration. For both sonorants and obstruents, [e] after breathy voiced consonants is significantly longer than [e] after plain consonants. This stands out: because part of the signal categorized by others as vowel was instead measured as PVI in the present work, many of the vowel-duration results that other researchers report were absent here. More specifically, while breathy voiced phonation has been associated with increased duration of subsequent vowels in Hindi (Dutta 2007), the pattern does not hold overall in Marathi data. Even when apportioning the entire breathy interval to the PVI, however, [e] after breathy voiced sounds is still longer than [e] after plain sounds. Thus while alveo-palatals as a whole lack some of the acoustic cues that distinguish between phonation types in the bilabials and dentals, we see that other acoustic parameters cue the distinction in alveo-palatals.

We might also note that the findings related to the duration of [e] after alveo-palatals is particularly interesting because [e] is precisely the vowel context wherein other cues are often weaker or absent. H1-A1c values in [a] after obstruents, for instance, show significant differences based on consonant phonation type for both male and female speakers; the difference in [e], however, is marginal for males, and is significant only in isolated windows of the vowel for females. For the males speakers, who showed marginally lower F0 values after breathy voiced obstruents as compared with values after plain voiced obstruents, differences are marginal for the first 80% of subsequent [a] vowels, but are marginal for only the first 40% of subsequent [e] vowels. In multiple ways, then, [e] is an environment wherein certain cues are absent or

weaker. Yet with regards to duration after alveo-palatals—a class of sounds which lack some of the cues present for other places of articulation—[e] yields a temporal cue that [a] lacks.

There are multiple additional instances of this kind of trade-off, all of which serve to underscore a critical point: phonation type contrasts are cued by multiple parameters. Cues are variable in numerous ways: they change from one position or context to the next. Cues that are present for female speakers may be absent for male speakers. Cues that are present in word initial position may be absent in word-medial position. The pattern observed in the low vowel context of [a] may be different, weaker, or absent in the higher-vowel context of [e].

These facts mean that two important points bear consideration. The more methodological of the two points is related to research design. The acoustic correlates of breathy phonation have proven varied and variable. Had the present study excluded female speakers, for instance, we would not have found that F0 differences do not cue phonation type contrasts in female speech, nor that H1-H2c differences in subsequent vowels cue the distinction for female speakers to a lesser degree than for male speakers. The implication for future production studies that seek to gain a comprehensive understanding of phenomena such as the acoustics of phonation type contrasts is that inclusion of both genders is absolutely critical. Study of one gender cannot be relied upon to yield a complete acoustic picture. The sex-based differences found in this study have important ramifications for future perception studies, as well, where F0 and H1-H2c are not predicted to aid listeners in identifying breathy voice produced by female talkers. Low-pass filtering of stimuli may seriously impair the ability of listeners' to distinguish plain from breathy phonation as produced by female talkers, for instance, whereas in theory male speech may still be distinguishable.

The more theoretical of the two points has to do with the nuanced picture revealed by the Marathi data. Phonation type contrasts in Marathi are cued by multiple acoustic parameters, and the interplay between these parameters is both subtle and profound. Different cues manifest in different contexts; the implication is that listeners are adept at tuning into different cues in different contexts. While this cannot be stated with any certainty until perception studies have been conducted, the implication from the present results is that listeners may be attuned to one set of cues for female talkers and a separate set of cues for male talkers. There is certainly some overlap: there are cues that cut across gender, vowel context, word position, and place of articulation. The ultimate message, though, is that the precise constellation of acoustic parameters that serve to cue breathy phonation is not absolute. Phonation type contrasts may be cued by many things—including but not limited to acoustic measures related to Open Quotient, speed of vocal fold vibration, speed of vocal fold closure, and a combination thereof—and listeners must therefore be sensitive to variation of these parameters.

Again, it is important to point out that while the phonetic data reported in this dissertation show a high degree of variability across speaker genders, vowel contexts, and word positions, the phonological component of the grammar does not show a similar sensitivity to these factors. Rather, the distribution of breathy voiced sonorants is not hampered by the acoustic differences presumably found in the different contexts.<sup>34</sup> This is illustrated in Table 17. Sumi data are from Harris (2009) and Camling data are from Ebert (1997, 2003).

---

<sup>34</sup> By this, I simply mean that we have data regarding the acoustic patterns in Marathi but do not have similar data regarding acoustic variability across distinct word and vowel positions in Sumi and Camling.

**Table 17 Breathy voiced sonorants in multiple word positions and vowel contexts in Marathi, Sumi, and Camling<sup>35</sup>**

	WORD-INITIAL	WORD-MEDIAL	LOW-VOWEL CONTEXT	HIGH-VOWEL CONTEXT
<b>Camling</b>	[l <sup>h</sup> amma] 'catch'	[lamma <sup>h</sup> a] 'dump'	[l <sup>h</sup> amma] 'catch'	[m <sup>h</sup> imma] <sup>36</sup> 'fight'
<b>Marathi</b>	[n <sup>h</sup> avi] 'barber'	[pun <sup>h</sup> a] 'again'	[r <sup>h</sup> as] 'deterioration'	[tum <sup>h</sup> i] 'you all'
<b>Sumi</b>	[n <sup>h</sup> a] 'close (dish)'	[an <sup>h</sup> a] 'mucus'	[l <sup>h</sup> a] 'flay'	[al <sup>h</sup> u] 'flea'

The variation across different contexts may yield diminished auditory distinctiveness overall, resulting in the contrast being absent from a language—recall, for instance, that while Camling shows phonemic breathiness among its sonorants, most of its close relatives have phonation type contrasts among their obstruents but not among their sonorants (Ebert 2003). This is an interesting point, and the topic of obstruent-sonorant differences is returned to in Section 8.4. For now, let us focus on this observation: given the data from Camling, Marathi, and Sumi, it seems that if a language includes phonation type contrasts in its sonorants, the contrast need not be positionally-restricted.

This presents a challenge, as much of the discussion herein has pointed to diminished auditory distinctiveness as a partial explanation for the typology of breathy voiced sonorants. Why, then, should it be the case that the phonology of these sounds in languages where they occur does not reflect positional restrictions based on the phonetic information brought to light in this dissertation? At issue here is the question of how to connect phonetics and phonology, and

<sup>35</sup> Note that transcription here is copied directly from Ebert, and the double [m] word-medially presumably indicates length.

<sup>36</sup> from *Languages of the World/Materials 103* (Ebert 1997). Available at [http://www.archive.org/stream/rosettaproject\\_rab\\_phon-1/rosettaproject\\_rab\\_phon-1\\_djvu.txt](http://www.archive.org/stream/rosettaproject_rab_phon-1/rosettaproject_rab_phon-1_djvu.txt)



this is no simple task. As Hayes (1999) points out in an article about phonetically-driven phonology, “phonetics involves gradient and variable phenomena, whereas phonology is characteristically categorial and far less variable” (p. 6). Furthermore, Hayes notes, “a research result in phonetics is not the same as a phonological constraint. To go from one to the other is to bridge a large gap.... In many cases, the phonetic research that explains the phonological pattern has been done very well and is quite convincing; it is only the question of how to incorporate it into a formal phonology that is difficult” (Hayes 1999: 5). This is precisely the difficulty raised by the Marathi data: how can we reconcile the phonetic findings presented herein with the phonology? One possible explanation that we may consider has to do with phonological simplicity.

At its core, an argument in favor of phonological simplicity as a shaping force is in line with the reasoning developed by Gordon (2002) about the phonetics and phonology of phenomena dealing with syllable weight, such as weight-sensitive stress. Gordon argues that the syllable weight distinctions found in natural languages reflect a compromise between phonetics and phonology: phonetically, the best syllable weight distinctions are those that are most auditorily distinct. Gordon argues that distinctions which prove overly complex in terms of phonology are disfavored, however, with the result that “languages eschew weight distinctions that are structurally too complex, even if they are phonetically more effective than certain simple distinctions” (p. 76). For example, Gordon (2002) includes an acoustic analysis of Chickasaw, a language with a three-way weight distinction where CVV syllables are heaviest, CVC syllables are heavy, and CV syllables are light (i.e.,  $CVV > CVC > CV$ ). Acoustic analysis indicates that different vowels are associated with quite distinct amounts of overall energy, however—the acoustic measure Gordon finds to correspond most accurately with the weight class divisions

found in languages with weight-sensitive phonological phenomena. In Chickasaw, syllables with long /a:/ and long /u:/ are phonetically quite distinct from other syllables containing long vowels. This indicates that a more complex distinction—whereby syllables with long /a:/ and/or long /u:/ rimes would form the heaviest weight class, followed by CVV > CVC > CV—would best reflect the phonetic properties of the Chickasaw inventory. The language does not employ such a distinction, however—because, Gordon argues, employing such system would “exceed the upper threshold of phonological simplicity” (p. 64).

Similarly, one possible explanation for the mismatch between the phonetic and phonological patterns of breathy voiced sonorants in Marathi—and presumably other languages—is that they are the result of a compromise between multiple linguistic imperatives. Specifically, while one linguistic imperative may well be that phonemic contrasts should be auditorily distinct, as claimed by dispersion theories (Flemming 1995; Liljencrants & Lindblom 1972; Lindblom 1986, 1990) and as argued in this dissertation, a pressure for phonological simplicity may also contribute to the ultimate shape of the phonological component of the grammar. In fact, when developing his formulation of Dispersion Theory, Flemming makes overt reference to the fact that “the selection of phonological contrasts is subject to three functional goals” (Flemming 1995: 24). These are listed as:

- i. Maximize the number of contrasts
  - ii. Maximize the distinctiveness of the contrasts
  - iii. Minimize articulatory effort
- (Flemming 1995: 24)*

One way to think about the fact that the phonetic variation we have seen across word positions and vowel contexts does not yield positional restrictions on breathy voiced sonorants in the Marathi phonology is to consider the fact that the patterns observed result from compromise

between the three functional goals described by Flemming and a pressure for phonological simplicity.

It is important here to note that the relationship between these factors is presumed to be complex, and multidirectional: a pressure for phonological simplicity may result in contrasts surfacing in positions where they are not maximally acoustically distinct, but the fact that these contrasts show acoustic variation based on word position and vowel context may also work to shape the overall grammar. Recall, for instance, that we have made the anecdotal observation that breathy voiced sonorants in the Marathi lexicon appear to occur far more often before [a] than before any other vowel.<sup>37</sup> Beyond this, phonation type contrasts are more common in obstruents than in sonorants typologically. The possible story, then, is that the phonetic reality of breathy phonation in sonorants eventually contributes to the statistical occurrence of these sounds in the lexicon—the predominance of breathy sonorants in the acoustically strong vowel context of [a] in Marathi, for instance—and perhaps even to a loss of the contrast from the phonology. At the same time, however, the pressure for phonological simplicity may dictate that when a contrast is present in the grammar, it holds across contexts that are subject to phonetic variation.

Again, it is critical to note that “the question of how to incorporate [phonetic results] into a formal phonology...is difficult” (Hayes 1999: 5). The explanation posited above is one possibility: the question certainly remains open, however, and will provide rich territory for future investigation.

---

<sup>37</sup> See Gordon (2002) for an in-depth analysis of the interplay between phonetics and phonology in determining weight classes in languages with phenomena such as weight-sensitive stress patterns.

Having addressed trading relations, we now turn to a more thorough discussion of gender. Speaker gender is one of the key sources of variation that we have seen in the data presented herein, and possible explanations for the variation found in the Marathi data are discussed.

### 8.3 Gender differences

Let us review the major points of variation seen in the male and female data herein. Some parameters show no or very little gender-based variation: these include the temporal measures of consonant duration, PVI, and vowel duration, as well as the spectral measures of H1-A2c and H1-A3c. For now, we will set those measures aside and focus on the parameters for which gender effects emerge. These include: **fundamental frequency**, **H1-H2c**, and **H1-A1c**, which were more reliable cues for breathy phonation for males than for females, and **Cepstral Peak Prominence**, which was a more reliable cue for females than for males. A review of the key points of relevance related to each of the above-mentioned measures is now provided.

F0 values plotted across the time course of full syllables appear in Figure 67. Male data for sonorants and voiced obstruents by phonation type is in (a.); female data is in (b.). The prediction for fundamental frequency was that breathy phonation would have a lowering effect on F0, and this turned out to be the case consonant-internally for male speakers: mean F0 values are lower in breathy voiced obstruents and sonorants than in their plain counterparts. No significant F0 differences based on phonation type appeared for females, however, and this was true both consonant-internally and in subsequent vowels. Differences in subsequent vowels are marginal for male speakers. Part of the way in which phonation type differences are cued in male speech, then, involves differences in the basic rate of vocal fold vibration; not so for the females, however.

Male speakers also use H1-H2c and H1-A1c values to cue phonation type differences. These data are in Figure 68 and 69. As consonant-internal spectral measures were taken only for the sonorants, mean values are plotted across sonorant-vowel syllables by gender and phonation type. H1-H2c plots are in Figure 68, and H1-A1c plots in Figure 69.

For the males, consonant phonation type triggers significant H1-H2c differences in vowels after both sonorants and voiced obstruents (shown in 68a.), and significant H1-A1c differences after voiced obstruents (shown in 69a.). Consonant-internal differences are significant for neither gender, but we can see when looking at the data that female values (in 68b. and 69b.) are opposite the expected pattern consonant-internally, with breathy values lower than plain values. This is true in both measures. The direction in which values diverge flips partway into the vowel, but even within vowels the male values are far more differentiated than the female values. H1-H2c values in subsequent vowels are marginal for the female speakers, but that is all.

If we take H1-H2c to reflect Open Quotient, the indication is that another piece of the story with regards to the cueing of phonation type differences in male speech involves differences in OQ. Furthermore, while H1-A1c is often grouped with H1-A2c and H1-A3c, and the three together are referred to as reflecting spectral tilt, it is the lowest-frequency of the three measures, and as such perhaps it is not surprising that it patterns with H1-H2c in the way discussed above. The overall message to take from these measures is that while phonation type triggers low-frequency spectral differences for males, the same is not true for females.

Females, however, show significant consonant-internal differences in mean CPP values, and this is true in both sonorants and voiced obstruents. These data appear in Figure 70. Males show differences in obstruents only; in terms of vowel-specific differences for males, vowels

after obstruents differ in Window three. Males show no use of CPP as a cue for breathy phonation within sonorants, in other words, while females do. In addition to the consonant-internal differences, females show significant differences in the first 20% of vowels after sonorants, and marginal overall differences in vowels after obstruents. Recall that CPP reflects periodicity and the strength of the harmonic components of the signal separated from the noise in the signal. The implication here, then, is that part of the way phonation type differences are cued in the female voice source involves differences in the amount of noise present in the signal. Modal sounds feature stronger harmonic components and less noise in the signal, while breathy sounds involve weaker harmonic components and more noise.

Figure 67 F0 across syllables by gender, phonation type, and obstruency

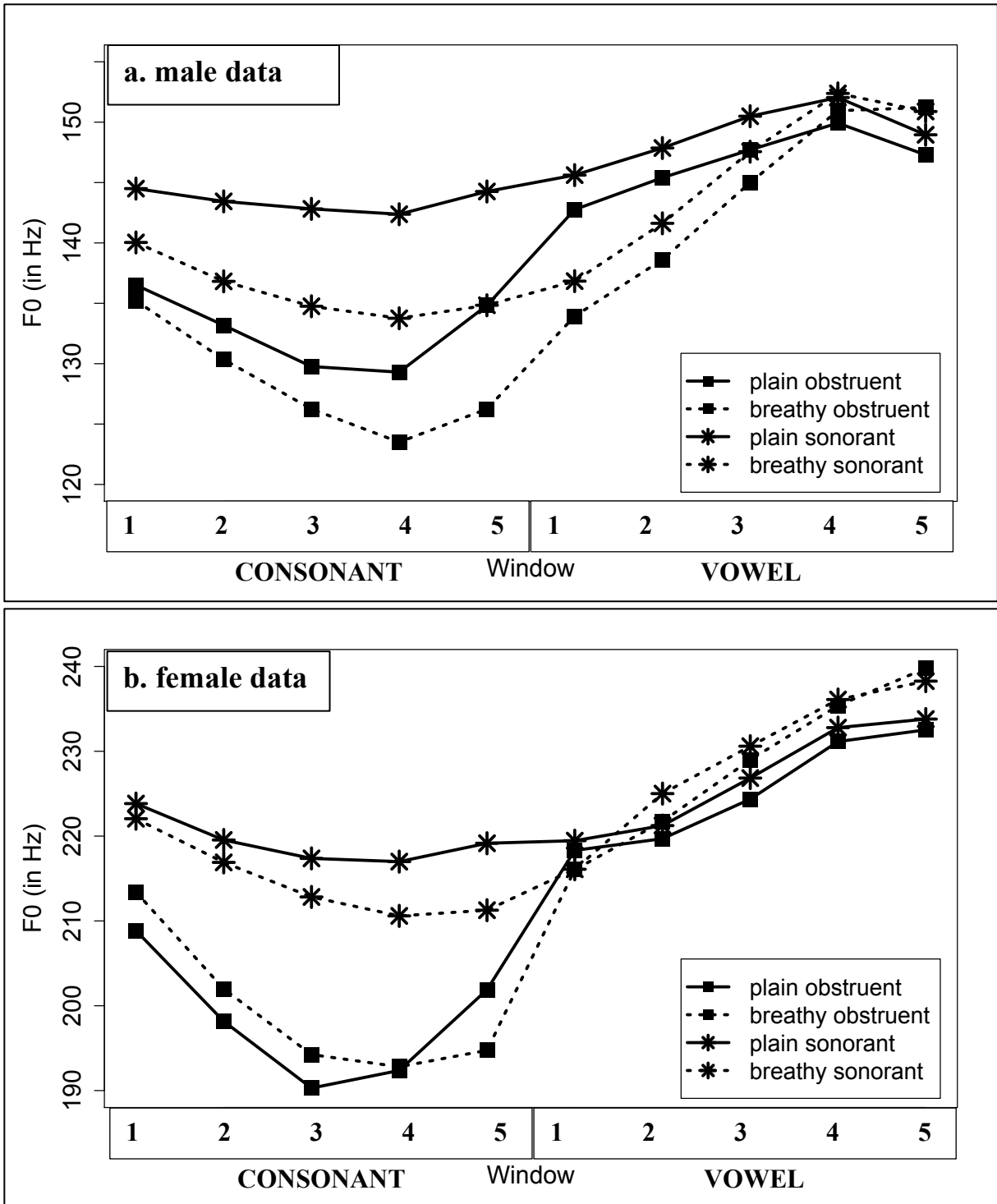


Figure 68 H1-H2c across syllables by gender (sonorant data only)

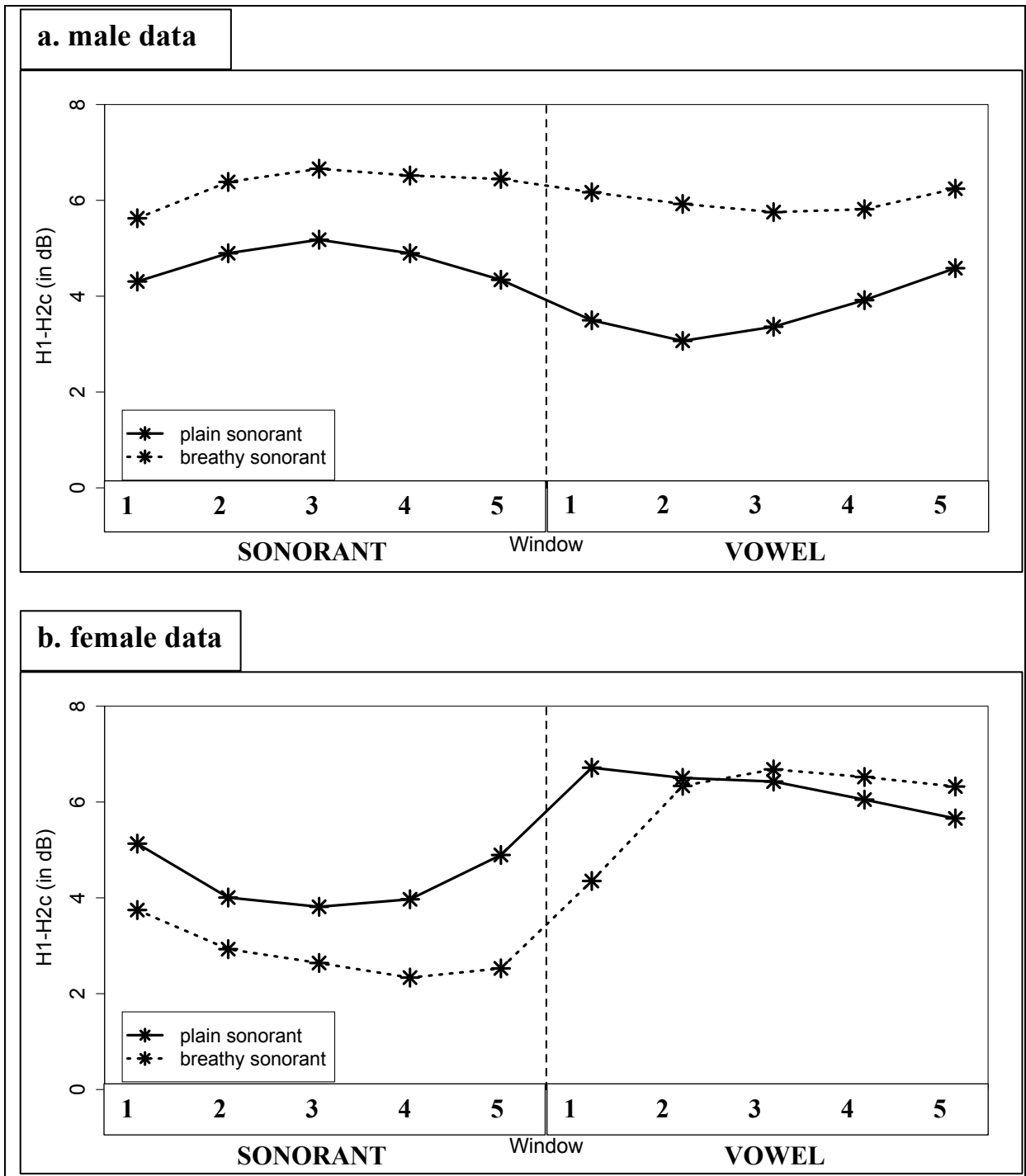




Figure 69 H1-A1c across syllables by gender (sonorant data only)

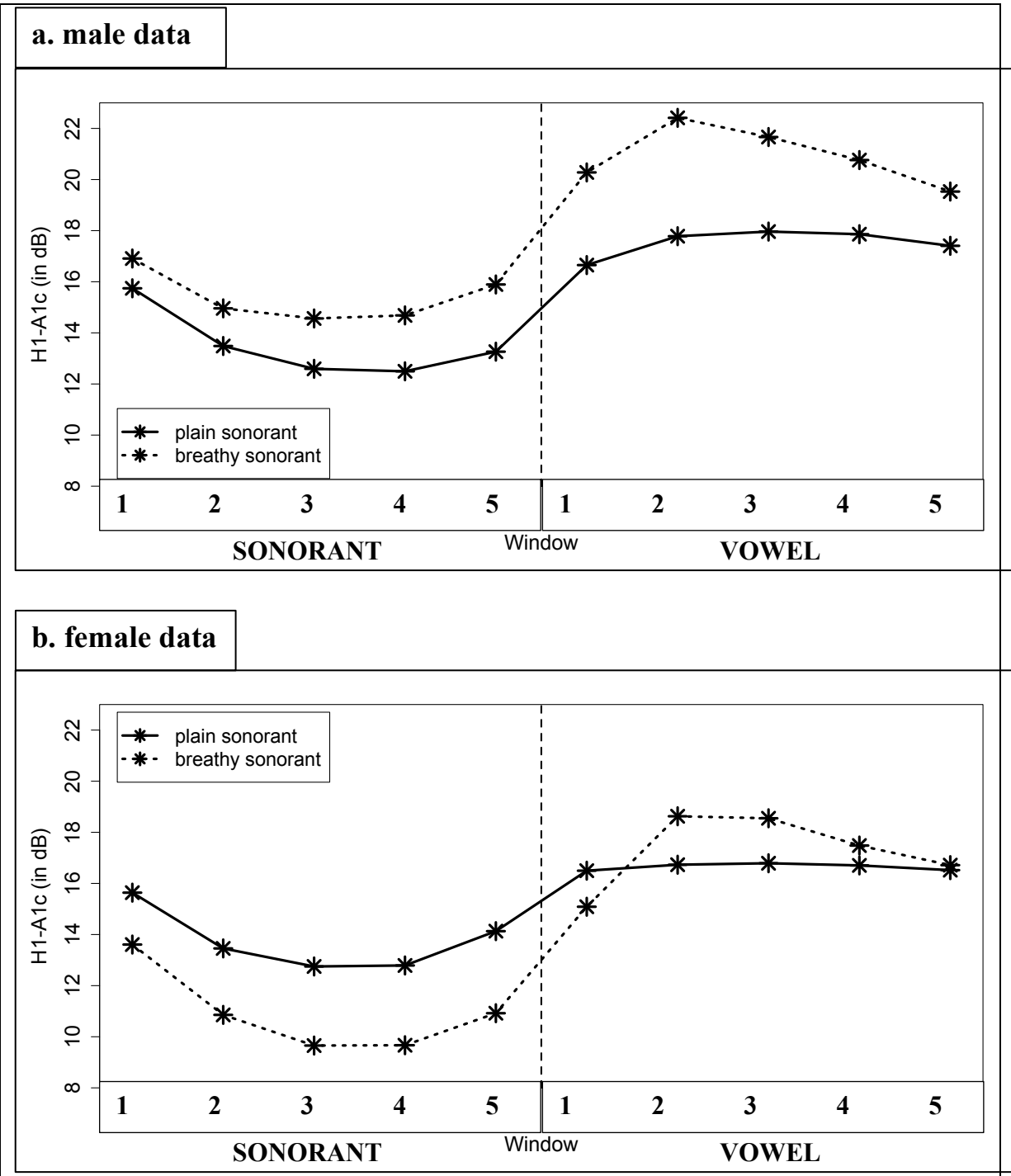
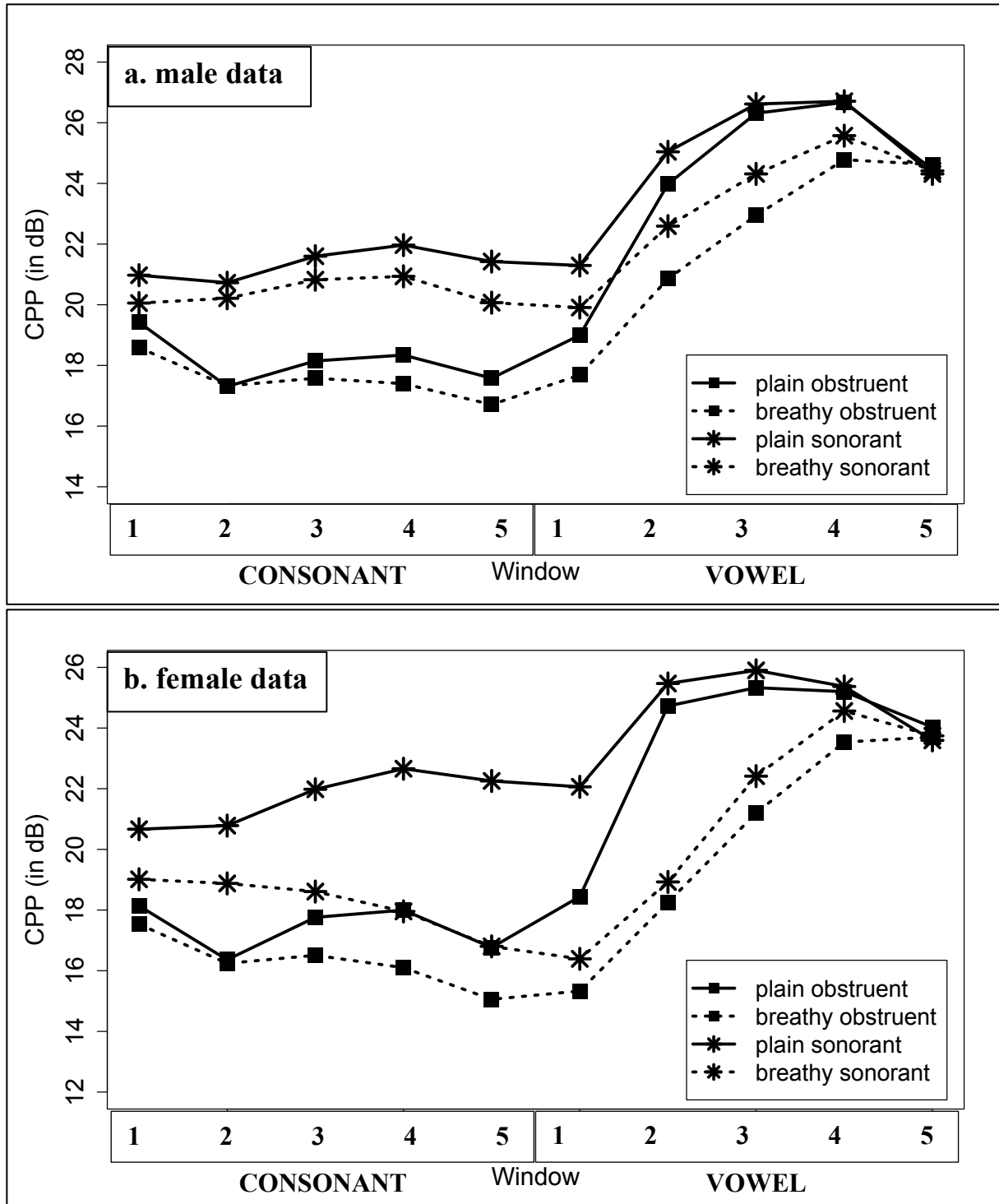


Figure 70 CPP across syllables by gender and obstruency



When we look to the existing literature on gender-based differences and the acoustic correlates of voice quality, we find ample evidence that the pattern observed in Marathi is representative. In an acoustic and perceptual study regarding voice quality that involved samples of real speech as well as synthesized tokens in modal and breathy voice, Klatt and Klatt (1990) observe that while increases in the relative amplitude of H1 are often perceived as an increase in breathiness in male speech, they may be perceived instead as increased nasality in female speech. They go on to state: “We tentatively conclude that either breathiness is signaled differently for men and women, or that the increases in the first harmonic observed in production data from women must be accompanied by other cues to be interpreted by the listener as cues to breathiness” (p.851). Without the presence of sufficient aspiration noise in the signal, an increase in H1 alone is not a reliable cue for breathiness in female speech. The present research strongly supports their tentative conclusion, for it reveals that the acoustic correlates of breathy voice in Marathi do indeed differ in male and female speech.

In order to make sense of this phenomenon, it is useful to refer to two lines of distinct yet related research. The first deals with the acoustic parameters found to be critical for the perception of breathiness. The second deals with gender-based physiological and production differences. Clearly, the two are interrelated: the way in which we produce distinct phonation types directly affects the acoustic correlates thereof. Furthermore, speakers must be sensitive to the cues which are most important to listeners if they want to successfully convey meaningful linguistic contrasts. We begin by looking to paired production and perception studies, in order to determine which parameters consistently prove critical in distinguishing modal from breathy voice.

In their extensive production, synthesis, and perception study, Klatt and Klatt (1990) analyzed the speech of ten female and six male speakers of American English<sup>38</sup>. Speakers were recorded producing sentences in two contexts, one of which encouraged a breathier production. Numerous acoustic analyses were conducted on these speech materials. After assessing pilot data from one female speaker, three parameters were predicted to be most influential in yielding perceptions of breathiness: these included increased energy in the first harmonic relative to other parts of the spectrum, increased aspiration noise in the spectrum, and extra resonances—or added poles—in the spectrum. The speech materials were then subjected to perceived-breathiness ratings by linguistically-savvy listeners, and correlations between breathiness ratings and acoustic parameters were investigated. Two of nine parameters were found to be significantly correlated with perceived breathiness. These included the strength of the first harmonic relative to other components of the spectrum, and the presence of aspiration noise in the signal.

Next, new stimuli were synthesized, using sentences produced by two female talkers as a reference. Numerous parameters were manipulated, including the amplitude of the fundamental frequency (increased by 6 dB and by 10 dB), the fundamental frequency, formant bandwidths, spectral tilt, and added aspiration noise. The results of perception tests with these data indicated that when judging synthesized data, single cues almost never elicited strong judgments of breathiness. Alone, none of the following yielded increased perceptions of breathiness: an increase in H1, increased formant bandwidths, a lowered F0, nor downward spectral tilt. The one exception was aspiration noise. When aspiration noise was added to the signal, judgments of breathiness increased significantly for many (but not all) listeners. It should also be added that when spectral tilt reflected a sharper drop-off in energy—as we have come to expect from the

---

<sup>38</sup> Two of the male subjects were Canadian.

spectrum of a breathy voiced sound—less added aspiration noise was needed to trigger the perception of breathiness. As may be expected, perceived breathiness was greatest when all cues were present.

Part of the issue that came to light during this perception study was that if H1 was increased in the absence of added aspiration noise in female speech, it often yielded judgments of increased nasality rather than increased breathiness. This is perhaps not surprising: Klatt and Klatt note that increased energy in the spectrum around 200 Hz—the general region where we may expect to find F0 for female speakers—is consistent with the addition of a nasal pole. Based on their acoustic findings, they further note that the tokens produced by female speakers contain more noise on average than those produced by male speakers.

The picture that begins to emerge, then, is that while both an increase in the amplitude of the fundamental frequency relative to the rest of the spectrum and the presence of aspiration noise in the signal may serve as clear cues for breathiness, an H1 increase alone is not a good cue for female speakers, as for them it may signal either breathiness or nasality. For male speakers, however, it is more clearly a cue for breathiness. The importance of aspiration noise in cueing breathy phonation is clear, however. It was the only parameter which, when present in the absence of any additional cues, correlated with increased perceptions of breathiness in Klatt and Klatt's synthesized stimuli, and has come to the fore as a critical component of breathy voiced sounds in other speech synthesis experiments as well (Kasuya & Ando, 1991). Perhaps the way to think about the Marathi data, then, is that while increases in H1 relative to, for example, H2 and A1 are not reliable cues for breathy phonation for females, control of the amount of noise in the signal is. This is exactly what the results would suggest, as CPP—the measure of noise in the

signal used in the present study—significantly distinguishes plain from breathy phonation for the female speakers.

Presented in this way, the implication is that for female speakers, CPP cues breathy phonation. We may wonder whether this effect serves to compensate for the fact that parameters such as H1-H2 are not as reliable in female speech as in male speech. There is some indication, however, that it may be important to consider the situation from a different angle. While increased H1 values may be poorer cues for female speakers because they may be interpreted by listeners as nasality rather than as breathiness, there is also some indication that increased aspiration noise may be more salient in female voices than in male voices.

In a study of 25 male and 20 female adult speakers of American English, Holmberg *et al.* (1988) measured inverse filtered airflow waveforms and estimated glottal air pressure and flow in the syllable [pæ] produced in loud, normal, and soft voices. While this study did not directly assess voice quality differences, it revealed aerodynamic features and likely articulatory correlates that are relevant to voice quality. Among the differences found between male and female glottal patterns in normal voice, males have lower Open Quotient values and higher peak flow rates, while females show greater leakage or flow during the so-called closed portion of the vocal fold vibration cycle.

This is something that has been mentioned previously: full closure may never be attained during the production of breathy voice; instead, a small inter-arytenoid opening may be maintained throughout with the result that there is continuous airflow (Stevens 1998). With regards to the observed pattern of leakage, or continuous airflow, Holmberg *et al.* note that it “serves as a turbulent noise source” and that it may “have a different effect on male and female source spectra, being less masked by the more rapidly dropping harmonic spectrum of the female

glottal waveform than that of the male glottal waveform. Thus the aspirated flow component may be perceptually more apparent in female normal and loud voices and contribute more to the voice quality for females than for males” (p. 524). In other words, because F0 tends to be higher for female speakers, harmonics are more spread out in the female spectrum than in the male spectrum, and the nonharmonic noise components that typify breathy voiced spectra are more apparent.

It seems as though we can consider this in both directions, then. A cue which is unmistakably tied to breathy phonation in male speech in terms of perception—namely, an increase in the amplitude of the fundamental component of the spectrum relative to the other harmonic components—cues breathy phonation in male speech. This cue is not perceptually unambiguous in the female voice source, however: it may be interpreted as nasality rather than breathiness, thereby leading to perceptual mistakes. Seen in this way, female reliance on CPP as a major phonation type cue may be seen as compensation.

The angle of vision may also be reversed, however. A cue which perception and synthesis studies have repeatedly found to be critically important in breathy phonation—namely, the presence in the signal of the random noise associated with aspiration—cues breathy phonation in female speech. It has been hypothesized that this cue is not as perceptually salient in male speech, however, and indeed, the measure which best reflects the presence of noise in the signal in this dissertation—CPP—proves highly significant for females, and decidedly less so for males. Seen this way, male reliance on increased H1 amplitude as a major phonation type cue may be seen as compensation.

I propose that the gender differences highlighted by this study are best understood as yet another example of a trading relation. While this is not the typical sense in which the term is

used, the scenario is very similar to what we have seen previously: where one cue is absent or weak, another cue is present or strong. The difference here is that instead of looking at the use of distinct cues in distinct vowel or word-position contexts, we are looking at pervasive, gender-based patterns of distinction.

It is also important to recall that CPP is not the only cue utilized by female speakers, nor are H1-H2c and H1-A1c the only cues utilized by male speakers: rather, they are members of a whole set of acoustic parameters, all of which work together to cue phonation type differences. More than two decades ago, Klatt and Klatt (1990) made the following statement: “A breathy voice quality is signaled by a surprisingly large number of diverse acoustic cues” (p. 852). The data presented in this dissertation truly illustrate that point. The acoustic correlates of phonation type differences in Marathi are many and varied.

One last point remains to be addressed, and that has to do with the cueing of breathy phonation in sonorants and in obstruents. We turn now to a discussion of the fact that differences that served to distinguish plain and breathy voiced obstruents were often marginal or absent for sonorants, and possible ramifications thereof.

#### **8.4 Sonorants and obstruents**

At the beginning of this dissertation, I set out the following broad questions with regards to phonation type:

- (1) What acoustic correlates are associated with different phonation types both crosslinguistically and language-specifically;
- (2) Why are some phonation types uncommon typologically; and
- (3) Can the acoustic correlates of phonation type shed light on the relative rarity of occurrence of a particular phonation type utilized in a particular way (i.e. breathy voice in sonorants)?



The bulk of the dissertation has focused on providing a detailed analysis of the acoustic correlates of phonation type in Marathi sonorants and obstruents. Having gained a large-scale picture of the ways in which phonation type and obstruency interact, we may now address Question (3). In doing so, we aim to get closer to answering Question (2), which is related but more broad.

There have been a number of effects, particularly in vowels, that were significant obstruents but marginal or absent for sonorants. An overview of these differences appears in Table 18.

**Table 18** A comparison of the measures that successfully distinguish phonation type in OBSTRUENTS versus SONORANTS

		MEASURE										
		CONSONANT			VOWEL							
		CONSONANT DUR.	F0	CPP	PVI	VOWEL DUR.	F0	H1-H2c	H1-A1c	H1-A2c	H1-A3c	CPP
SONORANTS	FEMALE	✓ (medial)		✓	✓					✓	✓	
	MALE	✓ (medial)	✓		✓		‡ [a]: W 1-2	✓ W 3,5		✓	✓	
OBSTRUENTS	FEMALE			✓	✓			✓ [a]		✓*	✓*	✓ INITIAL
	MALE		✓	✓	✓		‡ [a]: W 1-4	✓ MEDIAL	✓	✓*	✓*	✓ INITIAL

✓ = significant; ‡ = marginal; ✓\* indicates that the values show greater numerical differentiation than those in the other manner of articulation. Note that only those consonant measures which were gathered for both sonorants and obstruents appear in this table.

A consistent pattern emerges when we look at an overview of the significant results as they relate to obstruency. Some cues are strong regardless of obstruency: PVI, for instance, significantly

cues phonation type in both sonorants and obstruents (and for male and female speakers). Others, however, show variation, and there is only one cue that tips in favor of the sonorants. Sonorants have one temporal cue that the obstruents lack, although the cue is positionally limited: breathy sonorants are longer word-medially than plain sonorants, whereas voiced obstruents show no durational differences. In all other instances, however, when cues are present in one class of sounds and absent or reduced in the other, it is the voiced obstruents that come away with more distinguishing parameters than sonorants. Overall, what we see is the following:

- Mean CPP values in consonants distinguish phonation types in both sonorants and obstruents for females, but distinguish only obstruents for the male speakers.
- F0 differences in vowels after voiced consonants are marginal in both sonorants and obstruents, but they are marginal in 40% of the vowel after sonorants and in 80% of the vowel after obstruents.
- For male speakers, H1-H2c differences distinguish between phonation types in specific windows of vowels after sonorants, but differences are significant overall after medial voiced obstruents. For women, H1-H2c differences are never significant after sonorants; in [a] following voiced obstruents, however, H1-H2c successfully cues consonant phonation type distinctions.
- H1-A1c values in subsequent vowels are significantly different after only voiced obstruents produced by male speakers. No other H1-A1c values in vowels attain overall significance.
- For both male and female speakers, sonorants and voiced obstruents cue significant differences in H1-A2c and H1-A3c values in subsequent vowels. The differentiation after obstruents is numerically greater, however.

- Finally, CPP values in vowels after initial voiced obstruents successfully distinguish between phonation types. CPP differences after sonorants show no overall significance.

This is quite a list. In parameter after parameter, we see that the acoustic cues associated with phonation type differences are stronger after voiced obstruents than they are after sonorants. The message that must be taken from this compilation of data is that phonation type is not cued as well by sonorants as it is by voiced obstruents. More to the point, while breathy voiced obstruents trigger a multitude of acoustic differences in subsequent vowels, the differences triggered by sonorants simply do not match up. Even when considering a parameter that did not prove especially useful for female speakers, for instance—namely, the amplitude of the first harmonic relative to that of the second harmonic, or H1-H2c—we see that that it proves useful in at least one context after voiced obstruents (the [a] context).

These facts can be considered in light of two observations. First, recall that Lindblom and Maddieson (1988) assessed the obstruent-sonorant balance in the phonetic inventories of 381 languages (the number of languages contained in UPSID at that time) and found that obstruents compose between two thirds and three quarters of typical consonant inventories. Inventories regularly contain more obstruents than sonorants, in other words, and this holds true across distinct language families, language types, and geographical areas. As mentioned in Section 2.1, Lindblom and Maddieson take this typological data as evidence of an asymmetry between obstruents and sonorants: obstruents, they posit, are more perceptually (and perhaps articulatorily) rich than sonorants.

If this is in fact true, obstruents provide more room for auditory distinctiveness à la Flemming's formulation of Dispersion Theory. As mentioned previously, Flemming notes that

languages must balance three conflicting imperatives with regards to the establishment of phonemic contrasts. He elaborates the theory as follows.

The core of [Dispersion Theory] is the claim that the selection of phonological contrasts is subject to three functional goals:

- i. Maximize the number of contrasts
- ii. Maximize the distinctiveness of the contrasts
- iii. Minimize articulatory effort

These goals derive from language's function as a means for the transmission of information. The number of phonological contrasts should be maximized in order to enable us to differentiate a substantial vocabulary of words. The auditory distinctiveness of the contrasts should be maximized so that the differences between words can be easily perceived by a listener. The third requirement, that effort should be minimized, appears to be a general principle of human motor behavior, and is not specific to language. (Flemming 1995: 24).

The idea that there is more room for auditory distinctiveness in obstruents than in sonorants finds support in the present study: phonation type contrasts are better acoustically differentiated by obstruents than by sonorants. We do not yet have the perceptual data to confirm the tie between the acoustics reported herein and the auditory distinctiveness of phonation type contrasts in Marathi sonorants for listeners. Kawahara (2007), however, tied the typological observation that sonorants are less likely than obstruents to have phonemic length distinctions with the perceptual finding that it is more difficult for listeners to perceive segment duration differences in more sonorous sounds. Thus while perception research is a critical next step, it seems likely to posit that phonation type contrasts are less auditorily distinct in sonorants than in obstruents.<sup>39</sup>

This brings us to the second observation that we may consider when thinking about the interaction between obstruency and breathy phonation, which has to do with crosslinguistic patterns related to the retention versus the loss of phonemic breathy voice in obstruents versus

---

<sup>39</sup> The acoustic differentiation between modal and breathy sonorants is presumably better than the differentiation between singleton and geminate sonorants, however, due to the fact that there are a number of acoustic parameters that cue phonation type differences—not just duration.

sonorants. Here, we can point to the collection of Newari dialects and the family of Kiranti languages, both of which are found in Nepal. Newar, a Tibeto-Burman language, features a series of breathy voiced consonants in its phonetic inventory. While Newar used to contain both breathy voiced obstruents and sonorants, Kathmandu Newar appears to be the one dialect which has definitively retained breathy voiced sonorants (Genetti 2003, Hargreaves 2003). The major dialects spoken in the northeastern part of Nepal, meanwhile, retain only the breathy voiced obstruents (Genetti 2005). It should also be mentioned here that breathy voice seems to be disappearing completely from one of these dialects, Dolakha Newar, which retains breathy voiced obstruents in only a few words and from which breathy voiced sonorants have disappeared entirely (Genetti 2003, Hargreaves 2003). Furthermore, Genetti notes that even in the few lexical items where Dolakha seems to have retained breathy voiced obstruents, the distinction is neutralized in rapid speech (Genetti 2007).

A similar pattern of lost or absent breathy voiced sonorants is found in the group of Kiranti languages which are spoken in Eastern Nepal. A number of the 32 languages in this group feature breathy voiced obstruents, but only one—Camling—has been documented as containing breathy voiced sonorants (Ebert 2003). If the Camling and Kathmandu Newar examples are indicative of a larger pattern, then we can posit the following: phonation type contrasts are more common in obstruents than in sonorants and are more likely to be retained in obstruents than in sonorants.

There is in fact one more claim that can be posited, and this takes the form of an implicational universal. The majority of the crosslinguistic data presented in this dissertation has focused on breathy voiced sonorants as rare: Tibeto-Burman languages like Sumi, Camling, and Kathmandu Newar, and the Bantu language Tsonga, have been mentioned with reference to the

fact that they contain breathy voiced sonorants. We can consider this from another direction, however: instead of pointing to the presence of these sounds as unique, we can ask whether the presence of breathy voiced sonorants in a language's consonant inventory tells us anything about the obstruents in the inventory. The short answer is yes: the presence of phonation type contrasts in a language's sonorants tells us something about the presence of such contrasts in its obstruents, but the reverse is not true. Let us consider this claim in more detail.

We know that in Marathi, which contains breathy voiced sonorants, phonation type contrasts are also present in its obstruents. Consider the Tibeto-Burman languages next: Camling contains phonation type contrasts in its sonorants, and indeed, phonation type contrasts are present in its obstruents as well (Ebert 2003). The same is true of Kathmandu Newar, or Nepāl Bhāśā (Hargreaves 2003), and of Sumi (Harris 2009). We may also want to look for evidence to support this claim in very different language families: the Southern Bantu language Tsonga, or Changana, has breathy voiced sonorants, and indeed, it contains prenasalized breathy voiced obstruents as well (Janson 2001). Once again, then, phonation contrasts in sonorants tells us something about the obstruents contained in a language's inventory.

One case which may prove to complicate the claim that there is an implicational relationship between phonation type contrasts in sonorants and obstruents is the Edoid (Niger-Congo) language Isoko, which is reported by UPSID to contain a breathy voiced [v<sup>h</sup>] in its inventory in the absence of breathy voiced obstruents.<sup>40</sup> There are voicing contrasts among

---

<sup>40</sup> The story here is complicated: while some dialects of Isoko are reported to contain [v<sup>h</sup>], in others it is replaced by the voiceless bilabial fricative [ɸ]. Donwa-Ifode further reports that appearance of [v<sup>h</sup>] is very limited: it is often in free variation with [v]. Its status as a phoneme, then, seems less than stable. One hypothesis regarding the origin of this sound comes from Elugbe (1980). When commenting on the occasional occurrence of breathy voice in Edoid languages, Elugbe posits that breathy voice may occur as an intermediary step during development of phonemic voicing. In other words, it occurs during transition from voicelessness to voicing. More data regarding breathiness in Edoid languages is certainly necessary, but for now the questions regarding the status of breathy voice in these languages remain unresolved.

obstruents in Isoko—phonemic breathy voice in obstruents, however, is absent (Donwa-Ifode 1982, 1986). It seems, then, that an implicational statement must be phrased broadly rather than specifically—while we cannot claim that the presence of breathy voiced sonorants in an inventory implies the presence of breathy voiced obstruents, because of Isoko, we can certainly claim that the presence of phonation type contrasts in sonorants implies the presence of phonation type contrasts in obstruents.

A more general statement is beneficial, however, because it allows us to extend our field of inquiry: we have considered phonation type contrasts between modal and breathy voiced sounds, but what of other types of contrasts? Does the generalization hold across other categories as well? For information regarding voicing contrasts, we can look to Icelandic. Icelandic, from the North Germanic branch of the Indo-European family, is reported to have phonemic voicing contrasts among its sonorants, yielding minimal pairs like [ɛltɪ] ‘to cook’ and [ɛɫ̥tɪ] ‘to chase’ or [ni:ta] ‘to use’ and [ɲi:ta] ‘to knot’ (Bombien 2006: 1). Icelandic does not have voiced stops—it does indeed have phonation type contrasts in its obstruents, however, in the form of plain and aspirated voiceless phones. An example pair is [pan:a] ‘to prohibit’ and [p<sup>h</sup>an:a] ‘pan’ (Bombien 2006: 2). Similarly Burmese (or Myanmar Bhasa), another Tibeto-Burman language, contains voicing contrasts among its sonorants and a three-way distinction between voiced, voiceless aspirated, and plain voiceless obstruents (Wheatley 2003).

Given the cases outlined above, I propose the following implicational universal: if a language has phonation type contrasts in its sonorants, it also has them in its obstruents. Whether this is an absolute or a statistical universal remains an open question for now, but future investigation of sonorant and obstruent contrasts in languages from distinct geographic and genetic areas will allow for a more comprehensive understanding of this phenomenon.

In sum, then, the Marathi data shows us that phonation type contrasts are better acoustically differentiated in obstruents than in sonorants. Clearly, data from additional languages must be collected and analyzed before we can state with certainty that this is universally true. Perhaps other languages will reveal stronger acoustic correlates of phonation type in sonorants. To my knowledge, however, this work constitutes the most comprehensive assessment of breathy voiced phonation in sonorants that exists to date. As such, on the basis of these data, it seems likely that the typology of breathy voiced sonorants is tied to the acoustic correlates of phonation type in sonorants. These sounds are crosslinguistically rare because they are not well-cued. The implicational universal I have proposed may be considered in this line of reasoning as well: if a language considers plain and breathy voiced sonorants to be sufficiently auditorily distinct, then plain and breathy voiced obstruents—which are better differentiated acoustically—also differ sufficiently.



## 9 Conclusion

This dissertation presents a comprehensive acoustic analysis of phonation type distinctions in Marathi sonorants and obstruents. Nine acoustic measures were analyzed: sonorant constriction duration/obstruent closure duration, a temporal measure of the breathy interval referred to as the Pre-Vocalic Interval, vowel duration, F0, H1-H2c, H1-A1c, H1-A2c, H1-A3c, and Cepstral Peak Prominence. Major findings include the fact that male and female speakers cue breathy phonation quite differently, with males utilizing low-frequency measures like H1-H2c and H1-A1c, and females utilizing CPP, a measure which provides information about the presence of noise in the signal. Beyond these gender-related differences, numerous instances of trading have been found: where one cue is weak or absent, another cue is strong or present. Finally, we have also seen evidence that suggests that sonorants do not cue phonation type distinctions as well as obstruents do. The claim was made that the acoustic profile of these sounds contributes to their crosslinguistic rarity: phonemic contrasts between breathy and modal sonorants are uncommon because the two are not well enough differentiated acoustically.

Clearly, these data are only the beginning of the story. Now that that we understand the ways in which the acoustic parameters assessed contribute to phonation type distinctions in Marathi, we can move from production research into perception research. This is particularly important given that I argue for a connection between the acoustic profile of these sounds and their auditory distinctiveness: that particular cues prove reliable in the acoustic analysis does not mean that these are the cues to which listeners pay the most attention. As such, subsequent perception studies to determine the relative importance of each of the cues discussed herein are an important next step.

Initial tests could assess the importance of the temporal cues: word-medial breathy voiced tokens could be shortened, thereby eliminating a temporal cue which sonorants have and obstruents do not. Recall that Kawahara (2007) find it difficult to perceive the duration of sonorous segments: as such, this manipulation may not greatly affect perception. The PVI could be artificially manipulated, as well, to determine how crucial the presence of an audible breathy interval is to the perception of breathy voice. Given the relative importance of noise in the signal as a perceptual cue (Holmberg *et al.* 1988, Klatt and Klatt 1990), we may predict that the PVI provides an important perceptual cue. Manipulating the PVI may also shed light on whether the breathy interval is more crucial for the identification of breathy sonorants as compared with obstruents, since they have fewer spectral cues. Another perception experiment could involve low-pass filtering of speech; presumably, tokens produced by male speakers will still have critical information present in the signal even after low-pass filtering, as they showed greater differentiation between modal and breathy phonation at the low frequencies. In low-pass filtered female speech, on the other hand, phonation type distinctions may prove more difficult for listeners to perceive. Alternatively, if CPP is a very strong cue in and of itself—as may be predicted based on the findings in Klatt and Klatt (1990) and Holmberg *et al.* (1988)—the low-pass filtering may not interfere so drastically with perception of these contrasts in female speech.

Another important line of future research involves investigation of breathy voiced sonorants in additional languages—the Tibeto-Burman languages of Camling or Kathmandu Newar, for instance. Sumi may also provide rich ground for additional investigation, as Harris (2009) assessed the productions of only one female speaker. Other possibilities include Urdu or Parauk (Austro-Asiatic).

In the end, there are perhaps several major lessons with which to close. First, breathy phonation is associated with multiple acoustic cues, and these cues often engage in trading relations. This means that the cues found in one context may be replaced by alternate cues in another context, emphasizing the importance of including multiple vowel and word-position contexts in any experimental design. Gender differences also emerge, highlighting the importance of including both male and female speakers in any comprehensive study of phonation type distinctions among sonorants. Finally, the cues associated with breathy voice are weaker for sonorants than for obstruents. While I do not claim a direct, one-to-one correlation between the acoustics of a sound and its auditory distinctiveness, the two are clearly related and all of the evidence taken together supports the analysis that breathy voiced sonorants are crosslinguistically rare because they do not make for strong phonemic contrasts.

## References

- Andruski, J. & M. Ratliff. (2000). Phonation types in production of phonological tone: The case of Green Mong. *JIPA*, 30:37-61.
- Baayen, R.H. (2008). *Analyzing linguistic data: A practical introduction to statistics using R*. Cambridge: Cambridge University Press.
- Balasubramaniam, R.K., J. S. Bhat, S. Fahim III, & R. Raju III. (2011). Cepstral analysis of voice in unilateral adductor vocal fold palsy. *Journal of Voice*, 25.3: 326-329.
- Berkson, K. (2012). Capturing a four-way distinction: Durational measures of oral stops in Marathi. *Kansas Working Papers in Linguistics*, 33: 27-46.
- Benguerel, A. & T. Bhatia. (1980). Hindi stop consonants: An acoustic and fibroscopic study. *Phonetica*, 37:134-148.
- Blankenship, B. (1997). The time course of breathiness and laryngealization in vowels. PhD dissertation, UCLA.
- Blankenship, B. (2002). The timing of nonmodal phonation in vowels. *Journal of Phonetics*, 30: 163-191.
- Boersma, P. & D. Weenink. (2010). Praat: doing phonetics by computer [Computer program]. Version 5.1.32. Available at [www.praat.org](http://www.praat.org).
- Bombien, L. (2006). Voicing alterations in Icelandic sonorants—a photoglottographic and acoustic analysis. *ARIPUC, IPDS, Kiel*.
- Catford, J. (1977). *Fundamental problems in phonetics*. Edinburgh: Edinburgh University Press.
- Census of India. (2001). *Statement 1: Abstract of speakers' strength of languages and mother tongues*. Census of India Website: Office of the Registrar General & Census Commissioner, India. Online resource available at <[http://censusindia.gov.in/Census\\_Data\\_2001/Census\\_Data\\_Online/Language/Statement1.htm](http://censusindia.gov.in/Census_Data_2001/Census_Data_Online/Language/Statement1.htm)>. Accessed September 20, 2011.
- Cho, T., A. Jun, & P. Ladefoged. (2002) Acoustic and aerodynamic correlates of Korean stops and fricatives. *Journal of Phonetics*, 30: 198-228
- Davis, K. (1994). Stop voicing in Hindi. *Journal of Phonetics*, 22.2: 177-193.
- Deshpande, D. (1990). Palatal sibilant: Aspect of aspiration. *Osmania Papers in Linguistics*, 16: 1-4.
- Dhongde, R.V. & K. Wali. (2009). *Marathi*. John Benjamins Publishing Co.: Amsterdam.

- Donwa-Ifode, S. (1982). *The Sound System of Isoko*. PhD dissertation, University of Ibadan. Ibadan, Nigeria.
- Donwa-Ifode, S. (1986). Phonetic variation in consonants (Isoko). *Anthropological Linguistics*, 28.2: 149-160.
- Dutta, I. (2007a). Correlation between voicing lead time (VLT) and F0 lowering following breathy voiced stops in Hindi. *Papers from the 39<sup>th</sup> Annual Meeting of the Chicago Linguistic Society*, p. 405-422.
- Dutta, I. (2007b). Four-way stop contrasts in Hindi: An acoustic study of voicing, fundamental frequency and spectral tilt. Ph.D. dissertation, University of Illinois at Urbana-Champaign.
- Ebert, K. (1997). *Languages of the World/Materials 103*. LINCOM Europa: Miinchen, Newcastle. Available online at [http://www.archive.org/stream/rosettaproject\\_rab\\_phon-1/rosettaproject\\_rab\\_phon-1\\_djvu.txt](http://www.archive.org/stream/rosettaproject_rab_phon-1/rosettaproject_rab_phon-1_djvu.txt). Accessed February 27, 2013.
- Ebert, K. (2003). Kiranti languages: An overview. in G. Thurgood and R. J. LaPolla (eds.), *The Sino-Tibetan Languages*. Routledge: New York, NY, pp. 505-517.
- Elugbe, B. O. (1980). Reconstructing the lenis feature in Proto-Edoid. *Journal of African Languages and Linguistics*, 2: 36-67.
- Engstrand, O. & D. Krull. (2009). Durational correlates of quantity in Swedish, Finnish and Estonian: Cross-language evidence for a theory of adaptive dispersion. *Phonetica*, 51.1-3: 80-91.
- Esposito, C.M., S. Khan, & A. Hurst. (2005). Breathy nasals and /Nh/ clusters in Bengali, Hindi, and Marathi. *UCLA Working Papers in Phonetics*, 104: 82-106.
- Esposito, C.M. (2005). An acoustic and electroglottographic study of phonation in Santa Ana del Valle Zapotec. Poster presented at the 79<sup>th</sup> meeting of the Linguistic Society of America, San Francisco, CA.
- Esposito, C.M. (2006). *The Effects of Linguistic Experience on the Perception of Phonation*. Ph.D. dissertation, UCLA.
- Esposito, C. (2012). An acoustic and electroglottographic study of White Hmong tone and phonation. *Journal of Phonetics*, 40: 466-476.
- Fant, G. (1979). Glottal source and excitation analysis. *Speech Trans. Lab. Q. Prog. Stat. Rep.* 1. Royal Institute of Technology, Stockholm, 85-107.
- Fischer-Jørgensen, E. (1967). Phonetic analysis of breathy (murmured) vowels. *Indian Linguistics*, 28: 71-139.
- Flemming, E. S. (1995). *Auditory representations in phonology*. PhD dissertation, UCLA.

- Francis, A., V. Ciocca & J. M. Ching Yu. (2003). Accuracy and variability of acoustic measures of voicing onset. *JASA*, 113.2: 1025-1032.
- Garellek, M., & P. Keating. (2010). The acoustic consequences of phonation and tone interactions in Jalapa Mazatec. *UCLA Working Papers in Phonetics*, 108: 143-163.
- Genetti, C. (2003). Dolakhā Newār. in G. Thurgood and R. J. LaPolla (eds.), *The Sino-Tibetan Languages*. Routledge: New York, NY, pp. 355-370.
- Genetti, C. (2005). Preliminary notes on the Tauthali Dialect of Newar. Handout from the Himalayan Languages Symposium, December 2005. Bangkok, Thailand.
- Genetti, C. (2007). *A Grammar of Dolakha Newar*. Mouton de Gruyter. Berlin, Germany.
- Gordon, M. & P. Ladefoged. (2001). Phonation types: a cross-linguistic overview. *Journal of Phonetics*, 29: 383-406.
- Gordon, M. (2002). A phonetically driven account of syllable weight. *Language* 78.1: 51-80.
- Hargreaves, D. (2003). Kathmandu Newar (Nepāl Bhāśā). in G. Thurgood and R. J. LaPolla (eds.), *The Sino-Tibetan Languages*. Routledge: New York, NY, pp. 371-384.
- Harris, T. (2009). *An acoustic and articulatory study of sonorant phonation in Sumi*. Unpublished honours thesis; The University of Melbourne School of Languages and Linguistics. Melbourne, Australia.
- Hayes, B. (1999). Phonetically-driven phonology: The role of optimality theory and inductive grounding. in M. Darnell, E. Moravcsik, M. Noonan, F. Newmeyer, and K. Wheatley (eds.), *Proceedings of the 1996 Milwaukee Conference on Formalism and Functionalism in Linguistics*, Vol. 1. John Benjamins: Amsterdam, pp. 243–85. (manuscript, pages numbered 1-37, available at [http://www.blutner.de/signalling/hayes\\_grounding.pdf](http://www.blutner.de/signalling/hayes_grounding.pdf))
- Hayes, B. (2011). *Introductory phonology* (Vol. 36). Wiley-Blackwell: Malden, MA.
- Heman-Ackah, Y., R. Heuer, D. Michael, R. Otrowski, M. Horman, M. Baroody, J. Hillenbrand, and R. Sataloff. (2003). Cepstral peak prominence: a more reliable measure. *The Annals of Otology, Rhinology & Laryngology*, 112.4: 324-344.
- Hillenbrand, J., R. Cleveland, and R. Erickson. (1994). Acoustic correlates of breathy vocal quality. *Journal of Speech and Hearing Research*, 37: 769-778.
- Hillenbrand, J. & R. A. Houde. (1996). Acoustic correlates of breathy vocal quality: Dysphonic voices and continuous speech. *Journal of Speech and Hearing Research*, 39: 311-321.
- Holmberg, E., R. Hillman, & J. Perkell. (1988). Glottal airflow and transglottal air pressure measurements for male and female speakers in soft, normal, and loud voice. *JASA*, 84.2: 511-529.

- Holmberg, E., R. Hillman, J. Perkell, P. Guiod, & S. Goldman. (1995). Comparisons among aerodynamic, electroglottographic, and acoustic spectral measures of female voice. *Journal of Speech and Hearing Research*, 38: 1212-1223.
- Huffman, M. K. (1987). Measures of phonation in Hmong. *JASA*, 81: 495–504.
- Janson, T. (2001). Consonants in Changana/Tsonga. *South African Journal of African Languages*, 21.1: 16-32.
- Jha, A. (1977). An outline of Marathi phonetics. Deccan College Press: Pune, India.
- Kasuya, H. & Y. Ando. (1991). Acoustic analysis, synthesis, and perception of breathy voice. in J. Gauffin & B. Hammarberg (eds.), *Vocal Fold Physiology: Acoustic, Perceptual, and Physiological Aspects of Voice Mechanisms*. Singular Publishing Group: San Diego, CA, pp. 251-258.
- Kawahara, S. (2007). Sonorancy and geminacy. In L. Bateman, M. O’Keefe, E. Reilly, & A. Werle (eds.), *University of Massachusetts Occasional Papers in Linguistics 32: Papers in Optimality Theory III*. GLSA: Amherst, MA, pp. 145-186.
- Keating, P. & C. Esposito. (2007). Linguistic voice quality. *UCLA Working Papers in Phonetics*, 105: 85-91.
- Kelkar, A. (1958). *The phonology and morphology of Marathi*. PhD dissertation, Cornell University.
- Khan, S. D. & C. M. Esposito. (2011). Breathiness contrasts in consonants and vowels: A comparative study of Gujarati and White Hmong. *Proceedings from the 17<sup>th</sup> International Congress of Phonetic Sciences*, pp. 1050-1053.
- Khan, S. D. (2012). The phonetics of contrastive phonation in Gujarati. *Journal of Phonetics*, 40: 780-795.
- Klatt, D. H. & L. C. Klatt. (1990). Analysis, synthesis, and perception of voice quality variations among female and male talkers. *JASA*, 87.2: 820-857.
- Kreiman, J., B. Gerratt, & N. Antoñanzas-Barroso. (2007). Measures of the glottal source spectrum. *JSLHR*, 50: 595-610.
- Ladefoged, P. & I. Maddieson. (1996). *Sounds of the World’s Languages*. Blackwell Publishers, Inc.: Cambridge, MA.
- Laver, J. (1980). *The phonetic description of voice quality (Cambridge Studies in Linguistics 31)*. Cambridge University Press: London, UK.
- Liljencrants, J. & B. Lindblom. (1972). Numerical simulation of vowel quality systems. *Language*, 48: 839-862.

- Lindblom, B. (1986). Phonetic universals in vowel systems. in J.J. Ohala & J.J. Jaeger (eds.) *Experimental Phonology*. Academic Press: Orlando, FL.
- Lindblom, B. (1990). Phonetic content in phonology. *Phonetic Experimental Research at the Institute of Linguistics University of Stockholm*, 11: 100-118.
- Lindblom, B. & I. Maddieson. (1988). Phonetic universals in consonant systems. in L. M. Hyman & C. N. Li (eds.) *Language, speech and mind*. Routledge: London, UK, pp. 62-78.
- Lisker, L. & A. Abramson. (1964). A cross-language study of voicing in initial stops: Acoustical measurements. *Word*, 20: 384-422.
- Marchal, A. (2009). *From Speech Physiology to Linguistic Phonetics*. John Wiley & Sons, Inc.: Hoboken, NJ.
- Maryn, Y., N. Roy, M. De Bodt, P. Van Cauwenberge, & P. Corthals. (2009). Acoustic measurement of overall voice quality: A meta-analysis. *JASA*, 126.5: 2619-2634.
- Masica, C. (1991). *The Indo-Aryan languages*. Cambridge University Press: Cambridge, UK.
- MathWorks, Inc. (2011). MATLAB version 7.12.0.635. Natick, MA.
- Mikuteit, S. & H. Reetz. (2007). Caught in the ACT: The timing of aspiration and voicing in Bengali. *Language and Speech*, 50.2: 247-277.
- Moers, C., B. Möbius, F. Rosanowski, E. Nöth, U. Eysholdrt, & T. Haderlein. (2012). Vowel- and text-based cepstral analysis of chronic hoarseness. *Journal of Voice*, 26.4: 416-424.
- Murphy, P. (2006). Periodicity estimation in synthesized phonation signals using cepstral rahmonic peaks. *Speech Communication*, 48: 1704-1713.
- Ní Chasaide, A. & C. Gobl. (1995). Voice source variation. in W. Hardcastle & J. Laver (eds.), *The handbook of phonetic sciences*. Blackwell: Cambridge, MA, pp. 427-461.
- Ohala, M. & J. Ohala. (1972). The problem of aspiration in Hindi phonetics. *Annual Bulletin of the Research Institute of Logopedics and Phoniatics, Faculty of Medicine, University of Tokyo, Japan*. No. 6: 39-46.
- Ohala, M. (1979). Phonological features of Hindi stops. *Proceedings of the South Asian Languages Analysis Roundtable*, 1: 79-88.
- Padgett, J. (2003). Contrast and post-velar fronting in Russian. *Natural Language and Linguistic Theory* 21: 39-87.
- Padgett, J. & M. Tabain. (2005). Adaptive dispersion theory and phonological vowel reduction in Russian. *Phonetica*, 62.1: 14-54.
- Pandharipande, R. V. (1997). *Marathi*. Routledge: London, UK.



- Pandit, P. B. (1957). Nasalization, aspiration and murmur in Gujarati. *Indian Linguistics*, 17: 165-172.
- Patil, H., P. Dutta, & T. Basu. (2006). The wavelet packet based cepstral features for open set speaker classification in Marathi. *From Data and Information Analysis to Knowledge Engineering: Studies in Classification, Data Analysis, and Knowledge Organization*. Part 3: 134-141.
- Pennington, M. (2005). *The phonetics and phonology of glottal manner features*. PhD dissertation, University of Indiana.
- R Core Team. (2012). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. Vienna, Austria. available at <http://www.R-project.org>.
- Samlan, R. A. & B. A. Story. (2011). Relation of structural and vibratory kinematics of the vocal folds to two acoustic measures of breathy voice based on computer modeling. *JSLHR*, 54: 1267-1283.
- Schiefer, L. (1986). F0 in the production and perception of breathy stops: Evidence from Hindi. *Phonetica*, 43: 43-69.
- Schiefer, L. (1987). 'Voiced aspirated' or 'breathy voiced' and the case for articulatory phonology. *Forschungsberichte des Instituts für Phonetik und Sprachliche Kommunikation der Universität München*, 27: 257-278.
- Schwartz, J.-L., L.-J. Boë, N. Vallée, & C. Abry. (1997). The Dispersion-Focalization Theory of vowel systems. *Journal of Phonetics*, 25: 255-286.
- Shue, Y., P. Keating, C. Vicenik, & K. Yu. (2011). VoiceSauce: A Program for Voice Analysis. available at <http://www.ee.ucla.edu/~spapl/voicesauce/>.
- Silverman, D., B. Blankenship, P. Kirk, & P. Ladefoged. (1995). Phonetic structures in Jalapa Mazatec. *Anthropological Linguistics*, 37.1: 70-88.
- Simpson, A. P. (2012). The first and second harmonics should not be used to measure breathiness in male and female voices. *Journal of Phonetics*, 40: 477-490.
- Södersten, M., & P.-Å. Lindestad. (1992). A comparison of vocal fold closure in rigid telescopic and flexible fiberoptic laryngostroboscopy. *Acta Otolaryngologica*, 112.1: 144-150.
- Southworth, F. (2000). Review of *Marathi* by R. Pandharipande. *Lingua*, 110: 215-224.
- Steriade, D. (1997). Phonetics in phonology: The case of laryngeal neutralization. *UCLA Working Papers in Linguistics - Papers in Phonology*, 2.3: 25-146.
- Stevens, K. (1998). *Acoustic Phonetics*. MIT Press: Cambridge, MA.

- Trail, A. & M. Jackson. (1988). Speaker variation and phonation type in Tsonga nasals. *Journal of Phonetics*, 16.4: 385-400.
- UCLA Phonological Segment Inventory Database. [http://web.phonetik.uni-frankfurt.de/cgi-bin/upsid\\_sounds.cgi](http://web.phonetik.uni-frankfurt.de/cgi-bin/upsid_sounds.cgi). Accessed August 27, 2012; November 20, 2012.
- Watts, C. & S. Awan. (2011). Use of spectral/cepstral analyses for differentiating normal from hypofunctional voices in sustained vowel and continuous speech contexts. *Journal of Speech, Language and Hearing Research* 54: 1525-1537.
- Wayland, R. & A. Jongman. (2003). Acoustic correlates of breathy and clear vowels: The case of Khmer. *Journal of Phonetics*, 31: 181-201.
- Wheatley, J. (2003). Burmese. in G. Thurgood and R. J. LaPolla (eds.), *The Sino-Tibetan Languages*. Routledge: New York, NY, pp. 195-207.

## Appendix A – Demographic Data

<b>Female Participants</b>					
<b>Subject ID</b>	<b>W1</b>	<b>W3</b>	<b>W4</b>	<b>W5</b>	<b>W6</b>
<b>Age</b>	29	40	25	56	36
<b>Language spoken at home</b>	Marathi	Marathi	Marathi, Konkani	Marathi	Marathi
<b>Number of years Marathi was studied</b>	12	10	10	16	10
<b>Other languages spoken</b>	Hindi, English	Hindi, English, Kannada	Hindi, English, Konkani	Hindi, English	Hindi, English
<b>Hometown</b>	Pune	Wordha	Mumbai, Pune	Jalgaon/Pune	Pune
<b>Time spent in the US</b>	3 yrs	3 yrs	1 yrs	5 yrs	>5 yrs
<b>How much Marathi is spoken per day</b>	a lot, every day	30% of daily language use	50% of daily language use	50% of daily language use	30% of daily language use
<b>Parents' languages</b>	Marathi	Marathi	Konkani, Marathi	Marathi	Marathi

<b>Male Participants</b>					
<b>Subject ID</b>	<b>M1</b>	<b>M4</b>	<b>M5</b>	<b>M6</b>	<b>M7</b>
<b>Age</b>	30	34	43	64	56
<b>Language spoken at home</b>	Marathi, Hindi, English	Marathi	Marathi	Marathi	Marathi
<b>Number of years Marathi was studied</b>	6	10	12	16	11
<b>Other languages spoken</b>	Hindi, English, Bengali, French	Hindi, English	Hindi, English	Hindi, English	Hindi, English
<b>Hometown</b>	Mumbai	Pune	Pune	Asoda/Pune	Mumbai
<b>Time spent in the US</b>	2 years 8 months	2.5 yrs	3.5 years	5 years	>27 yrs
<b>How much Marathi is spoken per day</b>	15% of daily language use	10 min/week	30% of daily language use	25% of daily language use	2 - 4 hrs/day
<b>Parents' languages</b>	Marathi	Marathi	Marathi	Marathi	Marathi

## Appendix B – Wordlist

word initial							
<i>before [a]</i>				<i>before [e]</i>			
1	p	[pate]	पाते	blade of grass	[peʈi]	पेटी	small container
2	ph	[pʰaʈa]	फाटा	fork in road	[pʰeʈe]	फेटे	turbans
3	b	[badzu]	बाजू	side	[beʈe]	बेटे	islands
4	bh	[bʰatsa]	भाचा	nephew	[bʰegʌ]	भेग	crack (crevice)
5	m	[masa]	मासा	fish	[meʂ]	मेष	(a zodiac sign - Aries)
6	mh	[mʰatre]	म्हात्रे	surname	[mʰetʌr]	म्हेतर	a sweeper
7	v	[vara]	वारा	wind	[vedʌ]	वेड	passion, obsession
8	vh	[vʰave]	व्हावे	let it happen	[vʰegʌ]	व्हेग	an ancient tree*
9	t	[tapʰa]	ताफा	herd	[tetʰe]	तेथे	there
10	th	[tʰapa]	थापा	lies	[tʰe.ur]	थेऊर	place name
11	d	[dada]	दादा	elder brother	[dehʌ]	देह	body
12	dh	[dʰaga]	धागा	thread	[dʰenu]	धेनु	cow
13	n	[natu]	नातू	grandson	[nete]	नेते	leaders
14	nh	[nʰavi]	न्हावी	barber	[nʰeʈi]	न्हेटी	farming implement*
15	l	[lagu]	लागू	applicable (also, a last name)	[leki]	लेकी	daughters
16	lh	[lʰasa]	ल्हासा	capital of Tibet	[lʰete]	ल्हेते	wood-carving tool*
17	ch	[tʃaru]	चारू	girl's name	[tʃele]	चेले	apprentices, sidekicks
18	chh	[tʃʰati]	छाती	chest	[tʃʰedʌ]	छेडा	surname
19	j	[dʒagruk]	जागस्क	alert	[dʒetʰe]	जेथे	where (locative)
20	jh	[dʒʰapʌd]	झापड	slap	[dʒʰepʌ]	झेप	leap
21	r	[raja]	राजा	king	[reti]	रेती	sand
22	rh	[rʰas]	ऱ्हास	deterioration	[rʰekʌ]	ऱ्हेक	ancient style of planting*

word medial							
<i>before [a]</i>			<i>before [e]</i>				
23	p	[sopa]	सोपा	easy (masc)	[sope]	सोपे	easy (neuter)
24	ph	[gup <sup>h</sup> a]	गुफा	cave	[gup <sup>h</sup> e]	गुफे	caves
25	b	[dʌba]	डबा	box	[dʌbe]	डबे	boxes
26	bh	[ub <sup>h</sup> a]	उभा	standing (masc)	[ub <sup>h</sup> e]	उभे	standing (neut)
27	m	[rʌma]	रमा	(proper name)	[vime]	विमे	insurance policies
28	mh	[tum <sup>h</sup> a]	तुम्हा	to you	[tum <sup>h</sup> e]	तुम्हे	to you
29	v	[hʌva]	हवा	air	[hʌve]	हवे	wanted
30	vh	[giv <sup>h</sup> a]	गीव्हा	tongue	[nʌv <sup>h</sup> e]	नव्हे	is not
31	t	[neta]	नेता	leader	[nete]	नेते	leaders
32	th	[kʌt <sup>h</sup> a]	कथा	story	[tit <sup>h</sup> e]	तिथे	there
33	d	[gʌda]	गदा	mace (weapon)	[pʌde]	पदे	couplets
34	dh	[gʌd <sup>h</sup> a]	गधा	donkey	[gʌd <sup>h</sup> e]	गधे	donkeys
35	n	[mʌna]	मना	mind	[sone]	सोने	gold
36	nh	[pun <sup>h</sup> a]	पुन्हा	again	[pʌn <sup>h</sup> e]	पन्हे	mango drink
37	l	[pela]	पेला	glass/tumbler	[pele]	पेले	glasses/tumblers
38	lh	[dʒil <sup>h</sup> a]	जिल्हा	district	[dʒil <sup>h</sup> e]	जिल्हे	districts
39	ch	[tʌtʃa]	त्वचा	skin	[lotʃe]	लोचे	messes
40	chh	[pot <sup>h</sup> a]	पोछा	act of mopping (n)	[pot <sup>h</sup> e]	पोछे	acts of mopping (pl)
41	j	[mʌdʒa]	मजा	fun	[mʌdʒet]	मजेत	in fun
42	jh	[kidʒ <sup>h</sup> a]	किझा	Malaysian fruit*	[kidʒ <sup>h</sup> e]	किझे	Malaysian fruits*
43	r	[kʌra]	करा	do	[tare]	तारे	stars (pl.)
44	rh	[tʌr <sup>h</sup> a]	तन्हा	“the way” (of doing something)	[tʌr <sup>h</sup> e]	तन्हे	ways (of doing something)

\*Indicates a nonce word.