THE ENGLISH AND FOREIGN LANGUAGES UNIVERSITY

HYDERABAD, A.P.

MASTERS THESIS

Acoustic Classification of Velar Fricatives in Assamese

Author:
Charles H. Redmon

Supervisor:
Dr. Indranil Dutta

A thesis submitted in partial fulfillment of the requirements for the degree of Master of Arts

in

Linguistics

from the

School of Language Sciences

May 2014

Declaration of Authorship

I, Charles Redmon, declare that this thesis titled, 'Acoustic Classification of Velar Fricatives in Assamese' and the work presented in it are my own. I confirm that:

- This work was done wholly while in candidature for an M.A. degree at this University.
- Where I have consulted the published work of others, this is always clearly attributed.
- Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work.
- I have acknowledged all main sources of help.
- Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself.

Signed:

Date:

Acoustic Classification of Velar Fricatives in Assamese

by Charles H. Redmon

Abstract

Acoustic data was collected from twelve native speakers of Assamese – representing Jorhat and Nalbari varieties – to examine the velar fricative /x/ relative to acoustically similar posterior obstruents (/ɦ/, /k^h/, and /$g^ɦ$/) in word-initial, inter-vocalic, and word-final environments. Spectral moments from 20 ms windows at different locations in the consonant and preceding/following vowel were calculated and used alongside amplitudinal measures in a statistical classification task on the four target consonants. For classification, discriminant analysis was performed to determine the degree of separability between the velar fricative and posterior obstruents with similar acoustic features. Phonemic contrasts were examined for the three segmental environments in three subject groups: female speakers from Jorhat, male speakers from Jorhat, and male speakers from Nalbari. Overall classification accuracy of the model in the three subject groups was 79.8, 74.7, and 63.9%. Results proved highly variant depending on segmental context, particularly in pairwise comparisons. For example, accurate identification of /x/ word-initially was at 72, 69, and 63% for Jorhat Female, Jorhat Male, and Nalbari Male groups, respectively, as compared with 89, 83, and 88% word-finally.

As a test of the inter-relations between these groups, discriminant analysis was performed in cross-sex and cross-dialectal designs, using classifiers trained on the Jorhat Female and Nalbari Male sets to predict category distinctions in the Jorhat Male group. Results of the cross-sex design did not show any clear evidence for maintaining a gender distinction with respect to behavior of the posterior obstruent system; however, the cross-dialectal design did indicate that the Jorhat and Nalbari varieties are distinct relative to some contextually restricted obstruent contrasts. Thus we have been able to show that in studying the acoustics of velar frication in Assamese, maintaining distinctions between speech varieties is important, particularly when considering the system of phonemic contrast as a whole.

Acknowledgements

First and foremost I would like to thank my advisor, Indranil Dutta, for both his early guidance in teaching me the practice and philosophy of experimental science, and his specific contributions to the development of this thesis. Many intricacies of the design and its implications would have been lost without his input.

Additionally, the support of faculty at IIT Guwahati was invaluable; without it this experiment would have never gotten off the ground. Dr. Priyankoo Sarmah was instrumental in the development of the elicitation set, recruitment of speakers, and recording setup, and Dr. S. R. Mahadeva Prasanna was kind enough to allow the use of his lab and sound-attenuated booth for recording the material for this study.

To the friends who put up with my late nights and scattered mornings, I'm glad you never allowed me to lose perspective. And to Pamir I owe a particular debt of gratitude for answering n number of random questions about Assamese and advising me throughout the process.

Lastly I would like to thank my mother, for standing by as I traveled to the other side of the world to study linguistics and ultimately produce this research. She has been supportive the whole way, despite the difficulty of being separated from her son. This thesis is for her.

Contents

List of Figures

List of Tables

List of Symbols

f_i	frequency (in Hz) at bin i
A	root mean square (RMS) amplitude (in dB)
W_n	temporal window at position n
F_n	formant frequency n
F_p	spectral peak location (in Hz)
μ_n	the n^{th} central moment of the spectral energy distribution
A_n	normalized amplitude (in dB)
A_d	dynamic amplitude (in dB)
A_{r_n}	relative amplitude over Bark frequency range n (in dB)
F	Fisher test statistic
K	Kruskal-Wallis test statistic
$X_{S/P}$	random vector of parameters for subject S and environment P
x_i	random variable (acoustic parameter) in feature vector X
ω_i	category assignment from discriminant function
P_i	prior probability of category ω_i
p_i	conditional density function of category ω_i
Σ_i	variance-covariance matrix of category ω_i

For my mother

Chapter 1

Introduction and Literature Review

Within the sound inventory of Assamese, the velar fricative occupies a unique position. It is on the one hand one of the most ubiquitous features of the language, and is frequently invoked as a novelty among Indo-Aryan languages [Goswami, 1966, Kakati, 1941, Patgiri et al., 2013]. And yet at the same time very little is understood about its acoustics, phonotactic patterning, or interaction with other obstruents in the maintenance of phonemic contrast. The goal of this study is to begin to fill that gap. In this regard we investigate an array of spectral and amplitudinal measures of velar fricatives in different segmental environments by different speaker groups (representing gender and dialectal distinctions) in Assamese. The methodological paradigm adopted throughout the work will follow the perspective of statistical pattern classification, wherein the goal will be to distinguish the velar fricative from other obstruent sounds in Assamese on the basis of acoustic parameters alone. From this methodology we will also be able to determine which spectral and amplitudinal parameters are most significant in identifying the velar fricative relative to other acoustically similar sounds.[1]

1.1 Previous Work on Assamese

The first notable mention of velar frication in modern linguistic work on Assamese comes from G. A. Grierson, who in reference to the palato-alveolar fricative in Magadhi Prakrit notes that "in Assamese the sound has been further weakened to almost the sound of the *ch* in 'loch'." [Grierson, 1903, p. 6]. No further description of this sound was given

[1]The precise meaning of 'acoustically similar' in the context of this work will be given later in this chapter.

until Kakati [1941], whose analysis of linguistic structure and diachronic developments in Assamese remains the most thorough and expansive work on the language. Kakati's description of the consonant set establishes Grierson's *ch* as a voiceless velar fricative, further emphasizing that this sound, among Indo-Aryan languages, 'is peculiar to Assamese only' [Kakati, 1941, p. 60]. The remainder of the posterior obstruent set in Kakati's analysis comprises the plain voiced/voiceless velar plosives /k, g/, the aspirated/breathy velar plosives /k^h, $g^ɦ$/, and the voiced glottal fricative /ɦ/. Notably, Kakati mentions several alternations between members of this group of sounds, including aspiration of plain stops, lenition of /k^h/ word-finally, and fortition of /x/ to a 'soft' [k^h] inter-vocalically [Kakati, 1941, p. 220].

Kakati's focus is primarily diachronic, however, and his treatment of consonants is relatively sparse in accounting for phonotactic patterns. Goswami [1966] provides a more complete and phonologically formal treatment of the sound system in Assamese. In addition to echoing phenomena of spirantization and word-final lenition of aspirates, Goswami details finer distinctions in consonantal place and manner of articulation according to segmental and syllabic distributions. Regarding the velar fricative, he comments that 'the phonetic range of this phoneme is very large', and that its palatal variant adjacent to high/mid-front vowels is 'almost frictionless' [Goswami, 1966, p. 50]. Preceding and following the low vowel /a/, Goswami describes the place of articulation (POA) of /x/ as *mid-velar*; adjacent to back vowels POA becomes *back-velar* [Goswami, 1966].

To date the most extensive work on the subject of dorsal frication in Assamese is Sarma [2012], who analyzed both Barpeta and Standard dialects of Assamese, finding both exhibited a uvular fricative phonemically, with the velar being only an allophonic realization word-finally. Word-internally Sarma [2012] found that there was a dialectal difference in terms of the propensity for fortition in onset position. The Barpeta dialect exhibited this shift from fricative to plosive realization before all vowels, but in the standard dialect such cases could only be found with high vowels. Regarding spirantization, Sarma [2012] notes that the movement from /k^h/ to [x] occurs in the coda in both varieties, but in BA it is comprehensive; i.e. word-internally in addition to the word-final spirantization found in the standard dialect. Sarma admits, however, that this process may not be categorical but rather could be a gradient phenomenon. It may also be prosodically or morphologically conditioned. Sarma [2012] also describing the glottal fricative, which is typically claimed to be *voiced*, as voiceless, a fact which would have significant consequences for the system of contrast in Assamese.

Ultimately, Sarma's argument for the dorsal fricative's underlying status as a uvular comes from an acoustic description of the word-initial fricative spectrum, wherein the main acoustic resonance was found to be below that observed for the same consonant word-finally. In this regard, he states:

> "In the voiceless uvular fricative /χ/ of Assamese, the tongue body is not raised unlike the voiceless velar fricative /x/... Hence, /χ/ creates a larger space in the vocal tract... [and] less noise is noticed in /χ/ as a consequence of less friction" [Sarma, 2012, p. 24]

This larger space conceivably would also contribute to a lower main resonance, though whether or not the differences between onset and offset spectra suggest different places of articulation is unclear at this point. Velar-uvular alternations have been suggested in other recent work on Assamese [see Mahanta, 2012], but as yet such cases seem to be lexically idiosyncratic. In providing a thorough account of velar fricative acoustics in Assamese we hope to gain some insight into this issue, or at least what acoustic parameters would be involved in its solution.

Unfortunately acoustic phonetic work on Assamese is largely absent from the literature, and what information has been presented – as acoustic evidence for phonological processes in Mahanta [2007], for example – has generally focused more on the vowel system. One study on fricative identification from the computational perspective has been proposed in [Patgiri et al., 2013], though results of this study are not available at present. When complete, the study by Patgiri and colleagues should be a valuable companion to the work presented in this thesis.

1.2 Acoustics of Obstruent Consonants

Early work on obstruent acoustics – particularly those exhibiting substantial aperiodic noise, i.e. fricatives, affricates and aspirated plosives – identified spectral amplitude and its distribution over the frequency range as the fundamental baseline for distinguishing between consonants of this manner but differing in place or voicing characteristics. Hughes and Halle [1956] observed that the voiced fricatives in English are primarily defined by a strong amplitude component below 700 Hz (likely due to the effect of vocal cord vibration). Furthermore, place distinctions could be distinguished at least for the sibilants. The alveolar fricatives /s/ and /z/ were found to have spectral peaks consistently higher in frequency than their palato-alveolar counterparts. However, notable problems were encountered when attempting to characterize labio-dentals according to the same criteria. Strevens [1960] elaborated on these findings, establishing bounds on regions of visible energy in the spectrogram for a range of fricative types. His analysis included data on palatal, velar, uvular, and glottal fricatives, the latter three of which were notable for their formant-like structure and upper limits of energy around 6.5 kHz. These sounds were also found to exhibit a major spectral peak around 1.5 kHz, making them distinct from other classes but less clear as to their internal classification. Strevens [1960] notably used these acoustic results to extrapolate on the production characteristics which produced them, noting among other observations that the key variables which must be involved are *orifice shape*, *location*, and properties of the *flow* through the constriction.

With the grounding of acoustic phonetics in the source-filter theory of Fant [1960], studies on obstruent acoustics shifted emphasis toward a more exact understanding of the production processes involved in generating the descriptive patterns reported above. Most prominent among this work is that of Shadle [1985], who used a combination of mechanical and theoretical models and real speech samples define the noise sources involved in various types of fricative consonants. The primary distinction established in Shadle's thesis was that of *obstacle* versus *no-obstacle* configurations in turbulence production. Of note for this study is her finding that while velar fricatives have traditionally been thought of as characterized only by a wall source of friction, the direction of the airstream at the hard palate in many cases fits the *obstacle* configuration, thus placing the sound generated by these fricatives more closely in line with sibilants than with diffuse obstruents like /f/. Further refinements of these findings in Shadle [1990] led to the emphasis on velar and palatal fricatives as *distributed* rather than *point* sources. Finally, after Shadle [1985] it was later shown that in addition to production-based distinctions between obstacle and non-obstacle frication, the resulting acoustics could also be accounted for by way of the properties of the filter ahead of the constriction [Badin, 1989]. This result allowed for a grounding of obstruent acoustics in the shape of the spectrum, which follows directly from the filter function.

Thus far we have focused on pure descriptions of obstruent acoustics with the goal of understanding their underlying production mechanisms; however much of the work in the last three decades has adopted a slightly different paradigm. Here the aim has been to test an array of acoustic features on obstruent consonants and simply choose the set which best captures the distinctions in a given linguistic system. Kobatake and Ohtani [1987], for example, used a principal components analysis (PCA) of the stop spectra in voiceless labial, alveolar and velar stops to extract spectral shape and peak frequency parameters in the classification of this contrast. [Forrest et al., 1988] adopted a similar paradigm, only in their work obstruent spectra are first broken down into central moment-based parameters to account for the shape and energy distribution in the spectrum. Their results also indicated that when information from multiple locations throughout the consonant and vowel is included in the design, classification accuracy improves substantially. Further refinements to this spectral moment method can be seen in Shadle and Mair [1996], Jongman et al. [2000], [Tabain, 1998], and Jassem [1995], and this framework will also serve as the backbone of our analysis and classification of obstruents in Assamese.

1.3 Statistical Modeling of Speech Systems

The final major line of thinking which informs the present study is that of speech systems as understood through the lens of pattern recognition. Nearey [1997] presents this approach as a computational analogue of the perception process in human beings. Though human perception and statistical classification are not held to operate fundamentally in the same manner, the output of the machine learning process can provide some insight into which cues, if any, are likely to be latched onto by the auditory perception system. For empirical demonstration of this relation between statistical modeling and human perception results, Nearey synthesized /h*VC*/ nonce words across five back/central vowels and two alveolar stops, controlling for vowel and voice bar duration, and presented the stimuli to subjects for identification. The results of the perception experiment were then compared with results of logistic regressions with vowel and consonant category as dependent variables (in separate designs) and two types of acoustic parameters – voicing/vowel durations and formant frequencies – as predictors. In finding that the obtained coefficients from the logistic model are significantly consistent across subjects, Nearey [1997] demonstrates the degree of compatibility between the two methodologies (perception experiment and computational modeling via pattern classification), suggesting that the latter may be utilized as a robust means of extracting acoustic cues to linguistically significant phonetic categories.

In the context of the type of sounds which are the subject of this thesis, the work of Nossair and Zahorian [1991] is perhaps the most informative. Using an automatic recognition system, they are able to demonstrate computationally the importance of spectral shape cues over formant or amplitude-based measures extracted from obstruent consonants. They also note, echoing Forrest et al. [1988], that inclusion of temporal information in the model yields far superior performance than reliance on static cues. This finding is further substantiated in the work of Milner [1996], whose use of temporally stacked vectors in speech recognition has helped inform the methodology employed in this thesis.

Both Nearey [1997] and Nossair and Zahorian [1991] emphasize the correlation between cues in the multivariate classification model as critical to the design of recognition systems. One reason for this, as shown in McMurray and Jongman [2011], is that the relative importance of a particular acoustic cue is not static, but may be modulated by the prior or simultaneous occurrence of another cue or set of cues. This inter-relation between cues to category identity, where a particular cue may become more or less salient in a given context due to the heterogenous variance of the set cues as a whole, is often described in the literature on perception as *compensation*. More plainly, the ability of the perceptual system to compensate for noise in a subset of cues by prioritizing another subset is fundamental to the categorization process employed by humans. McMurray and Jongman [2011] confirm this presupposition by showing that among a variety of different classification models, the one incorporating cue compensation – specifically, the Computing Cues Relative to Expectations (C-CuRE) model – was the only one which patterned similarly to listener judgements from a perception experiment. While this compensatory model is not adopted in the present work, McMurray and Jongman's demonstrated parallelism between human perception and statistical classification validates our adopted perspective in the study of Assamese velar fricative acoustics.

Having reviewed the fundamental points of consideration taken from the literature on Assamese, general phonetics, and statistics, the organization of the rest of the thesis is as follows. Chapter 2 presents an analysis of the acoustic descriptors used to define velar frication in Assamese. Because we consider the acoustics of velar frication as a classification problem, a target set of consonants was identified to serve as referents in the model. In this regard we chose to focus on those sounds with the greatest acoustic similarity to the velar fricative; that is, we narrowed the problem to classification within a set of obstruents with a turbulent sound source located in the posterior region of the vocal tract. Thus the focal consonant /x/ was examined relative to the two aspirated/breathy velar plosives, /k^h, $g^ɦ$/, and the glottal fricative /ɦ/. From the acoustic description in Chapter 2, we turn to Chapter 3 where the acoustic parameters established in Chapter 3 are used in a classification problem between the target obstruent set. Finally, Chapter 4 concludes the work and offers areas for extension/refinement of the model.

Chapter 2

Acoustic Analysis

The goal of this chapter is to formulate and assess a set of acoustic parameters which may be used both as general descriptors of the target posterior obstruent set, and as predictors – or 'cues' – in a statistical classification task. As introduced in the previous chapter, by *target posterior obstruents* we mean the set of sounds corresponding to the phonemic categories /x, ɦ, k^h, $g^{ɦ}$/ in Assamese. Measures are presented for two regional varieties – Jorhat and Nalbari – of which the Jorhat group comprises both male and female data.[1]

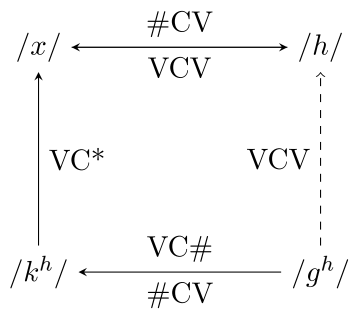


Figure 2.1: Neutralization potential among posterior obstruents

Returning to the problem of triangulating velar fricative acoustics in the Assamese speech system, our aim in the next two chapters will be to account for phonemic contrast and its corresponding loss/neutralization in a quantitative manner. Figure 2.1 diagrams the posterior obstruent system in Assamese, with arrows indicating direction of mapping from one category onto another (solid for examples cited in the literature, dashed for suggested but not substantiated pathways) and segmental environment noting the context of reported neutralizations [Goswami, 1966, Kakati, 1941, Mahanta,

[1]Though questions of comparison between the different speech varieties presented in this thesis are interesting and given some treatment in our discussion of results, it should be noted that in developing this design, addressing sociolinguistic issues is not our primary intent; rather, the breakdown into gender and dialect-based groups is simply a necessary control on the particular linguistic system being modeled.

2012]. In presenting this diagram as the guiding picture of the problem we seek to tease apart, we note that throughout our ivestigation of a range of acoustic measures within different speaker groups and segmental environments, accounting for the velar fricative within this system is our ultimate aim. For this reason contrasts of greater relevance to acoustic identification of the velar fricative will often be emphasized over other contrasts within the same system. This fact will be made clearer in the design of classification tasks in Chapter 3.

2.1 Materials and Methods

2.1.1 Participants

Twelve native speakers of Assamese were recruited from two localities in Assam: Jorhat, representing the Eastern Assamese variety, and Nalbari, representing a variety of the Western Assamese dialect group. The Jorhat group comprised 4 female speakers (median age: 33.5) and 4 male speakers (median age: 35.5), while the Nalbari group was limited to 4 male speakers (median age: 41.5), as female speakers from Nalbari could not be recruited at the time of data collection. All speakers were educated in Assamese medium up to class 10 and exhibited no known speech or hearing difficulties. Subjects were compensated for their participation in the study.

2.1.2 Stimuli

Three real speech tokens were recorded for each of the 16 target sequences (4 consonants, /x, ɦ, k^h, $g^ɦ$/, × 3 segmental environments, {#C[a], [ɔ]C[ɔ], [a]C#}; plus 2 consonants, /x, k^h/, × 2 segmental environments, {#C[i], [i]C#}). The latter set comprising the two voiceless velars in positions adjacent to the high-front vowel was used only for descriptive purposes to address the question of vowel-dependent allophony (relative to #C[a] and [a]C# environments); our primary concern in evaluating acoustic measures and implementing the resulting parameters in a classification task will be restricted to the former set of /x, ɦ, k^h, $g^ɦ$/ fixed in word-initial (#C[a]), inter-vocalic (VCV) and word-final ([a]C#) positions.[2] The 48 tokens were recorded in isolated and utterance-framed contexts, where the frame utterance was *moi tat* ___ *buli likha dekhilu* (‘I saw ___ written there’), and repeated in 3 randomized blocks, yielding 288 stimuli per speaker, or 3456 total productions across the three subject groups. This number was later reduced to 1610 as items were held out from further analysis as a result of one of three factors: (1)

[2]Unless specified otherwise, *CV* and *VC* will often be used as shorthand for the low vowel variant of word-initial or word-final environments

word-initial and word-final tokens in the utterance-framed context could not be isolated from coarticulatory effects of preceding and following word boundaries; (2) four words initially recorded to complete the VCV and $VC\#$ sets (/g^{ɦ}ɔmɔg^{ɦ}ɔntɔ/, /nak/, /rag/, and /tik^{h}a/) were later determined to be categorically different in their production patterns from the remainder of the set; and (3) general errors in production of target stimuli.[3]

The 48-word target set was then padded with 20 *dummy* words, plus an additional 12-word set of voiceless velars in word-initial and word-final environments adjacent to the high-back vowel /u/, which was not analyzed further for reasons discussed below. The dummy set was chosen so as to provide a variety of phonetic types out of sync with the turbulent sound-based target set.

2.1.3 Recording Setup

All speech samples were recorded in a sound-attenuated booth in the Electro-Medical and Speech Technology Laboratory at the Indian Institute of Technology in Guwahati, Assam. Subjects were seated at a table in the middle of the room, in front of a computer monitor and a fixed AudioTechnica 2020 cardoid condenser microphone, positioned approximately 10 cm from the subject's mouth, angled slightly out of axe to reduce effects of wind noise.[4] Stimuli were presented on the computer monitor in a three-slide sequence for each token. First, an English translation of the target word was given to cue the lexical meaning of the following word (and mitigate simple reading of the orthography). Subjects were instructed that they were not to produce the English word out loud, and that the English translation merely served as an aid telling them what Assamese word to expect on the next slide.[5] On the second slide the target word was presented in Assamese orthography; subjects were instructed to read this word aloud as naturally as possible. Finally, a third slide was presented with the question: *Apuni tat ki buli likha dekhile?* ('What did you see written there?'), to which subjects were instructed to respond with the frame *moi tat* ___ *buli likha dekhilu*, with the word on the previous slide inserted in the blank. This question-answer methodology was adopted to reduce the influence of reading prosody. All stimuli were randomly arranged in three blocks with a 1 minute break between blocks. In the first block 10 additional dummy words were

[3]The most common problem encountered across subjects, but particularly in the male groups, was approximant-like productions of /g^{ɦ}/ and /ɦ/, which could not be analyzed along the same acoustic dimensions as the turbulent sounds on which this study is based

[4]Despite this preventive measure, initial recordings of C[u] and [u]C sequences had to be removed from consideration due to wind noise produced as a result of excessive expulsion of air from the rounded lip configuration for these vowels.

[5]All English-Assamese pairs were given to subjects prior to running the experiment to ensure that subjects were both familiar with all the Assamese words and comfortable with the English translations.

presented at the start to allow subjects time to adjust to the experimental procedure. Stimulus presentation was not timed, but rather subjects were allowed to navigate the slides from a keyboard, and thus could move through the experiment at their own pace. Though this greater freedom allowed less control over the speed and constancy of production, speakers were found to adopt a relatively constant pace early-on and maintain that pace throughout the experiment. Moreover, the overall speed of completion was not markedly different between subjects, with a few exceptions. In such cases subjects were removed from the final set for analysis.[6]

Audio signals were input via USB to a MacBook Pro and digitized in Praat at 44.1 kHz with 16-bit quantization. Speech samples were further filtered with a high-pass Hann band of 50 Hz cutoff frequency to correct for the presence of low-frequency noise resulting from the recording room's location adjacent to a computer lab and AC equipment. Similar conditions and filtering methods were reported in Shadle [1985] and Forrest et al. [1988].

The four target consonants – /x, ɦ, k^h, $g^ɦ$/ – comprising a total of 1610 separate speech samples covering three target positions (CV, VCV, and VC) were then examined with a range of measures along both spectral and amplitudinal dimensions. As noted at the beginning of this chapter, the ultimate aim in selecting these measurements will be to produce a stable set of cues for use in distinguishing posterior obstruent contrasts in different vocalic and syllabic contexts.

2.1.4 Spectral Measures

Two notable frameworks for analysis of noise spectra in obstruent consonants may be identified from the literature discussed in the previous chapter. First, a more deterministic approach can be seen in measures like spectral peak frequency as an indicator of place of articulation. This measurement standard has its ultimate theoretical grounding in the work of Fant [1960], where peak frequencies are related to vocal tract resonances and consequently the locus of constriction where the obstruent is produced; however, it has been used descriptively in earlier work by Hughes and Halle [1956], and remains a mainstay for baseline assessments of production characteristics [see Shadle, 1990]. The second approach is utilized much more in classification tasks where multiple obstruent categories must be distinguished within a give phonemic system. In this framework the aperiodic source in such consonants (i.e. *noise*) is foregrounded. Shadle [1985] uses this fundamental property in her justification of treating the description of obstruent

[6] In total 18 subjects were recorded: 4 female speakers from Jorhat, 7 male speakers from Jorhat, 1 female speaker from Nalbari, and 6 male speakers from Nalbari. Twelve were retained in the final analysis, primarily to balance the other subject pools with the Jorhat Female group.

acoustics as a statistical problem, leading namely to the work of Forrest et al. [1988], and by extension Jassem [1995], Jongman et al. [2000], and others. We adopt the first approach briefly in a review of peak frequency in the two velar obstruents – /x/ and /k^h/ – in both high and low-vowel environments. The second approach, however, will form the basis for the rest of the analysis presented in this chapter, as well as the statistical classification task described in Chapter 3.

Peak frequency, or *spectral peak location* was calculated as the frequency of maximum amplitude from 0.5 to 17 kHz. Spectra for this measurement were generated from a 512 point FFT of a 20 ms Hann-windowed segment at the midpoint of the noise interval in /x/ and /k^h/ preceding and following the low vowel /a/ and the high front vowel /i/ in both word-initial and word-final environments. This measurement is not used in the general classification procedure, but rather serves as an initial descriptor of the velar fricative; calculations from the voiceless plosive counterpart serve as a necessary referent when considering questions of place of articulation as derived from acoustic resonance.

The rest of the spectral measures reported in this study – those which will be utilized as predictors of category membership between the four target obstruents – are essentially spectral shape parameters, calculated following the procedure in Forrest et al. [1988]. Their approach is to treat the noise spectrum as a random probability distribution whose form may then be characterized by an ensemble of central moments, where the n^{th} moment about the mean of the distributions is defined as

$$\mu_n = E\{(\mathbf{x} - \eta)^n\} = \int_{-\infty}^{\infty} (x - \eta)^n f(x) dx \tag{2.1}$$

where η represents the first moment, μ_1, which is the distribution *mean* $E\{\mathbf{x}\}$, and higher moments (n = 2, 3, ...) follow from 2.1 [Papoulis and Pillai, 2002]. In the discrete case for a 512 point FFT (i.e. that derived by extending a 20 ms, or 400 point Hann-window with zeros to the next higher power of 2), this equation becomes

$$\mu_n = \sum_{i=1}^{256} (f_i - \mu_1)^n A(i) \tag{2.2}$$

for $n > 1$, where $A(i)$ is the RMS amplitude at frequency bin i. Since this definition is recursive, the first central moment is defined separately as

$$\mu_1 = \sum_{i=1}^{256} f_i A(i) \tag{2.3}$$

The first moment is also commonly referred to as the *center of gravity*, and essentially represents the ‘balancing’ point of the spectrum. This measure is utilized over spectral peak location for categorization among posterior obstruents because while the two should theoretically be aligned in spectra possessing a single main resonance, μ_1 is more robust to temporal variation (particularly as the consonant moves into the vowel) and furthermore is applicable to diffuse spectra and spectra exhibiting multiple resonances which decrease in amplitude with frequency as predicted from source-filter theory [Fant, 1960, Stevens, 1998].

The second moment about the mean – i.e. the *variance* (σ^2) – captures the relative dispersion of the distribution. Standard deviation (σ), is commonly used in place of the second moment, and is simply the square root of the variance. Here we will refer to the standard deviation computed from the spectrum as the *spectral dispersion*, and denote it μ_2 for simplicity and continuity with the *spectral mean* parameter. Higher moments are then scaled by μ_2, or the standard deviation σ, to produce ‘dimensionless’ measures [Newell and Hancock, 1984]. Spectral *skewness*, capturing the degree of asymmetry in the spread of energy over the spectrum, is represented in the third moment, μ_3, which is scaled as $\mu_3^* = \mu_3/\sigma^3$. Spectral *kurtosis*, or the relative ‘peakedness’ of the distribution, corresponds to the fourth moment, scaled as $\mu_4^* = \mu_4/\sigma^4$. While we emphasize here that as with the standard deviation, skewness and kurtosis are not raw measures of the third and fourth moments about the mean, we will use the simpler notation of μ_3 and μ_4 for skewness and kurtosis, respectively, throughout the paper.

The first four moments were calculated from a 20 ms Hann window centered at multiple locations in the consonant and following/preceding vowel. In the word-initial environment, spectra were calculated at three points in the consonant – referred to as the initial, medial, and final positions (W_{C_i}, W_{C_m}, W_{C_f}) – one centered at the consonant-vowel boundary (W_{CV}), and one spanning the initial 20 ms of the following vowel (W_{V_2}). Word-finally, the same five window locations are used, in linear-temporal sequence: W_{V_1}, W_{VC}, W_{C_i}, W_{C_m}, W_{C_f}. Inter-vocalically, spectra are computed at seven points; i.e. the word-final sequence plus W_{CV} and W_{V_2} to capture the transition into the following vowel. For all obstruents the transition between consonant and vowel was identified from both the appearance of periodicity in the signal and the onset of first formant (F_1) energy in the spectrogram. Where there was an overlay of noise on the periodic signal, transition points could usually be identified as the offset/onset (per CV/VC context) of high frequency energy in the spectrogram. As for the window locations over the consonant interval, separate procedures were used for fricatives and plosives. Because all word-initial and word-final samples were taken from the isolated productions, fricative onset could not reliably be identified, often exhibiting a gradual transition from silence to full fricative noise. For this reason ‘initial’ and ‘medial’ locations were not stable positions for the fricative consonants, and C_i/C_m were therefore defined relative to the vowel boundary instead. As C_f covered the final 20 ms before vowel onset, C_m was defined from a 10 ms shift from C_f into the vowel, with C_i shifted a further 10 ms. Therefore, $C_i - C_f$ span a 40 ms interval in the fricative consonants. With the plosives, consonant onset could reliably be identified from the burst, so C_i and C_m were defined at onset and midpoint of the noise interval for /$\mathrm{k^h}$/ and /$\mathrm{g^ɦ}$/. Spirantized plosives were analyzed according to the procedure used for the fricatives. See Figures 2.2 – 2.4 for sample spectra and window locations on the corresponding signals.

The frequency range used for calculation of spectral moments in Forrest et al. [1988], Jongman et al. [2000], and others is $0 - 10$ kHz, though other ranges have been tested to determine whether or not inclusion of higher frequencies is necessary for discriminating certain classes of obstruents [Shadle and Mair, 1996, Tabain, 1998]. As a preliminary test of the effects of manipulating the frequency range for spectral moment measurement, $\mu_1 - \mu_4$ were calculated in word-initial position for a pilot set of speakers from the Jorhat Female set over four frequency ranges: (1) 50 Hz $-$ 10 kHz, (2) 50 Hz $-$ 17 kHz, (3) 200 Hz $-$ 10 kHz, and (4) 200 Hz $-$ 17 kHz. Of the four frequency ranges, the 50 Hz $-$ 10 kHz range exhibited the least variance in the four spectral moments over all window locations. Thus the range used in Forrest et al. [1988] was adopted for spectral measurements in this study.[7]

2.1.5 Amplitudinal Measures

Three measures of amplitude were implemented: *normalized amplitude*, *dynamic amplitude*, and *relative amplitude*. Normalized amplitude is adopted as a way of measuring the overall noise amplitude in the consonant while controlling for global differences between subjects, as well as local effects of prosody and intonation [Jongman et al., 2000]. Denoted A_n, we define normalized amplitude as the mean RMS amplitude between 50 Hz and 17 kHz in a 20 ms window at the midpoint of the consonant, minus the same amplitude calculation at the midpoint of the preceding or following vowel. Thus $A_n = \bar{A}_C - \bar{A}_V$, where in the inter-vocalic case $\bar{A}_V$ is the mean of $\bar{A}_{V_1}$ and $\bar{A}_{V_2}$. Relative amplitude is calculated in a similar manner, only instead of taking the mean amplitude over the full frequency range we measure relative to the vowel in specific frequency bins. Stevens [1998], Jongman et al. [2000], and others report relative amplitudes approximately centered at the corresponding frequencies of F_3, F_4, and F_5 in the vowel. For the present study we chose to make this formulation more exact and define specific frequency ranges for each of the third through fifth formant frequencis. Because auditory bands represent a natural binning of the audible frequency range, the regions $F_3 - F_5$ were

[7]The 50 Hz lower bound on the range comes from the initial high-pass filtering described in the Recording Setup.

partitioned according to the 16th through 18th Bark frequency bands; i.e. 2675 – 3125 Hz, 3125 – 3675 Hz, and 3650 – 4350 Hz. Calculations of consonant amplitude relative to the vowel were then taken as the difference between the maximum RMS amplitude in each bin in the final and initial 20 ms windows at CV/VC boundary. Thus $A^{cv}_{r_{16}}$ represents the maximum amplitude between 2675 and 3125 Hz in the C_f window minus the maximum amplitude over the same range in the V_2 window. Relative amplitude in word-initial frames and word-final frames where the vowel offset is at the onset of the noise interval of the consonant (i.e. word-final fricatives and spirantized plosives) was captured in three parameters: $A_{r_{16}}$, $A_{r_{17}}$, and $A_{r_{18}}$. Inter-vocalically for both the VC transition and the CV are represented, yielding six relative amplitude parameters.

The final amplitudinal measure is dynamic amplitude, which unlike the previous two measures is not defined relative to the vowel but rather refers to internal amplitude distinctions between different frequency ranges in the noise spectrum. Dynamic amplitude, or A_d, is presented in Shadle and Mair [1996] as the difference between the maximum amplitude over the range 0.5 to 17 kHz and the minimum amplitude from 50 to 2000 Hz. This measure was derived by Shadle and Mair from two separate measures, A_T and A_0, presented in Shadle [1985], where A_T was the amplitude over the total frequency range and A_0 represented the low-frequency amplitude, i.e. the amplitudinal range from 500 Hz to the frequency of peak amplitude. Shadle and Mair [1996] use dynamic amplitude as a measure of the 'noise source characteristics' in the consonant, noting its primary use in distinguishing *obstacle* and *no-obstacle* configurations in fricative production. For the present study A_d was calculated over the same window as that used for consonant amplitude in A_n; i.e. the middle 20 ms of the noise interval.

2.2 Results

Each of the 36 acoustic parameters (4 spectral moments $\times$ 7 window locations, plus 8 amplitudinal measures) was tested for significant main effects of consonant type in word-initial, inter-vocalic, and word-final environments. One-way repeated measures ANOVAs were performed for the word-initial context, where the four-way contrast among posterior obstruents exhibited approximately equal sample sizes. In post-vocalic environments, because only spirantized plosives were studied relative to the velar and glottal fricatives, these designs were inherently unbalanced (due to the fact that only a portion of the velar plosives would be spirantized). In this case the assumption of equal category variances could not be met and therefore analysis of variance was not valid in inter-vocalic or word-final tests. Thus we adopted the non-parametric Kruskal-Wallis rank sum test for evaluation of consonant category effects in these two environments [R Core Team,

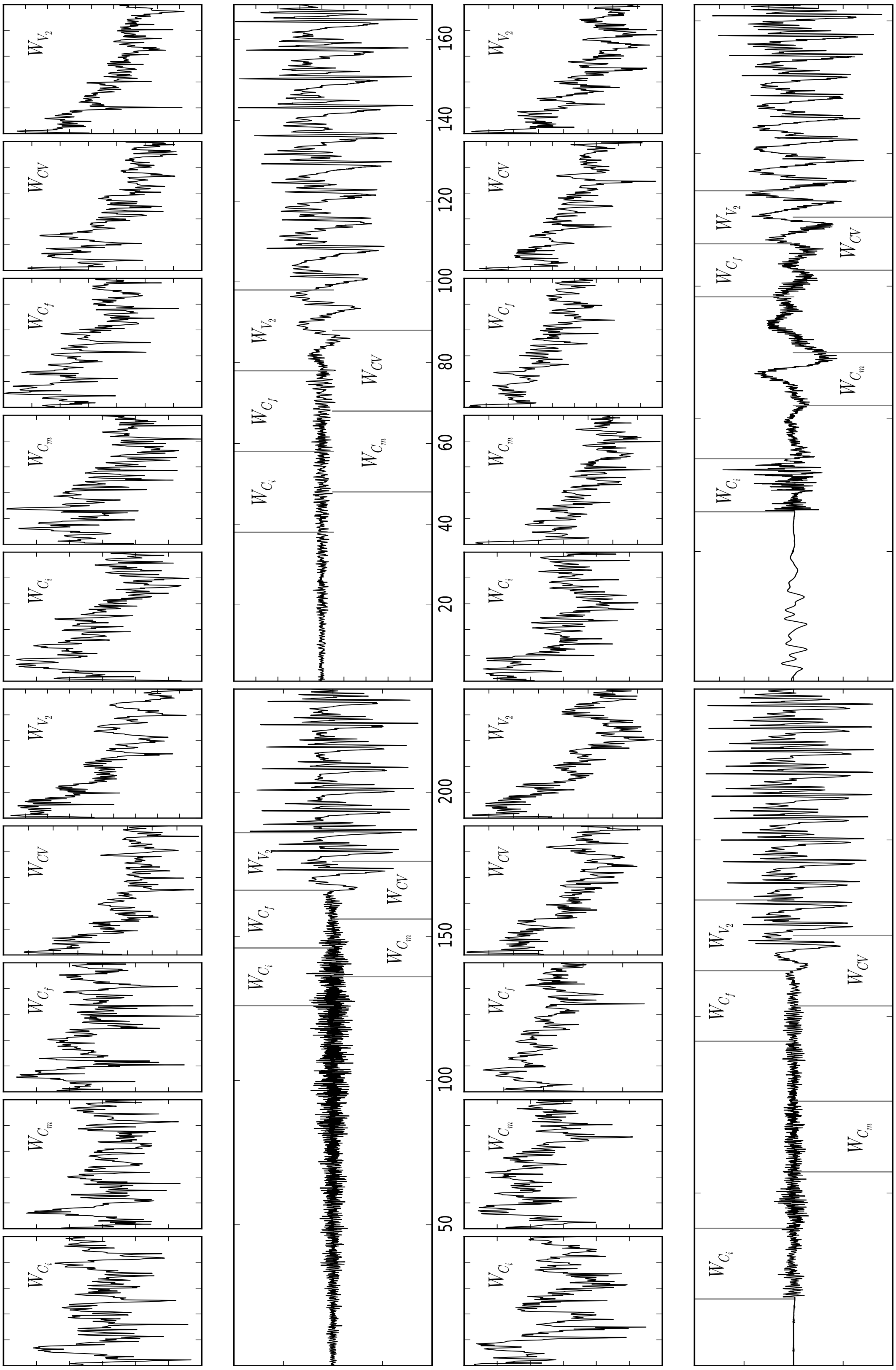


Figure 2.2: Speech samples and computed spectra from word-initial obstruents produced by speaker JM02. *Left-to-Right, Top-to-Bottom.* Extracted portions of target words *xat*, *ɦat*, k^h*at*, and $g^ɦ$*at*. Window locations are indicated in the waveform plots (x-axis in ms), with corresponding computed spectra (below 10 kHz) plotted above.

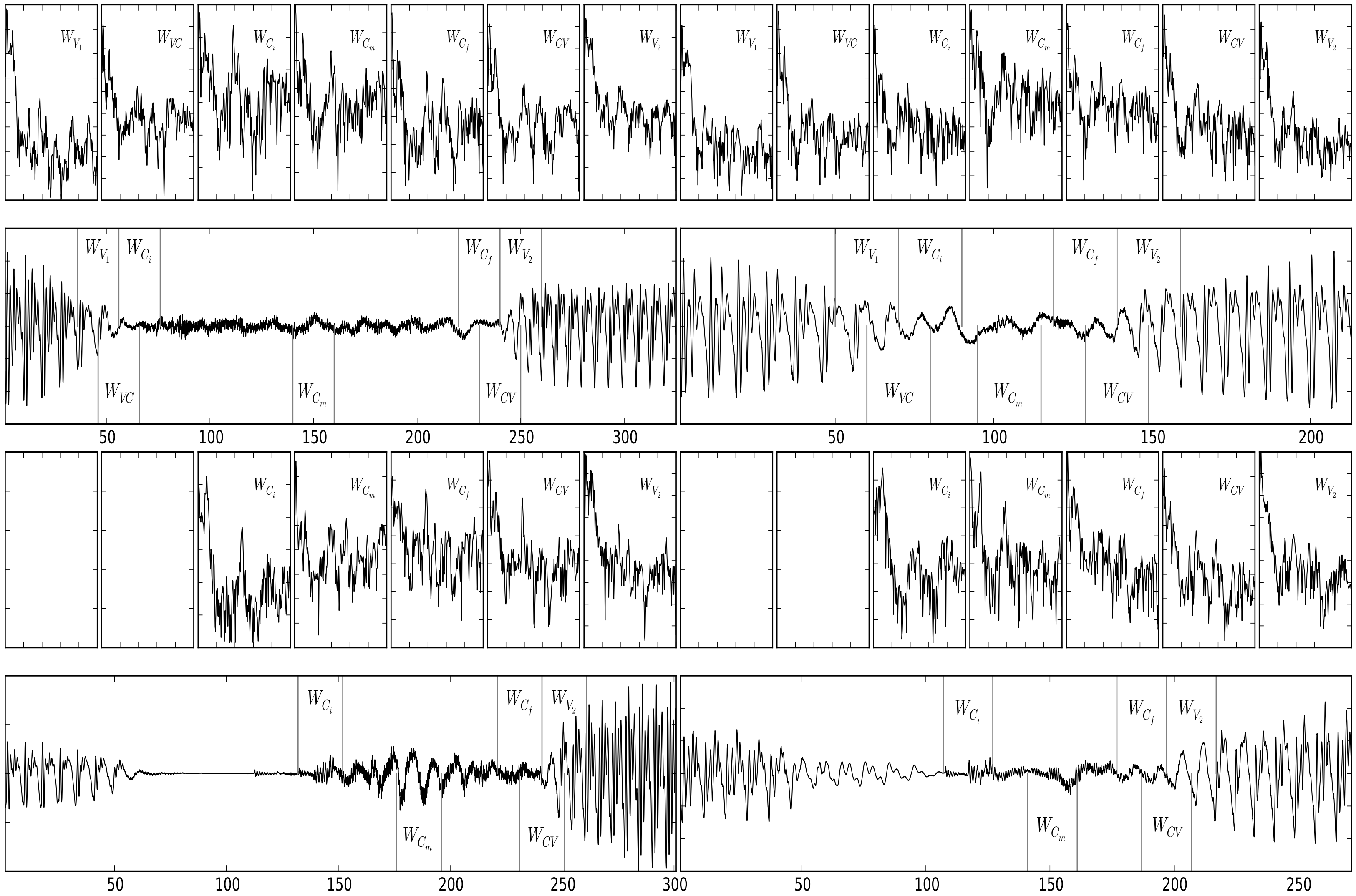


FIGURE 2.3: Speech samples and computed spectra from inter-vocalic obstruents produced by speaker JM02. *Left-to-Right, Top-to-Bottom.* Extracted portions of target words *ɔxɔm*, *bɔɦɔl*, *sɔk*^{h}*ɔn*, and *sɔg*^{ɦ}*ɔr*. Window locations are indicated in the waveform plots (x-axis in ms), with corresponding computed spectra (below 10 kHz) plotted above.

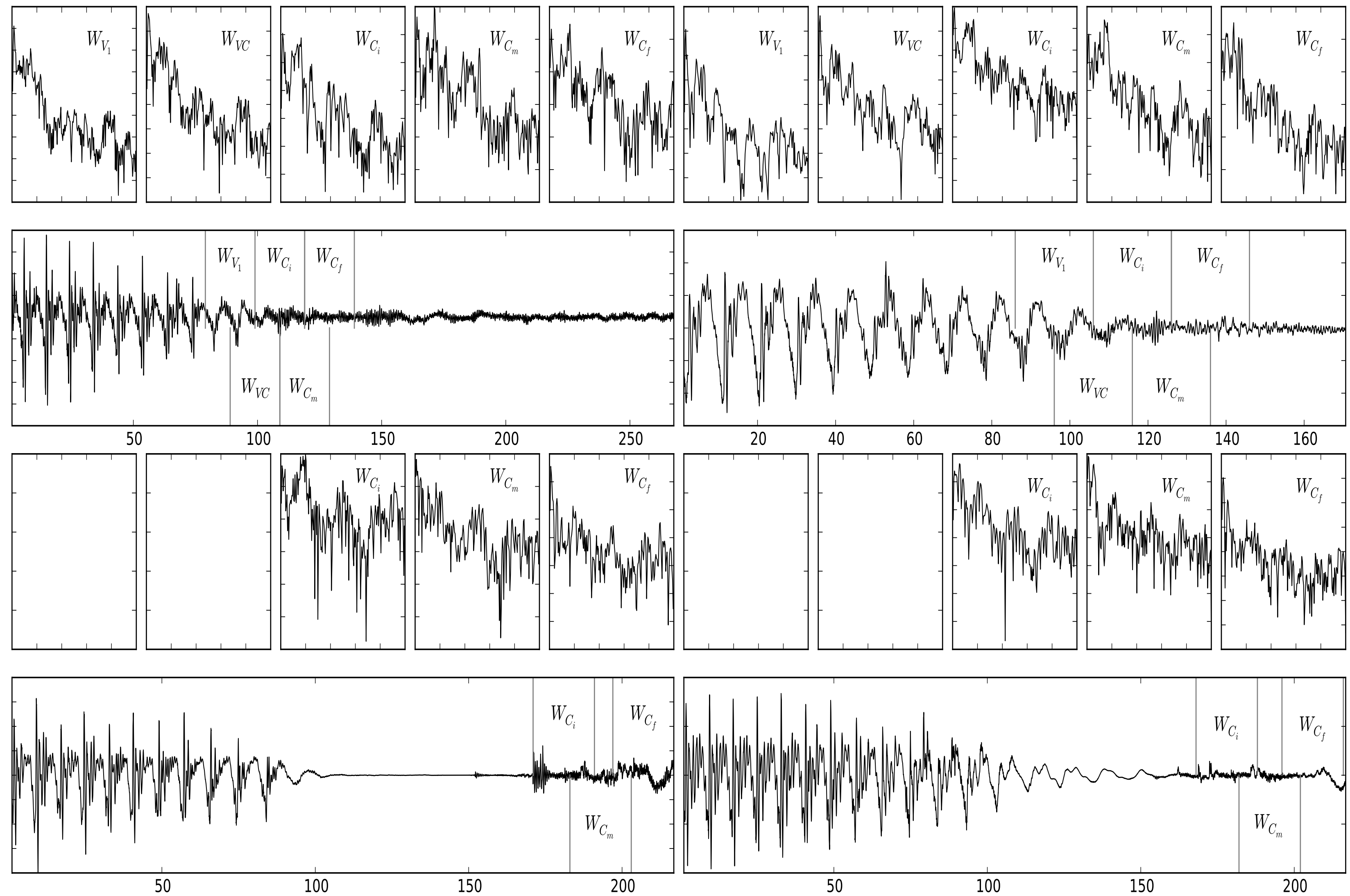


FIGURE 2.4: Speech samples and computed spectra from word-final obstruents produced by speaker JM02. *Left-to-Right, Top-to-Bottom.* Extracted portions of target words *rax*, *safi*, *lak*^{h}, and *bag*^{fi}. Window locations are indicated in the waveform plots (x-axis in ms), with corresponding computed spectra (below 10 kHz) plotted above.

		JF		JM		NM	
Consonant		/a/	/i/	/a/	/i/	/a/	/i/
CV	/x/	1934	4184	1432	2835	1704	4284
		(892)	(1617)	(444)	(634)	(767)	(595)
	/k^{h}/	1583	3685	1446	2919	1668	3710
		(221)	(1209)	(497)	(1004)	(702)	(896)
VC	/x/	1930	3335	1487	2568	1887	3006
		(1421)	(1091)	(201)	(770)	(997)	(1584)
	/k^{h}/	1600	2961	1777	2904	1442	3676
		(617)	(1934)	(1746)	(1522)	(234)	(1670)

Table 2.1: Spectral peak location of voiceless velars by vocalic environment

2013]. Normality assumptions were tested prior to each statistical test via a combination of Shapiro-Wilk test results and visual inspection of Q-Q plots [R Core Team, 2013]. All main tests were subsequently evaluated in multiple comparisons using Holm-Bonferroni correction of pairwise t-tests for the ANOVA model, and non-parametric Tukey HSD for the Kruskal-Wallis model [Fox and Weisberg, 2011, Konietschke, 2012].

2.2.1 Spectral Peak Location

Before we assess parameter significance for the later classification models presented in Chapter 3, we address the question of velar allophony with respect to the fricative and aspirated plosive in our target set. By looking at the main resonances of each consonant in two vocalic positions – preceding/following the low vowel /a/ and preceding/following the high front vowel /i/ – we get a sense of the relation between place of articulation of the two. This will not only serve to ground subsequent models in the basic production characteristics of the sounds being studied, but will also begin to address a debate raised in Sarma [2012] regarding the place of articulation of the dorsal fricative in Assamese as either velar or uvular.

Table 2.1 shows means and standard deviations of /x/ and /k^{h}/ in word-initial and word-final environments where *V* comprises the set /i, a/. The results for the Jorhat Male group are then plotted in Figure 2.5, where consonant fronting before the high-front vowel (as derived from the higher peak frequency, and therefore shorter front cavity producing the main resonance) is evident for both the fricative and the plosive.[8] This is not conclusive evidence that the dorsal fricative is fundamentally *velar* in its place of articulation, but it does suggest that any model which would suggest a uvular place of articulation for the voiceless dorsal fricative will have to account for this mirrored

[8]For the remainder of the chapter all plots are shown for Jorhat Male data only

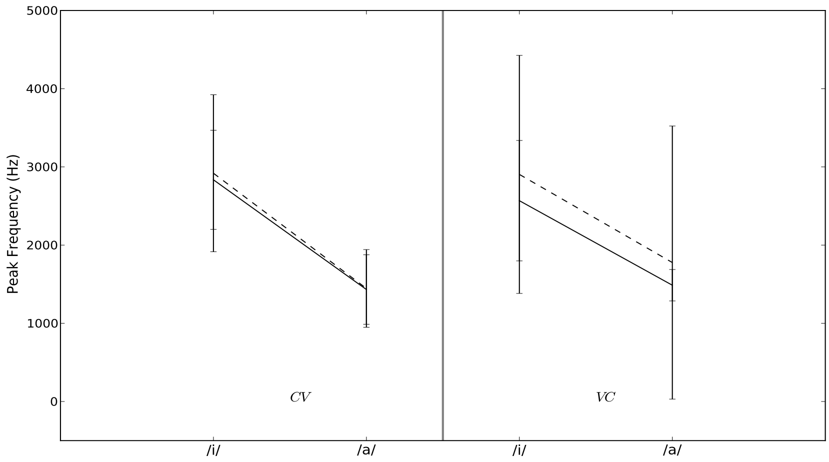


FIGURE 2.5: Spectral peak location of voiceless velars by vocalic environment. /x/ is plotted with the solid line, /k^h/ with the dashed line.

behavior of /x/ and /k^h/ in different vocalic environments. At present we find no compelling reason to consider this fricative anything other than categorically a velar, which is why we have referred to it as such throughout this work.

2.2.2 Spectral Moments

Mean

Results of one-way repeated measures ANOVAs on spectral mean in the Jorhat Female group revealed significant main effects of consonant type at C_i/C_m [$25.7 < F(3,130) < 31.0$, $p < 0.001$] and $C_f - V_2$ [$13.1 < F(3,130) < 14.8$, $p < 0.001$]. In the Jorhat Male group significant effects were found at C_f/C_m [$42.3 < F(3,124) < 70.2$, $p < 0.001$] and C_i [$F(3,124) = 15.3$, $p < 0.01$]. Among Nalbari males consonant category was also significant on μ_1 at C_m [$F(3,129) = 87.3$, $p < 0.001$] as well as at C_i/C_f [$11.1 < F(3,129) < 15.7$, $p < 0.01$]. Figure 2.6 displays these results for the Jorhat Male group, showing a clear separation between consonant categories over the three consonantal windows, then rapid convergence toward a single μ_1 value in the vowel.

Inter-vocalically, a similar pattern emerges between the three continuant obstruents: /x/, /ɦ/, and spirantized /k^h/. Results of Kruskal-Wallis tests on μ_1 in the Jorhat Female group revealed a significant main effect of consonant category over windows $VC - C_f$ [$28.7 < K(2) < 55.2$, $p < 0.001$], a result which was replicated in the Nalbari Male group [$16.8 < K(2) < 30.9$, $p < 0.001$]. For Jorhat males significant main effects of consonant category were found at $C_i - C_f$ [$23.4 < K(2) < 33.0$, $p < 0.001$], V_1 [$K(2) = 10.0$, $p < 0.01$], and at VC/CV [$7.4 < K(2) < 7.5$, $p < 0.05$]. Distributions of μ_1 by consonant category over window locations in inter-vocalic position are shown

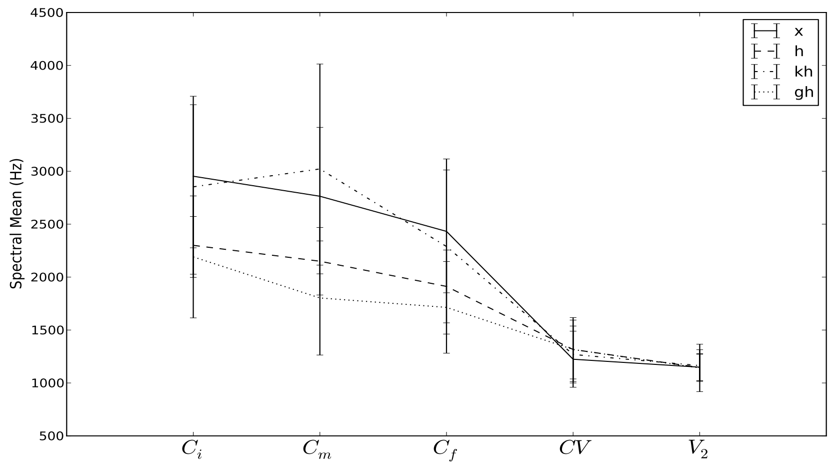


FIGURE 2.6: Spectral mean by window location: #C[a]

in Figure 2.7, again emphasizing the stability of the mid-consonant region for spectral mean distinctions.

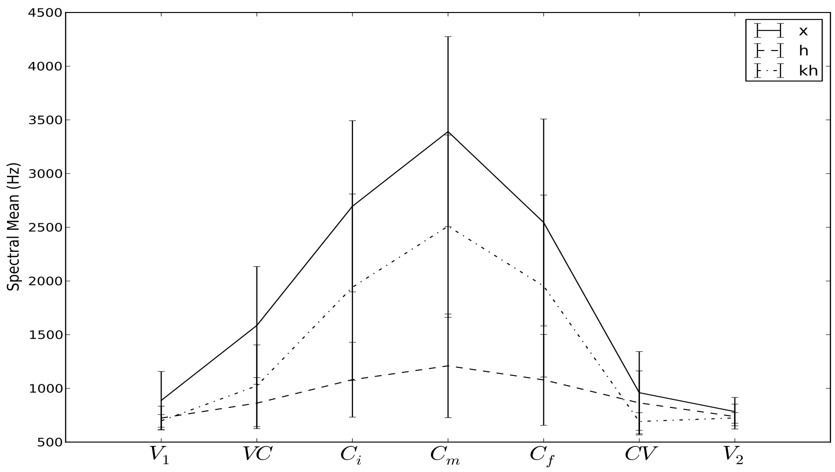


FIGURE 2.7: Spectral mean by window location: [ɔ]C[ɔ]

In word-final position, results of Kruskal-Wallis tests revealed a significant main effect of consonant on μ_1 over windows V_1, $C_i - C_f$ [$13.9 < K(2) < 51.5$, $p < 0.001$] in the Jorhat Female group. Among Jorhat males consonant category was again significant at $C_i - C_f$ [$33.1 < K(2) < 41.7$, $p < 0.001$] as well as at VC [$K(2) = 6.5$, $p < 0.05$]. In the Nalbari Male group results mirrored those of the Jorhat Male group, with consonant category significant at $C_i - C_f$ [$38.8 < K(2) < 44.1$, $p < 0.001$] and VC [$K(2) = 13.0$, $p < 0.01$]. Values of spectral mean over the five windows in word-final environment are shown in Figure 2.8, which more-or-less mirrors the picture from the word-initial set. Continuant obstruents are most highly separated in the middle portion of the consonant interval, and taper off toward a single μ_1 value in the preceding vowel.

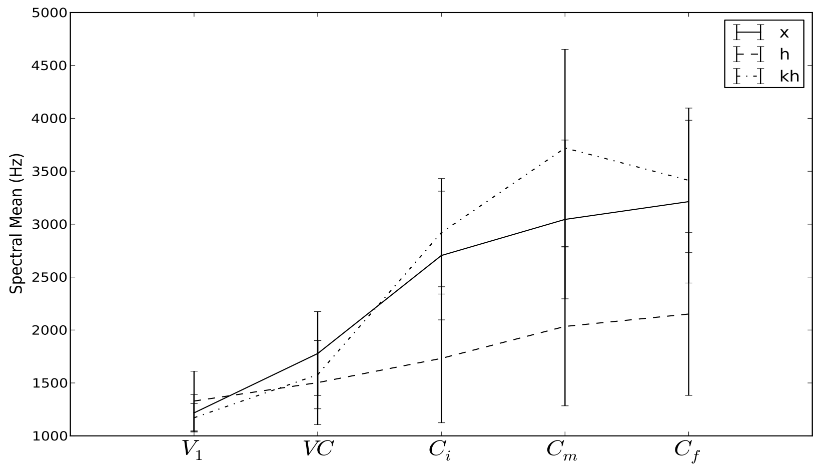


FIGURE 2.8: Spectral mean by window location: [a]C#

Aggregate results of significance testing on the μ_1 parameter are displayed in Table C.1. Significance size and results of multiple comparisons are presented for each subject group, segmental environment, and window position. Empty cells indicate window positions which were not applicable to a given segmental environment.

Dispersion

Beginning with the word-initial environment, results of one-way repeated measures ANOVAs on μ_2 revealed a significant main effect of consonant category at $C_i - C_m$, $CV - V_2$ [$30.0 < F(3,124) < 244$, $p < 0.001$] and C_f [$F(3,124) = 22.8$, $p < 0.01$]. In the Jorhat Male group, consonant category was significant on μ_2 at C_i, $C_f - V_2$ [$41.9 < F(3,124) < 266$, $p < 0.001$] and C_m [$F(3,124) = 12.6$, $p < 0.01$]. For male speakers from Nalbari, results were significant at $C_i - C_m$ [$94.6 < F(3,129) < 99.3$, $p < 0.001$], V_2 [$F(3,129) = 13.5$, $p < 0.01$], and CV [$F(3,129) = 5.4$, $p < 0.05$]. The distribution of spectral dispersion from Jorhat Male data can be seen in Figure 2.9. Here, unlike for μ_1, greater separation appears at the margins, though C_m remains a key window location.

Inter-vocalically, results of Kruskal-Wallis tests in the Jorhat Female group revealed a significant main effect of consonant category on μ_2 at $VC - C_f$, V_2 [$15.9 < K(2) < 53.0$, $p < 0.001$] and CV [$K(2) = 7.0$, $p < 0.05$]. For Jorhat males, significant results were obtained at window locations $VC - C_f$, V_2 [$17.2 < K(2) < 31.6$, $p < 0.001$], V_1 [$K(2) = 12.8$, $p < 0.01$], and CV [$K(2) = 6.8$, $p < 0.05$]. In the Nalbari Male group, consonant category was significant at $VC - C_f$ [$17.7 < K(2) < 32.8$, $p < 0.001$]. Inter-vocalic data from the Jorhat Male group are displayed in Figure 2.10, where again it

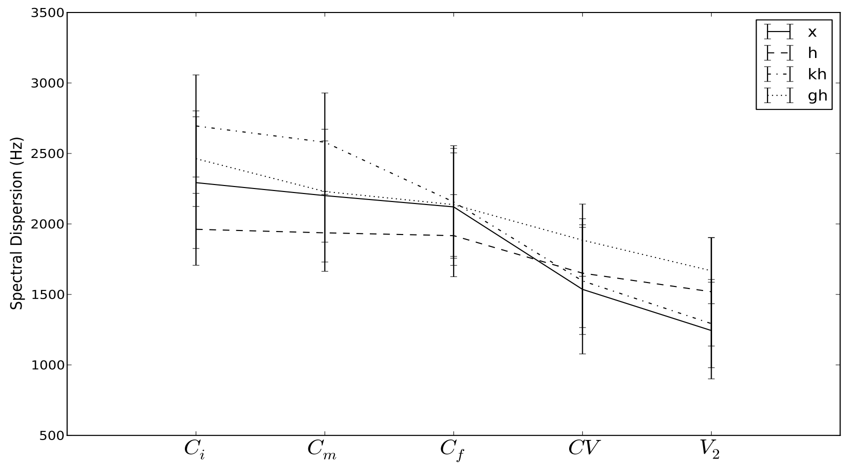


FIGURE 2.9: Spectral dispersion by window location: #C[a]

can be seen that the middle portion of the consonant interval in VCV sequences is more stable for μ_2 contrast than at the boundaries.

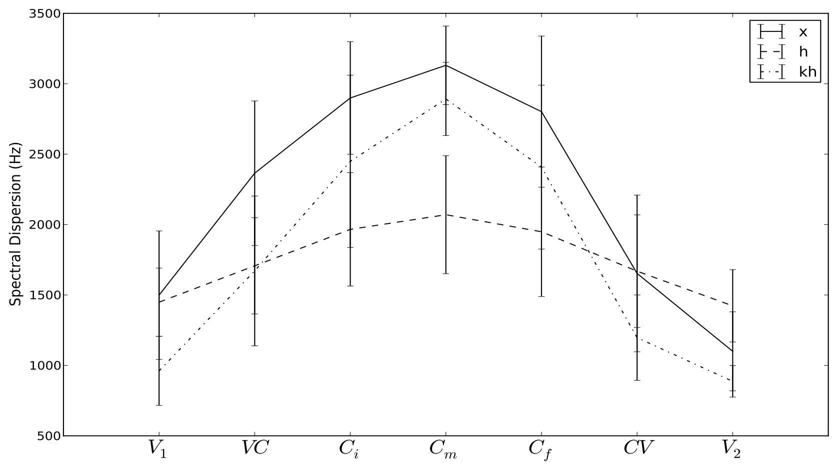


FIGURE 2.10: Spectral dispersion by window location: [ɔ]C[ɔ]

In word-final position, results of Kruskal-Wallis tests revealed a significant main effect of consonant category on μ_2 at V_1, $C_i - C_m$ [$19.7 < K(2) < 36.0$, $p < 0.001$] in the Jorhat Female group. Jorhat males exhibited similar results, with consonant category significant at $C_i - C_f$ [$29.3 < K(2) < 34.9$, $p < 0.001$] and V_1 [$K(2) = 13.7$, $p < 0.01$]. In the Nalbari Male group, consonant category was also significant over windows C_i–C_f [$38.9 < K(2) < 44.1$, $p < 0.001$] as well as at VC [$K(2) = 13.0$, $p < 0.01$]. This distribution in word-final context is shown in Figure 2.11, again showing greater convergence of values at the vowel-consonant boundary. Parameter significance for spectral dispersion is summarized in Table C.2.

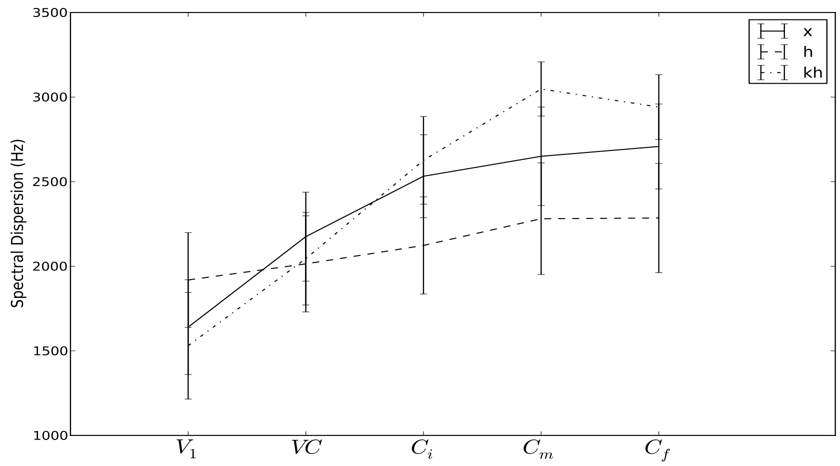


FIGURE 2.11: Spectral dispersion by window location: [a]C#

Skewness

Results of one-way repeated measures ANOVAs on μ_3 in word-initial position revealed a significant main effect of consonant category at CV/V_2 [$26.1 < F(3,130) < 46.7$, $p < 0.001$] and $C_i - C_m$ [$16.3 < F(3,130) < 17.1$, $p < 0.01$] among Jorhat females. In the Jorhat Male group, consonant category was significant at C_i/C_f [$34.3 < F(3,124) < 40.5$, $p < 0.001$], C_m/V_2 [$14.6 < F(3,124) < 16.9$, $p < 0.01$], and CV [$F(3,124) = 9.0$, $p < 0.05$]. For male speakers from Nalbari, consonant category was significant at window locations $C_i - C_m$ [$26.5 < F(3,129) < 180$, $p < 0.001$], C_f [$F(3,129) = 17.7$, $p < 0.01$], and CV [$F(3,129) = 7.2$, $p < 0.05$]. The distribution of skewness values over the /x/–/ɦ/–/$\mathrm{k^h}$/–/$\mathrm{g}^{ɦ}$/ set is shown for Jorhat males in Figure 2.12.

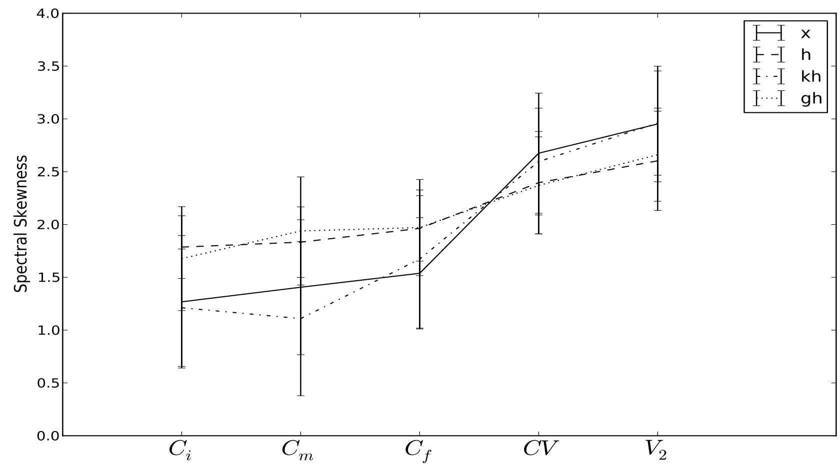


FIGURE 2.12: Spectral skewness by window location: #C[a]

In inter-vocalic position results of Kruskal-Wallis tests in the Jorhat Female group revealed a significant main effect of consonant category on μ_3 over window locations $VC{-}CV$ [$15.8 < K(2) < 51.8, p < 0.001$] and V_2 [$K(2) = 13.5, p < 0.01$]. Among Jorhat males, consonant category was significant over the range $V_1 - C_f$ [$14.9 < K(2) < 33.3$, $p < 0.001$] and at V_2 [$K(2) = 11.2$, $p < 0.01$]. In the Nalbari Male group, consonant category was significant over $VC - C_f$ [$15.1 < K(2) < 28.4$, $p < 0.001$]. The distribution of skewness values in inter-vocalic context for the Jorhat Male group are shown in Figure 2.13. As expected, spectra become more skewed as they approach the boundaries, and are least skewed in the middle portion of the consonant. Notably, /ɦ/ resists this trend to a great degree.

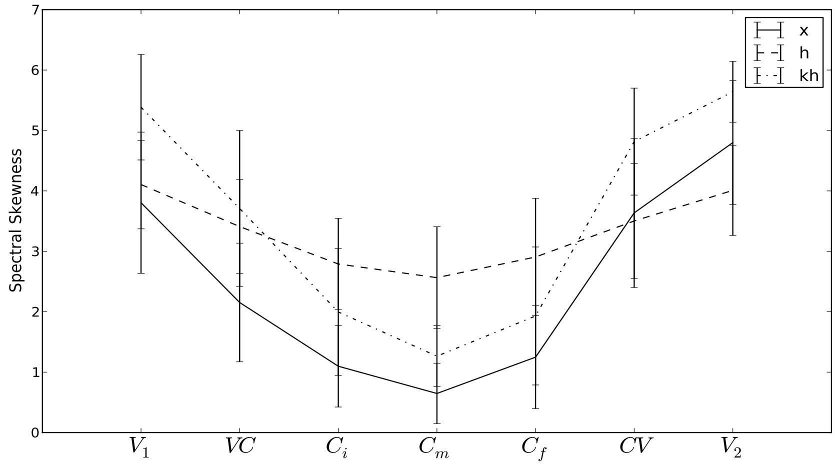


FIGURE 2.13: Spectral skewness by window location: [ɔ]C[ɔ]

Word-finally, results of Kruskal-Wallis tests on μ_3 revealed a significant main effect of consonant category at V_1, $C_m - C_f$ [$18.2 < K(2) < 32.2$, $p < 0.001$] and C_i [$K(2) = 11.0$, $p < 0.01$] in the Jorhat Female group. For Jorhat males, consonant category was significant over $C_i - C_f$ [$41.4 < K(2) < 44.1$, $p < 0.001$] and at VC [$K(2) = 10.3$, $p < 0.01$]. This pattern was repeated among Nalbari males, where consonant category was significant at window locations $C_i - C_f$ [$31.5 < K(2) < 39.8$, $p < 0.001$] and VC [$K(2) = 11.3$, $p < 0.01$]. Figure 2.14 exhibits these trends, where skewness shows greater separation between the two velars and the glottal fricative into the consonant.

Kurtosis

Results of one-way repeated measures ANOVAs on μ_4 word-initially in the Jorhat Female group revealed a significant main effect of consonant category at C_i, $CV - V_2$ [$27.1 < F(3,130) < 40.1$, $p < 0.001$] and C_m [$F(3,130) = 23.6$, $p < 0.01$]. For the Jorhat Male group, similar results were obtained, with consonant category significant at C_i, $CV - V_2$ [$41.1 < F(3,124) < 46.6$, $p < 0.001$], C_f [$F(3,124) = 20.1$, $p < 0.01$], and C_m [$F(3,124) = 6.2$, $p < 0.05$]. Among Nalbari males, consonant category was significant on μ_4 word-initially at $C_i - C_m$ [$69.9 < F(3,129) < 360$, $p < 0.001$] and $C_f - CV$ [$5.0 < F(3,129) < 6.0$, $p < 0.05$]. Spectral kurtosis values for Jorhat males over window

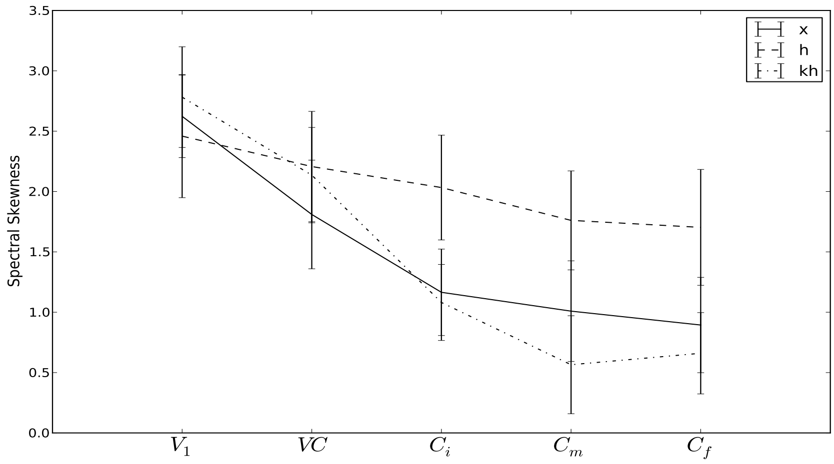


FIGURE 2.14: Spectral skewness by window location: [a]C#

locations in word-initial position are shown in Figure 2.15. The pattern of distribution across windows and separation between consonants is less clear in this picture, though the voiceless velars do notably separate from the glottal fricative and breathy velar at the vowel, likely a reflection of the different phonatory properties at vowel onset.

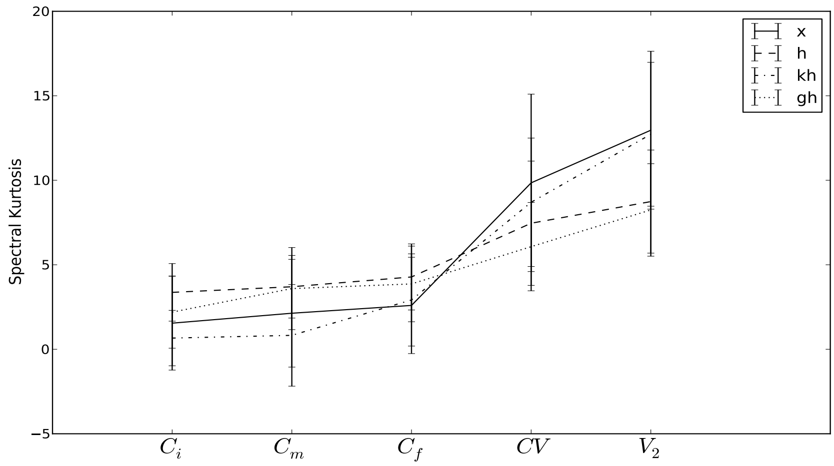


FIGURE 2.15: Spectral kurtosis by window location: #C[a]

Inter-vocalically, results of Kruskal-Wallis tests of consonant category on μ_4 were significant for the Jorhat Female group at $VC - C_f$, V_2 [$14.2 < K(2) < 51.7$, $p < 0.001$] and CV [$K(2) = 13.3$, $p < 0.01$]. Among Jorhat males, consonant category was significant over all windows; namely, $V_1 - C_f$ [$14.7 < K(2) < 33.0$, $p < 0.001$], V_2 [$K(2) = 13.0$, $p < 0.01$], and CV [$K(2) = 6.0$, $p < 0.05$]. In the Nalbari Male group significant results were obtained over windows $VC - C_f$ [$15.7 < K(2) < 29.1$, $p < 0.001$]. Jorhat Male data for spectral kurtosis is displayed in Figure 2.16, where notable separation between the three categories is evident at the margins.

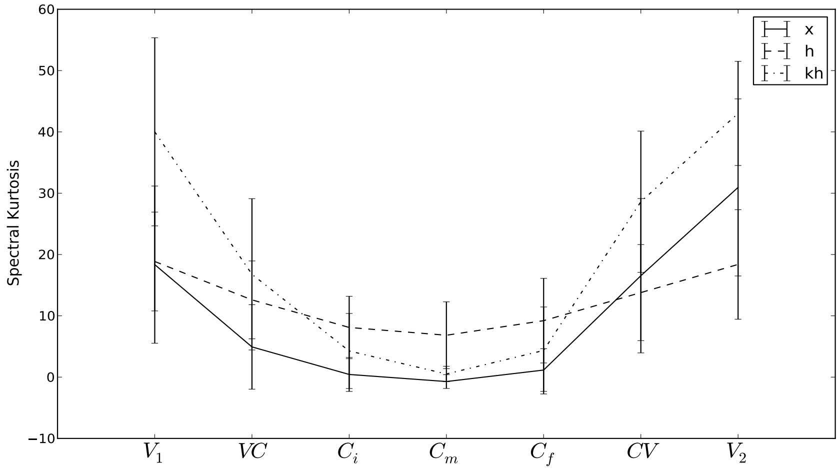


FIGURE 2.16: Spectral kurtosis by window location: [ɔ]C[ɔ]

Finally we present results from Kruskal-Wallis tests of consonant category on μ_4 word-finally. In the Jorhat Female group, a significant main effect of consonant category was revealed at V_1, $C_m - C_f$ $[16.4 < K(2) < 43.6$, $p < 0.001]$ and C_i $[K(2) = 11.0$, $p < 0.01]$. For Jorhat males, category significance was obtained at $C_i - C_f$ $[36.0 < K(2) < 43.1$, $p < 0.001]$ and VC $[K(2) = 11.1$, $p < 0.01]$. In the Nalbari Male group, consonant category was significant over $C_i - C_f$ $[39.7 < K(2) < 42.3$, $p < 0.001]$ and $V_1 - VC$ $[9.6 < K(2) < 10.5$, $p < 0.01]$. From the display of kurtosis values in the Jorhat Male set we see that /x/ and /k^h/ are well-aligned throughout the transition, and distinct from the pattern observed in the glottal.

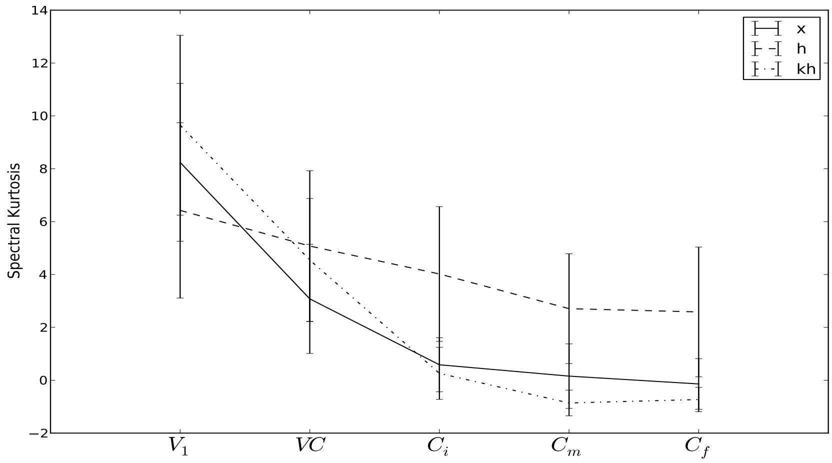


FIGURE 2.17: Spectral kurtosis by window location: [a]C#

Table C.4 summarizes the results of significance testing on the spectral kurtosis parameter across subject, window, and positional distinctions.

2.2.3 Normalized Amplitude

Results of one-way repeated measures ANOVAs on normalized amplitude (A_n) word-initially revealed a significant main effect of consonant category in all three subject groups: Jorhat Female [$F(3,124) = 149$, $p < 0.001$], Jorhat Male [$F(3,130) = 31.3$, $p < 0.001$], and Nalbari Male [$F(3,129) = 62.9$, $p < 0.001$]. Inter-vocalically, Kruskal-Wallis tests produced a similarly robust effect across subject groups: Jorhat Female [$K(2) = 27.3$, $p < 0.001$], Jorhat Male [$K(2) = 19.3$, $p < 0.001$], and Nalbari Male [$K(2) = 19.3$, $p < 0.001$]. Word-finally, however, only the Nalbari Male group showed a significant main effect of consonant category on A_n [$K(2) = 14.4$, $p < 0.001$]. Normalized amplitude in the three segmental conditions is shown for the Jorhat Male group in Figure 2.18. Notably, normalized amplitude of the glottal fricative (suggesting closer approximation of the amplitude of the vowel) is significantly reduced inter-vocalically, producing separation from the velar that was not present in word-initial position.

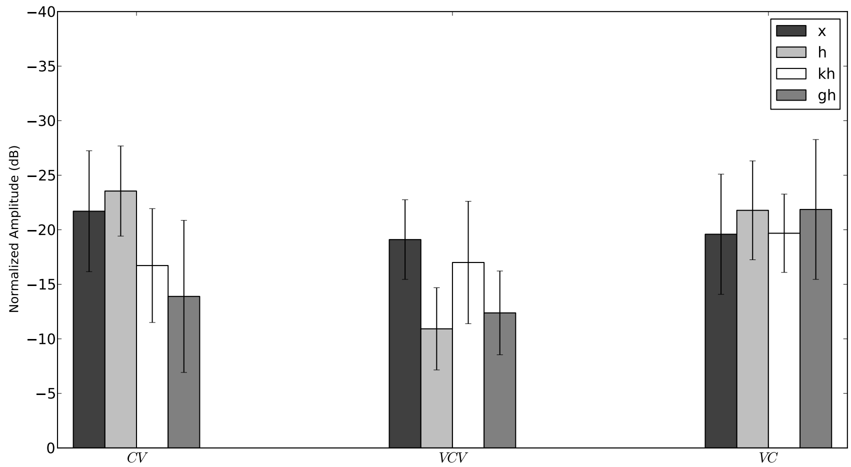


Figure 2.18: Normalized amplitude by position

2.2.4 Dynamic Amplitude

Across subjects, dynamic amplitude (A_d) was less consistently distinguished according to consonant category than was shown with normalized amplitude. In the Nalbari Male group, no significant effects of consonant category on A_d were found for any segmental environment. Among Jorhat females similar results were found, though in the word-initial condition consonant category was marginally significant [$F(3,130) = 5.1$, $p < 0.05$]. In the Jorhat Male group, significance was obtained in all three positions: word-initially [$F(3,124) = 19.6$, $p < 0.01$], inter-vocalically [$K(2) = 11.2$, $p < 0.05$], and word-finally [$K(2) = 11.0$, $p < 0.01$]. Figure 2.19 displays these results.

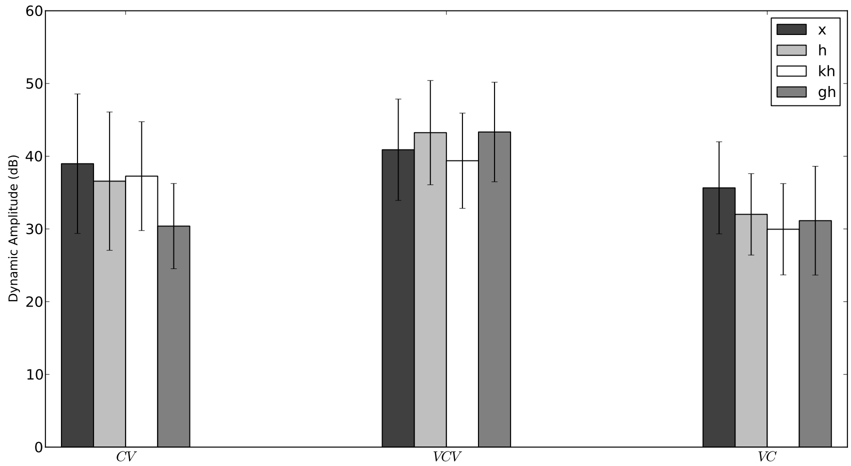


FIGURE 2.19: Dynamic amplitude by position

2.2.5 Relative Amplitude

Results of one-way repeated measures ANOVAs on relative amplitudes in word-initial position for the Jorhat Female group revealed a significant main effect of consonant category on $A_{r_{16}}$ and $A_{r_{17}}$ [$5.3 < F(3, 130) < 8.8$, $p < 0.05$]. For Jorhat males similar results were obtained, only for $A_{r_{16}}$ and $A_{r_{18}}$ [$5.6 < F(3, 124) < 8.8$, $p < 0.05$]. In the Nalbari Male group, consonant category was significant on $A_{r_{18}}$ [$F(3, 129) = 18.3$, $p < 0.01$] and $A_{r_{17}}$ [$F(3, 129) = 8.9$, $p < 0.05$]. Results for Jorhat males are displayed in Figure 2.20. As the error bars indicate, these measures were not very reliable in word-initial position.

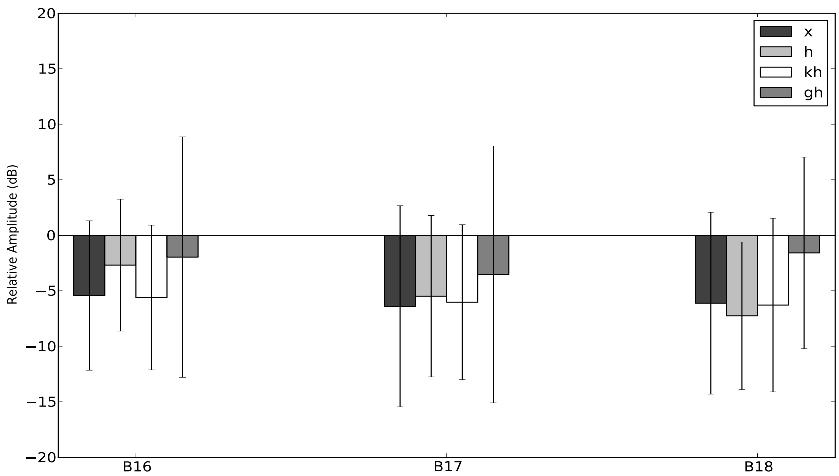


FIGURE 2.20: Relative amplitudes at Bark 16, 17, and 18 in word-initial position

Inter-vocalically, results of Kruskal-Wallis tests in the Jorhat Female group revealed a significant main effect of consonant category on $A_{r_{17}}$ between V_1 and C [$K(2) = 7.4$, $p < 0.05$]; between C and V_2, significant effects were obtained for $A_{r_{16}}$ and $A_{r_{17}}$ [$15.4 < K(2) < 17.1$, $p < 0.001$], and $A_{r_{18}}$ [$K(2) = 11.4$, $p < 0.01$]. In the Jorhat

Male group, relative amplitudes over all three bark intervals were significant between V_1 and C [$11.1 < K(2) < 13.4$, $p < 0.001$], and between C and V_2 consonant category was significant for $A_{r_{18}}$ [$K(2) = 11.2$, $p < 0.01$]. Relative amplitudes in inter-vocalic position are presented for Jorhat Male data in Figure 2.21.

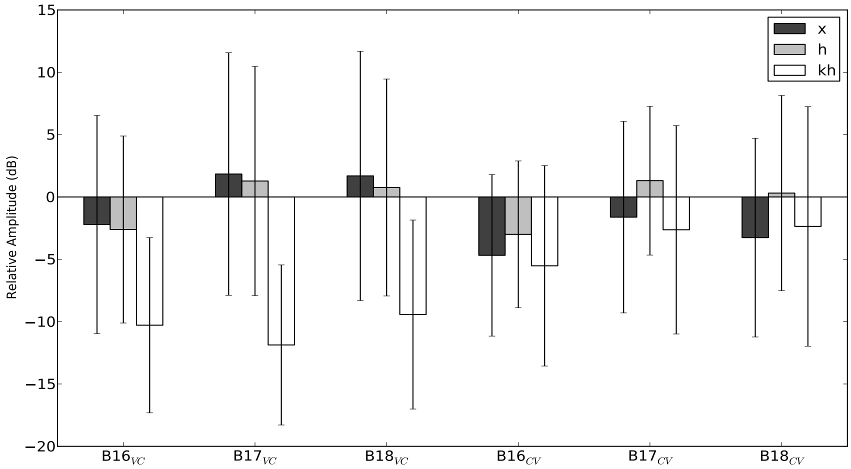


Figure 2.21: Relative amplitudes at Bark 16, 17, and 18 in inter-vocalic position

Finally, in word-final position we find significant main effects of consonant category on relative amplitude only over the Bark 16 range, with Kruskal-Wallis tests significant in the Jorhat Female [$K(2) = 14.6$, $p < 0.001$] and Jorhat Male [$K(2) = 6.1$, $p < 0.05$] groups. Relative amplitude values for Jorhat males are presented in Figure 2.22.

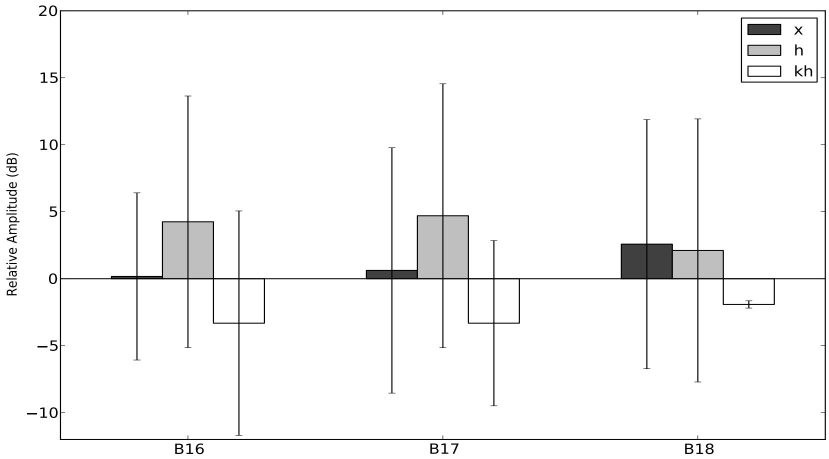


Figure 2.22: Relative amplitudes at Bark 16, 17, and 18 in word-final position

Results of significance testing on all amplitudinal parameters are presented below in Table C.5.

2.3 Acoustic Feature Summary

Acoustic features used as input parameters in a classification task were chosen as those from the initial set which reached statistical significance for main effects of *Consonant* within word position (CV, VCV, and VC) and *Speaker* subgroups, and which showed at least one contrast distinguished in post-hoc multiple comparisons. To maintain some congruity between the spectral shape parameters, and to simplify the task and its presentation, spectral moments are taken as an ensemble and not selected individually according to the above criteria. Therefore, if one of the four moments meets these requirements the whole set is taken and input into the classifier.

In the Jorhat Female group, the acoustic feature vector for word-initial position is defined as

$$X_{JF/cv} = \begin{bmatrix} \mu_{1/W_n} & \mu_{2/W_n} & \mu_{3/W_n} & \mu_{4/W_n} & A_n & A_d & A_{r_{16}} & A_{r_{17}} \end{bmatrix}^T$$

where W_n is the set of window locations from which the four spectral moments are calculated, defined for the word-initial classification task in the Jorhat Female group as

$$W_n = \left\{ W_{c_i}, W_{c_m}, W_{c_f}, W_{cv}, W_{v_2} \right\}$$

Thus all five windows are incorporated for spectral information, and all further parameters but the relative amplitude between C and V_2 in the Bark 18 frequency range ($A_{r_{18}}$) are used for amplitudinal information. Inter-vocalically, the feature vector becomes

$$X_{JF/vcv} = \begin{bmatrix} \mu_{1/W_n} & \mu_{2/W_n} & \mu_{3/W_n} & \mu_{4/W_n} & A_n & A^{cv}_{r_{16}} & A^{cv}_{r_{17}} & A^{cv}_{r_{18}} \end{bmatrix}^T$$

$$W_n = \left\{ W_{cv}, W_{c_i}, W_{c_m}, W_{c_f}, W_{cv}, W_{v_2} \right\}$$

Here spectral shape parameters from window V_1 are excluded, as well as dynamic amplitude (A_d) and relative amplitudes between V_1 and C at Bark 16, 17, and 18 ($A_{r_{16}}$, $A_{r_{17}}$, $A_{r_{18}}$). This result underscores the acoustic instability of the V_1 region in our particular set of measurements. Finally, for the Jorhat Female group, the feature vector for word-final position is defined below.

$$X_{JF/vc} = \begin{bmatrix} \mu_{1/W_n} & \mu_{2/W_n} & \mu_{3/W_n} & \mu_{4/W_n} & A_{r_{16}} \end{bmatrix}^T$$

$$W_n = \left\{ W_{v_1}, W_{c_i}, W_{c_m}, W_{c_f} \right\}$$

Unlike the inter-vocalic set, V_1 spectral information is incorporated in the model, while moments measured at the VC boundary are excluded. Among the amplitudinal parameters only the relative amplitude between C and V_2 in the Bark 16 range is retained.

For the Jorhat Male group, similar results obtain for the word-initial set, with the only exception being inclusion of $A_{r_{18}}$ and the exclusion of its Bark 16 and 17 counterparts.

$$X_{JM/cv} = \Big[\mu_{1/W_n} \quad \mu_{2/W_n} \quad \mu_{3/W_n} \quad \mu_{4/W_n} \quad A_n \quad A_d \quad A_{r_{18}}\Big]^T$$

$$W_n = \big\{W_{c_i}, W_{c_m}, W_{c_f}, W_{cv}, W_{v_2}\big\}$$

In inter-vocalic position, spectral moments are again incorporated across all possible window locations, though relative amplitudes are asymmetrically utilized, with VC appearing the more stable relation than CV. As with half of the feature vectors established so far, dynamic amplitude (A_d) does not prove viable for the VCV model.

$$X_{JM/vcv} = \Big[\mu_{1/W_n} \quad \mu_{2/W_n} \quad \mu_{3/W_n} \quad \mu_{4/W_n} \quad A_n \quad A^{vc}_{r_{16}} \quad A^{vc}_{r_{17}} \quad A^{vc}_{r_{18}} \quad A^{cv}_{r_{18}}\Big]^T$$

$$W_n = \big\{W_{v_1}, W_{vc}, W_{c_i}, W_{c_m}, W_{c_f}, W_{cv}, W_{v_2}\big\}$$

Word-finally, however, the opposite obtains for amplitudinal measures. Dynamic amplitude is incorporated at the exclusion of the rest. As for the spectral parameters, we again find the full set of window locations utilized.

$$X_{JM/vc} = \Big[\mu_{1/W_n} \quad \mu_{2/W_n} \quad \mu_{3/W_n} \quad \mu_{4/W_n} \quad A_d\Big]^T$$

$$W_n = \big\{W_{v_1}, W_{vc}, W_{c_i}, W_{c_m}, W_{c_f}\big\}$$

Finally, we examine the feature vectors established for the Nalbari Male group. In the word-initial set, only spectral moments from the three consonantal windows (i.e. C_i, C_m, and C_f) were included. In addition, normalized amplitude and relative amplitude between C and V_2 at Bark 17 and 18 were found viable in initial significance testing.

$$X_{NM/cv} = \Big[\mu_{1/W_n} \quad \mu_{2/W_n} \quad \mu_{3/W_n} \quad \mu_{4/W_n} \quad A_n \quad A_{r_{17}} \quad A_{r_{18}}\Big]^T$$

$$W_n = \big\{W_{c_i}, W_{c_m}, W_{c_f}\big\}$$

Inter-vocalically the picture is similar. Consonantal windows (with the inclusion of VC) are utilized for spectral parameters, and normalized amplitude is again incorporated as a robust discriminator in the model. It is notable that only the relative amplitude across the VC boundary in the Bark 18 range passed initial significance testing.

$$X_{NM/vcv} = \Big[\mu_{1/W_n} \quad \mu_{2/W_n} \quad \mu_{3/W_n} \quad \mu_{4/W_n} \quad A_n \quad A^{vc}_{r_{18}}\Big]^T$$

$$W_n = \big\{W_{vc}, W_{c_i}, W_{c_m}, W_{c_f}\big\}$$

In word-final position the picture is almost identical to the inter-vocalic set, with the one exception being that no relative amplitudes were incorporated in this model.

$$X_{NM/vc} = \begin{bmatrix} \mu_{1/W_n} & \mu_{2/W_n} & \mu_{3/W_n} & \mu_{4/W_n} & A_n \end{bmatrix}^T$$

$$W_n = \{W_{vc}, W_{c_i}, W_{c_m}, W_{c_f}\}$$

With the baseline acoustic feature vectors now defined, we turn now to the problem of *classification*. The next chapter will examine the process of phonemic contrast discrimination from the set of cues established above. This procedure will allow both an assessment of the interaction among the various acoustic parameters examined independently thus far, and a broad assessment of the system of contrast in the three subject pools and segmental environments elicited.

Chapter 3

Classification

Having explored a broad set of spectral and amplitudinal properties of posterior obstruents in Assamese, and established their relative discriminability as independent measures, the goal of this chapter is then to examine the behavior of the set of parameters as a whole. Namely, we take the *Subject* $\times$ *Position* feature vectors as an ensemble of cues in the task of identifying a given consonant from the acoustic signal. To this end we implement a pattern classifier as a computational approximation of the human linguistic perception process, using the classification results in the determination of both the relative significance of extracted acoustic parameters and the degree of category separability between target phonemes.

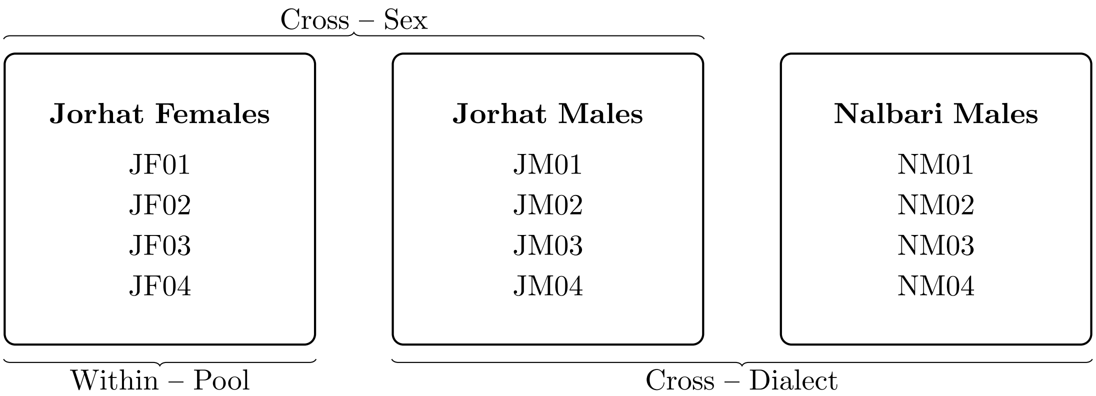


FIGURE 3.1: Classification scheme by subject group

As with the parameter evaluation procedure in the previous chapter, the classification task is broken down into separate tasks according to word position (i.e. CV, VCV, and VC). On a wider scale the classification procedure is organized according to three designs: the *within-pool* design, where phonemic discrimination is examined separately for each subject pool (i.e. Jorhat Females, Jorhat Males, and Nalbari Males); the *cross-sex* design, where a single classifier is tested across male and female speakers of the same dialect; and the *cross-dialectal* design, where a single classifier is tested across male speakers of different dialects. The latter two designs are meant to test the degree of compatibility between different linguistic subsystems in Assamese, helping to evaluate whether or not we are justified in considering phonemic contrast among the /x, ɦ, $\mathrm{k^h}$, $\mathrm{g^ɦ}$/ set as characteristically different across sex and dialect distinctions. Furthermore, the *cross-sex* design in particular allows us to test a major result from the Forrest et al. [1988] study demonstrating the distribution of spectral moments in word-initial fricatives in American English is invariant between male and female speakers.

3.1 Methods

The classification procedure implemented below is organized in two stages: (1) a classifier with the aforementioned input parameters is applied to the data, with category identification accuracy and misclassification proportions output for the given consonant set; (2) pairwise logistic regressions mirroring the classifier design are performed to yield the relative significance of individual parameters in the classification task.

3.1.1 Discriminant Analysis

Given a set of continuous predictor variables and a categorical response variable, the discriminant function, as defined by R. A. Fisher in 1936, is the linear combination of predictors whose coefficients minimize internal variance within each response group while maximizing the difference between the means of the groups [Lachenbruch, 1975]. Determining which group an individual observation is assigned to is typically done via the Bayesian decision rule:

$$P_i p_i(X) \underset{\omega_i}{\overset{\omega_j}{\lessgtr}} P_j p_j(X) \tag{3.1}$$

where X is a vector of predictor variables for a given observation – in our case the acoustic feature vectors defined in Chapter 2 – P_i is the prior probability of class ω_i, and $p_i(X)$ is the conditional density function of class ω_i; i.e. the probability of X given ω_i [Fukunaga, 1990, p. 52]. The prior probabilities P_i are typically assumed to be equal due to experimental control over sample sizes, but there is reason to question the assumption of equiprobable categories in linguistic studies, particularly when lexical frequency is not controlled in the design of word lists for elicitation. This issue will be revisited in Chapter 4, but for the present study, with no thorough way of determining category priors, we adopt the simpler model. This choice allows considerable room for experimental bias, but nevertheless will serve as a baseline for initial classification of velar and glottal obstruents in Assamese.

Fisher's discriminant function does not assume any prior distribution, though multivariate normality is expected for robust results. To this end we applied Henze-Zirkler's test for multivariate normality to each feature vector prior to classification. Violations of this assumption are noted when present, though discriminant analysis was still performed in such cases [Korkmaz and Goksuluk, 2014]. A more critical assumption in discriminant analysis is that of homogeneity in variance-covariance matrices (Σ_n), though it is not applicable for all discriminant techniques. Because each phonemic category is expected to be differentially affected by segmental and positional factors, we do not maintain this assumption. For this reason, quadratic classifiers were used in place of the more typically employed linear classifier. The two are remarkably similar in application, however, with linear discriminant analysis (LDA) a specialized case of quadratic discriminant analysis (QDA) with equal variance-covariance matrices. When $\Sigma_i \neq \Sigma_j$, QDA is a more robust method [Fukunaga, 1990]. Quadratic classifiers were employed in all tasks initially, though in some cases LDA was required when a category sample size was too small for the QDA algorithm to operate over [Venables and Ripley, 2002]. The switch from quadratic to linear classification is noted in each instance where the backoff was required (see coordinate axes of cluster plot), and includes the word-final environment for male and female Jorhat speakers, and the inter-vocalic environment for Nalbari and Jorhat males. As variance-covariance matrices are still assumed to be inherently unequal, classification results from the linear design should be interpreted with caution.

Each classification design (see Figure 3.1) contained both training and prediction components, though the methodology varied between the internal (*within-pool*) and crossed designs. For within-pool classification, leave-one-out cross-validation (also known as the 'jackknife' method) was used to draw predictions from subsets of the training set itself. In the crossed designs (i.e. *cross-sex* and *cross-dialect*) the subject group with the smaller feature vector was used as a training set, and the other as the prediction set. This strategy had the result that in both designs the Jorhat Male group became the prediction set, with the Jorhat Female and Nalbari Male groups serving as the training set. Discriminant analysis was performed on the full obstruent set for the word-initial condition, and on the continuant obstruent set (/x/, /ɦ/, and spirantized /k^h/) for inter-vocalic and word-final positions.

3.1.2 Logistic Regression

While discriminant analysis is a useful procedure for assessing the degree of separability between categories from a given set of predictor variables, it does not provide any indication as to the relative contribution of individual variables to the classification result. In other words, we wish to determine not only the distribution of category assignments, but also which parameters were most 'useful' in distinguishing any given pair of categories. For this measure we turn to logistic ('logit') regression, where the response variable is a binary category and the predictors are continuous variables (i.e. the same set as our feature vector X for discriminant analysis). Logistic regression gives a measure of statistical significance (from the t-distribution) for each predictor variable, and thus will allow a post-hoc assessment of the relative contribution of individual variables to the classification result. In the word-initial set we apply the logit model pairwise to the fricative (/x/–/ɦ/) and plosive (/k^h/–/$g^ɦ$/) contrasts because they account for the majority of errors in classification tasks. Inter-vocalic and word-final classification focused on only three categories (/x/, /ɦ/, and spirantized /k^h/) so all contrast permutations could be captured in two distinctions: /x/–/ɦ/ and /x/–/k^h/.

3.2 Results

3.2.1 Within-Pool Design

Jorhat Females

From the outset our expectation from the Jorhat Female group was that the velar-glottal fricative contrast would show the greatest category overlap word-initially. This hypothesis was based primarily on impressionistic data and intuitions of a native speaker from the same Upper Assam dialect region that the two phonemes are often hard to distinguish and may be confused in certain contexts.[1] The results of the classification task on the word-initial obstruent set do in fact meet this expectation, with /x/ and /ɦ/ exhibiting the greatest pairwise confusion probabilities among the four target consonants (Table 3.1).

	/x/	/ɦ/	/k^h/	/$g^ɦ$/
/x/	0.72	0.25	0	0.03
/ɦ/	0.24	0.62	0	0.15
/k^h/	0.03	0.03	0.92	0.03
/$g^ɦ$/	0	0.11	0.06	0.83
Acc.	0.72	0.62	0.92	0.83

Table 3.1: Confusion matrix of obstruents in word-initial position (JF)

Overall the classification accuracy in this frame was 77.5%, with the velar plosives being the most accurately identified (at 92 and 83%, respectively). The only other major

[1] Pamir Gogoi, personal correspondence

misclassification was a recorded overlap of breathy velars and glottal fricatives in 13% of tokens, on average, but this result was also expected given the two categories' shared feature of breathiness at the onset of the following vowel.

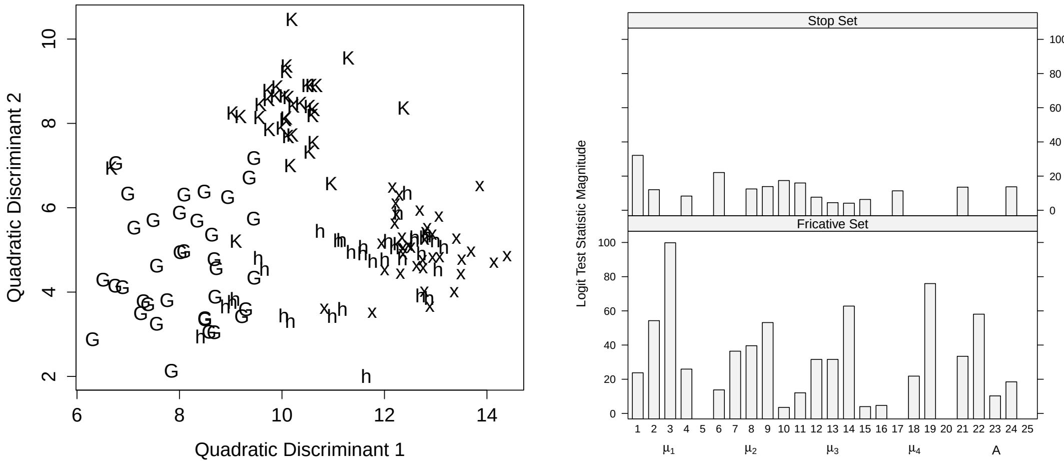


FIGURE 3.2: Classification results for word-initial obstruents (JF). *Left.* Clustering solution of the four consonant categories (/kʰ/ and /gʱ/ designated "K" and "G", respectively) in word-initial position. *Right.* Relative contribution of acoustic parameters in pairwise discrimination of stops and fricatives. The five parameters spanned in each μ_n correspond to windows $C_i - V_2$; amplitudinal parameters spanned by A are in the order: $[A_n, A_d, A_{r_{16}}, A_{r_{17}}, A_{r_{18}}]$.

Figure 3.2 shows the clustering solution in the word-initial set along the first two discriminant dimensions, as well as the relative significance of individual parameters in this result. All but the spectral kurtosis at W_{V_2} were significant in the fricative subset, with the most effective parameters (in terms of Logit test statistic magnitude) being μ_1 at W_{C_m} and W_{C_f}, μ_3 and μ_4 at the consonant-vowel boundary, and the dynamic amplitude of the two fricatives.[2] In the stop subset kurtosis was only significant in the middle of the consonant, and three of the five amplitudinal features were not significant in the logit model. Among the stops the most effective parameters were μ_2 at W_{C_i}, W_{CV}, and W_{V2}, and both μ_1 and μ_3 at consonant onset.

	/x/	/ɦ/	/kʰ/
/x/	0.96	0.01	0.03
/ɦ/	0.07	0.92	0.02
/kʰ/	0.81	0.08	0.11
Acc.	0.96	0.92	0.11

TABLE 3.2: Confusion matrix of continuant obstruents in inter-vocalic position (JF)

[2]All reported 'most effective' logit parameters are significant at $p < 0.001$ unless otherwise specified.

In inter-vocalic position a different pattern emerges among the continuant obstruents: /k^h/ is almost completely mapped onto /x/, and the earlier misclassification density between the two fricatives in word-initial position drops out almost completely. The overall accuracy of the quadratic classifier in this context was 76.3%, the higher overall figure relative to the poor accuracy for the velar aspirate being accounted for primarily due to the smaller sample size of spirantized /k^h/ inter-vocalically.

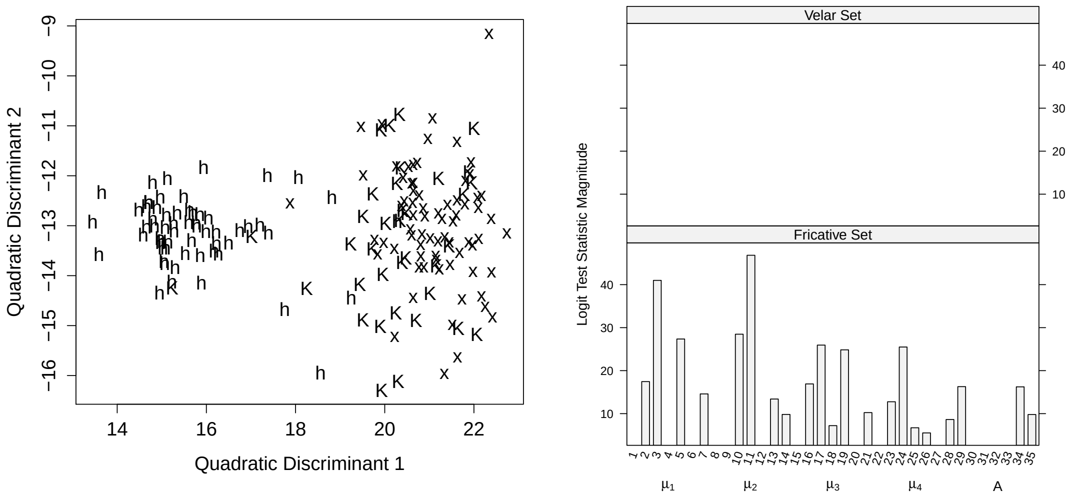


FIGURE 3.3: Classification results for inter-vocalic continuant obstruents (JF). *Left.* Clustering solution of the three consonant categories (/k^h/ designated "K") in inter-vocalic position. *Right.* Relative contribution of acoustic parameters in pairwise discrimination of fricatives (/x/–/ɦ/) and velars (/x/–/k^h/). The seven parameters spanned in each μ_n correspond to windows $V_1 - V_2$; amplitudinal parameters spanned by A are in the order: $[A_n,\ A_d,\ A^{vc}_{r_{16}},\ A^{vc}_{r_{17}},\ A^{vc}_{r_{18}},\ A^{cv}_{r_{16}},\ A^{cv}_{r_{17}},\ A^{cv}_{r_{18}}]$.

Among the velars no parameters were significant in the logit model, as expected from the classifier performance which is visually demonstrated in the cluster plot in Figure 3.3. Here we find the /k^h/ and /x/ regions completely overlapped in QD1-QD2 discriminant space. The fricative set, on the other hand, is clearly separated, with the greatest contribution coming from all four spectral moments in the initial portion of the consonant, as well as μ_2 medially and μ_1 finally. All relative amplitudes except those at 17 and 18 Bark at the CV boundary were insignificant.

	/x/	/ɦ/	/k^h/
/x/	0.89	0.09	0.03
/ɦ/	0.11	0.89	0
/k^h/	0.17	0.17	0.67
Acc.	0.89	0.89	0.67

TABLE 3.3: Confusion matrix of continuant obstruents in word-final position (JF)

Word-finally the classifier yields improved separation between the velars, with /k^{h}/ accuracy 67%, up from 11% in the inter-vocalic frame. Misclassification of /k^{h}/ as /ɦ/ increased to 17% in this context, however. Overall accuracy of continuant obstruent classification in word-final position is 85.5%.

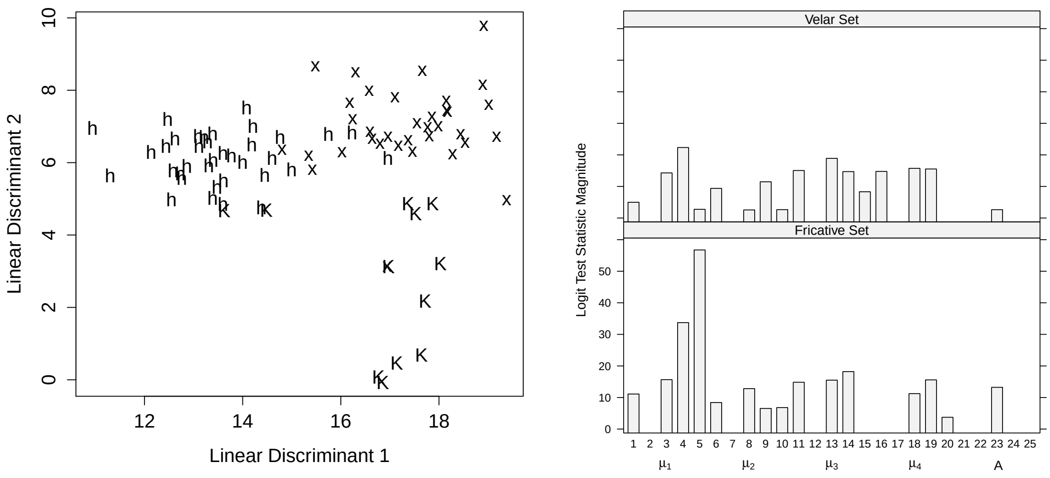


FIGURE 3.4: Classification results for word-final continuant obstruents (JF). *Left.* Clustering solution of the three consonant categories (/k^{h}/ designated "K") in word-final position. *Right.* Relative contribution of acoustic parameters in pairwise discrimination of fricatives (/x/–/ɦ/) and velars (/x/–/k^{h}/). The five parameters spanned in each μ_n correspond to windows $V_1 - C_f$; amplitudinal parameters spanned by A are in the order: $[A_n, A_d, A^{vc}_{r_{16}}, A^{vc}_{r_{17}}, A^{vc}_{r_{18}}]$.

From the pairwise logit model we find /x/ and /ɦ/ primarily distinguished according to μ_1 over windows $W_{C_i} - W_{C_f}$, with μ_3 and μ_4 major determinants over the middle window of the fricative. Amplitudinal measures (aside from $A_{r_{16}}$) and spectral moments at the VC boundary were not significant. Among the velars spectral skewness and kurtosis at consonant onset (C_i) and spectral mean at W_{C_m} were most effective at separating the two categories. As with the fricative subset, the VC window proved an unstable region for spectral measures.

Jorhat Males

Unlike with the Jorhat Female group, we had no prior expectations as to how the data would pattern in the Jorhat Male group. In this sense the classification task was more exploratory, aimed at determining whether or not the there was a notable gender distinction in the determination of contrast in the Jorhati variety of Assamese. For the four-way contrast in word-initial position, similar results were obtained for the /x/–/ɦ/ distinction (i.e. mean confusion between the two categories was at 27%); however, classification of /ɦ/ was in general poor and nearly half (47%) of the breathy velar plosives were misclassified as voiceless aspirates (Table 3.4). Overall accuracy of the classifier in word-initial position was 59.6%.

	/x/	/ɦ/	/k^h/	/$g^ɦ$/
/x/	0.69	0.14	0.17	0
/ɦ/	0.40	0.27	0.17	0.17
/k^h/	0.08	0	0.83	0.08
/$g^ɦ$/	0	0	0.47	0.53
Acc.	0.69	0.27	0.83	0.53

Table 3.4: Confusion matrix of obstruents in word-initial position (JM)

Results from the Jorhat Male group at word onset indicate that while similarity in the acoustics of /x/ and /ɦ/ (as interpreted through a quadratic classifier) is not unique to female speakers, the considerable overlap observed between breathy and voiceless aspirated velar plosives in among Jorhat males is characteristically different from the previous distribution shown for female speakers.

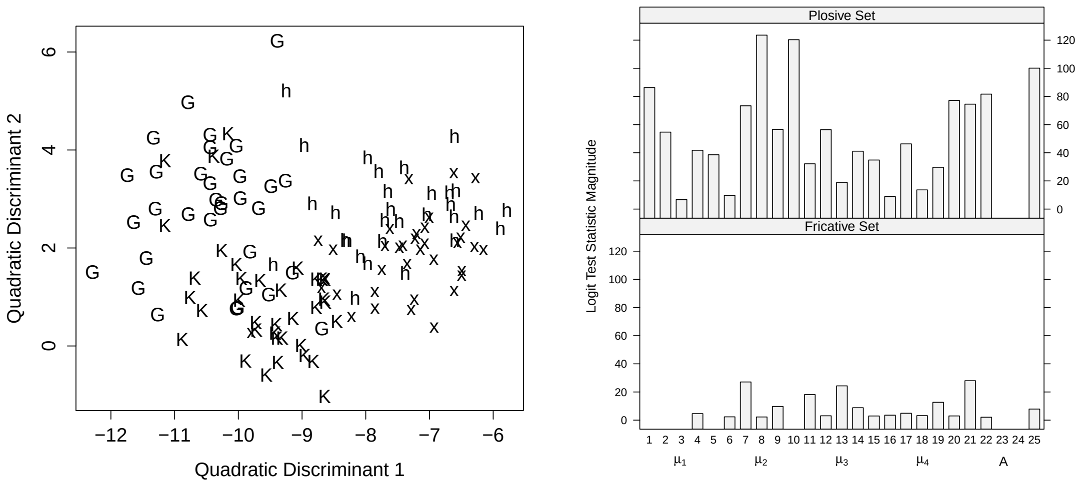


Figure 3.5: Classification results for word-initial obstruents (JM). *Left.* Clustering solution of the four consonant categories (/k^h/ and /$g^ɦ$/ designated "K" and "G", respectively) in word-initial position. *Right.* Relative contribution of acoustic parameters in pairwise discrimination of stops and fricatives. The five parameters spanned in each μ_n correspond to windows $C_i - V_2$; amplitudinal parameters spanned by A are in the order: $[A_n, A_d, A_{r_{16}}, A_{r_{17}}, A_{r_{18}}]$.

Pairwise logit models indicate, however, that within the fricative and plosive subsets the majority of parameters in the acoustic feature vector are in fact significant. This discrepancy between QDA and logistic regression results is likely due to the fact that while misclassification between both fricatives and plosives was relatively high, their distribution was asymmetric, so that while mean confusion between /x/ and /ɦ/, and /k^h/ and

/gɦ/ was at 27 and 27.5%, respectively, in both cases over 40% of the misclassifications were restricted to one direction of confusion.

Contributing most to the separation of the fricative consonants were normalized amplitude (A_n), μ_2 at W_{C_m}, and spectral skewness over the initial and final windows in the consonant interval. Relative amplitudes at Bark 16 and 17 and spectral mean over all but the CV boundary position were not significant. In the plosive subset μ_2 at W_{C_f} and W_{V_2}, spectral mean at consonant onset, and relative amplitude at Bark 18 exhibited the highest logit model scores. As in the fricative test, relative amplitudes at 16 and 17 Bark were not significant. Inter-vocalically the picture is considerably different, but

	/x/	/ɦ/	/k^h/
/x/	0.83	0.03	0.14
/ɦ/	0.06	0.94	0
/k^h/	0.67	0	0.33
Acc.	0.83	0.94	0.33

Table 3.5: Confusion matrix of continuant obstruents in inter-vocalic position (JM)

mirrors results obtained for the Jorhat Female group. Overall accuracy was 79.5%, with the greatest errors occurring between the velar fricative and spirantized velar plosive (Table 3.5).

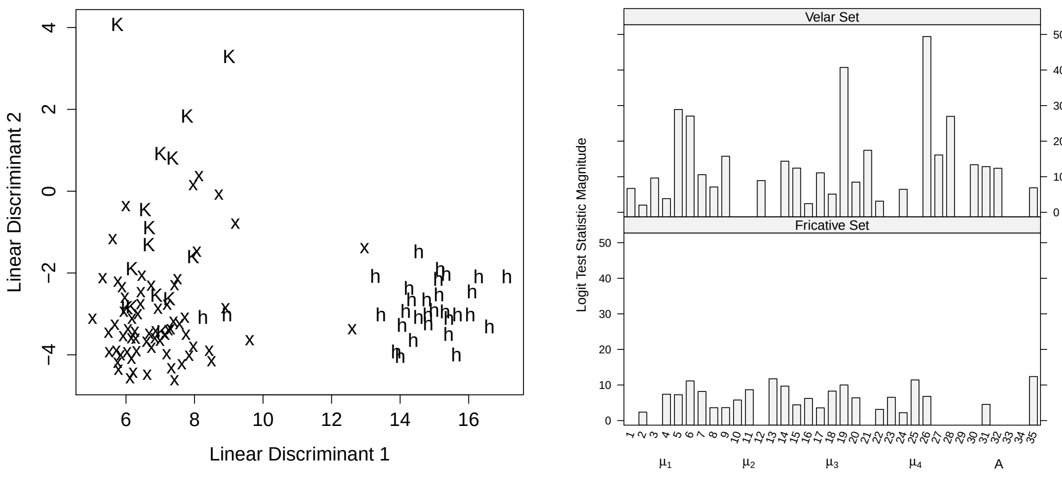


Figure 3.6: Classification results for inter-vocalic continuant obstruents (JM). *Left.* Clustering solution of the three consonant categories (/k^h/ designated "K") in inter-vocalic position. *Right.* Relative contribution of acoustic parameters in pairwise discrimination of fricatives (/x/–/ɦ/) and velars (/x/–/k^h/). The seven parameters spanned in each μ_n correspond to windows $V_1 - V_2$; amplitudinal parameters spanned by A are in the order: [A_n, A_d, $A^{vc}_{r_{16}}$, $A^{vc}_{r_{17}}$, $A^{vc}_{r_{18}}$, $A^{cv}_{r_{16}}$, $A^{cv}_{r_{17}}$, $A^{cv}_{r_{18}}$].

Again we find considerable overlap between /x/ and /kʰ/ categories in the clustering solution in Figure 3.6. The glottal fricative, however, is clearly separated, suggesting again that syllabic position is a significant factor in determining the production characteristics – and consequently acoustic characteristics – of this set of posterior obstruents.

From the pairwise logit model comparisons we find relative amplitude at Bark 18 the strongest predictor of category membership among the fricatives, with spectral mean and dispersion at CV boundary and kurtosis at W_{C_m} similarly prominent. Dynamic amplitude and relative amplitudes at Bark 16 and 17 were not significant. In the velar subset, spectral mean, skewness and kurtosis in the final window of the consonant interval (W_{C_f}) contributed most to category discrimination. As with the fricatives, dynamic amplitude and relative amplitudes at Bark 16 and 17 were not significant.

	/x/	/ɦ/	/kʰ/
/x/	0.83	0.08	0.08
/ɦ/	0.07	0.93	0
/kʰ/	0.27	0	0.73
Acc.	0.83	0.93	0.73

TABLE 3.6: Confusion matrix of continuant obstruents in word-final position (JM)

As with the Jorhat Female group, classification performance improved in word-final position, though /x/–/kʰ/ misclassification was still a notable trend. Overall accuracy for the linear classifier in this context was 85.1% (Table 3.6).

From the clustering solution in Figure 3.7 we see that in addition to exhibiting greater overlap with /x/ than is found for the glottal, /kʰ/ is much more diffuse in the linear discriminant coordinate system. The velar and glottal fricatives, by contrast, are much more tightly distributed. Logit predictions suggest spectral moment parameters (μ_1, μ_2, μ_3) at fricative onset, kurtosis at W_{C_m}, and dispersion in the final window on the vowel were most critical in separating glottal and velar fricatives word-finally. In the /x/–/kʰ/ contrast spectral mean at VC boundary and consonant onset was most prominent in the model, followed by spectral dispersion over the first two windows on the consonant. Amplitudinal parameters, with the exception of A_d, were insignificant for both fricative and velar subsets.

Nalbari Males

Classification results for word-initial obstruents in the Nalbari Male group were similar to those obtained for Jorhat males, with one critical exception being that in the Nalbari case, errors between /x/ and /ɦ/, and /kʰ/ and /gʱ/ were evenly balanced between both

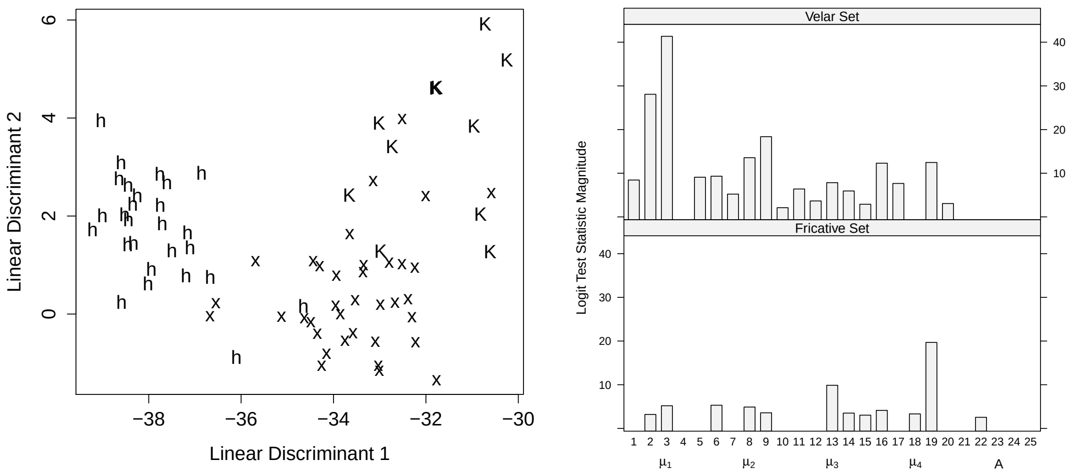


FIGURE 3.7: Classification results for word-final continuant obstruents (JM). *Left.* Clustering solution of the three consonant categories (/k^h/ designated "K") in word-final position. *Right.* Relative contribution of acoustic parameters in pairwise discrimination of fricatives (/x/–/ɦ/) and velars (/x/–/k^h/). The five parameters spanned in each μ_n correspond to windows $V_1 - C_f$; amplitudinal parameters spanned by A are in the order: $[A_n,\ A_d,\ A^{vc}_{r_{16}},\ A^{vc}_{r_{17}},\ A^{vc}_{r_{18}}]$.

directions of misclassification. This distinction could suggest alternations within plosive and fricative subsets are more stable in the Nalbari variety than in Jorhati Assamese, however these speculations are left to be elaborated in the Conclusion for now.

	/x/	/ɦ/	/k^h/	/$\text{g}^\text{ɦ}$/
/x/	0.63	0.29	0.09	0
/ɦ/	0.35	0.62	0.03	0
/k^h/	0.03	0	0.72	0.25
/$\text{g}^\text{ɦ}$/	0	0.03	0.31	0.66
Acc.	0.63	0.62	0.72	0.66

TABLE 3.7: Confusion matrix of obstruents in word-initial position (NM)

Overall accuracy of the quadratic classifier on the four-way contrast in word-initial position was 66%, similar to results from the Jorhat Male group, both of which are poor relative to performance of the classifier on Jorhat females.

From Figure 3.8 the distinction between symmetric and asymmetric confusion (relative to Nalbari and Jorhat Male classification results, respectively) within fricative and plosive subsets can be seen more clearly. Category dispersion and overlap is more balanced between /x/ and /ɦ/, and /k^h/ and /$\text{g}^\text{ɦ}$/ than for the Jorhat males (Figure 3.5). Additionally, the subgroups themselves (fricative and plosive) are more clearly demarcated along the QD1 dimension.

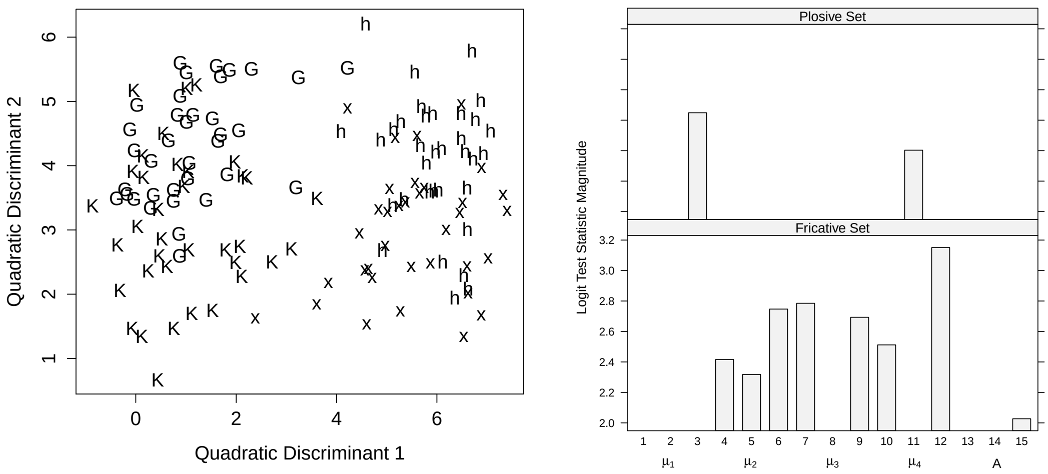


FIGURE 3.8: Classification results for word-initial obstruents (NM). *Left.* Clustering solution of the four consonant categories (/kʰ/ and /gʱ/ designated "K" and "G", respectively) in word-initial position. *Right.* Relative contribution of acoustic parameters in pairwise discrimination of stops and fricatives. The three parameters spanned in each μ_n correspond to windows $C_i - C_f$; amplitudinal parameters spanned by A are in the order: $[A_n, A_{r_{17}}, A_{r_{18}}]$.

Pairwise tests in the logit model for the fricative subset revealed prominent effects of μ_2, μ_3, and μ_4 at W_{C_f}, and μ_3 and μ_4 at W_{C_i}. These effects were significant at $p < 0.01$. Spectral mean over all window positions and normalized amplitude were not significant. Among plosives only two parameters were significant, μ_1 at W_{C_f} and μ_4 at W_{C_m} [$p < 0.05$].

	/x/	/ɦ/	/kʰ/
/x/	0.60	0.03	0.37
/ɦ/	0.28	0.72	0
/kʰ/	0.35	0.02	0.63
Acc.	0.60	0.72	0.63

TABLE 3.8: Confusion matrix of continuant obstruents in inter-vocalic position (NM)

In inter-vocalic position notable errors occurred between the velar fricative and spirantized velar plosive, again balanced along both directions of misclassification. This result fits the errors obtained in inter-vocalic classification in the Jorhat Male, but again with a more symmetric distribution than the latter. Unlike in the Jorhat Male or Female groups, however, a significant proportion of glottal fricatives (28%) were misclassified as velars. This result is surprising given the otherwise high accuracy of /ɦ/ classification inter-vocalically (89 and 94% for JF and JM, respectively). Overall accuracy of the linear classifier for inter-vocalic continuant obstruents in the Nalbari Male set was 62.8%.

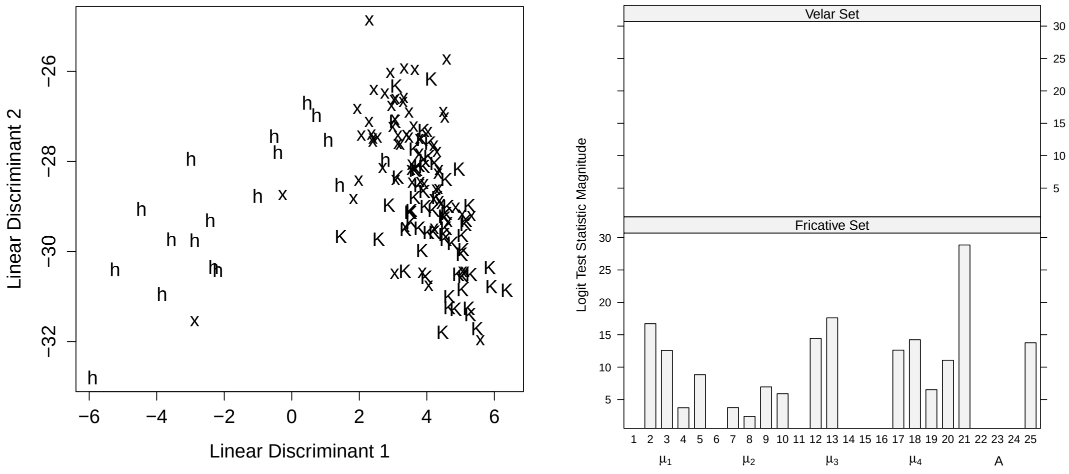


FIGURE 3.9: Classification results for inter-vocalic continuant obstruents (NM). *Left.* Clustering solution of the three consonant categories ($/\text{k}^{\text{h}}/$ designated "K") in inter-vocalic position. *Right.* Relative contribution of acoustic parameters in pairwise discrimination of fricatives (/x/–/ɦ/) and velars (/x/–$/\text{k}^{\text{h}}/$). The five parameters spanned in each μ_n correspond to windows $V_1 - C_f$; amplitudinal parameters spanned by A are in the order: $[A_n,\ A_d,\ A^{vc}_{r_{16}},\ A^{vc}_{r_{17}},\ A^{vc}_{r_{18}}]$.

Tests of predictor variable contribution in logit models revealed no significant parameters in categorization among the velars. For the glottal-velar fricative contrast, normalized amplitude was the most prominent parameter, followed by μ_3 and μ_4 at W_{C_i}, and μ_1 and μ_3 at W_{VC}. Spectral moments over the V_1 window were not significant, nor was dynamic amplitude.

	/x/	/ɦ/	$/\text{k}^{\text{h}}/$
/x/	0.88	0.03	0.09
/ɦ/	0.24	0.76	0
$/\text{k}^{\text{h}}/$	0.95	0.05	0
Acc.	0.88	0.76	0

TABLE 3.9: Confusion matrix of continuant obstruents in word-final position (NM)

The continuant obstruent contrast in word-final position in the Nalbari Male data exhibited the most dramatic display of neutralization of any classification task reviewed thus far. Not a single $/\text{k}^{\text{h}}/$ token was accurately classified. Overall performance of the quadratic classifier in this context was 63.1%, however, owing to the relativel high accuracy of /x/ and /ɦ/ categories.

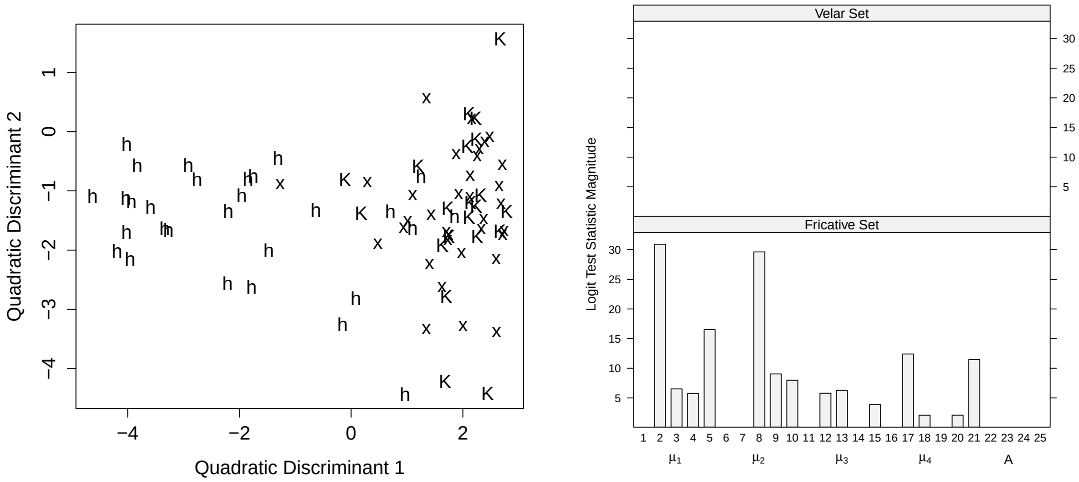


FIGURE 3.10: Classification results for word-final continuant obstruents (NM). *Left.* Clustering solution of the three consonant categories (/k^h/ designated "K") in word-final position. *Right.* Relative contribution of acoustic parameters in pairwise discrimination of fricatives (/x/–/ɦ/) and velars (/x/–/k^h/). The five parameters spanned in each μ_n correspond to windows $V_1 - C_f$; amplitudinal parameters spanned by A are in the order: [A_n, A_d, $A^{vc}_{r_{16}}$, $A^{vc}_{r_{17}}$, $A^{vc}_{r_{18}}$].

As with the logit model results from inter-vocalic classification, no significant parameters were obtained among the velars. In the fricative subset, μ_1 and μ_4 at the VC boundary, and normalized amplitude were prominent predictors. As in the intervocalic case, all spectral moments at W_{V_1} were insignificant.

3.2.2 Cross-Sex Design

We now examine the relative compatibility between the Jorhat Male and Female posterior obstruent systems, considering each syllabic context in turn. For each context a classifier trained on the Jorhat Female data was tested on the Jorhat Male set, with the parameters chosen as the intersection of those applied to each group in the within-pool design. In the word-initial condition this parameter set corresponds to the $X_{JF/cv}$ feature vector minus relative amplitudes. Performance of a quadratic classifier on the crossed design was generally poor (52.9% overall accuracy), where the greatest category errors occurred from the velar fricative /x/ and the breathy plosive /$g^ɦ$/. Misclassifications of /x/ were evenly distributed over the other three categories, while errors from /$g^ɦ$/ were entirely classified as the glottal fricative /ɦ/. Given that the /x/–/ɦ/ alternation was robust in the Jorhat Female within-pool design, and that internal classification of Jorhat Male data exhibited misclassifications of /k^h/ as /x/ and /$g^ɦ$/, the distributed nature of /x/ misclassifications in the cross-sex design is perhaps not surprising.

	/x/	/ɦ/	/k^h/	/g^ɦ/
/x/	0.36	0.28	0.22	0.14
/ɦ/	0.10	0.73	0	0.17
/k^h/	0.08	0.14	0.64	0.14
/g^ɦ/	0	0.58	0	0.41
Acc.	0.36	0.73	0.64	0.41

TABLE 3.10: Confusion matrix of obstruents in word-initial position (Cross-Sex)

Performance of the inter-vocalic model was better overall, at 70.5%, though the pattern of misclassification matched neither the within-pool result from the Jorhat Female group nor the result from the Jorhat male group. All spirantized velar plosives were misclassified, with 67% categorized with the glottal, and 33% categorized with the velar fricative. Additionally, 21.2% of the velar fricatives were classified as glottals, a result not mirrored in either of the two subject pools. These results may not correspond to any real differences between the speech patterns of male and female speakers of Jorhati Assamese, however, but rather could simply be a consequence of the complex mapping of one /x/–/k^h/ alternation onto another.

	/x/	/ɦ/	/k^h/
/x/	0.72	0.21	0.06
/ɦ/	0	1	0
/k^h/	0.33	0.67	0
Acc.	0.72	1	0

TABLE 3.11: Confusion matrix of continuant obstruents in inter-vocalic position (Cross-Sex)

Word-finally overall accuracy was on par with the inter-vocalic case at 68.9%, though in this context the majority of misclassifications of /k^h/ were categorized as /x/. This is not a symmetric alternation, however, as only 8% of /x/ tokens were misclassified as /k^h/, as opposed to 28% which were incorrectly classified as the glottal fricative. As in the inter-vocalic environment, /ɦ/ shows a high classification accuracy, one feature which remains constant between the intra-group results and the crossed design.

Our examination of obstruent classification across male and female Assamese speakers from Jorhat has yielded mixed results. On the one hand, the patterning of results – particularly in word-initial and inter-vocalic environments – was widely deviant from the distributions observed in internal classifications of the Jorhat Male and Female groups. Yet given that the two groups were already misaligned in classification performance (namely with respect to /k^h/ categorization) a crossed design would be expected to be

	/x/	/ɦ/	/k^h/
/x/	0.64	0.28	0.08
/ɦ/	0	0.96	0.04
/k^h/	0.73	0.09	0.18
Acc.	0.64	0.96	0.18

Table 3.12: Confusion matrix of continuant obstruents in word-final position (Cross-Sex)

similarly idiosyncratic. Furthermore, while the mean overall accuracy in the cross-sex design (64.1%) was lower than that for within-pool classification on the Jorhat Female and Jorhat Male sets (79.8 and 74.7%, respectively), the difference is not substantial enough to warrant firm conclusions on the relation between speech systems in the two varieties. Neither did we find the high classification accuracy between female and male speakers as reported in Forrest et al. [1988], nor do we find clear evidence for incompatibility of the two systems. At present the crossed procedure as applied to the intra-dialectal gender distinction cannot be utilized as a robust indicator of either affinity or disaffinity between male and female speech varieties.

3.2.3 Cross-Dialectal Design

Building a crossed classifier between dialectal varieties is in some respects a more ambitious task than training and testing between male and female groups from the same locality. This view is especially apt in the context of velar fricatives and their relation to breathy/aspirated velar plosives, because not only have dialectal distinctions been noted with respect to the nature of velar frication in Assamese [see Sarma, 2012], but the phonatory distinction between plosives – particularly velars – has also been cited as significantly variant between Eastern and Western varieties, of which Jorhat and Nalbari are two exemplars [Goswami, 1966, Kakati, 1941].

	/x/	/ɦ/	/k^h/	/$g^ɦ$/
/x/	0.39	0.17	0.36	0.08
/ɦ/	0.2	0.77	0	0.03
/k^h/	0.03	0.06	0.67	0.25
/$g^ɦ$/	0.29	0.26	0.06	0.38
Acc.	0.39	0.77	0.67	0.38

Table 3.13: Confusion matrix of obstruents in word-initial position (Cross-Dialect)

In word-initial position, overall classification accuracy is 54.4% when training on the Nalbari Male group and testing on Jorhat Male data. These results are compatible with the performance of the cross-validated within-pool classifier. Notable deviations include the even distribution of misclassifications of /$g^{ɦ}$/ over the remainder of categories, and the decline in accuracy of /k^h/ classification. Regarding the former, classification of /$g^{ɦ}$/ in the within-pool design for the Jorhat Male group was similarly poor, however in that case misclassifications were entirely limited to /k^h/. The poor performance of the crossed classifier on the voiceless aspirated velar is possibly due to the substantial acoustic overlap between /k^h/ and /$g^{ɦ}$/ in the Nalbari Male set, biasing the classifier toward a near-chance distinction between velar plosives.

	/x/	/ɦ/	/k^h/
/x/	0.11	0.02	0.88
/ɦ/	0.19	0.74	0.06
/k^h/	0.13	0	0.87
Acc.	0.11	0.74	0.87

TABLE 3.14: Confusion matrix of continuant obstruents in inter-vocalic position (Cross-Dialect)

Inter-vocalically, classifier performance overall is just above chance level, at 38.4%, yielding our first clear case of incompatibility between the obstruent systems of Nalbari and Jorhati Assamese. When trained and cross-validated internally, classification accuracy in the Jorhat Male group for inter-vocalic consonants /x/, /ɦ/, and /k^h/ was 79.5%, a two-fold improvement over the crossed design. Misclassification of /x/ as /k^h/ was the most prominent contributor to this result, an expected outcome given the stable alternation between the velar fricative and spirantized velar plosive inter-vocalically in the Nalbari data.

	/x/	/ɦ/	/k^h/
/x/	0.83	0.11	0.06
/ɦ/	0.22	0.78	0
/k^h/	1	0	0
Acc.	0.83	0.78	0

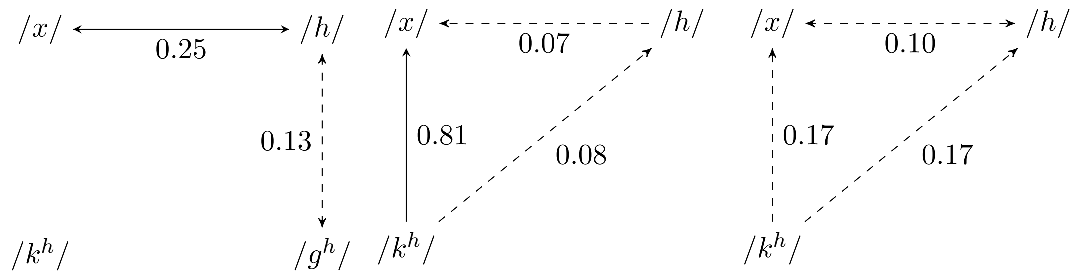
TABLE 3.15: Confusion matrix of continuant obstruents in word-final position (Cross-Dialect)

Overall performance of the crossed classifier on word-final continuant obstruents was comparable to (though reduced from) that of the cross-validated linear classifier in the within-pool design—68.9% as compared with 85.1% among Jorhat males. The pattern of misclassifications in the crossed design, however, deviated substantially in /k^h/ category assignment. Here the complete mapping of /k^h/ onto /x/ mirrors the word-final classification result for within-pool classification of Nalbari Male data.

Considering the results of cross-dialect classification in all three positions (CV, VCV, and VC), there does appear to be preliminary evidence that the hypothesis of equality of posterior obstruent systems across dialects in Assamese is untenable. Word-initially, overall performance and misclassification patterns were not markedly different between the Nalbari Male and Jorhat Male groups, nor was the crossed design notably different from the within-pool result. Inter-vocalically and word-finally, however, training on one dialect and testing on the other produced prominent errors, particularly with respect to classification of the spirantized velar plosive. Thus if the cross-dialectal design may be used to pinpoint the locus of difference between the two speech systems, it would be at the /x/–/k^h/ category boundary post-vocalically.

3.3 Discussion

In the preceding section we presented classification results from a range of subject groups, segmental environments, and design types. We now return to the *neutralization potential* diagram from Chapter 2 (Figure 2.1) and use this schema to generate diagrams of *contrast loss probabilities* for each subject group and segmental environment. Thus rather than considering pure neutralization as a categorical phenomenon to be explained via deterministic phonological rules, we view the posterior obstruent system through the lense of potential for loss of contrast. As in the classification task, only /x/, /ɦ/, and spirantized /k^h/ are considered post-vocalically.[3]

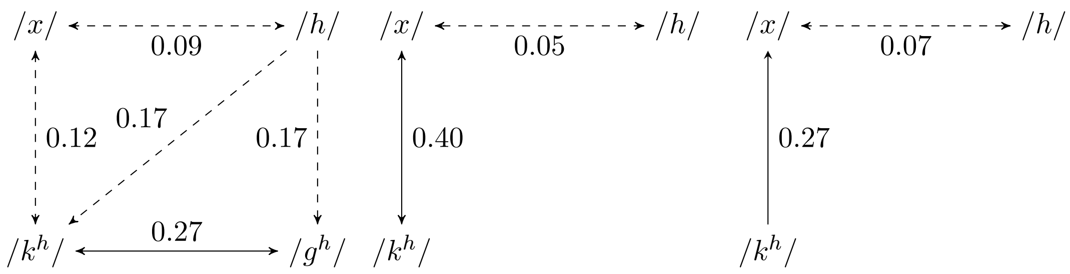


(A) VCV (B) VC#

FIGURE 3.11: Contrast loss probabilities by position (JF). Reported probabilities represent mean misclassification proportion for a given contrast. Arrows indicate direction and magnitude of misclassification (solid : $p > 0.25$, dashed : $0.05 < p < 0.25$).

In the Jorhat Female group the primary axes of alternation are /x/–/ɦ/ word-initially, and /k^h/ to /x/ (i.e. unidirectional movement) inter-vocalically. Word-finally all three categories alternate with relatively even probabilities, though the direction of movement is skewed toward /x/ and /ɦ/ and away from /k^h/.

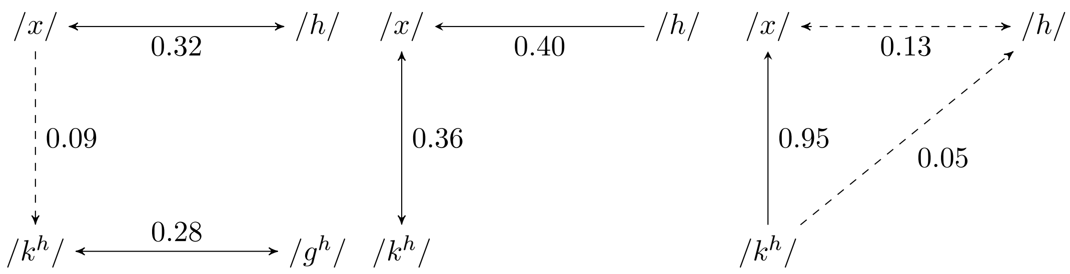
[3]This procedure will naturally bias the contrast loss potential in such cases. A full picture would require incorporation of spirantization probabilities as well as lexical priors for target words



(A) #CV (B) VCV (C) VC#

FIGURE 3.12: Contrast loss probabilities by position (JM). Reported probabilities represent mean misclassification proportion for a given contrast. Arrows indicate direction and magnitude of misclassification (solid : $p > 0.25$, dashed : $0.05 < p < 0.25$).

Among Jorhat males, nearly all axes of contrast show some degree of contrast loss potential, though most critical is the /k^{h}/–/g^{ɦ}/ distinction. Minor bi-directional observations are the velar and glottal fricatives, and the voiceless velars. Inter-vocalically the primary source of classification error is between /x/ and /k^{h}/, with minor alternations between the two fricatives. Word-finally a similar system is obtained, though contrast loss between the two velars is unidirectional in this case.



(A) #CV (B) VCV (C) VC#

FIGURE 3.13: Contrast loss probabilities by position (NM). Reported probabilities represent mean misclassification proportion for a given contrast. Arrows indicate direction and magnitude of misclassification (solid : $p > 0.25$, dashed : $0.05 < p < 0.25$).

In the Nalbari Male system word-initial alternations are more constrained and stable than was found for Jorhat males. With the exception of minor misclassifications of the velar fricative as its aspirated plosive counterpart, this subsystem is defined by two major alternations: that between the glottal and velar fricatives, and that between voiceless aspirated and breathy voiced velar plosives. Inter-vocalically the velar contrast is similar to that in the Jorhat Male group; however, the Nalbari system is also defined by a prominent unidirectional mapping of the glottal onto the velar fricative. Word-finally the Nalbari Male group deviates from Jorhat males only in magnitude, with a

95% unidirectional mapping of /k^h/ onto /x/, a result emphasized previously in the cross-dialectal classification task.

In exploring a variety of classification tasks on the set of posterior obstruents identified as most likely to be acoustically similar to the velar fricative in Assamese, this chapter has provided a fine-grained dissection of the system of contrast among these consonants. The next chapter will combine these results with the acoustic descriptions from Chapter 2 to develop a clearer picture of what we know at present about this obstruent system in Assamese and what further research is needed to address still unanswered questions.

Chapter 4

Conclusion

In Chapter 2 and Chapter 3 we presented data on an array of spectral and amplitudinal properties of velar and glottal obstruents, as well as results from the application of those parameters to a pattern classification task. We now consolidate these findings with two aims. First, we address the question of whether or not there are stable acoustic features from the measured set which may reliably be used to identify a given consonant in a particular syllabic or vocalic environment. The second question concerns the system of contrast in Assamese as a whole, focusing in particular on positional constraints and differences between speech varieties.

4.1 Summary of Results

From the general acoustic description presented in Chapter 2 we can see a number of larger patterns in the spectral moment and amplitudinal data across consonants and segmental environments. For one, spectral measurements at the middle and final portions of the consonant appear to be the most stable overall. In the former case this is likely to do with the relative isolation of the C_m window from adjacent vowels or onset effects when produced in isolation. The window at C_f, on the other hand, is more vulnerable to contextual effects, but also can be precisely defined relative to the vowel in all tokens. Nevertheless, across the posterior obstruent system it is clear that the four moments offer different ideal separation regions and thus the inclusion of a wide set of cues from different temporal locations seems justified. Additionally, contrary to the findings of Forrest et al. [1988], results from the Assamese system studied in this thesis show no cause for broad exclusion of any of the four spectral moments. Among the amplitudinal parameters normalized amplitude remains the simplest and most effective measure at distinguishing among categories in this set. Dynamic amplitude was less reliable, as was relative amplitude, which both exhibited points of critical value in some subsets of the classification task, but overall were not as robust as normalized amplitude or any of the spectral shape parameters.

From the classification task in Chapter 3 we find that while the model offers good discrimination of certain consonants in certain environments, there were no clear overall areas of success (that is, categories where classification accuracy was high across all contexts and subject groups). But when the speech variety is controlled, and the problem is broken down into specific segmental environments clear patterns do emerge from the classification results. Notably, word-initial discrimination of velar and glottal fricatives is poor for all groups, as is word-final discrimination of the velar fricative and spirantized velar plosive. Among male speakers the phonation contrast among the plosives also exhibits a high degree of alternation word-initially, though it remains to be seen if the same effects are present in the word-final position. This is partly a difficult question to answer because aspirated and breathy stops are in general not common word-finally. Lastly, while the cross-sex classification design did not produce any clear evidence for the treatment of male and female speakers from Jorhat as exhibiting fundamentally different systems of contrast among the target posterior obstruents, the cross-dialectal design does seem to suggest that very conclusion for Nalbari and Jorhat varieties of Assamese.

4.2 Implications and Future Work

In order to achieve greater precision in our model of posterior obstruent acoustics in Assamese we have had to produce a set of acoustic and classification results which are finely differentiated according to segmental environment and subject group; thus the broad implications of this seemingly fragmented solution may be less clear upon initial parse. Before examining what the results imply about the linguistic system in Assamese, however, we first examine the model itself, assessing its validity and areas for refinement.

Discriminant analysis is the primary statistical classification model employed in work on fricative acoustics [see Forrest et al., 1988, Jassem, 1995, Jongman et al., 2000, Shadle and Mair, 1996], and it has been used in the present work primarily for its simplicity and transparency in evaluating predictors and category identification accuracy. Yet because linear discriminant analysis (LDA) relies on the assumption of homogeneity of variances, an assumpion which is not met in the present data and which should not be expected to be met in any linguistic data of this sort (due to contextual variability resulting primarily from the influence of coarticulation), quadratic classifiers are the only statistically valid models of the discriminant type. For this reason we have employed quadratic discriminant analysis (QDA) as the baseline classification procedure in this study.[1] The use of a common model across classification tasks has allowed us to make clear comparisons between results in different segmental environments and subject groups; however, because only one type of classification procedure was employed, evaluation of the general efficiency of the model (as in McMurray and Jongman [2011]) was not possible. Thus it is unclear at present whether discriminant analysis is the best model with which to evaluate the system of posterior obstruent contrasts in Assamese. Furthermore, our calculation of the discriminant function has relied on the experimental distribution of category exemplars in the determination of priors. Since the elicitation set was designed to be as balanced as possible between the four target categories, the prior probabilities incorporated in the model are notably artificial. That is, they cannot be assumed to reflect the natural distribution of category contrasts in the Assamese lexicon. For example, some of the misclassifications discussed in Chapter 3 – such as the common identification of spirantized /k^h/ as /x/ word-finally – would likely be reduced or made statistically insignificant with the inclusion of corpus-based phonotactic and lexical frequencies. In future work we plan to model these contextual asymmetries by extracting type and token-based segmental frequencies from a text corpus of Assamese.

The extension of classification results to a descriptive evaluation of the linguistic system in Assamese is a more vexed issue, and one which has not been addressed in any of the literature reviewed in Chapter 1. In all of these studies the category contrast has been assumed to be clear and absolute; that is, in such predominantly English-based studies there was no theoretical or descriptive impetus for researchers to model full or contextually restricted perceptual overlap between phonemic categories. As a result of this fact, the classification task in these studies served as a means of evaluating the cues necessary to distinguish the contrast. Description of the salience of the contrast itself was not the aim; however, in our investigation of velar fricatives within the posterior obstruent system in Assamese this question of contrast salience under different phonotactic and speaker conditions was of major interest. As such we ought to consider whether the conclusions of this sort discussed in the previous section are empirically valid. We believe our framework is indeed valid in this regard for two reasons: (1) since the same baseline set of acoustic parameters was used in each classification task we can conclude that major differences in the performance of the model in different subject groups and segmental conditions reflect differences in the salience of contrast in these contexts; (2) the fact that some categories were identified with perfect or near-perfect accuracy suggests that the set of acoustic parameters is not generally insufficient for the data set

[1] As noted in Chapter 3, LDA had to be used as a fallback in some cases due to limitations in the QDA algorithm employed in R [R Core Team, 2013].

and classification task, and thus high misclassification of certain contrasts likely reflects cases of neutralization or free variation in the linguistic system itself.

In future work we plan to conduct a longitudinal study of similar structure to the present design, where gender and dialect distinctions can be studied as a dynamically evolving system. If the Jorhat and Nalbari varieties are in fact distinct with regard to phonemic contrast among posterior obstruents (as suggested in Chapter 3), they should exhibit not only static differences but should follow different trajectories in their change over time. This framework has not been employed in historical linguistic work as yet, however, so its productivity in this regard is uncertain at present.

Appendix A

Stimulus Design

#C[a]

/x/	/ɦ/	/kʰ/	/gʱ/
xat	ɦat	kʰat	gʱat
xap	ɦal	kʰap	gʱap
xar	ɦar	kʰam	gʱam

#C[i]

/x/	/kʰ/
xit	kʰin
xil	kʰili
xir	kʰir

[ɔ]C[ɔ]

/x/	/ɦ/	/kʰ/	/gʱ/
ɔxɔm	zɔɦɔr	nɔkʰɔn	nɔgʱɔr
rɔxɔna	rɔɦɔn	sɔkʰɔn	sɔgʱɔr
bɔxɔntɔ	bɔɦɔl	zɔkʰɔla	–

[a]C#

/x/	/ɦ/	/kʰ/	/gʱ/
kax	saɦ	lakʰ	bagʱ
nax	maɦ	rakʰ	magʱ
rax	zaɦ	–	–

[i]C#

/x/	/kʰ/
bix	likʰ
dix	nirikʰ
xirix	–

Appendix B

Elicited Words

Target	
Assamese	*English*
bag^{ɦ}	tiger
bix	pain
bɔɦɔl	wide
bɔxɔntɔ	Spring
buxɔn	ornament
dix	direction
duk^{h}	sorrow
g^{ɦ}am	sweat
g^{ɦ}ap	stroke of dao blade
g^{ɦ}at	river port
g^{ɦ}ɔmɔg^{ɦ}ɔntɔ	chaos
ɦal	plough
ɦar	bone
ɦat	arm/hand
kax	beside
k^{h}am	envelope
k^{h}ap	level
k^{h}at	bed/cot
k^{h}ili	nail
k^{h}in	thin
k^{h}ir	milk rice dish
k^{h}ub	very
k^{h}un	murder (N)
k^{h}ur	razor

lakʰ	one-hundred-thousand
likʰ	write (IMP)
magʱ	month of Assamese year
maɦ	month
mukʰ	face
nak	nose
nax	destroy
nirikʰ	rate
nɔgʱɔr	nine houses
nɔkʰɔn	nine-CL
ɔxɔm	Assam
purux	man
rag	musical arrangement
rakʰ	keep (IMP)
rax	a festival
rɔɦɔn	hue
rɔxɔna	taste-bud/tongue
saɦ	tea
sɔgʱɔr	six houses
sɔkʰɔn	six-CL
tikʰa	steel
uxa	morning
xap	snake
xar	manure/fertilizer
xat	seven
xil	stone
xir	vein
xirix	tea garden tree
xit	Winter
xukʰ	happiness
xun	hear/listen (IMP)
xur	music/voice
xut	interest on a loan
zaɦ	immerse (IMP)
zɔɦɔr	poison
zɔkʰɔla	ladder

Dummy

Assamese	*English*
am	mango
bal	male child
beli	sun
bɛl	wood apple
b^{ɦ}al	good
bil	lake
bol	let’s go
bɔr	big
bud^{ɦ}bar	Wednesday
bul	a person’s name
bʊl	color
dada	elder brother
diŋi	neck
do͡ɪ	curd
dɔlɔŋ	bridge
gɔrɔm	hot
gɔru	cow
moɦ	mosquito
mur	head
nemu	lemon
nodi	river
pani	water
pasi	basket
p^{h}ɔl	fruit
pɔrijal	family
rɔŋ	color
sowali	girl
suli	hair
tan	hard/difficult
tita	bitter

Appendix C

Univariate Parameter Significance

	JF			JM			NM		
Window	*CV*	*VCV*	*VC*	*CV*	*VCV*	*VC*	*CV*	*VCV*	*VC*
V_1		n.s.	***		**	n.s.		n.s.	n.s.
		$\emptyset$	(2)		(2)	$\emptyset$		$\emptyset$	$\emptyset$
VC		***	n.s.		*	*		***	**
		(2)	$\emptyset$		(1)	(1)		(2)	(2)
C_i	***	***	***	**	***	***	**	***	***
	(4)	(2)	(2)	(4)	(1)	(2)	(4)	(2)	(2)
C_m	***	***	***	***	***	***	***	***	***
	(4)	(2)	(2)	(4)	(2)	(2)	(4)	(2)	(2)
C_f	**	***	***	***	***	***	**	***	***
	(3)	(2)	(2)	(4)	(2)	(2)	(1)	(3)	(2)
CV	**	n.s.		n.s.	*		n.s.	n.s.	
	(3)	$\emptyset$		$\emptyset$	(1)		$\emptyset$	$\emptyset$	
V_2	**	n.s.		n.s.	n.s.		n.s.	n.s.	
	(3)	$\emptyset$		$\emptyset$	$\emptyset$		$\emptyset$	$\emptyset$	

TABLE C.1: Results of significance testing for main effects of *Consonant* on μ_1 across window locations, speaker groups, and positions. For each combination ($Speaker \times Position \times Window$), statistical significance is indicated in the first row (n.s. : $p > 0.05$, * : $p < 0.05$, ** : $p < 0.01$, *** : $p < 0.001$), with the number of contrasts distinguished in multiple comparisons in the second.

	JF			JM			NM		
Window	*CV*	*VCV*	*VC*	*CV*	*VCV*	*VC*	*CV*	*VCV*	*VC*
V_1		n.s.	***		**	**		n.s.	n.s.
		∅	(2)		(2)	(2)		∅	∅
VC		***	n.s.		***	n.s.		***	**
		(2)	∅		(2)	∅		(2)	(1)
C_i	***	***	n.s.	***	***	***	***	***	***
	(6)	(2)	∅	(5)	(1)	(2)	(4)	(3)	(2)
C_m	***	***	***	**	***	***	***	***	***
	(4)	(3)	(2)	(5)	(2)	(3)	(4)	(3)	(2)
C_f	**	***	***	***	***	***	n.s.	***	***
	(3)	(2)	(2)	(0)	(2)	(3)	∅	(3)	(2)
CV	***	*		***	*		*	n.s.	
	(3)	(1)		(2)	(1)		(0)	∅	
V_2	***	***		***	***		**	n.s.	
	(4)	(1)		(4)	(3)		(0)	∅	

Table C.2: Results of significance testing for main effects of *Consonant* on μ_2 across window locations, speaker groups, and positions. For each combination ($Speaker \times Position \times Window$), statistical significance is indicated in the first row (n.s. : $p > 0.05$, * : $p < 0.05$, ** : $p < 0.01$, *** : $p < 0.001$), with the number of contrasts distinguished in multiple comparisons in the second.

	JF			JM			NM		
Window	CV	VCV	VC	CV	VCV	VC	CV	VCV	VC
V_1		n.s.	***		***	n.s.		n.s.	n.s.
		$\emptyset$	(2)		(2)	$\emptyset$		$\emptyset$	$\emptyset$
VC		***	n.s.		***	**		***	**
		(2)	$\emptyset$		(2)	(1)		(2)	(2)
C_i	**	***	**	***	***	***	***	***	***
	(3)	(2)	(2)	(4)	(1)	(2)	(4)	(2)	(2)
C_m	**	***	***	**	***	***	***	***	***
	(3)	(2)	(2)	(4)	(2)	(2)	(4)	(2)	(2)
C_f	n.s.	***	***	***	***	***	**	***	***
	$\emptyset$	(2)	(2)	(2)	(2)	(2)	(1)	(3)	(2)
CV	***	***		*	n.s.		*	n.s.	
	(3)	(2)		(0)	$\emptyset$		(0)	$\emptyset$	
V_2	***	**		**	**		n.s.	n.s.	
	(4)	(1)		(4)	(2)		$\emptyset$	$\emptyset$	

Table C.3: Results of significance testing for main effects of *Consonant* on μ_3 across window locations, speaker groups, and positions. For each combination ($Speaker \times Position \times Window$), statistical significance is indicated in the first row (n.s. : $p > 0.05$, * : $p < 0.05$, ** : $p < 0.01$, *** : $p < 0.001$), with the number of contrasts distinguished in multiple comparisons in the second.

	JF			JM			NM		
Window	*CV*	*VCV*	*VC*	*CV*	*VCV*	*VC*	*CV*	*VCV*	*VC*
V_1		n.s.	***		***	**		n.s.	n.s.
		∅	(2)		(2)	(2)		∅	∅
VC		***	n.s.		***	**		***	**
		(2)	∅		(2)	(1)		(2)	(2)
C_i	***	***	**	***	***	***	***	***	***
	(4)	(2)	(1)	(3)	(1)	(2)	(4)	(2)	(2)
C_m	**	***	***	*	***	***	***	***	***
	(3)	(2)	(2)	(2)	(2)	(3)	(4)	(2)	(2)
C_f	n.s.	***	***	**	***	***	*	***	***
	∅	(2)	(2)	(1)	(2)	(2)	(0)	(3)	(2)
CV	***	**		***	*		*	n.s.	
	(4)	(1)		(3)	(1)		(0)	∅	
V_2	***	***		***	**		n.s.	n.s.	
	(3)	(1)		(4)	(2)		∅	∅	

Table C.4: Results of significance testing for main effects of *Consonant* on μ_4 across window locations, speaker groups, and positions. For each combination ($Speaker \times Position \times Window$), statistical significance is indicated in the first row (n.s. : $p > 0.05$, * : $p < 0.05$, ** : $p < 0.01$, *** : $p < 0.001$), with the number of contrasts distinguished in multiple comparisons in the second.

	Parameter	JF			JM			NM		
		CV	*VCV*	*VC*	*CV*	*VCV*	*VC*	*CV*	*VCV*	*VC*
	A_n	***	***	n.s.	***	***	n.s.	***	***	***
		(6)	(2)	∅	(4)	(2)	∅	(3)	(2)	(2)
	A_d	*	n.s.	n.s.	**	*	**	n.s.	n.s.	n.s.
		(0)	∅	∅	(3)	(0)	(1)	∅	∅	∅
VC	$A_{r_{16}}$		n.s.	***		**	*		n.s.	n.s.
			∅	(1)		(2)	(0)		∅	∅
	$A_{r_{17}}$		*	n.s.		**	n.s.		*	n.s.
			(0)	∅		(2)	∅		(0)	∅
	$A_{r_{18}}$		n.s.	n.s.		**	n.s.		*	n.s.
			∅	∅		(2)	∅		(2)	∅
CV	$A_{r_{16}}$	*	***		*	n.s.		n.s.	n.s.	
		(1)	(1)		(0)	∅		∅	∅	
	$A_{r_{17}}$	*	***		n.s.	n.s.		*	n.s.	
		(1)	(2)		∅	∅		(2)	∅	
	$A_{r_{18}}$	n.s.	**		*	**		**	n.s.	
		∅	(1)		(1)	(2)		(3)	∅	

TABLE C.5: Results of significance testing for main effects of *Consonant* on A_n, A_d, and A_r across speaker groups and positions. For each combination (*Speaker* × *Position*), statistical significance is indicated in the first row (n.s. : $p > 0.05$, * : $p < 0.05$, ** : $p < 0.01$, *** : $p < 0.001$), with the number of contrasts distinguished in multiple comparisons in the second.

Bibliography

Badin, P. (1989). Acoustics of voiceless fricatives: production theory and data. *Quarterly Progress and Status Report, STL-QPSR*, 89(3):33–55.

Fant, G. (1960). *Acoustic theory of speech production.* Mouton.

Forrest, K., Weismer, G., Milenkovic, P., and Dougall, R. N. (1988). Statistical analysis of word-initial voiceless obstruents: Preliminary data. *The Journal of the Acoustical Society of America*, 84(1):115–123.

Fox, J. and Weisberg, S. (2011). *An R Companion to Applied Regression.* Sage, Thousand Oaks CA, second edition.

Fukunaga, K. (1990). *Introduction to statistical pattern recognition.* Academic Press.

Goswami, G. C. (1966). *An introduction to Assamese phonology.* Spicer College Press.

Grierson, G. A. (1903). Linguistic survey of india: Vol. v, part i, indo-aryan family. eastern group. specimens of the bengali and assamese languages. *Calcutta: Superintendent of Government Printing, India.*

Hughes, G. W. and Halle, M. (1956). Spectral properties of fricative consonants. *The Journal of the Acoustical Society of America*, 28(2):303–310.

Jassem, W. (1995). The acoustic parameters of polish voiceless fricatives: an analysis of variance. *Phonetica*, 52(3):251–258.

Jongman, A., Wayland, R., and Wong, S. (2000). Acoustic characteristics of English fricatives. *The Journal of the Acoustical Society of America*, 108(3):1252–1263.

Kakati, B. (1941). *Assamese, its formation and development.* Government of Assam.

Kobatake, H. and Ohtani, S. (1987). Spectral transition dynamics of voiceless stop consonants. *The Journal of the Acoustical Society of America*, 81(4):1146–1151.

Konietschke, F. (2012). *nparcomp: Perform multiple comparisons and compute simultaneous confidence intervals for the nonparametric relative contrast effects.* R package version 2.0.

Korkmaz, S. and Goksuluk, D. (2014). *MVN: Multivariate Normality Tests*. R package version 3.3.

Lachenbruch, P. A. (1975). *Discriminant analysis*. Wiley Online Library.

Mahanta, S. (2007). Directionality and locality in vowel harmony. *Unpublished doctoral dissertation. Utrecht Institute of Linguistics.*

Mahanta, S. (2012). Assamese. *Journal of the International Phonetic Association*, 42(02):217–224.

McMurray, B. and Jongman, A. (2011). What information is necessary for speech categorization? harnessing variability in the speech signal by integrating cues computed relative to expectations. *Psychological Review*, 118(2):219.

Milner, B. (1996). Inclusion of temporal information into features for speech recognition. In *Spoken Language, ICSLP 96 Proceedings*, volume 1, pages 256–259. IEEE.

Nearey, T. M. (1997). Speech perception as pattern recognition. *The Journal of the Acoustical Society of America*, 101(6):3241–3254.

Newell, K. and Hancock, P. (1984). Forgotten moments: A note on skewness and kurtosis as influential factors in inferences extrapolated from response distributions. *Journal of Motor Behavior*, 16(3):320–335.

Nossair, Z. B. and Zahorian, S. A. (1991). Dynamic spectral shape features as acoustic correlates for initial stop consonants. *The Journal of the Acoustical Society of America*, 89(6):2978–2991.

Papoulis, A. and Pillai, S. U. (2002). *Probability, random variables, and stochastic processes*. Tata McGraw-Hill Education.

Patgiri, C., Sarma, M., and Sarma, K. K. (2013). Recurrent neural network based approach to recognize assamese fricatives using experimentally derived acoustic-phonetic features. In *Emerging Trends and Applications in Computer Science (ICETACS), 2013 1st International Conference on*, pages 33–37. IEEE.

R Core Team (2013). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Sarma, M. K. (2012). *Aspects of the phonology of Assamese and Assamese English*. Unpublished Ph.D. dissertation, The English and Foreign Languages University, Hyderabad.

Shadle, C. H. (1985). *The acoustics of fricative consonants*. Unpublished Ph.D. dissertation, Massachusetts Institute of Technology, Cambridge.

Shadle, C. H. (1990). *Articulatory-acoustic relationships in fricative consonants*, pages 187–209.

Shadle, C. H. and Mair, S. J. (1996). Quantifying spectral characteristics of fricatives. In *ICSLP 96, Proceedings*, volume 3, pages 1521–1524. IEEE.

Stevens, K. N. (1998). *Acoustic Phonetics*. MIT Press, Cambridge, MA.

Strevens, P. (1960). Spectra of fricative noise in human speech. *Language and Speech*, 3(1):32–49.

Tabain, M. (1998). Non-sibilant fricatives in english: Spectral information above 10 khz. *Phonetica*, 55(3):107–130.

Venables, W. N. and Ripley, B. D. (2002). *Modern Applied Statistics with S*. Springer, New York, fourth edition. ISBN 0-387-95457-0.