REGIONAL VARIETIES OF NEPALI: AN ACOUSTIC ANALYSIS OF VOWELS

Thesis submitted to the Jawaharlal Nehru University
in partial fulfilment of the requirements
for the award of the Degree of

DOCTOR OF PHILOSOPHY

BY
NISHANT LOHAGUN

Under the supervision of
PROF. VAISHNA NARANG

Centre for Linguistics
School of Language, Literature and Culture Studies
Jawaharlal Nehru University
New Delhi – 110067
India
2018

Centre for Linguistics
School of Language, Literature & Culture Studies
Jawaharlal Nehru University
New Delhi-110067, India

Dated: 27/04/2018

CERTIFICATE

This thesis titled "Regional Varieties of Nepali: An Acoustic Analysis of Vowel", submiited by Nishant Lohagun, Centre for Linguistics, School of Language, Literature and Culture Studies, Jawaharlal Nehru University, New Delhi, for the award of the degree of **Doctor of Philosophy**, is an original work and has not been submitted so far in part or in full, for any other degree of diploma of any other University or Institution.

This may be placed before the examiners for evaluation for the award of the degree of Doctor of Philosophy.

27.4.18

(PROF. VAISHNA NARANG)
Supervisor

Prof. Vaishna Narang
Centre for Linguistics
School of Language, Literature & Culture Studies
Jawaharlal Nehru University, New Delhi-110067

27/4/18

(PROF. RIZWANUR REHMAN)
Acting Chairperson

Acting Chairperson
Centre for Linguistics
School of Language, Literature & Culture Studies
Jawaharlal Nehru University, New Delhi-110067

Dated 27/04/2018

DECLARATION BY THE CANDIDATE

This dissertation titled "**Regional Varieties of Nepali: An Acoustic Analysis of Vowels**" submitted by me for the award of the degree of **Doctor of Philosophy**, is an original work and has not been submitted so far in part or in full, for any other degree of diploma of any other University or Institute.

Nishant Lohagun
Centre for Linguistics
School of Language, Literature and Culture Studies
Jawaharlal Nehru University
New Delhi – 110067

Acknowledgement

At the very outset, I would like to express my heartfelt gratitude to my thesis supervisor Prof. Vaishna Narang. This thesis has taken its current shape and form only through her guidance, support and encouragement. Her long-standing experience in the field, expertise in the subject matter and exemplary commitment to teaching and research will always be a source of inspiration. I am indebted to her for moulding my scholarship and shaping my interest in this area of research and supervising this thesis.

Research for this thesis was supported by the Fulbright-Nehru Doctoral Research Fellowship 2015-16. I thank J.W.Fulbright Foreign Scholarship Board, United States Department of State, the Bureau of Cultural Affairs, the United States-India Educational Foundation and the Institute for International Education for administering the grant and for providing me the opportunity to pursue my research at North Carolina State University. I also thank the University Grants Commission, India for supporting my research through the Junior Research Fellowship (2010-2015).

At North Carolina State University, I am grateful to Prof. Erik Thomas for taking me under his wings and supervising my research. I thank him for providing me the exposure to state-of-the-art cutting edge research in sociolinguistics and sociophonetics. The exposure gained from learning from one of the leading figures in the field of sociophonetics will surely help me in my research career. I also express my gratitude to Dr. Robin Dodsworth for introducing me to programming in R and conducting statistical analyses. Similarly, I am thankful to Prof. Jeff Mielke for teaching me how to adapt the P2FA aligner for Nepali speech data. I am grateful to the Office for International Services, staff at the Department of English the library staff for their help and support during my stay at North Carolina State University.

My sincerest thanks also to Dr. Katie Drager at University of Hawaii, Dr. Dan McCloy at University of Washington, Dr. Paul DeDecker at Memorial University and Dr. Bodo Winter, University of Birmingham for their e-mail responses to my queries. I would also like to thank Prof. Madhav Pokharel for allowing me to use the data from his study to generate vowel plots. My sincerest thanks to Dr. Rajesh Khatiwada as well for his suggestions on my research project.

I would like to thank all of the Nepali speakers from the Darjeeling, Nepal, Dooars and Sikkim who participated in the study. Some welcomed me to their homes while some readily agreed to meet in other locations for recording of samples. I would also like to thank Dr. Kabita Lama, Dr. Sangmu Thendup and Dr .Samar Sinha at Sikkim University for helping me recruit speakers from Sikkim and letting me use facilities at their disposal for data collection. Similarly I express my gratitude to Mr. Pradip Lohar and family, residents of Rahimabad T.E in Alipurduar district for facilitating my field work in the Dooars region.

My heartfelt gratitude also extends to all the faculty members and office staff at the Centre for Linguistics, School of Language, Literature and Culture Studies at Jawaharlal Nehru University. I thank my colleagues at the Linguistic Empowerment Cell and Neurocognitive Linguistics Labs at Jawaharlal Nehru University for their constant support and encouragement over the course of writing this dissertation.

I would like to express thanks to my family and friends for their love and being my pillars of strength all throughout.

I apologize if I have inadvertently failed to acknowledge the efforts of anyone who has helped me in conducting this research. Any errors and mistakes are solely my own.

Dedicated to my father.

Table of Contents

List of Figures

List of Tables

This page is intentionally left blank

Chapter 1

Introduction and Review of Literature

1.1 Introduction

Language is a heterogeneous entity and approaches to the study of language, synchronic or diachronic should consider it as one. The structural perspective of language should not be associated with homogeneity and proficiency in variants is not mere performance but a component of linguistic competence. A linguistic model must accommodate language variation conditioned by stylistic and social features for a proper account of linguistic competence. Variation in languages is not random but there is a structure to the heterogeneity. A theory of language change must strive for "describing orderly differentiation in a language serving a community" (Weinreich, Labov, & Herzog, 1968).

This dissertation is an attempt to analyze variation in vocalic phenomena in four regional varieties of Nepali. This study presents a comparative analysis of the spectral and temporal properties of oral monophthongs in four regional varieties of Nepali. Differing from earlier impressionistic and instrumental accounts which are based on a single standard dialect, this study provides 'local' descriptions of vowel quality in terms of steady state formant frequencies, intrinsic pitch and vowel duration to observe the effect of speakers' region of origin on these acoustic measures. Additionally, the acoustic vowel spaces for the four different regions is analyzed to study vowel dispersion patterns in a comparative framework and the vowel space area metric is analyzed as a function of region.

This study draws on quantitative data from a speech corpus created by undertaking fieldwork in the four regions and applying the latest techniques in speech segmentation and quantitative analysis. This chapter will provide a brief overview of this thesis with respect to the rationale and scope, the language and regions in context, research objectives, a review of literature on the topic together with related issues, a concise note on method adopted, a brief description of software and tools and the structure of this thesis.

1.2 Rationale and scope

Research on regional varieties of American English and Dutch indicate that descriptions of a 'general' or 'standard' variety are further enriched when their varieties are also taken into consideration. Extant literature, while acknowledging the presence of different regional dialects of Nepali, provides little insight on Nepali spoken by speakers other than the ones who speak the standard dialect. For proper descriptive accounts of languages, it is essential to understand the variation which exists in languages. It is equally important to document linguistic variation precisely because of the insights it provides on language populations and their movements across boundaries and development of variants. Linguistic variation across regions is very important to understand how languages change and this will have implications for language maintenance and language shift.

The motivation for this study stems from the lack of variation-based accounts of the Nepali vowel system considering the vast geographical stretch and variety of social settings where the Nepali speech community resides. Nepali is spoken in small pockets throughout the Himalayan region in South Asia. It is the national language of Nepal and in India, it is spoken in a vast area stretching throughout the Himalayan belt from the state of Jammu and Kashmir in the north all the way to Mizoram in the north-east. It is also spoken in small pockets in Myanmar and southern Bhutan. Phonological descriptions of Nepali, however, have focused almost entirely on the eastern dialect of Nepali spoken in areas in and around the Kathmandu valley and eastern Nepal. Descriptive (Bandhu, Dahal, Holzhausen, & Hale, 1971; Acharya, 1991) as well as instrumental (Pokharel, 1989; Khatiwada, 2009) analyses of the Nepali speech segments are based on data from talkers speaking the standard variant. A dialect-oriented survey and analysis has been conducted by the Language Division of the Office of the Registrar General of India as part of the Linguistic Survey of India (Office of the Registrar General, India, 2011), hereafter referred to as the LSI. Preliminary drafts based on impressionistic transcriptions indicate variation in vowel inventories in Nepali as spoken in the states of West Bengal (hereafter WB), Sikkim and Himachal Pradesh (hereafter HP) where the survey was conducted. Under-description, especially non-impressionistic, constitutes the major rationale for this study.

Ideally, a proper phonetic description of a language (or its varieties, as in this case) must consider the articulatory, acoustic and auditory aspects of production and perception of the sounds in the language. The same aspects, however, provide motivation for limiting the scope of this proposed study to the domain of acoustics. In contrast to any researcher’s impressionistic, but well-informed, judgments about articulatory aspect of sounds in languages, acoustic phonetics provides an objective, replicable and an empirically verifiable means of providing an accurate description and a comparative analysis. The development of the sound spectrograph at the height of World War II had major implications for linguistic research especially in studies concerning speech production and perception. Spectrographic analysis presented a significant alternative to impressionistic transcription because of its empirical and objective character. Ever since the publication of the results of the Peterson and Barney (1952) study, instrumental techniques have been widely employed by phoneticians and sociolinguists in the study of segmental and prosodic phenomena. The sub-discipline of sociophonetics has emerged on the shoulders of instrumental techniques with a constantly growing body of literature on the interface of phonetics and sociolinguistics. Therefore, this study will attempt to describe the regional variation in Nepali vowels in acoustic terms.

The scope of this study is limited to the vowel sounds of Nepali. More specifically, only oral monophthongs have been included for analysis. Dialects of English are greatly distinguished by their vowels. Acoustic studies of vowels have received scholarly attention for decades resulting in the development of standard procedures of elicitation and analysis which are employed by researchers the world over. Though the primary reason for the exclusion of nasalized vowels and diphthongs is lack of a uniform phonetic context in the corpus, it is to be noted that the corpus was created in field conditions making ambient noise an inevitable. Nasalized vowels are characterized by anti-formants which, at times, carry the propensity to complicate analysis of laboratory recordings. Other factors for delimiting the scope to only vowels and excluding the other major classes of speech sounds - consonants, liquids and glides – are considerations of time and the enormity of scale. Taking 20 participants each from four different regions, the number of tokens for spectrographic analysis is 1440 (80 speakers X 6 vowels X 3 repetitions). Considering the current state of extant literature and availability of language resources, the sample size should suffice in contributing to our understanding of the regional vowel variation in Nepali.

Another point to be noted is that Nepali is spoken throughout the Himalayan belt stretching from Jammu and Kashmir in the north all the way to Mizoram in the northeast. It is also spoken in small pockets in Myanmar and southern Bhutan and Nepal, of course, where it is the national language. In India, it is spoken in Jammu and Kashmir, Himachal Pradesh, Delhi, Uttarakhand, West Bengal, Sikkim, Assam, Manipur, Meghalaya, Nagaland, Arunachal Pradesh and Mizoram. This study involves participants from Nepal, Sikkim and West Bengal in India. The selection of these regions is based on three factors – demography, sociolinguistic setting and practical considerations. The selected regions are inhabited by the largest concentration of Nepali speakers, the figures of which is appended. Moreover, in all the regions varying degrees of multilingualism exist. Lastly, ease of access to participants in these regions was also a factor.

1.3 Language and regions in context

Nepali belongs to the Indo-Aryan branch of the Indo-European language family and is spoken in majority of the areas in Nepal; the Darjeeling district of West Bengal; Sikkim; Assam; Arunachal Pradesh; Bihar; Haryana; Himachal Pradesh; Uttar Pradesh; Uttarakhand; Manipur; Mizoram; Nagaland; Meghalaya, Tripura. Generally considered to have been developed from Sanskrit, the works of Grierson (1916) and Turner (1931, 1966) suggest Nepali may have developed from the Northwestern variety of Prakrit. Like most Indo-Aryan languages spoken in areas along the Himalayan and Aravalli ranges, Nepali is considered to be a member of the Outer group of Indo-Aryan languages (Hoernle,1880).

Khas Kura or language of the Khas people who lived in areas around the Himalayan mountain ranges, Parbatiya, Gorkhali and Pahari- these are various nomenclatures with which Nepali has been associated. Historically, warfare and economic activities have brought the Khas people in constant interaction with other cultures. In due course of time, the Devanagiri script was adopted for Nepali orthography. After the disintegration of the Khas empire, many loans from Persian and Arabic have pervaded into the language of the Khas people.

According to the Ethnologue, an encyclopedic reference guide to the known languages of the world (Lewis, 2009), there is an estimated total of 13,875,700 Nepali speakers across the globe. As per the Census of Nepal (2001), there are 11,100,000 native speakers of Nepali in Nepal. According to the Census of India (2001) there are 2,871,749 speakers of Nepali in India. In Darjeeling and Jalpaiguri districts of West Bengal, there are about 1,400,000 Nepali speakers. It is one of the languages which enjoys constitutional status following its inclusion into the Eighth Schedule of the Indian Constitution in 1992.

J.A.Ayton was the first to present a preliminary account of the grammar of Nepali in 1820 and it is perhaps the earliest known work in Nepali. Numerous studies have followed since such as Turnbull (1888), Dixitacharya (1913), Dahal (1974), Pokharel (1989), Acharya (1991), etc. which have described different aspects related to the language. Turnbull (1888) is the only study which has been conducted on the Darjeeling variety of Nepali. The eastern dialect is considered to be the source for the standardized dialect. However, even within these three broad categories, there are further variations.

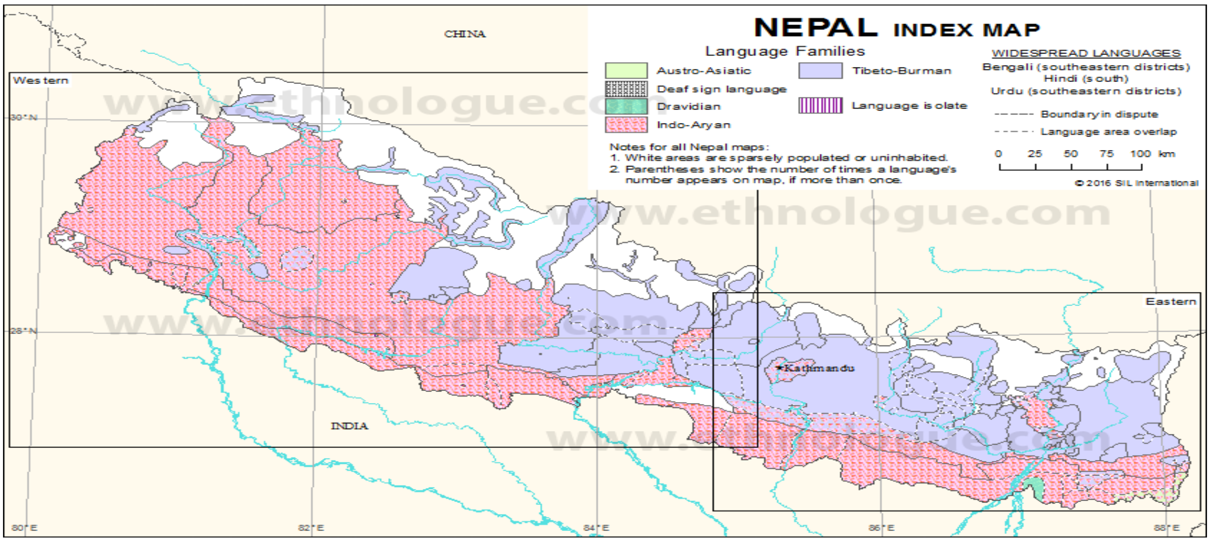


Figure 1. 1 Map of Nepal

(accessed from https://www.ethnologue.com/map/NP_x)

Nepali is the official language in Nepal where it is spoken by 11,100,000 people (Census of Nepal 2001). According to Acharya (1991), there are many social variants of Nepali. Within Nepal, there are three dialects – Western, Central and Eastern, depending upon geographical factors as well as social hierarchy. The eastern dialect is considered to be the source for the standardized dialect. However, even within these three broad categories, there are further variations. Acharya (1991, p6) notes the Darjeeling variety of Nepali to be another distinct variety.

Figure 1. 2 Map of Darjeeling district in West Bengal (accessed from Google Maps)

The Darjeeling district is the northernmost district of West Bengal. Four sub-divisions namely Darjeeling Sadar, Kalimpong, Kurseong and Siliguri constitute the district out of which the first three are located in the hilly regions while Siliguri sub-division falls in the Terai region at the foothills. The district shares two international boundaries with Nepal in the west and Bhutan in the east. The state of Sikkim lies in the north of the Darjeeling district. The southern side is bound by the Jalpaiguri district of West Bengal. The Nepali speaking population in the Darjeeling district has been living in the area for more than two centuries now and has since come in contact with languages like Hindi, Bangla, Tibetan, Lepcha, Santhali, Munda, Oraon, Rajbanshi and several other dialects. Nepali has, however, established itself as the lingua franca of the three sub-divisions namely Darjeeling Sadar, Kurseong and Kalimpong along with certain portion of the Terai region.

Figure 1. 3 Map of Sikkim (accessed from Google Maps)

Sikkim is another state in the laps of the Eastern Himalayas where Nepali is spoken by majority of the population. There are other speech communities such as the Bhutias and the Lepchas but Nepali is widely used across the state in the spheres of education, formal and informal official communication, newspapers and periodicals, legislative deliberations and judicial functions. Unlike in the Darjeeling district, Nepali language in Sikkim is largely in contact with the languages belonging to the Tibeto-Burman family.

Figure 1. 4 Map of the Dooars region in the Alipurduar distict of West Bengal

The Dooars region comprising areas in the districts of Jalpaiguri and Alipurduar in West Bengal. The linguistic landscape of this region is dominated by Bangla along with its dialects such as Rajbansi or Bahe, as it is locally known. The region is however home also to numerous aboriginal communities such as Coches, Meches and Ravas who speak languages belonging to the Tibeto-Burman family. In addition, there are the Oraons (Dravidian) , Mundas and Santhals (Austro-Asiatic) who use their mother tongues within their groups but developed a pidgin which we now know as Sadri or Madesia for communication among themselves. Sadri has now been creolized and serves as a mother-tongue for many. The Nepali speech community completes this diverse linguistic landscape where it is bound to be influenced by other languages from different language families with which it is in contact.

1.4 Research objectives

The objectives of this study are three fold:

1. To provide a descriptive account of vowel dispersion in four regional varieties of Nepali and contrast the vowel acoustic space areas of the regional varieties in question.
2. To observe the effect of region on F1 and F2.
3. To observe the effect of region on intrinsic pitch denoted by fundamental frequency.
4. To observe the effect of region on vowel duration across vowel categories.

1.5 Review of Literature

The goal of this chapter is to trace the development of acoustic phonetics and its application in contemporary linguistic analysis. The chapter begins with a discussion on the historical development of methods in acoustic phonetics followed by its relevance in linguistic theory. The acoustic properties of vowel sounds are discussed next along with descriptions of acoustic correlates of vowel quality such as formants, fundamental frequency and vowel duration. The next section discusses the concept of vowel acoustic space in light of theoretical considerations and its application in various areas such as linguistic description, language pathology, sociolinguistics, sociophonetics and regional dialectology.

1.5.1 Historical development of acoustic phonetics

Acoustic phonetics is concerned with the study of physical properties of speech. Adoption has experimental methods has significantly expanded the scope of phonetics. Ohala (1979) observes how experimental methods have accelerated achievements in science. In the context of the language sciences Ohala believed that phonology stands greatly to benefit by applying the principles of physics. Studies such as Stevens (1972) and Liljencrants and Lindblom (1972) show how the principles of acoustics have been very instrumental in accounting for sound patterns of various languages. Labov (1963, 1966) adopted experimental methods and analyzed acoustic parameters such as formant frequencies to explain sociolinguistic issues such as sound change.

Although initial enquiries had already been made on the nature of speech sounds and their acoustic properties, it was not until the 1940s when major developments in acoustic phonetics began to take place. The invention of the sound spectrograph marked the beginning of a new era in acoustic phonetic research. Applying the electric circuit analogy to the production of speech sounds, Ralph Potter and his colleagues had developed the sound spectrograph at Bell Laboratories. This was a major development considering the fact that spectral properties such as frequency and intensity along with temporal aspects could now be easily observed, something which in the time of Helmholtz's experiments was a daunting task. The spectrograph played an influential role in the development of Pattern Playback by Cooper and his associates at Haskins Laboratories. Pattern Playback was used to synthesize speech using information from spectrographs and it served as major tool for speech perception research.

Chiba and Kajiyama (1942) had also begun their efforts to introduce principles of natural sciences in the study of the language sciences in Japan. The idea that the acoustic properties of vowels is determined by the shape of the vocal tract was already proved in Chiba and Kajiyama (1942) who used X-ray imaging and other technologies to measure the area of the three-dimensional shape of the vocal tract. Their seminal work “The Vowel: Its Nature and Structure” published in 1942 contained the seeds of the acoustic theory of speech production which was later elaborated by Fant (1960).

The relationship between F1 and F2 of vowels and how they affected vowel positioning in the cardinal vowel chart developed by Daniel Jones was demonstrated in the works of Joos (1948) and Delattre (1951). Their observations were consolidated in Peterson and Barney (1952) study which showed how formant patterns varied between speakers on the based on age and gender while remaining remarkably stable for a speaker.

The acoustic theory of speech production (Fant, 1960) exposes the dynamics of the relation between the vocal tract and the resultant acoustic output. The Source-Filter theory of sound production is a framework which can be used to describe the production of vowels in acoustic terms. Like most sounds, vowels sounds have their source in the vibrations created by the vocal folds. The articulators like the tongue body and lips alter the shape of the vocal tract. This altered shape of the vocal tract results in variations in the dimensions of the resonating cavities consequently changing the amplification and reduction of the sound source across a range of frequencies. In the case of vowels, these changes in resonances create a series of peaks of acoustic energy which reflects as formants on a spectrogram. The location of formant peaks in the frequency range does not occur arbitrarily but is determined in an orderly way by the shape of the mouth when forming the vowel. These formant frequency patterns serve as acoustic cues in speech perception and help in distinguishing between different vowels.

1.5.2 Acoustic Phonetics in Linguistic Theory

In the development of linguistic theory, acoustic phonetics was introduced in the work of Jakobson, Fant and Halle (1952). It was the first instance when linguistic theory was supported by acoustic notions. They identified the smallest amount of features required to discriminate between utterances present in languages of the world. Using sound spectrographs, they defined these features in acoustic terms. They identified twelve binary oppositions (pp. 40) and presented their acoustic correlates. The focus was more on acoustic features because of their importance in speech perception.

The publication of The Sound Pattern of English (SPE) by Chomsky and Halle (1968) was the next major development in theoretical linguistics where acoustic notions made further inroads. They proposed a larger set of features consisting of twenty four binary characteristics for contrasting speech segments based on both articulatory and acoustic properties. The SPE was focused more on explaining the sound patterns in one language and addressing core phonological concerns such as the representation of lexical contrasts, constraints in syllable formation and phonological alterations.

K.N.Stevens' (1972) Quantal Theory of Speech attempts to explain why certain sounds are favoured crosslinguistically. Stevens argues that contrasts in a language are due to differences between 'quantal' regions and all quantal areas define contrastive sounds. Using nomograms, he observes that the relation between the articulatory mechanisms and the resultant acoustic output is non linear and demonstrates the acoustic output when a single articulatory parameter such as the place of constriction is taken into consideration. According to Stevens, there are certain regions, which he calls 'quantal' regions, where changes in articulation do not have serious effect on the speech/acoustic output. He states that these stable acoustic regions are areas where two formants meet each other. This leads him to conclude that [a], [i] and [u] are quantal vowels as they exploit quantal regions and they are present in many languages. However the theory has been criticized on the grounds that it does not account for cross-linguistic contrasts. Moreover, Stevens' theory does not provide sufficient account of the distinctions based on multiple articulatory parameters. The Quantal Theory, thus, does provide some leeway for articulatory sloppiness in speech but it is not deemed to be necessary.

1.5.3 Vowel Acoustics

A vowel sound is the combination of different pitches and these overtone pitches give a vowel its distinct quality. The vocal tract acts as an amplifier for different frequencies which are produced every time the vocal cords open and close setting the air in the vocal tract as well as the resonating cavities into vibration. These vocal cord vibrations create harmonics which are high frequency vibrations over and above the fundamental frequency. Depending upon the shape of the mouth, some harmonics are amplified more than the others. Acoustic information about the quality of a vowel can analyzed through its formant structure. Ladefoged (1975) defines formants as "resonances of the vocal tract." A vowel is characterized by three formants – F1, F2 and F3. These formants are the result of the different shapes of the vocal tract and are distinct from the fundamental frequency, f_0 or pitch, which is determined by the rate of vocal cord vibrations per second. The values of the formants can be obtained when a speech sound is analyzed through a spectrogram. In a spectrogram the formants are reflected as dark bands of acoustic energy.

Formant values are affected by the positions of the articulators such as lips, tongue and jaws. According to Ladefoged (1962), there are three factors which affect all the formant frequencies –"the position of the point of maximum constriction in the vocal tract (which is controlled by the backward and forward movement of the tongue); the size or cross-sectional area of the maximum constriction (which is controlled by the movements of the tongue towards and away from the roof of the tongue); and the position of the lips."

There exists an inverse relation between the degree of oral constriction and F1. A high vowel such as /i/ and /u/ is characterized by low F1 values whereas a low vowel such as /a/ is characterized by higher F1 values. F2 values for front vowels such as /i/ are higher while F2 values for back vowels such as /u/ are low. The amount of lip rounding reduces all the formant frequencies as is reflected in the low F1, F2 and F3 values of rounded vowels such as /u/ and /o/ as compared to unrounded vowels like /i/ and /e/ which return higher F1, F2 and F3 values. The characteristic tendencies of the first two formants enable us to distinguish between vowels. The first formant or F1 gives an indication of the open/close dimension as it is observed that open vowels have a high F1 while closed vowels have low F1 values. The second formant or F2 gives an indication of the front/back dimension as front vowels have high F2 while back vowels have low F2.

It must be noted that formant values of vowel sounds are not constant and the reasons for its variability are phonological and physiological in nature. Vowels sounds very rarely occur in isolation. They are in most cases preceded and followed by a consonant sound. The acoustic properties of the vowel sounds are affected by the phonological environment in which they occur. Physiological factors such as age, sex and shape and size of the vocal tract also influence formant values. This can be noticed in the high frequency formants of women and children compared to the low formant frequencies of adult male speakers. However, the acoustic properties of vowels are still crucial in their identification in spite of inconsistencies, as Strange (1999) notes that the acoustic cues in speech perception provided by vowels remains more or less constant.

1.5.4 Pitch

The perceptual or auditory correlate of the acoustic feature fundamental frequency or F0 is known as pitch. Speech and language, to a major extent, depends upon the fundamental frequency of the vocal cords. Fundamental frequency is the number of times the vocal cords open and close per second. Pitch of a voice is dynamic and this dynamicity is reflected in the inter-speaker and intra-speaker variations in pitch. Low range of fundamental frequencies can be noticed in men while children's speech is characterized by high range of fundamental frequencies. The range of fundamental frequencies for women is higher than that of men but lower than that of children placing them in the intermediate zone. The average fundamental frequency for men, women and children are 120 Hz, 225 Hz and 265 Hz respectively. Among several factors which affect the pitch of a voice, the crucial determinant is the tension of the vocal cords. Stretched vocal cords produce high pitch. Humans regularly manipulate the tension in the vocal cords which can be observed in the pitch variations in their speech. Another determinant of pitch is the amount of eggressive pulmonic air. Speakers usually apply extra breath force in speech to give the effect of stress which increases the pitch of the voice.

1.5.5 Vowel Duration

Vowel duration, or vowel length, is one acoustic property of vowels which hasn't received much scholarly attention. It is calculated by measuring the distance between the onset and offset of the vowel but determining them can be difficult at times. Thomas (2011) provides some pointers to ease the problems which might arise in measuring vowel length. In word-initial position one should measure onset after the pause. Duration of vowel after a glottal stop should be measured from the first vocal vibration or when the vocal fold vibration becomes stronger and more regular. If a vowel comes after a stop, the burst can be used as a reference point. If the preceding stop is aspirated, the beginning of voicing should be taken as the onset. For preceding voiceless and voiced fricatives, one should measure from the point where the F2 becomes very clear. Vowels before approximants such as [w], [j] and [ɹ] can be difficult to measure as there is no clear boundary. Laterals, taps and trills are somewhat easier. Vowel duration in such preceding contexts can be studied through steady states and transitions or by splicing the acoustic signal. Contrastive length is a linguistic process which affects vowel duration. Some languages have phonologically long and short vowels. This opposition is usually subsumed within the tense/lax distinction in English. Long vowels, however, may only be about 50% longer than their shorter counterparts but that is enough to signal a perceptual difference.

Lisker (1974) notes that the dependence of the duration of the vowel on two factors: the degree of opening of the vowel, and, the nature of the following consonant. According to Lehiste (1970), “the greater length of low vowels is due to the greater extent to the articulatory movements involved in their production”. This statement can be analysed in the light of F1 values, which is associated with the open/close dimension of vowels, to understand the relation between vowel duration and opening.

Duration of vowels and their differences can be used as cues to identify segment. Lower vowels show longer duration than high vowels. Vowel duration also acts a cue to identify following voiced or voiceless consonants. Duration of vowels is longer before voiced consonants than before voiceless ones. In the prosodic domain, stressed vowels are longer than unstressed ones but this depends upon the language in consideration. However, the rate of speech may have to be normalized to examine contextual vowel length.

1.5.6 Vowel Acoustic Space

Acoustic space is a tool which is employed to show how the formant frequencies help to objectively define the vowel space in a language. It helps us to empirically test the hypotheses of the Quantal theory and the Dispersion theory. The acoustic space tool helps to arrive at phonological distinctness and also to consolidate notions about vowels.

The Dispersion Theory or the Theory of Adaptive Dispersion (Liljencrants and Lindblom 1972, Lindblom 1986) was tries to account for the sound patterns present in a language. Its basic hypothesis is that in a language, the sounds are selected in a manner which serves to provide the element of maximal, or rather, sufficient perceptual contrast. To test the dispersion hypothesis, empirical methods have to be used. The acoustic space model has to be created to verify the notion of contrast. Liljencrants and Lindblom created such a model by plotting F1 values against F2 values. The perceptual contrast between two vowels was measured as the linear distance between them in mel units. This enabled researchers to fairly predict optimal vowel inventories of various sizes (Disner, 1983).

According to Ladefoged (2001), “the acoustic vowel space can be considered to be an area bounded by the possible ranges for the frequencies of the first two formants.” An adherence to a limit or space in the oropharyngeal cavity is observed by Catford (1988) in the production of vowels which defines the acoustic space of vowels in a language. A slight deviance from the space results in the production of an approximant type of sound. This idea of vowel limit or vowels space has influenced the Cardinal Vowel Chart designed by Daniel Jones.

Kewley-Port et. al. (1996) conduct an acoustic analysis of vowels to demonstrate its applicability in language pedagogy, more specifically in L2 learning. Using acoustic data of American English vowels produced by Japanese speakers, the authors show the variations in vowel intelligibility depending on the vowels present in the native language or L1.

Vorperian and Kent (2007) investigate relation between developmental changes in the vocal tract anatomy and its impact on the acoustics of speech. This study uses formant specification to establish a systematic relationship between formant specification and vowel articulation. It is well known that the oral and pharyngeal cavities which constitute the vocal tract of an individual go through several changes in terms of its shape and its size in the course of maturation. Using imaging methods such as magnetic resonance imaging (MRI) and computed tomography (CT) to demonstrate changes in the anatomy of the vocal tract, the authors provide an account of the development of vowel space in males and females in the course of development from infancy to adulthood. The primary data for the study were the formant frequencies and vocal fundamental frequency observed in the vowels of both male and female speakers and speakers of various age groups.

Narang and Misra (2010) conducted an empirical study on the basis of data collected from eight Thai speakers to determine the acoustic space of Bangkok Thai. The authors also attempt to examine the durational contrast and centralization of vowels in the language. Thai is a tonal language with five contrasting tones – mid, low, falling, high and rising. The data collected was analyzed to study F1 correlation with the F0 and the correlation of F3 with F2 and F1. Though the study was conducted on a small database, the results reveals that durational contrast was reflected more in the case of peripheral vowels such as / i, u, a / and gradually decreasing to centralized vowels. The authors note that the use of mid/level tone with all the 18 vowels shows that F0 values are inversely proportional to F1. The acoustic space reflects no remarkable differences between short and long vowels as there is no noteworthy difference in their F1 or F2 values but the two peripheral vowels, /i/ and /a/, which show maximal durational contrast are placed differently in the acoustic space. The authors observe that /a:/ is more open than /a/ and /i:/ is more front and high than /i/. On the correlation of F3 with F2 and F1, the authors observe that F3 is inversely proportional to F1 if the front and the back vowels are examined separately. Lower F3 values were observed for back vowels and higher values were observed for their front counterparts.

1.5.7 Sound Change

The nature and cause of sound change has been studied with different approaches- one which makes a distinction between linguistic and social factors; and another which argues against this separation. The proponents of the first approach believe linguistic factors are responsible for the origin of sound change while social factors are responsible for their spread. Teleological models of sound change which focus on ease of articulation and dispersion (Lindblom et al. 1995) as primary motivations for sound change. Phonetic models of sound change such as Ohala's misperception-based model theorizes sound change as non-teleological. Blevins' model (2004) incorporates ideas from Ohala and Lindblom and provides a more elaborated theory. None of the above mentioned approaches, however, consider, the social factors involved in sound change. They all assume that linguistic factors and social factors are separable with only linguistic factors causing sound change and social factors spreading that change.

Central to teleological approaches to sound change is the notion that language change takes place towards an end or for a purpose. The purpose usually is to ease articulation and processes such as deletion, assimilation and monophthongization of diphthongs lend credence to this argument. But processes such as aspiration and affrication provide counterexamples. Moreover, the notion of ease of articulation motivating sound change loses further ground when articulatory gestures such as frication of approximants or diphthongization of a monophthong are considered. Hyperarticulation is also considered as a cause of sound change with the clarity of meaning being the end rather than the economy of effort. Lindblom et al. (1995) stated that variation occurred because of hyperspeech and hypospeech depending upon the communicative need of the listeners.

The clarity is brought about by maximal dispersion of sounds which contrast in the perceptual space. This notion of maximal dispersion works well for vowel contrasts and it has been successful in predicting vowel configurations (Liljencreants and Lindblom, 1972) and Lindblom (1986). Maximal dispersion also accounts for chain shifts. The weakness of the maximal dispersion hypothesis is evident in phonological mergers as witnessed in situations of language contact or dialect mixture (Harold, 1997). Secondly, sounds such as [f] and [θ] contrast minimally. Thirdly, although secondary articulations such as breathiness, creakiness, nasalization or pharyngealization could be used to contrast vowels in a language, it is rarely seen. Therefore, sufficient contrast, rather than maximal contrast is used for phoneme differentiation.

1.5.8 Regional Dialectology

Application of acoustic methods in sociolinguistics has provided great insights to account for the dialectal variations observed in languages. From Hagiwara (1997) we can assume that languages are constantly changing and vary considerably by geographic location. According to Carver (1998), a dialect is "a variety of language distinguished from other varieties by a set of grammatical, phonetic and lexical features (pp.5). Even though syntactic and lexical properties of a language are considered as the core parameters in the identification of a dialect, researchers have commonly focused on the phonetic differences, especially in the context of dialects of American English (Clopper and Pisoni, 2004a, 2004b, 2006, 2007; Labov,1991 ; Clopper, Pisoni and deJong, 2005).

Clopper and Pisoni (2004b) have examined several acoustic measures to identify factors that might be used to distinguish regional dialects across speakers of American Englsih. They find that features such as fricative voicing and duration, rhotacization, backness, diphthongization are all found to be very significantly depending on the regional dialect of the speaker.

Jacewicz, Fox and Salmons (2007a) also note how the differences in the rate and magnitude of the vowel formant frequency change, vowel duration and a speakers acoustic vowels space account for the regional variations in the context of American English. According to Clopper and Pisoni (2004a), vowels are majorly involved in the perceptual differences among dialects. Jacewicz, Fox and Salmons (2007b) have used vowel quadrilaterals to show significant differences among the Northern, Southern and Midland varieties of American English. They also report that while the exact values of formant frequencies differ among dialects, speakers of different dialects seem to use a similar amount of space when producing vowels.

Labov (1991, 1998) says that dialects are formed by the systematic changes in vowel production. The production of vowel patterns and their usage are greatly influenced by various factors such as acceptance in society, culture, geographical location, educational and economic benefits. Traditionally, the patterns of vowel production which cause differences in dialects can be described in terms of chain shifts, mergers and changes in the acoustic vowel space.

Chain shift occurs when formant frequencies of many values systematically change while each vowel maintains its perceptual distinctness. In phonetics, mergers take place when two vowels combine to form a single phonemic category. Two vowels become indistinguishable, both perceptually and acoustically. Mergers often initiate chain shifts because when two vowels combine, an opening is created in the vowel system which allows a new vowel to take that space.

1.5.9 Nepali Vowel Phoneme Inventory

There are 6 oral vowels and 5 nasalized vowels. Out of the 6 oral vowels, [i] and [e] are front vowels while [u], [o], [a] and [ʌ] are back vowels. Khatiwada (2009) states that all vowels except [o] have a corresponding nasal counterpart which is distinctive in nature. This view has also been shared by Pokharel (1989,p.34). Stress, pitch and length are not phonemic in Nepali unlike nasalization and intonational pitch difference.The fact that the six oral vowels constitute the contrasting vowels can be conclusively established using the minimal pair test which is illustrated in tables given below.

Table 1. 1 Minimal pairs in the context of cVcv.

Vowel	Word	Gloss
/i/	[tsili]	Having bitten by a mosquito or stung by a bee. –i is a perfective marker
/e/	[tseli]	Female disciple, devotee or follower
/ɑ/	[tsɑli]	Having driven a vehicle, walked by foot, rowed a boat; a female with an attitude in colloquial
/ʌ/	[tsʌli]	Continuous playful

		activity; moving; blowing like the wind
/o/	[tsoli]	Blouse worn by Nepali women
/u/	[tsuli]	Filled to the brim usually of a vessel, mountain top.

1.5.10 Vowel Nasalization

Nepali vowels can be nasalized and they are distinctive. Barring back vowel [o], there exists a phonemic contrast between the five oral vowels and their nasalized versions as is illustrated in the table 1.2 below.

Table 1. 2 Phonemic contrasts between oral and nasal vowels

Vowel	Oral	Gloss	Nasalized	Gloss
/i/	[i]	These	[ ]	right here
/e/	[dek e]	(they) saw	[dek]	(I) Saw
/a/	[kati]	having cut	[k ti]	nail
/ʌ/	[gʌu]	Sing	[gʌu]	village, wheat
/u/	[d ua]	Make somebody wash something	[d a]	smoke

1.5.10 Diphthongs

According to Ladefoged (2001), diphthongs are “sounds that have a change in vowel quality during the course of a syllable.” Pokharel (1989: 37–38) identifies ten diphthongs in Nepali. They are [ui], [ei], [oi], [ʌi], [ai],[iu], [eu], [ou], [ʌu] and [au]. In the Darjeeling variety of Nepali examples of all the ten diphthongs, except one [ou], can be observed. The following table lists some examples.

Table 1. 3 Diphthongs

/ui/	/d̪uita/ - ‘two’
	/kuiyeko/ - ‘rotten’
	/kuina/ - ‘elbow’
/ei/	/t̪ei/ - ‘that’
	/kei/ - ‘some’
/oi/	/coita/ - ‘fragment’
	/koila/ - ‘coal’
/ʌi/	/ʌina/ - ‘mirror’ or ‘glass’
	/bʌɪni/ - ‘younger sister’
	/pʌit̪ala/ - ‘sole of the feet’
/ai/	/mait̪a/ - ‘a girls parent’s house’
	/sait̪/ - ‘a special day or time chosen for religious or cultural practices’
	/saĩla/ - ‘third eldest brother or son’
/iu/	/iu/ - ‘stone embedded on a ring’
	/jiud̪o/ - ‘living or alive’
	/ciura/ - ‘flattened rice’
/eu/	/euta/ - ‘one’
	/deuta/ - ‘God’
	/beura/ - ‘behaviour’ or ‘manners’
/ʌu/	/ʌula/ - ‘may come’

	/ʌ̃ula/ - ‘fingers’ or ‘toes’
	/tʌuko/ - ‘head’
	/pʌuɽi/ - ‘swimming’
/au/	/au/ - ‘come’
	/ghau/ - ‘wound’
	/mau/ - ‘mother’ (usually of an animal newborn)
	/cauri/ - wrinkle

Though Pokharel (1989: 65) gives an example of a word with the vowel sequence [ou] as [dhou] meaning “Wash !”, this word is generally not observed in the Darjeeling variety of Nepali. For the same token to have the same meaning, the word d u] or [d uwʌ] is used.

1.5.11 Studies on Nepali

The vowel system of Nepali is traditionally characterized by the presence of 11 contrastive phonemes. It has six oral (/i, e, ɑ, ʌ, o, u/)and five nasal(/ĩ,ẽ,ɑ̃,ʌ̃,ũ /) vowels. Though nasalized [õ] appears sporadically in words such as [õʈʰ '*lips*', hõt͡so '*low*', kʰõt͡s '*remoteorfar-flung*', kʰõʈe '*irregular*'], it bears no phonological contrast with /o/. The central-mid vowel /ə/ has been subject to different interpretations with Bandhu et al.(1971) and Acharya (1991) describing it as a ‘schwa’ whereas instrumental accounts by Pokharel (1989), Khatiwada (2007) and Lohagun (2016) characterize it as a low-mid back rounded vowel represented by a‘wedge’/ʌ/.

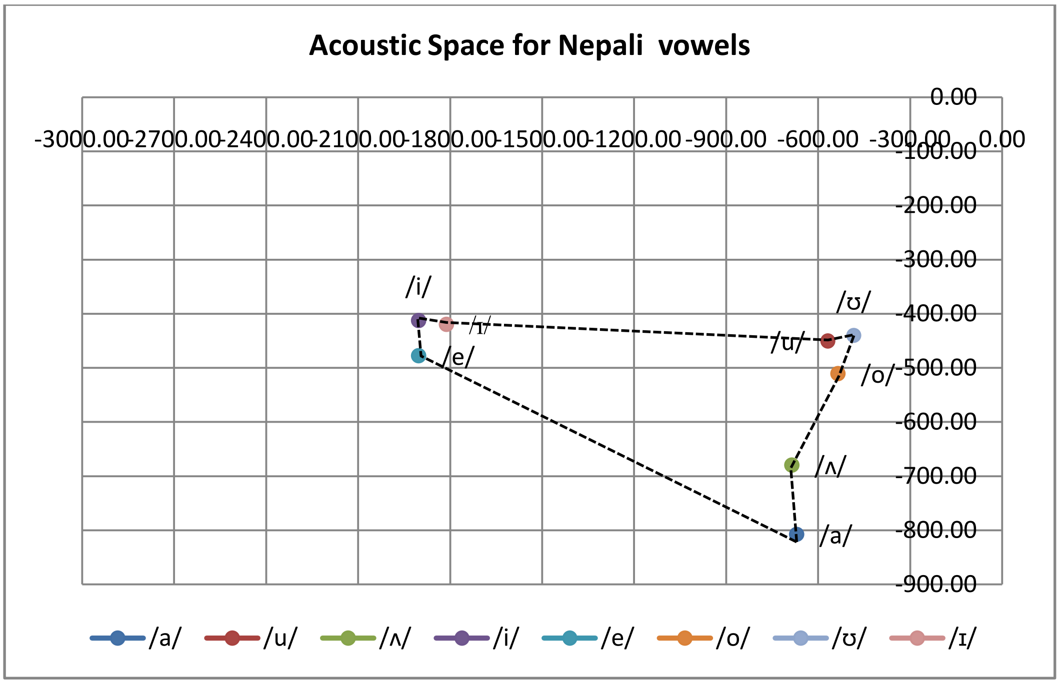


Figure 1. 5 Acoustic space of vowels in the Darjeeling variety (cf.Lohagun 2012)

LSI reports them as /ə/ for the WB, Sikkim and the HP varieties. The LSI notes six oral vowels for WB Nepali but seven for the Sikkim and HP varieties.

	Front	Central	Back
High	*i*		*u*
Mid	*e*	*ʌ*	*o*
Low		*a*	

Figure 1. 6 Nepali Vowel Phonemes in the Darjeeling variety

(figure taken from Srivastava's LSI Report on Nepali in West Bengal, pg.157)

For Sikkim, apart from the six vowels found in traditional descriptions, the mention the presence of /ɔ/ though it could be another variant of /ə/ as pointed out by Khatiwada (2009, p.338).

	Front	Central	Back
High	*i*		*u*
High Mid	*e*		*o*
Low Mid		*A*	*O*
Low		*a*	

Figure 1. 7 Nepali Vowel Phonemes in the Sikkim variety

(figure taken from Nakkeerar's LSI Report on Nepali in Sikkim, p.32)

The LSI report for Nepali in HP, while documenting the regular six vowels including /ə/, also indicates the presence of /ɛ/. This could possibly be an artifact resulting from language contact with Hindi and other local languages spoken in the region. Most Nepali speakers are multilingual and speak the local language(s) of the region.

Vowels	Front	Central	Back
High	*i*		*u*
Mid	*e*	*A*	*o*
Low	*E*	*a*	

Figure 1. 8 Nepali Vowel Phonemes in the Himachal Pradesh variety

(figure taken from Baskaran's LSI Report on Nepali in Sikkim, p.238)

1.5.12 Pokharel's generalizations

Though an impressionistic account of the location of vowels in a cardinal chart had already been provided by Bandhu (1968, 1973), the first attempt to locate the vowels in the vowel quadrilateral in the form of a formant chart can be found in Pokharel (1989). His doctoral thesis entitled Experimental Analysis of Nepali Sound System was a pioneering study on the sound system of Nepali using experimental methods. His study was conducted with data collected from nine informants from various parts of midland Nepal.

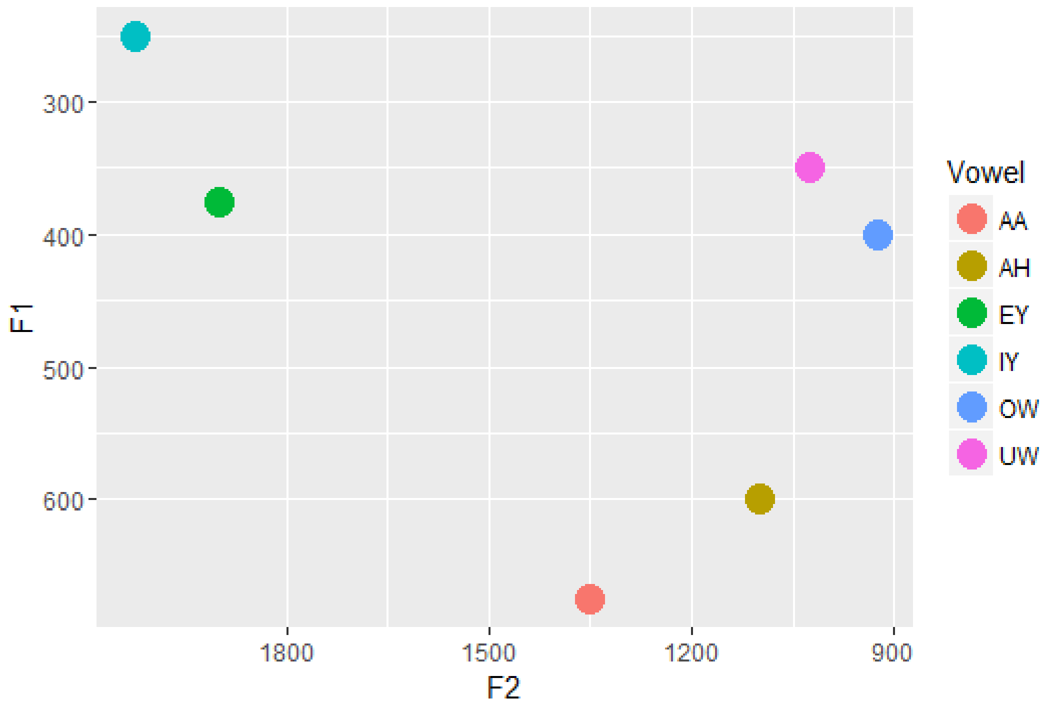


Figure 1. 9 Acoustic vowel space of standard Nepali

(cf. Pokharel 1989)

Pokharel's generalizations on vowels on the basis of formant analysis can be listed as follows:

a. Pokharel states that [i] and [e] are front vowels while [ʌ], [a], [o] and [u] are back vowels. According to him, the vowel [ʌ] is also a back vowel though traditionally it could be considered to be a central vowel phonetically but it is a short variant of the vowel [a] and since [a] us is the backmost vowel in the vowel quadrilateral, [ʌ] is also considered to be a back vowel.

b. Back high vowel [u] is fronter than the round back non high vowel [o]. Similarly[a] which is low back non round low tense vowel in fronter than low back non round low tense vowel [ʌ]. The tense high and low back vowels are fronter than their non tense counter parts on the axis of rounding.

c. Acoustic space formed by vowels of the open syllable is greater than those formed by vowels of the closed syllable.

d. Open syllable vowels are more tensed than closed syllable vowels.

e. The formant chart suggests that rounding of vowels in Nepali is related to the distance of the vowel from the point of origin.

f. Back round vowels like [u] and [o] are lower than the corresponding front vowels such as [i] and [e] in the case of closed syllables.

g. Progression from [i] to [u] reveals than front vowels reflect increasing F1 and decreasing F2 while the back vowels show diminishing F1.

h. Similarly in progression from [i], [e], [a], [ʌ], [o] to [u] the round vowels reflect steadily falling F2.

i. For front vowels the difference between F2 - F1 is greater that the difference between F3- F2. For back vowels the value of F2-F1 is less than the value of F3-F2.

j. The value of F3- F1 signifies quantal vowels. They are reported to be least for the central vowel [ʌ] and it gradually increases as we move to the left or right and is at its highest at the ends. Pokharel notes (p.60) that from [ʌ], slowly decreasing value of F3-F1 denotes progression towards the front vowels while the increasing value marks progressions towards the back vowels.

k. The value of F3-F1 is lowest for a low vowel signifying a direct correlation between the difference of the third and the first formants and the height of the vowel.

l. The front vowels have higher F2 than the back vowels.

This section has summarized the growth and the development of acoustic phonetics, its footing in linguistic theory and highlighted how instrumental techniques have aided linguistic analysis. The inconsistencies in impressionistic analysis as exemplified in the different studies can best be addressed through objective and scientific methods which lies at the core of experimental phonetics. There are no studies till date on the Nepali that is spoken in the Dooars region of West Bengal. Considering the linguistic landscape characterized by the number of languages spoken in the area, it provides fertile ground for linguistic research. The dearth of variationist accounts of Nepali necessitates an empirical investigation, the methods of which are discussed in the following chapter.

1.6 Hypothesis

Based on a review of literature and a pilot study, details of which are described in the next chapter, the hypotheses which are tested with the help of statistical models in chapter can be listed as follows:

1. Region affects F1.
2. Region affects F2.
3. Region affects intrinsic pitch.
4. Region affects vowel duration.
5. Region affects acoustic vowel space area.

1.7 Concise note on method

The method adopted for this study can broadly be divided into two parts – data elicitation and analytical procedures. Data elicitation involves identification of regions and participants, compilation of wordlist, wordlist style structured elicitation and recording. Four geographically contiguous regions with the biggest concentration of Nepali speakers are present were identified on the basis of census data. A wordlist was compiled containing bisyllabic words with target vowels in stressed positions. Significant attention was paid to ensure that the target vowels were enveloped in a neutral consonantal phonetic context. A wordlist style structured elicitation approach was adopted as citation-form speech produces heavily stressed and longer tokens and they are more likely to reach their phonetic targets. All interviews were recorded in the Waveform Audio File (.wav) format with 16-bit quantization and sampling frequency of 44.1 kHz using a Zoom H1 Handy Recorder. Every word was repeated thrice. The recordings were made in field settings and every effort was made to mitigate background noise.

Analytical procedures involve token segmentation, determination of measurement points, measurements, vowel normalization, corpus creation, modelling variation, plotting and statistical analysis. Automatic segmentation of tokens was carried out using the Montreal Forced Aligner. The alignments were hand-checked and manually corrected for errors. A Praat script was used to extract vowel duration and formant measurements 50% through the course of the vowel. Vowel normalization was carried out in order to eliminate variation in measurements caused due to physiological differences in the vocal tracts of males and females. Another motivation behind carrying out normalization was to preserve dialectal differences in vowel quality (Thomas, 2011: 161). For this study, formant measurements (F1 and F2) were normalized using the Lobanov normalization algorithm through NORM (Thomas & Kendall, NORM: The vowel normalization and plotting suite, 2007). A Praat script was used to extract vowel duration and formant measurements 50% through the course of the vowel. Normalized formant frequencies (F1 and F2) were then plotted to visualize the acoustic space in the statistical program R.

Linear mixed effects models were used for statistical analysis using R. In order to resolves non-independencies of data because of repetitions of the same token for a vowel category by every speaker it was essential that a mixed effect approach was adopted which accounts for the variability arising out of both the fixed and random effects. Therefore, region, phonetic environment and repetition were kept as fixed effects while speaker was retained in the model as a random effect. For each acoustic measure (F1, F2 and vowel duration), three models were built: a full model with all the fixed and random effects; and two reduced models without one of the fixed effects. The full model and the reduced model was then compared using a maximum likelihood ratio test to gauge the significance of the fixed effects on the model.

1.8 Tools and software

Various open-source tools such as Praat, Montreal Forced Aligner and R have been used in this study. All of them are available for free and have a vibrant support community on the internet. Despite being freely available and relatively easy to install with sufficient documentation and support on forums, they are incredibly powerful in their functions and are widely used by researchers across the globe.

Praat is free software which is used widely by linguists and researchers for the analysis of speech. Designed by Paul Boersma and David Weenink of the University of Amsterdam, this software offers a great deal of operational flexibility as it can run on a wide range of operating systems such as Microsoft Windows (95, 98, NT, ME, 2000, XP, Vista, Windows 7), different Unix versions, Linux and Mac. It has an in-built sound recorder but also provides the functionality of using sounds recorded by other means. Praat enables users to save the recorded files in different formats such as .wav, .flac, .aifc, .aiff, etc. It can record and analyze mono and stereo signals. It facilitates spectrographic analysis of sound waves and indicates aspects such as intonation, intensity, formants, pitch contour, amplitude and other details. Praat can support speech synthesis as well as articulatory synthesis. Praat was used for spectrographic analysis and script was employed for measurement of tokens.

The Montreal Forced Aligner (McAuliffe, Socolof, Mihuc and Wagner ,2016) was used for automatic segmentation of speech data. Forced alignment requires an audio file, its orthographic transcription and a pronunciation dictionary to provide time-aligned textgrids. Unlike other forced alignment tools like the Penn Phonetics Forced Aligner (P2FA) (Yuan and Liberman, 2008), FAVE-align (Rosenfelder et al. 2014) and Prosodylab-aligner (Gorman, Howell and Wagner, 2011) which are based on the HTK toolkit, the MFA is based on Kaldi to do the alignment. Like Prosodylab, the MFA can be used to train acoustic models thereby facilitating development of language resources.

R (R Core Team, 2013) is a software environment used for statistical computing. Compared to similar programs, it is freely available under the GNU General Public Licence. In linguistics, other than hypothesis testing, R can be used for pattern discovery and building statistical models. In this dissertation, R has been used for statistical modeling, hypothesis testing, data organization and data visualization.

1.9 Structure of thesis

Chapter 1 positions the current study with reference to the current state of research on the topic. It puts focus on oral vowels as the subject of the investigation and provides justifications for delimiting the overall scope. It also places the language in the context of this study and provides a description of the regions where the Nepali speech community can be located. It highlights the objectives of this study, specifies the acoustic parameters analyzed and briefly outlines method which is adopted. Finally, the chapter ends with descriptions of tools and software used in this research. The section on literature review provides an account of the extant literature on core aspects related to this study.

Chapter 2 highlights the methodological issues and theories associated with them. It will elaborate on the data elicitation tools, techniques and analytical procedures.

Chapter 3 will focus on the tabulation of data after spectrographic analysis. The core acoustic parameters is analyzed for the Nepal, Darjeeling, Sikkim and Dooars varieties and they will be discussed in detail.

Chapter 4 – Summary and Conclusion: The final chapter summarizes the results obtained after analysis in the preceding chapters and presents an overall account of the regional variation in Nepali. Contrasts will be drawn between the acoustic spaces of different regions. Acoustic factors which characterize a regional variant will be clearly outlined. The chapter will conclude with difficulties faced at any phase, if any, while conducting the study and the steps taken to handle them. In the end, an overview of what could possibly done in the future will be presented.

This concludes this chapter on introduction and review of literature. The next chapter elaborates on the method adopted for the current study.

CHAPTER 2

METHOD

2.1 Introduction

This goal of this chapter is to situate the topic under investigation against macro-level frameworks dominating the current discourse in the study of dialectal variation with reference to vocalic phenomena. In addition, the chapter describes the micro-level techniques employed in data collection and analysis. The chapter begins with an overview of studies which shows how methods in acoustic phonetics are increasingly adopted in studies of dialectal variation. Direct data collection and issues related to it is discussed next along with tools and strategies employed for data elicitation and to address ethical considerations. The next section provides a description of the development of a Nepali forced-alignment tool for automatic segmentation of speech which is a pioneering effort. Subsequently, strategies adopted in measurement of tokens, data base design and analytical procedures are discussed in the sections that follow. The chapter concludes with the description and results of a pilot study and the outlines the final procedure adopted for this study.

The fact that acoustic data could be reliably used for studying dialectal variation was established by Labov, Yaeger and Steiner (1972). Since then acoustic phonetic methods have found increasing use in both production and perception studies involving dialectal vowel variation. English vowels, especially, has received a lot of attention from researchers studying dialectal variation using instrumental methods. Studies such as Labov et al.(1972), Hagiwara (1997), Thomas (2001,2010), Fridland (2000), Clopper, Pisoni and de Jong (2005), Clopper and Pisoni (2006) and Jacewicz, Fox and Salmons (2007, 2011) have all employed methods in acoustic phonetics for a comparative analysis of speech across geographical regions. Apart from English, dialectal variation in languages such as Dutch, Spanish, Albanian and Swedish has also been analysed using spectrographic methods. Vocalic variation in Dutch spoken in the Netherlands and the Flanders region has been studied using formant values by Adank, van Hout and van de Velde (2004) while Spanish spoken in south west United States is compared to peninsular Spanish and other regional varieties such as Mexican Spanish on the basis of formant frequencies by Willis (2007). Standard and regional dialects of Albanian have been described by Moosmuller and Granser (2006). Similarly the SweDia project (Erikkson 2004; Leinonen 2012) acoustically examined Swedish spoken in communities across Sweden and selected regions in Finland where a variety of Swedish is spoken. The increasing rate of adoption of acoustic phonetic methods in the field of dialectology highlights its efficacy in the analysis of the phonetic variation found in the speech community and observing correlations with demographic variables.

Although consonants are primary studied in articulatory terms, acoustic methods are gaining ground in studies involving consonantal variation. In England, Docherty and Foulkes (1999) have compared /t/ between different phonological contexts and different regions by spectrographically analysing glottal pulses. Aspects of consonants such as preaspiration in dialects of Swedish (Tronnier 2002; Wretling et al. 2002); voice contrast in Wisconsin English (Purnell et al., 2005 a,b); VOT (Docherty el al, 2011; Syrdal 1996; Takada and Tomimori, 2006); "light" and "dark" laterals in Catalan (Recasens, 2004; Recasens and Espinosa, 2005); rhotics in dialects of Spanish(Willis 2006; Bradley and Willis, 2012), and assibilation of /r/ and /j/ in varieties of Argentinian Spanish (Colantoni, 2006) have all been studied using acoustic techniques.

The growing popularity of acoustic phonetic methods in studies of spatial variation can be attributed to its advantage over auditory coding methods which have been traditionally employed. Acoustic phonetic methods enable a scientific investigation with greater precision and enables replication of results.

2.2 Identification of regions and participants

2.2.1 Regions

Pokharel (2009) citing the work of Timilsina (2050 BS), Subedi (2051 BS), Acharya (2053 BS) and Dhungana (2054 BS) points out that although there are 12 or 13 dialects of Nepali, the eastern dialect has been standardised and used in contemporary media, literature, academia and administration. The eastern dialect is spoken from the regions around the Bheri river in Nepal to Dibrugarh in Assam in north-eastern India (ibid, pp.330). The four regions chosen for the study namely Nepal, Darjeeling, Dooars and Sikkim are geographically contiguous and form a part of the eastern dialect continuum.

The standard variety spoken in Nepal was chosen as it serves as a point of reference while comparing other regional varieties and the analysis is aided by the availability of impressionistic as well as instrumental accounts. Pokharel (1989) and Khatiwada (2009) are the two instrumental accounts for standard Nepali. The following is a plot of the Nepali acoustic vowels space reproduced from Pokharel's data.

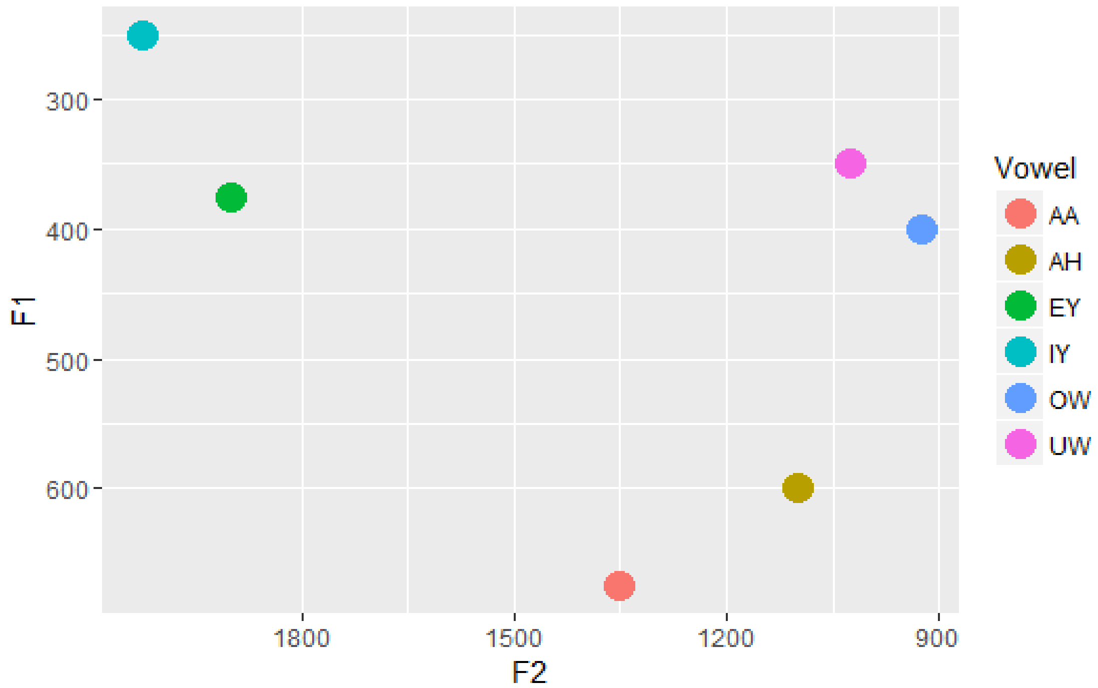


Figure 2. 1 Acoustic vowel space of standard Nepali (cf. Pokharel 1989)

Pokharel identifies 6 distinct oral vowels in the standard variety of Nepali and the same has also been mentioned by Khatiwada (2009). Khatiwada also proposed [ʌ] as the basic norm with other variants such as [ɔ, ɘ] realized according to differing speakers and consonantal contexts.

The second region, Darjeeling, was chosen as a field site primarily because it is home to the largest concentration of Nepali speakers outside of Nepal. The speakers in this region are mostly multilingual. Other Indo-Aryan languages such as Hindi and Bangla are spoken in the region. The only instrumental account of the vowels in the Darjeeling variety has been provided by Lohagun (2016) which is primarily a descriptive account of vowel quality in terms of acoustic measures such as fundamental frequency, formants, vowel duration and acoustic space of vowels. However, a comparative analysis between the standard and the regional variant was not presented. This study differs from the Lohagun(2016) with respect to its analytical procedures such as automatic segmentation through forced alignment, script-based measurements and statistical testing.

The third region, Sikkim was also chosen because of lack of empirical description of the Nepali spoken in the region. Nakeerar's (2011) report for the LSI is the only phonological description of Nepali spoken in the region which mentions another vowel /ɔ/ along with the 6 widely attested in literature. Along with Nepali, Tibeto Burman languages such as Bhutia and

Lepcha are spoken in the region. Preliminary field visits also indicated a lesser degree of multilingualism as compared to the Darjeeling district. Therefore, this study would be the first to provide an empirical and comparative account of Nepali vowels in the Sikkim variety, the results of which may serve as a background for further studies on language contact, language variation and change.

Similarly, the fourth region Dooars was chosen as a field site because of the lack of any linguistic description of the variety of Nepali spoken there. The society is a composite mix of people of different ethnic stocks such as the Adivasis, Bengalis, Biharis and Nepali. The socio-economic status of the inhabitants is relatively poor compared to other regions but the levels of multilingualism is high in the region. Sadri (popularly known as Adivasi), Hindi, Bangla and Nepali are used on a daily basis in different domains. Field interviews were conducted in Rahimabad Tea Estate in Kumargram block of Alipurduar district.

The four regions thus chosen for this study provide an opportunity to enrich linguistic description of Nepali vowels and provide novel descriptions for varieties which have not been studied so far. In all the four regions there are varying degrees of multilingualism and the language is in contact with different languages from different language families.

2.2.2 Participants

The sample size for this study is relatively larger than studies conducted so far (N = 78; 46 males and 32 females). Pokharel (1989) interviewed 9 males for his study while Khatiwada's (2009) study is based on data from 5 male speakers. The corpus for this study is based on data from male and female speakers from the four regions. A total of 8 female speakers from each of the four regions were interviewed. 12 male speakers each from Darjeeling, Dooars and Sikkim participated in the study while 10 male speakers from Nepal were interviewed. The participants in each of the regions fall in the age bracket of 18-55. The participants from Nepal, Sikkim and Darjeeling are mostly university students whereas the participants from Dooars are mostly engaged in primary education and agriculture.

The most important criteria for subject inclusion is literacy. Since a wordlist in the Devanagari script was administered as a tool of data elicitation it was necessary that the participants had attained some level of education. The second criterion for inclusion is related to the quality of recordings. The interviews were collected in field settings where despite all efforts to mitigate ambient noise, some recordings were unusable for acoustic analysis and treated as outliers in the data.

2.3 Ethical considerations

To address ethical issues arising out of research involving human participants, informed consent was taken from all the participants who were interviewed for the study. The proforma for informed consent form is appended. The goals of the research along with their involvement with respect to time commitments, activity such as reading out of a wordlist and potential risks and benefits of the study were explained to the participants. The participants were all consenting adults competent to take autonomous decisions. It was specified that their participation was voluntary and they were made aware that they had the right to discontinue whenever they choose. Recording of speech samples was done with prior permission and no inducement, financial or otherwise, was provided. The participants were assured of confidentiality and adequate steps were taken to code participant information for analysis.

2.4 Data Elicitation

The method adopted in this study is significantly informed by methods in Labovian variationist sociolinguistics in so far as they are grounded in bottom-up approaches in order to formulate a hypothesis that regional background affects vowel quality. A bottom-up empirical approach is also preferred by experimental phoneticians but unlike sociolinguists who prefer naturalness in speech samples by placing more emphasis on the collection of conversational data, experimental phoneticians prefer citation-form speech samples usually elicited through thoughtfully designed wordlists or target segments in carrier phrases. Citation-form speech has been elicited using a wordlist for this study because it offers two main advantages over casual conversational speech. Although at the expense of naturalness, wordlist style elicitations produce heavily stressed and longer tokens which approximate their phonetic targets (Veatch, 1991). A second advantage is that it allows for replicability and validity of results by allowing to control for factors such as phonetic environment.

A concern with citation-form wordlist style elicitation is coarticulation effects of neighbouring segments especially on formant values. Coarticulation influences both spectral and durational properties of vowels (Lehiste, 1970). The effects can be strong enough to alter formant values for the same vowel. Therefore, it was necessary to control for phonetic environment while preparing a wordlist for this study. A neutral phonetic context such as preceding /h/ which requires the least amount of supralaryngeal activity is considered best and is used in many studies of vowels in American English. Taking this approach, however, would have resulted in more nonsense words in the wordlist. To circumvent this issue, the next best preceding context to minimize coarticulatory effects, preceding alveolar and bilabial voiceless stop consonants was chosen while compiling the wordlist.

Nepali has 6 oral vowels which occurs in word initial, medial and final positions as shown in table 2.1 below.

Table 2. 1 Occurrence of vowel phonemes in word positions

Vowel	Word Initial	Gloss	Word Medial	Gloss	Word Final	Gloss
/i/	[it̪ru]	small	[pip]	Pus	[keti]	Girl
/e/	[euta]	one	[tek]	Step or stamp on something-IMP	[d̪ekʰe]	saw (past)
/a/	[ayo]	came	[kat]	Cut -IMP	[pʌkka]	surely, firm
/ʌ/	[ʌrko]	another	[pʌt]	Onomatopoeic root for a manner of breaking	[sukʰʌ]	Prosperity
/o/	[oɖal]	cave	[tok]	Bite! -IMP	[oralo]	downhill
/u/	[umkyo]	escaped	[kut]	Hit! Beat! -IMP	[goru]	ox

Although vowel length does not lead to any contrast, there is contextual shortening of vowels especially before geminates in words such as [kʊkkʊr ~ *dog*, pʌtti ~ *bandage*, tʰɪkkʌ ~ *exact*]. Since the short vowels aren't phonemic as has been attested in Pokharel (1989) and Khatiwada (2009), they have not been included for analysis. The final wordlist which was used as an elicitation tool consisted of bisyllabic words with the target vowel between alveolar and voiceless stop consonants as is shown in table 2.2 below.

Table 2. 2 Wordlist used for data elicitation

Vowel	Word	IPA	Gloss
/i/	टिपी	[tipi]	Pick - IMP
/e/	पेटी	[peti]	belt

/a/	टापा	[tapa]	Clothes hanger
/ʌ/	पटि	[pʌti]	towards
/u/	टुटी	[tuti]	Break – IMP

2.5 Recording and Repetitions

Data for this study was collected through field interviews in each of the four regions. In Rahimabad, help from a local member of the community was solicited to reach out to participants and also to minimize the interviewer effect. Contrary to studies in experimental phonetics which relies on laboratory recordings, a facility as such was not available in all the regions. In Rahimabad for instance, it appeared that visitors were usually hosted at the porch. Although recording was done in field settings - usually at the participant's residence, vacant classrooms and student housing rooms, every effort was made to mitigate background noise in order to ensure that the recordings could be used for acoustic analysis. . All interviews were recorded in the Waveform Audio File (.wav) format with 16-bit quantization and sampling frequency of 44.1 kHz using a Zoom H1 Handy Recorder. A structured elicitation approach was adopted and the participants were asked to read out from a wordlist. Every word was repeated thrice by all the participants. This allowed the inclusion of more tokens for analysis.

2.6 Forced Alignment

An acoustic examination of the spectral and durational properties of vowels requires that the phonetic segments are annotated by time-aligning the words and marking phoneme boundaries. Schiel and Draxler (2003) estimate that it takes 800 times the real time if segmentation is done manually. Advances in automatic speech recognition (ASR) technology have been incorporated by linguists as a methodological tool and toolkits such as the Hidden Markov Models Toolkit (HTK) and Kaldi are increasingly employed to develop tools such as force aligners. Forced alignment automates the process of phonetic segmentation which significantly reduces the time taken to conduct an acoustic analysis. Labov et al.(2013) observed that using a forced aligner enabled them to increase their token for analysis by a whopping 2900 percent from each interview. Forced alignment tools have facilitated the analysis of large speech corpora. It also enables consistency in segmentation since there hardly ever is any inter-annotator agreement when labelling in done manually (Chuchiarini, 1993).

According to McAuliffe et al. (2016), "forced alignment is a technique to take an orthographic transcription of an audio file and generate a time-aligned version using a pronunciation dictionary to look up phones for words." Boundaries between phonetic segments are determined on the basis of acoustic models using a computer algorithm. Acoustic models are usually built on the Hidden Markov Model (HMM) platform. HMMs takes a sequence of acoustic vectors and assigns it a label by modelling each label as sequence of "hidden" states. Training of acoustic models takes place through cepstral coefficients (MFCCs) and perceptual linear predictive (PLP) components. In brief summation, forced aligners operate by matching orthographic transcriptions to items in the dictionary to create a sequence of phones and using HMMs of phones along with the acoustics to determine segment boundaries.

There are several forced alignment tools such as the Penn Phonetics Forced Aligner or the P2FA (Yuan and Liberman, 2008), Prosodylab – aligner (Gorman, Howell and Wagner, 2011), Munich Automatic Segmentation System or MAUS (Schiel 1999, 2004) and Montreal Forced Aligner or MFA (McAuliffe, Socolof, Milhuc and Wagner, 2016). With the exception of MFA, all the above mentioned tools are based on the HTK. While aligners such as the P2FA and its adaptations such as the Forced Alignment and Vowel Extraction, commonly known as FAVE (Rosenfelder, Fruehwald, Evanini and Yuan, 2011) are language specific and work solely for English, MAUS supports European languages such as German, Portuguese, Icelandic, Italian, Estonian, Hungarian, Spanish and Dutch in addition to varieties of English. Prosodylab – aligner can be used for any language as it allows for training of acoustic models provided that sufficient audio data (at least one hour), word-level transcriptions and a pronunciation dictionary is present. Similarly the MFA can also be used on arbitrary data to train acoustic models but it uses the Kaldi ASR toolkit for alignment. Although, HTK also supports training of acoustic models, Kaldi offers ease of access in local installation without compromising its accuracy and efficiency.

For the purposes of this study, the MFA based on the Kaldi toolkit was used for phonetic segmentation. Currently there is no forced alignment system for Nepali. Development of such a tool carries the potential to analyze large speech corpora. While it is beyond the scope of this study and it might as well be incorrect to call it a forced aligner for

Nepali as the dictionary only contained words in the wordlist, a preliminary attempt at developing a forced alignment for Nepali (FANE) was made. An orthographic transcript was created first for the audio files for all the speakers. A TextGrid (as in figure 2.2) format was used for saving the transcripts which specified orthographic transcriptions for short (below 30 seconds) intervals of speech.

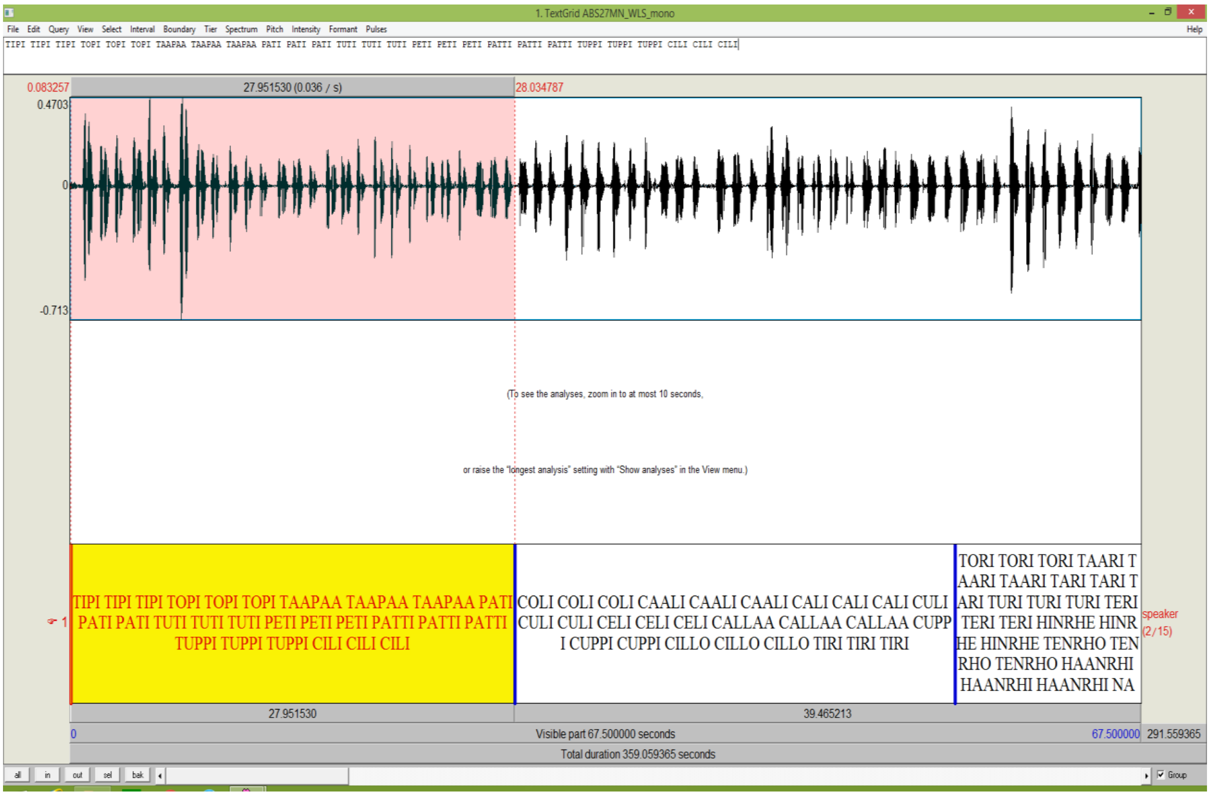


Figure 2. 2 Annotation textgrid for training acoustic models for forced alignment

A pronunciation dictionary containing all the words in the wordlist was then compiled. The scheme followed in the CMU Pronouncing dictionary of mapping words to their pronunciations coded in ARPAbet phoneset was retained while preparing the dictionary as shown in figure 2.3.

nepali.dict - Notepad

File Edit Format View Help

```
TIPI T IY1 P IY1
TOPI T OW1 P IY1
TUPI T UW1 P IY1
TAAPAA T AA1 P AA1
PATI P AH1 T IY1
TUTI T UW1 T IY1
PETI P EY1 T IY1
PATTI P AH1 T T IY1
TUPPI T UH1 P P IY1
CILI CH IY1 L IY1
COLI CH OW1 L IY1
CAALI CH AA1 L IY1
CALI CH AH1 L IY1
CULI CH UW1 L IY1
CELI CH EY1 L IY1
CALLAA CH AH1 L L AA1
CUPPI CH UH1 P P IY1
CILLO CH IH1 L L OW1
TIRI T IY1 R IY1
TORI T OW1 R IY1
TAARI T AA1 R IY1
TARI T AH1 R IY1
TURI T UW1 R IY1
TERI T EY1 R IY1
```

Figure 2. 3 Sample of the pronunciation dictionary

The 39 phonemes in the phoneset sufficiently encompass the monophthong inventory of Nepali. An acoustic model for Nepali was trained using approximately 6 hours of speech data from 78 speakers. Figure 2.4 shows the duration of speech sample for every participant used for training the models.

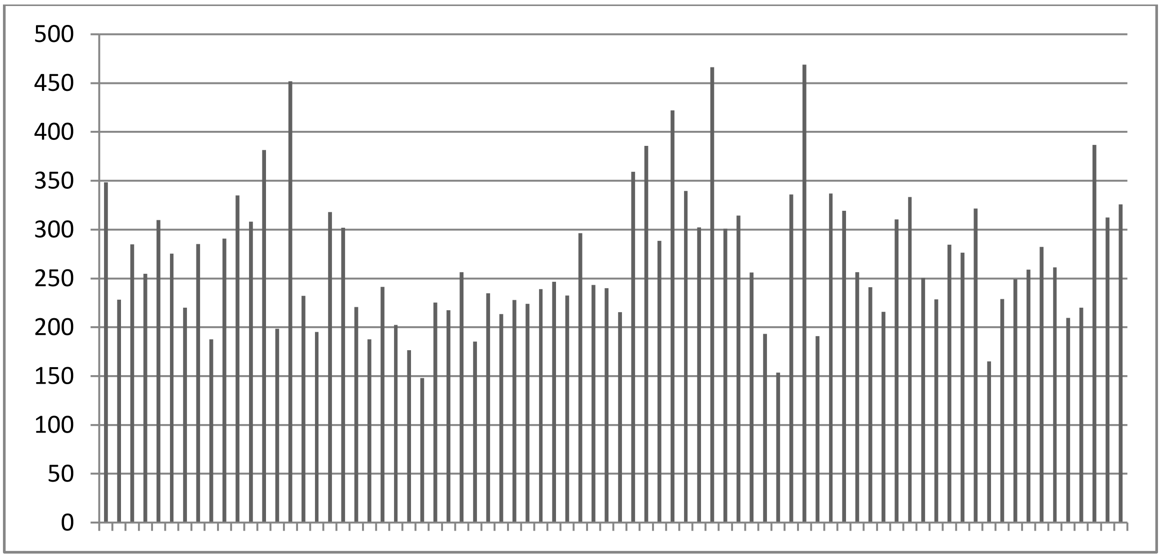


Figure 2. 4 Duration of speech sample for every participant used for training the models

With the acoustic models for Nepali in place, forced alignment was done on all the words in the wordlist. The alignments from the resultant output as shown in figure 2.7 .were then manually inspected for any errors in boundary placement.

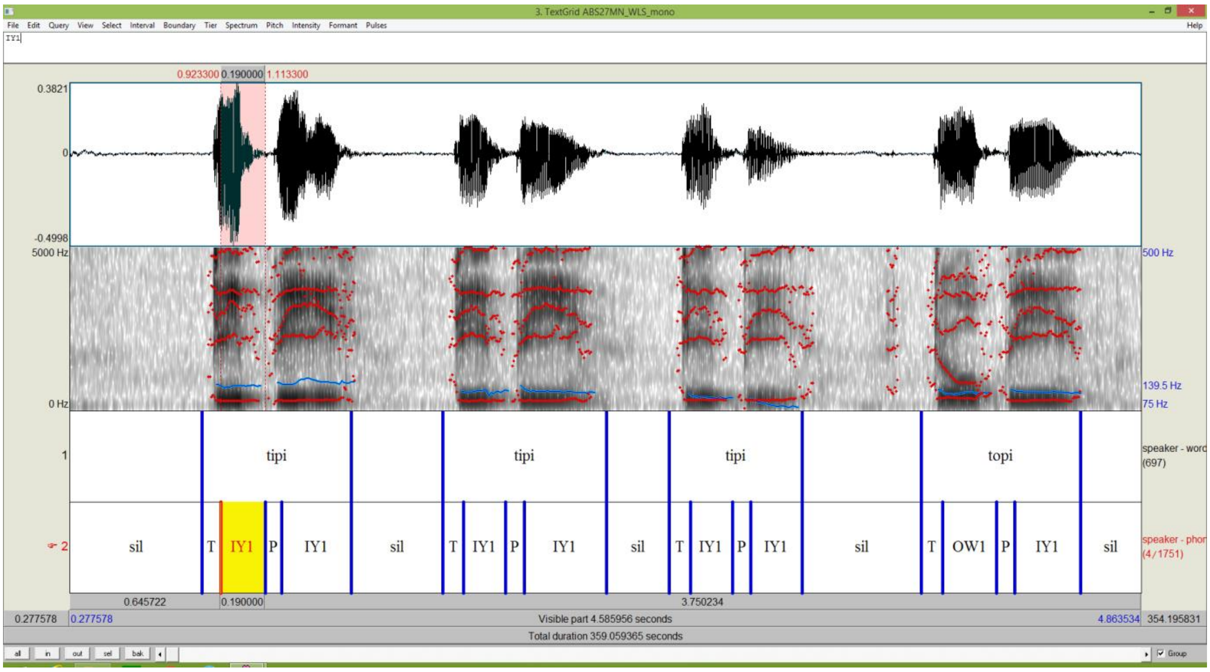


Figure 2. 5 Sample force aligned tier on a spectrograph

2.7 Measurement points

In the discussion on the scope of this study in the earlier section 1.2, tensed monophthongs in Nepali was chosen as the object of analysis for mostly practical reasons. Monophthongs are characterised by a single steady state pattern signifying canonical formant values for the vowels. Steady states appear as regions of stability where the difference between formants values across time points is almost negligible. Formants can be measured at different time points and each approach has its own rationale, the choice of which depends upon the objectives of the study. Two common approaches are the single point of measurement approach which is based on the notion of the vowel "target" and taking measurements at multiple points which accounts for vowel-inherent spectral change. In single-point measurement approaches, measurements are taken at either the midpoint – the absolute centre of the vowel, or at the point of maximal displacement where one or more formants change direction. The advantage of the single-point measurement approach is that the measurements are taken at a point where they show the least coarticulation with the neighbouring segments. This approach is suitable for monophthongs but for diphthongs taking measurements at multiple points is best as they have multiple steady states corresponding to the nucleus and the glide. For diphthongs, a trajectory analysis of formants is preferred and measurements are usually made at specified distances from the onset or offset; at intervals of specified distances such as 10 ms intervals; and taking measurements at fractions or percentages of the distance through the course of the vowels.

Vowel duration, or vowel length, is one acoustic property of vowels which is calculated by measuring the distance between the onset and offset of the vowel but determining them can be difficult at times. Thomas (2011) provides some pointers to ease the problems which might arise in measuring vowel length. In word-initial position one should measure onset after the pause. Duration of vowel after a glottal stop should be measured from the first vocal vibration or when the vocal fold vibration becomes stronger and more regular. If a vowel comes after a stop, the burst can be used as a reference point. If the preceding stop is aspirated, the beginning of voicing should be taken as the onset. For preceding voiceless and voiced fricatives, one should measure from the point where the F2 becomes very clear. Vowels before approximants such as [w], [j] and [ɹ] can be difficult to measure as there is no clear boundary. Laterals, taps and trills are somewhat easier. Vowel duration in such preceding contexts can be studied through steady states and transitions or by splicing the acoustic signal. In the current study, the target vowels are all preceded and followed by voiceless stop consonants (see table 2.2). Therefore, the onset of the vowel is the point immediately following the stop burst while the offset is the point where the vocal pulsing fades (see fig. 2.6).

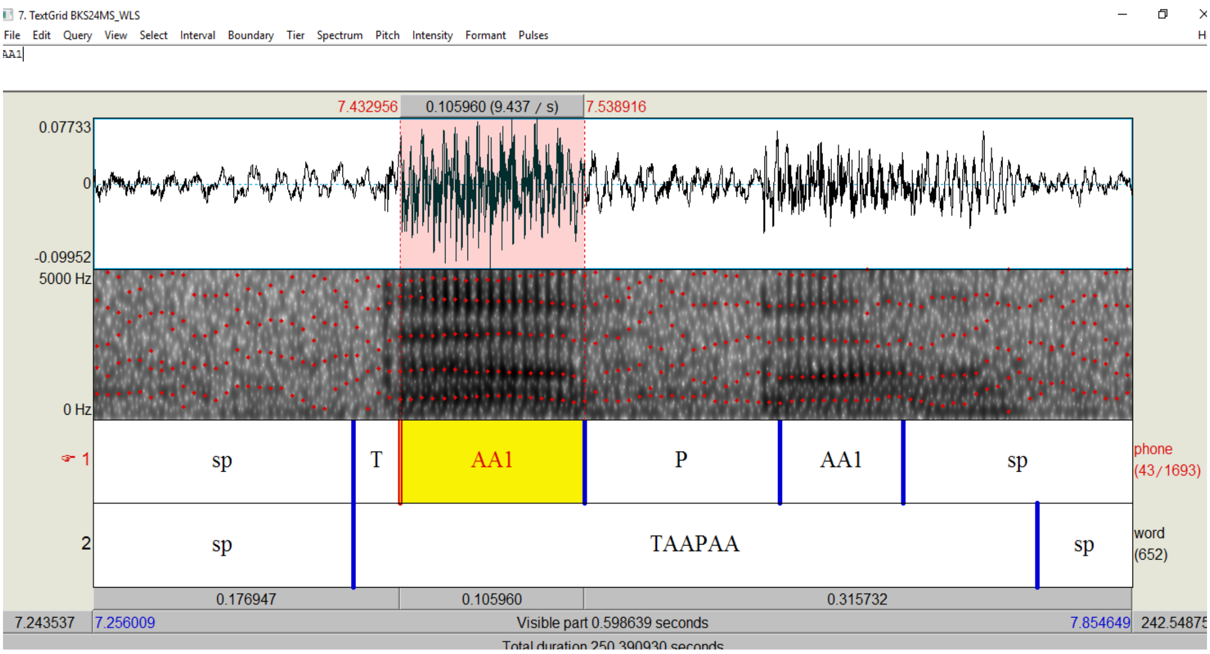


Figure 2. 6 Spectrographic showing selection of onset and offset for [ɑ]

In the present study, an interval approach was adopted and formants measurements were extracted using a Praat script which took measurements at 35 ms into onset, midpoint and 35 before offset time points into the course of the vowel. Although the scope of this study is delimited to monophthongs, the interval approach used mostly for analysis of diphthongs is adopted with an eye for future analysis of dynamic aspects of vowel quality without compromising the effect of coarticulation from the neighbouring segments by only taking measurements at 50 ms time-point for this analysis. The single scripts extracted measurements for the relevant parameters for this study namely: f_0, F1, F2 and vowel duration.

2.8 Outlier rejection

The next step in the process was examining the data set for outliers. Outliers are values in the data set which differ from the general pattern due to errors in experiment design or variability in the dataset. Removing outliers is necessary to ensure the reliability of acoustic measures given the potential they hold to skew comparisons. Possible reasons for the presence of outliers in the current data set are background noise, as most recordings were done on the field; and formant tracking errors.

Figure 2.7, 2.8 and 2.9 shows the distribution of normalized F1, normalized F2 and log transformed vowel duration for [ɑ], respectively. From the boxplot in figure 2.7 , it can be seen that there are six data points which diverge from the general trend in the data set for normalized values of F1 for [ɑ] . These six data points were excluded from the analysis.

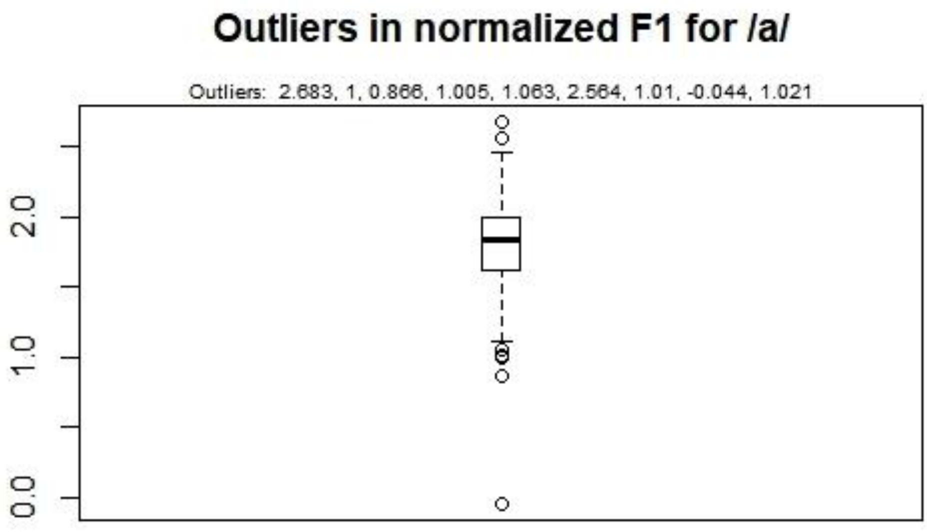


Figure 2. 7 Outliers in normalized F1 for [ɑ]

Similarly, figure 2.8 shows the boxplot for normalized F2 for the vowel [ɑ] from the sample. There are four points below the minimum value in the dataset. These suspected outliers were removed from the data set for analysis.

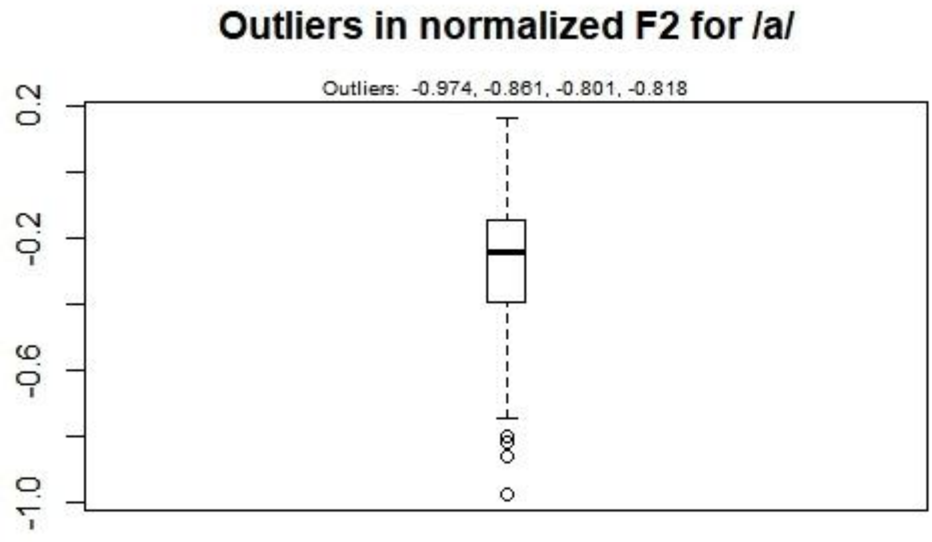


Figure 2. 8 Outliers in normalized F2 for [ɑ]

Figure 2.9 shows a boxplot for the log transformed values for vowel duration for [ɑ]. There are six data points in which are outliers. The utterance and acoustic parameters corresponding to these divergent values were excluded from the final sample for analysis.

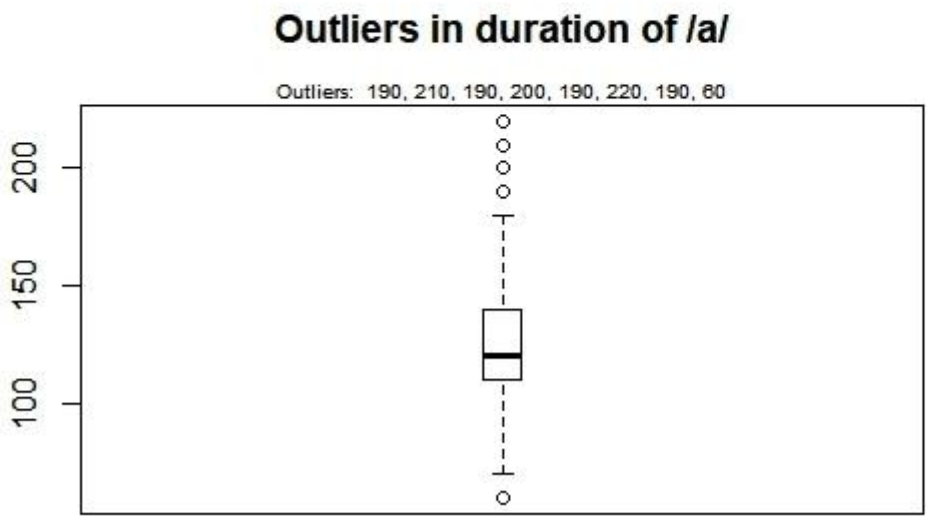


Figure 2. 9 Outliers in vowel duration for [ɑ]

The analysis for fundamental frequency (f0) was carried out only on the second repetitions. This was considered necessary to eliminate readings which arise due to the effects of list intonation. No normalization procedure was performed on pitch values, therefore, separate datasets were created on the basis of gender in order to facilitate a comparison based on gender between the different regions. The outlier identification and rejection procedure remained uniform. In figure 2.10 the distribution of f0 values for the vowel [ɑ] from male speakers is shown with the help of a boxplot. There are three points in the data set which exceed the maximum value. One value in particular, 396Hz for a male speaker is uncharacteristic and definitely is an outlier in the data.

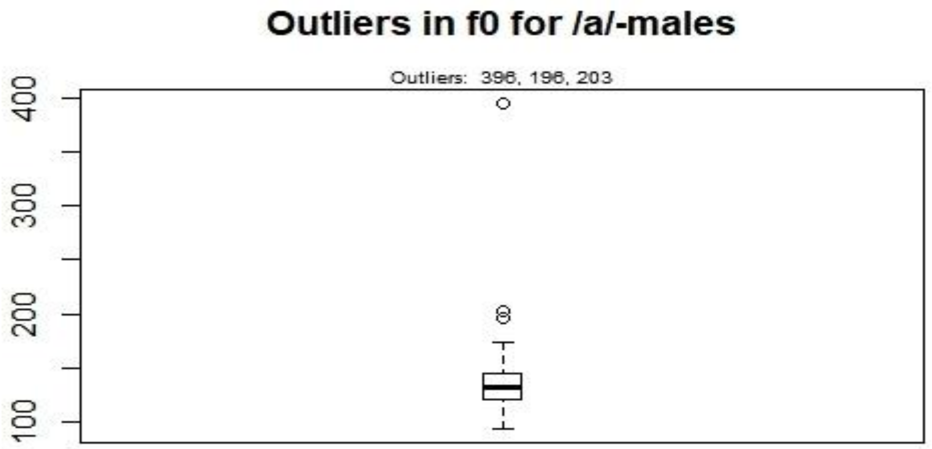


Figure 2. 10 Outliers in fundamental frequency for male speakers for [ɑ]

Similarly, a boxplot in figure 2.11 displays the distribution of f0 values the vowel [ɑ] from female speakers. There are two outliers here which differ significantly from all the other values.

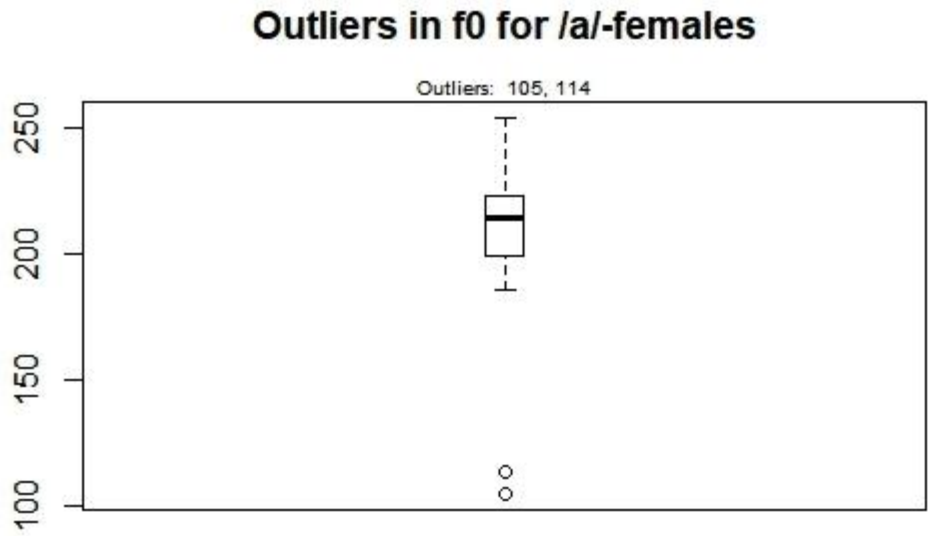


Figure 2. 11 Outliers in fundamental frequency for female speakers for [ɑ]

A similar method was used for the identification and exclusion of outliers from all acoustic measures under investigation for all the vowels (/i , e, ɑ , ʌ, o , u/).

2.9 Normalization

In studies such as the present one where the sample is constituted of acoustic data involving male and female participants, there is another methodological step known as normalization which needs to be performed on the acoustic measures before any further analysis is conducted. Four general goals of normalization have been put forward by Disner (1980) and Thomas (2002a):

a) Eliminating variation caused by physiological differences among speakers;
b) Preserving sociolinguistic/dialectal/cross-linguistic differences in vowel quality;
c) Preserving phonological distinctions among vowels
d) Modelling the cognitive processes that allow human listeners to normalize vowels uttered by different speakers.

Differences in the size of the vocal tract between children, adult males and females cause differences in formant resonances. Therefore, for formant data to be made comparable across speakers and groups in an acoustic investigation of vowels, it is essential that any acoustic differences in vowel production due to anatomical differences between speakers must be objectively factored out through the process of normalization.

There are two broad categories into which vowel normalization algorithms generally fall: vowel-intrinsic and vowel-extrinsic. Vowel intrinsic methods are based upon information contained in a single vowel with no reference to other. Syrdal and Gopal (1986) is the most popular vowel-intrinsic normalization technique. The advantage with vowel-intrinsic methods is that it is practical in the sense that it doesn't need measurements of all vowels from all speakers. Secondly, as Thomas (2011) notes, that difference in phonological inventories of dialects has no bearing on the results. The drawbacks of this method are that it is since it reliance on f_0 or F3 and their propensity to distort the shape of the vowel space. Vowel extrinsic methods are dependent on information from a range of vowels produced by the same person. They don't require F0 or F3 data relying solely on F1 and F2 values and work better with the speaker's entire vowel system. The disadvantages of this method are that it requires more data causing practical concerns; and it may not work well when languages with different vowel systems are compared.

For this study, formant measurements (F1 and F2) were normalized using the Lobanov normalization algorithm through NORM (Thomas & Kendall, 2007).The Lobanov formula ($z=(f-\mu)\div\sigma$) (Lobanov, 1971) is a vowel-extrinsic and speaker intrinsic technique which calculates a z-score of each formant for a speaker by dividing the difference between the raw Hertz values of a formant (f) and its mean value (μ) for all the vowels by the standard deviation (σ) for that formant across vowels for that speaker. Though vowel extrinsic methods may not be suitable for comparing two or more languages with different vowel inventories (Disner, 1980), they – the Lobanov method in particular -have been known to perform better than vowel-intrinsic procedures in preserving social and regional information for a single language while eliminating variation due to physiological factors (Adank et al., 2004, Clopper, 2009, Flynn and Foulkes, 2011). Figure 2.12 presents a comparison of vowel plots based on normalized and unnormalized F1 and F2. The plots in the top row are based on normalized values whereas the plots in the bottom row are based on unnormalized values. It is evident from the plots below that normalization significantly minimizes the effect of gender on formant frequencies.

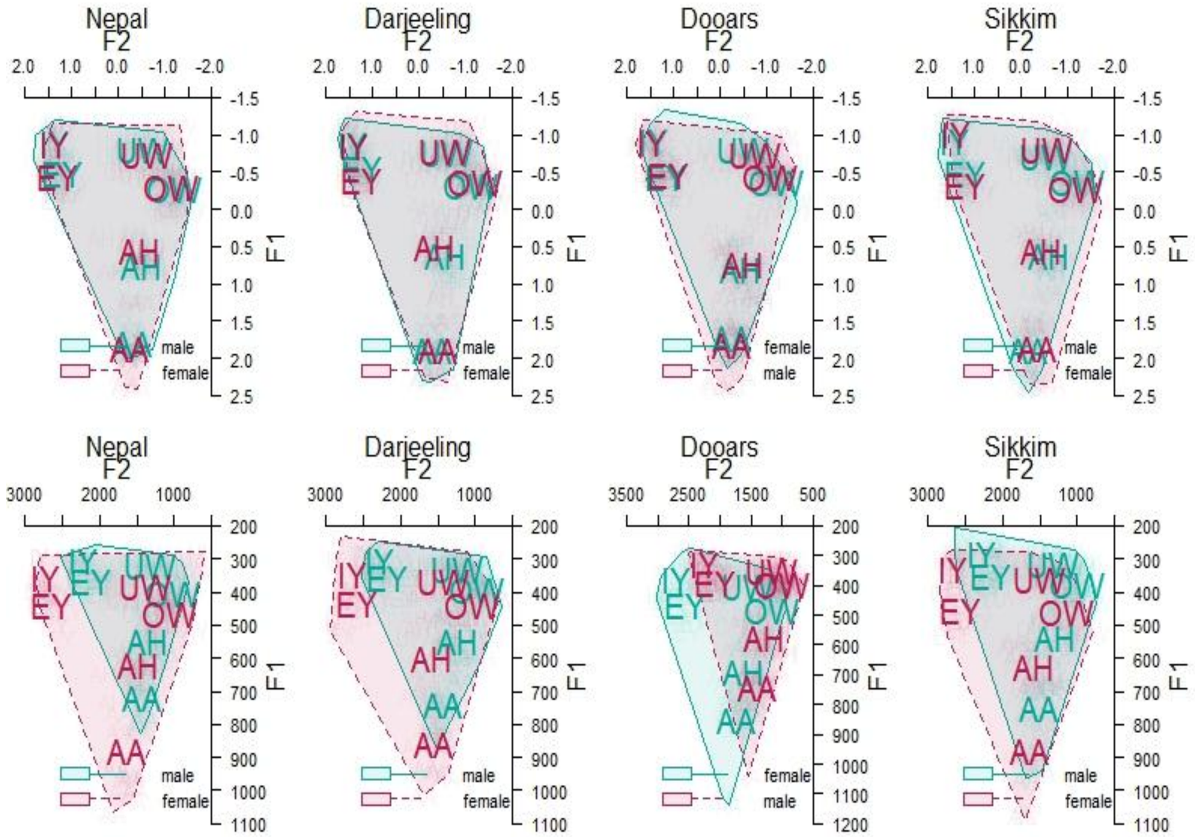


Figure 2. 12 Vowel plots based on normalized and unnormalized formant values

2.10 Hypothesis Testing and Statistical Tests

Central to any quantitative analysis is hypothesis testing using statistical procedures. The choice of a statistical test is determined the overall methodological design of the study. It is essential to delineate the dependent variable under investigation and the independent variable which accounts for any variation in the dependent variable. This study seeks to analyse the effect of region on acoustic parameters such as F0, F1, F2 and vowel duration.

Therefore, these acoustic measures are the dependent variables whereas region is the independent variable or the fixed effect. The dependent variable is continuous in nature while the independent variable is categorical with four different levels corresponding to the number of regions included in this study.

Linear mixed effects models were used for statistical analysis using the *lme4* (Bates, Maechler & Bolker, 2015) in R (R Core Team, 2013). The advantage of mixed-effects modelling lies in its ability to include random effects which are "groupings in the data that are not generalizable to the wider population" (Hay, 2011). A random effect carries the potential to have an idiosyncratic or unpredictable impact on the data. In order to resolves non-independencies of data because of repetitions of the same token for a vowel category by every speaker it was essential that a linear mixed effects models was adopted which accounts for the variability arising out of both the fixed and random effects.

A mixed effects analysis of the relationship between the acoustic parameters such as F0, F1, F2 and vowel duration, and region was performed. Region and repetition were included in the model as a fixed effect. Speaker was entered as a random effect in the model. For each acoustic measure (F1, F2 and vowel duration), three models were built: a full model with all the fixed and random effects; and two reduced models without one of the fixed effects. A one way ANOVA was then performed on the full model and the reduced model to arrive at p-values to gauge the significance of the fixed effects on the model using a maximum likelihood ratio test.

2.11 Quantifying Acoustic Space/Vowel Space Area

The final parameter assessed in the current study is vowel space area or the acoustic space of the four different regional varieties. Traditionally, acoustic space or vowel space area has been theorized as an area on a two-dimensional plane corresponding to the first and second formant frequency, bounded by peripheral vowels. Various approaches have been developed over time for obtaining metrics for vowel space area. The traditional approach of calculating the area of vowel triangle encompassing the corner vowels /i, ɑ, u/ or the area of the vowel quadrilateral with /i, u, ɑ, æ/ as the corner vowels has been used widely in clinical and linguistic research, especially for American English. Fox and Jacewicz (2017) observe that while the vowel space areas computed on the basis of corner vowels are well suited for the analysis of speech development in children and adolescents (Al-Tamimi and Ferragne, 2005), stylistic variation (Bradlow et al., 1996; Fourakis, 1991) and speech pathology

(Higgins and Hodge, 2002; Weismer et al., 2001), the vowel triangle or the vowel quadrilateral best accounts for within-speaker variation. For analyses of between-speaker variation, studies such as Fox and Jacewicz (2008) and Jacewicz et al. (2007) find inadequacies in vowel space area estimates computed through the traditional techniques. Another approach adopted in the estimation of vowel space area involves calculating the area of a convex hull created by the perimeter of all vowels in the vowel inventory of a language. According to Sandoval et al. (2013), the area of the convex hull provides a better estimation of the vowel space area compared to metrics derived from a vowel triangle or a vowel quadrilateral. It is important to point out that vowel space area estimation from vowel quadrilaterals and convex hulls are usually done on the basis of formant data measured at vowel mid-point. The most recent method of vowel space area estimation has been put forward by Fox and Jacewicz (2017) who propose a change of nomenclature to "formant space" as their method focuses on the distribution of formant points and their frequency distribution irrespective of vowel category. Their approach to the estimation of formant space relies on multiple measurements from the formant trajectory at regular intervals rather than mid-point measurements, upon which the vowel polygon and the convex hull approaches rely upon. From a study of three different dialects of American English, they demonstrate using spectral density maps that even though the working vowel space for different dialects may be the same, dialects can be differentiated on the basis of internal distribution of spectral density regions.

Although Fox and Jacewicz (2017) have proposed that analysis of the formant space computed with measurement of formant trajectories and analysis of the distribution of formant points gives a better estimate of the vowel space area as a function of dialect, in this study the estimation of acoustic vowel space is based on the convex polygon approach. There are two reasons for this: firstly, the convex hull method is still an improvement over the traditional methods based on the vowel triangle and the vowel quadrilateral; secondly, the formant space approach is fairly recent and relies on measurements of formant trajectories. The current study was conceptualized much prior to the development of the formant space method and although the measurement script collects formant data at three temporal points (25%, 50% and 75%) of the vowel, it is not well suited to capture information of the formant trajectory. All analysis has been performed on the basis of measurements taken at vowel mid-point therefore it is only apt if a method known to work well with such measures is adopted.

2.12 Pilot study

2.12.1 Objectives

A pilot study with data was conducted at the incipient stages of this research to compare and contrast the acoustic spaces and vowel duration in two major regional varieties of Nepali : the standard dialect with speakers from Nepal; and the Darjeeling variety which is the largest demographic group which speaks Nepali other than in Nepal. The primary objectives were to empirically account for cross dialectal differences in terms two key acoustic parameters of vowel quality i.e. steady state formant frequency and vowel duration; and to observe the effect on these acoustic measures as a function of speakers' region of origin.

2.12.2 Method

11 participants (6 from Darjeeling and 5 from Nepal) were interviewed and a wordlist was administered for data elicitation. Bisyllabic words presented in Devanagari script ([tipi], [peti], [tapa], [pʌti], [topi] and [tuti]) with stress on the target vowel were read out three times by all the participants. The total number of tokens used for acoustic analysis was 198 (11 X 6 X 3). A transcript of the wordlist was created and the sound files were automatically segmented using the Penn Phonetics Lab Forced Alignment Tool (Yuan & Liberman, 2008). The alignments were hand-checked and manually corrected for errors. A Praat script was used to extract vowel duration and formant measurements at 50% through the course of the vowel. Since both male and female participants had been interviewed, Normalization was deemed necessary in order to eliminate variation in measurements caused due to physiological differences in the vocal tracts of males and females. Another motivation behind carrying out normalization was to preserve dialectal differences in vowel quality (Thomas, 2011, p.161). For this study, formant measurements (F1 and F2) were normalized using the Lobanov normalization algorithm through NORM (Thomas & Kendall, NORM: The vowel normalization and plotting suite, 2007).

Linear mixed effects models were used for statistical analysis using R. In order to resolves non-independencies of data because of repetitions of the same token for a vowel category by every speaker it was essential that a mixed effect approach was adopted which accounts for the variability arising out of both the fixed and random effects. Therefore, region, phonetic environment and repetition were kept as fixed effects while speaker was retained in the model as a random effect. For each acoustic measure (F1, F2 and vowel duration), three models were built: a full model with all the fixed and random effects; and two reduced models without one of the fixed effects. The full model and the reduced models were then compared using a maximum likelihood ratio test to gauge the significance of the fixed effects on the model.

2.12.3 Results

2.12.3.1 F1

Table 2.3 summarizes the effect of fixed effects and mixed effects of region on F1 of vowels across all categories.

Table 2. 3 Summary of model estimates and results of likelihood ratio tests for F1

Effect of region on F1 All values are relative to the Nepal variety						
Fixed Effects						Mixed Effects
Vowel category	Estimate	Std.Error	Df	T-value	P-value	ANOVA using MLR
/i/	-0.50001	0.44438	11.00000	-1.125	0.2845	(χ^2 (1) = 1.1983, p = 0.2737
/e/	0.05722	0.48532	11.00000	0.118	0.908	(χ^2 (1) = 0.139, p = 0.9062
/ɑ/	1.1956	0.3595	11.00000	3.325	0.006767**	(χ^2 (1) = 7.6536, p = 0.005666**
/ʌ/	0.51154	0.47873	11.00000	1.069	0.3082	(χ^2 (1) = 1.0863, p = 0.2973
/o/	-0.2289	0.3905	11.00000	-0.586	0.5696	(χ^2 (1) = 0.3383, p = 0.5608
/u/	-0.3503	0.3671	11.00000	-0.954	0.360	(χ^2 (1) = 0.875, p = 0.3496

2.12.3.2 F2

Table 2.4 summarizes the fixed effects and mixed effects of region on F2

Table 2. 4 Summary of model estimates and results of likelihood ratio tests for F2

Effect of region on F2 All values are relative to the Nepal variety	
Fixed Effects	Mixed Effects

Vowel category	Estimate	Std.Error	Df	T-value	P-value	ANOVA using MLR
/i/	0.6203	0.4882	11.00000	1.271	0.230	(χ^2 (1) = 1.5065, p = 0.2197
/e/	0.8648	0.4656	11.00000	1.857	0.090218.	(χ^2 (1) = 3.0006, p = 0.08324 .
/ɑ/	-0.2247	0.5086	11.00000	-0.442	0.6672	(χ^2 (1) = 0.1935, p = 0.6601
/ʌ/	0.1205	0.4965	11.00000	0.243	0.8128	(χ^2 (1) = 0.0587, p = 0.8085
/o/	-0.9896	0.3598	11.00000	-2.751	0.0189*	(χ^2 (1) = 5.7575, p = 0.01642 *
/u/	-0.41436	0.51183	11.00000	-0.810	0.43435	(χ^2 (1) = 0.6366, p = 0.4249

2.12.3.3 Vowel Duration

Table 2.5 summarizes the fixed effects and mixed effects of region on F2

Table 2. 5 Summary of model estimates and results of likelihood ratio tests for vowel duration

Effect of region on duration All values are relative to the Nepal variety						
Fixed Effects						Mixed Effects
Vowel category	Estimate	Std.Error	Df	T-value	P-value	ANOVA using MLR
/i/	-1.2451	0.2648	11.00000	-4.702	0.000648***	(χ^2 (1) = 12.122, p = 0.0004984 ***
/e/	-1.23560	0.42523	11.00000	-2.906	0.0143*	(χ^2 (1) = 6.2658, p = 0.01231 *
/ɑ/	-1.0868	0.4495	11.00000	-2.418	0.03415*	(χ^2 (1) = 4.6878, p = 0.03038 *
/ʌ/	-1.4564	0.2869	11.00000	-5.076	0.000357***	(χ^2 (1) = 13.274,

						p = 0.000269 ***
/o/	-1.0648	0.4244	11.00000	-2.509	0.0291*	(χ^2 (1) = 4.9769, p = 0.02569 *
/u/	-1.12265	0.45154	11.00000	-2.486	0.0302*	(χ^2 (1) = 4.9054, p = 0.02677 *

2.12.4 Discussion

With respect to the steady state formant frequencies, results indicated two points of statistically significant contrasts: Darjeeling /a/ is lower relative to the Nepal variant; and, Darjeeling /o/ is relatively back-er in the vowel space than the Nepal variant. In other words, F1 of /a/ ($p<0.001$) and F2 of /o/ ($p<0.005$) were significantly affected as a function of region. It was also observed that the dispersion of vowels in the acoustic space of the Darjeeling variety was in a relatively larger space as compared to the standard variety.

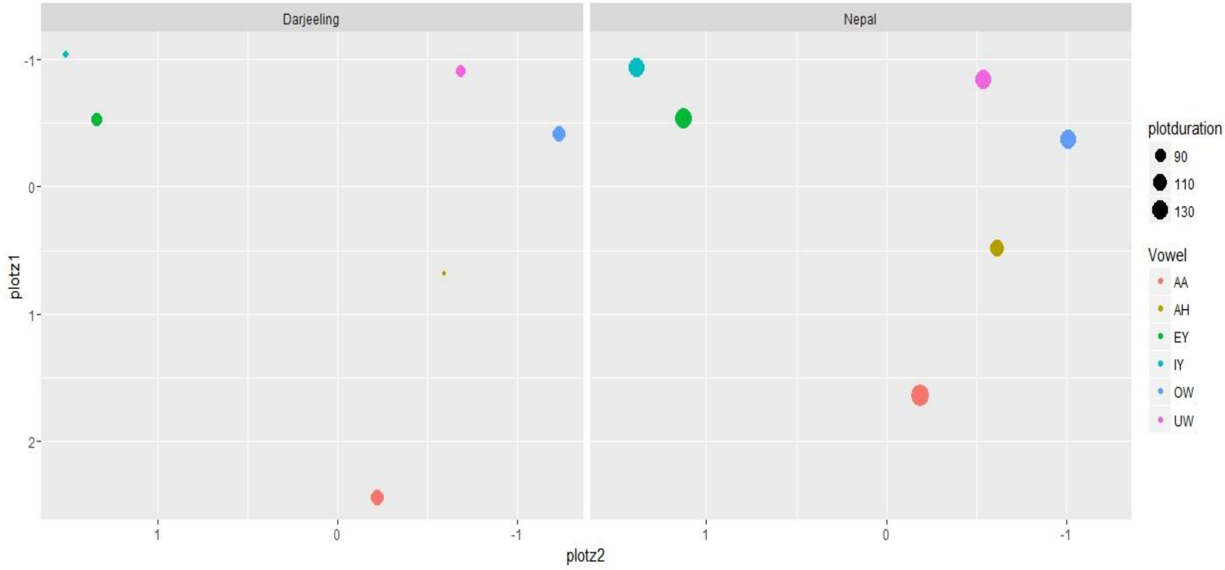


Figure 2. 13 Vowel dispersion in Nepal and Darjeeling varieties

The analysis of vowel duration revealed significant insights into what could possibly account for native speaker intuitions regarding regional phonetic variation. Duration across vowel categories was significantly affected as a function of speakers' region of origin perhaps serving as perceptual cue for dialect differentiation and native speaker intuitions which needs to be tested further.

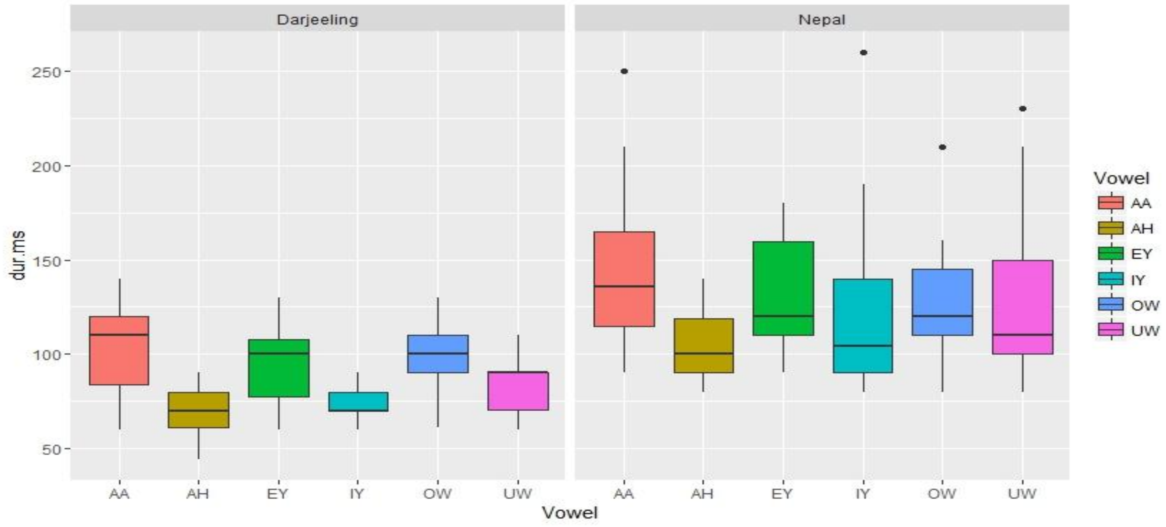


Figure 2. 14 Vowel duration in Darjeeling and Nepal varieties

2.13 Hypothesis revisited

From the results of the pilot study, two statements of tentative hypothesis can be formulated:

1. A speakers region of origin affects spectral aspects of vowels quality especially with reference to [ɑ] and [o].
2. A speakers region of origin significantly affects temporal aspects of vowel quality across all categories.

2.14 Tools and Instruments revisited

A major revision in the tools includes the development of a forced aligner for Nepali (FANEP). Although output from the P2FA can yield favourable results, it is desirable that efforts towards the development of a language resource for under-resourced such as Nepali carries the potential for better alignments as they are trained on the language itself and can be subsequently used for further studies.

2.15 Final procedure

The final procedure there can be listed as follows:

1. Identification of vowel categories
2. Identification of regions and participants
3. Determining inclusion criteria
4. Data elicitation through WLS
5. Recording
6. Cleaning recordings through editing without any loss of data, wherever possible.

7. Compiling a transcript
8. Compiling a pronunciation dictionary
9. Training acoustic models
10. Forced alignment
11. Checking alignment manually; adjustments, if required
12. Script-based measurement
13. Tabulation and coding of variables
14. Examining outliers
15. Statistical testing
16. Plotting
17. Reporting results

This concludes the chapter on method adopted for this study. The next chapter presents the data and analysis for this study.

CHAPTER 3

DATA AND ANALYSIS

3.1 Introduction

The preceding chapters of this dissertation provide the context for this chapter on data and analysis. In chapter 1, a case was made for the significance of the study considering the lack of variationist and experimental accounts of regional variation in Nepali. The rationale and scope of the study was highlighted along with a brief summary of the methods adopted. The vowel inventory and phonemicity was examined based on existing literature and a minimal pair test. A sociolinguistic profile of the four regions examined in this dissertation provided information on geographical and social setting. Chapter 1 also positioned the topic of current study as a lacuna in existing research. Tracing the historical evolution of acoustic phonetics, the second chapter discusses the relevance of acoustic phonetics in linguistic theory and discusses the major acoustic parameters which have been analyzed in this chapter. Some major studies which reflect the re-emergence of regional dialectology due to advances in methods in experimental phonetics have been discussed. The first chapter ends with a review of findings of some impressionistic and instrumental studies which have been conducted for the standard variety of Nepali. Chapter 2 elaborates on the method used for this study. The chapter discusses the rationale for the identification of regions, sampling participants and strategies adopted to overcome ethical issues. It also discusses the best practices in segmentation and approaches in measuring acoustic data. The chapter describes the statistical tools and procedures employed and results of a pilot study from a smaller dataset involving two of the four regions selected for this study are presented along with the recapitulation of the research hypothesis.

The goal of the current chapter is to present the data and an acoustic analysis of the parameters outlined for this study. This chapter begins with a description of the corpus gathered for this study. Post removal of outliers, the final sample for analysis has been tabulated. The chapter then goes on to describe the statistical testing procedure employed for the analysis of the acoustic parameters selected for this study. F1, F2, vowel duration and intrinsic pitch as denoted by fundamental frequency for males and females are analyzed separately for every vowel to see the effect of region on these measures. A summary of findings for each vowel is provided at the end of the analysis for each vowel category. Finally the acoustic space or the vowel space area, computed as convex hull covering the perimeter vowels, as a function of region is analyzed.

3.2 Sample

A total of 78 speakers (Darjeeling: n = 20, Dooars: n = 20, Sikkim: n= 20, and Nepal: n=18) were interviewed for the study. A wordlist style data elicitation approach was adopted because it offers certain advantages in comparison to data from other elicitation approaches such as conversational data or reading passages. Administering a wordlist carefully designed ensures the entire vowel inventory is elicited and that there are enough tokens for all the vowels. Additionally it allows for control over phonological environment and it produces longer stressed tokens which are suitable for acoustic analyses. Three repetitions of six bisyllabic words containing the target vowel in the word-medial position were elicited from every participant. One participant from Nepal ignored the stimulus for the [ɑ] vowel. Mid-point values of f_0, F1 and F2 along with vowel duration were measured for each vowel token with the help of a Praat script for analysis. Table 3.1 below presents the number of vowel tokens collected for every region and every vowel token.

Table 3. 1 Distribution of tokens in the sample

Distribution of tokens in the sample							
	[ɑ]	[ʌ]	[e]	[i]	[o]	[u]	Total
Darjeeling	60	60	60	60	60	60	360
Dooars	60	60	60	60	60	60	360
Sikkim	60	60	60	60	60	60	360
Nepal	51	54	54	54	54	54	321
Total	231	234	234	234	234	234	**1401**

The measurements were taken from recordings which were conducted in field settings in most cases. The field was chosen for interviews primarily to enlarge the sample as it would not have been possible to recruit enough speakers from the Dooars region in Delhi where a laboratory could possibly have been used for recordings. Controlling for background noise is difficult in field settings due to social and cultural aspects. Therefore, the data had to be analyzed for outliers and remove them from the data set. From a total of 1400 tokens, 113 tokens (8.07%) were rejected. After outlier rejection, the final data set consisted of 1287 tokens. Table 3.2 shows the distribution of tokens in the final dataset.

Table 3. 2 Distribution of tokens in the final dataset

Distribution of tokens in the final dataset							
	[ɑ]	[ʌ]	[e]	[i]	[o]	[u]	Total
Darjeeling	54	58	52	56	56	56	332
Dooars	57	56	54	55	58	57	337
Nepal	47	52	49	47	49	45	289
Sikkim	53	57	54	54	54	57	329
Total	211	223	209	212	217	215	**1287**

The next section presents an analysis of the acoustic measures outlined in the study which are f_0, F1 , F2 and vowel duration after all the outliers were removed from the dataset. Separate analyses are presented for every vowel in light of these acoustic parameters. F1 and F2 values were normalized using the Lobanov formula by using the phonR package in R to eliminate variation emanating from physiological differences between speakers. A log transform was performed on duration of vowel tokens to account for skew in the data. With regard to f_0, only the second repetition was included to eliminate readings which could possibly arise from list intonation.

3.3 Statistical tests

For F1, F2 and vowel duration, a multivariate regression model with random effects was used as the statistical method. Linear regression is used to analyze the effect one or more independent variable on a continuous dependent variable. One of the most vital assumptions for linear regression analysis is that there should be independence of data. In the current data for each vowel category, there were three repetitions which were elicited. This violates the assumption of independence of data as readings from the same speaker for any particular vowel category are not independent of each other. Instead of following a bootstrapping approach by averaging readings from the three repetitions, adding it as a random effect helps to align the data with the independence assumption by assigning a coefficient to each speaker. The significance of the model remains unaffected even though the speakers vary randomly among themselves (Hay 2013).Furthermore, using such an approach results in more tokens for analysis. Region, gender and log transformed vowel duration were used as fixed effects for F1 and F2 while speaker was added to the model as a random effect.

Vowel duration was analyzed by considering it as dependent variable and adding region and normalized values of F1 and F2 as fixed effects along with speaker identity as a random effect. Data for the analysis of fundamental frequency does not violate the independence assumption. Therefore, a simple linear regression model was created with f_0 values as the dependent variable and region as the independent variable.

The analysis begins with the analysis of sample density for each of the acoustic parameters to check for normal distribution of data followed by results from statistical testing procedures to check the effect of region on each of the acoustic measures. The overall significance of the models is presented along with contrasts between different regions on the basis of estimates provided by the fixed effects estimates.

3.4 Vowel [i]

This section presents the data and its analysis for vowel [i]. A total of 212 token were analyzed for [i]. The first step in the process is to observe if the data is normally distributed. The mean of the distribution is -0.905 and the median is -0.9001. The mean and the median are close to one another suggesting that the data is normally distributed. Figure 3.1 shows the sample density distribution of normalized F1 for [i] for each of the different regions. The bell shaped curves for each of the regions suggest that there is normal distribution of data. In the context of formant data, Hay (2013) notes that the tails in the curve may not be symmetric due to the nature of following speech segment.

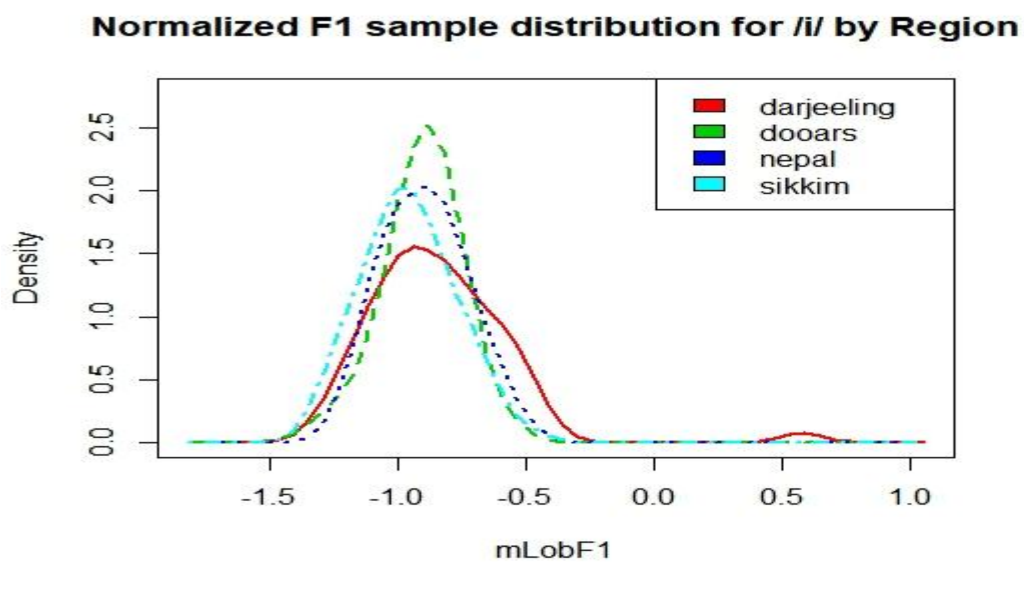


Figure 3. 1 Sample density distribution of F1 for [i] by region

To observe the effect of region on the height of vowel [i], two multivariate regression models were fit to the data. The first model was a full model with normalized F1 as the dependent variable predicted by region, gender, vowel duration and repetition as fixed effects and speaker as a random effect. The second null model fit was identical to the first with the exception that region was left out from the model as a fixed effect. A one way ANOVA was performed on the two models to check the effect of region on the dependent variable with statistical significance being estimated through maximum likelihood ratio.

3.4.1 F1 for [i]

Results of the one way ANOVA between the full and null models fail to observe a significant effect of region on normalized F1 for [i]. In the data for this study, region did not affect F1 for [i] ($\chi 2$ (3) = 5.4801, p = 0.1398). These statistics are based on the complete model with all the fixed and random effects.

Table 3.3 presents the estimates of fixed effects of the effect of independent variables on the normalized values of F1. To recapitulate, the relation between the height of a vowel and its F1 value is inversely proportionally. This implies that high vowels have lower F1 values and vice versa.

Table 3. 3 Fixed effects estimates of normalized F1 for [i]

Fixed effects estimates of normalized F1 for [i]					
	Estimate	Std.Error	df	t-value	Pr(>|t|)
(Intercept)	-0.94161	0.325248	170.95	-2.895	0.00429**
relevel(Region,nepal)darjeeling	0.057671	0.049113	78.61	1.174	0.24384
relevel(Region,nepal)dooars	-0.00562	0.048966	78.05	-0.115	0.90892
relevel(Region,nepal)sikkim	-0.05461	0.049576	80.16	-1.101	0.27398
relevel(Region, "darjeeling")sikkim	- 0.112279	0.047251	76.140000	-2.376	0.02000 *
relevel(Region, "darjeeling")dooars	- 0.063291	0.047260	76.010000	-1.339	0.18449
relevel(Region, "dooars")sikkim	- 0.048988	0.047702	77.370000	-1.027	0.30763
Repetition second	0.013783	0.028271	142.93	0.488	0.62663
Repetition third	0.077309	0.028334	147.49	2.728	0.00714**
log.dur.ms	0.00272	0.06874	171.53	0.04	0.96848

Gendermale	-0.00208	0.034688	77.59	-0.06	0.95227

The fixed effects estimates output provides estimates of the intercept and associated standard error and degrees of freedom, the t-value and the p-values. The intercept indicates the expected mean value of normalized F1 for [i] when all the independent variables are 0. In the data set, region is a categorical variable with four levels corresponding to the regions. Therefore, estimates for the intercept become unimportant to contrast groups.

The estimate for the F1 of Darjeeling [i] with respect to Nepal speakers is 0.057671 which is positive indicating that [i] for Darjeeling speakers is lower in the vowel space than speakers from Nepal. However, the p-values indicate that this difference is not statistically significant. Figure 3.2 below is a scatter plot with normalized F2 values on the x-axis and normalized F1 values on the y-axis. The plot displays the relative positioning of individual values and mean values of F1 and F2 for [i]. The scales are reversed to approximate the vowel acoustic space.

In comparison to standard variety, the F1 for Dooars speakers is in the negative which indicates that [i] is located higher in the vowel space. This difference in height between the Nepal and the Dooars varieties isn't statistically significant. Similarly the negative estimate for F1 in the Sikkim variety in relation to speakers from Nepal indicates that it is higher in the vowel space but the difference again is not statistically significant. However, there is a statistically significant difference in the height of [i] between the Darjeeling and the Sikkim varieties. The F1 for Sikkim speakers is lower than that of speakers from Darjeeling as the estimate of -0.112279 indicates that the Sikkim [i] is positioned significantly higher in the vowel acoustic space. No statistically significant difference was found between speakers from Darjeeling and Dooars, and Sikkim and Dooars as far as F1 was concerned. It can be observed nonetheless that Darjeeling [i] is lower relative to the Dooars variant which is positioned lower than the Sikkim variant.

There are no significant effects of either vowel duration or gender on F1 for the vowel [i]. The estimates for repetition reflect an expected outcome with the third repetition showing significant effects on F1. Values for the third repetition are higher suggesting minimization of effort and articulatory control on the final utterance.

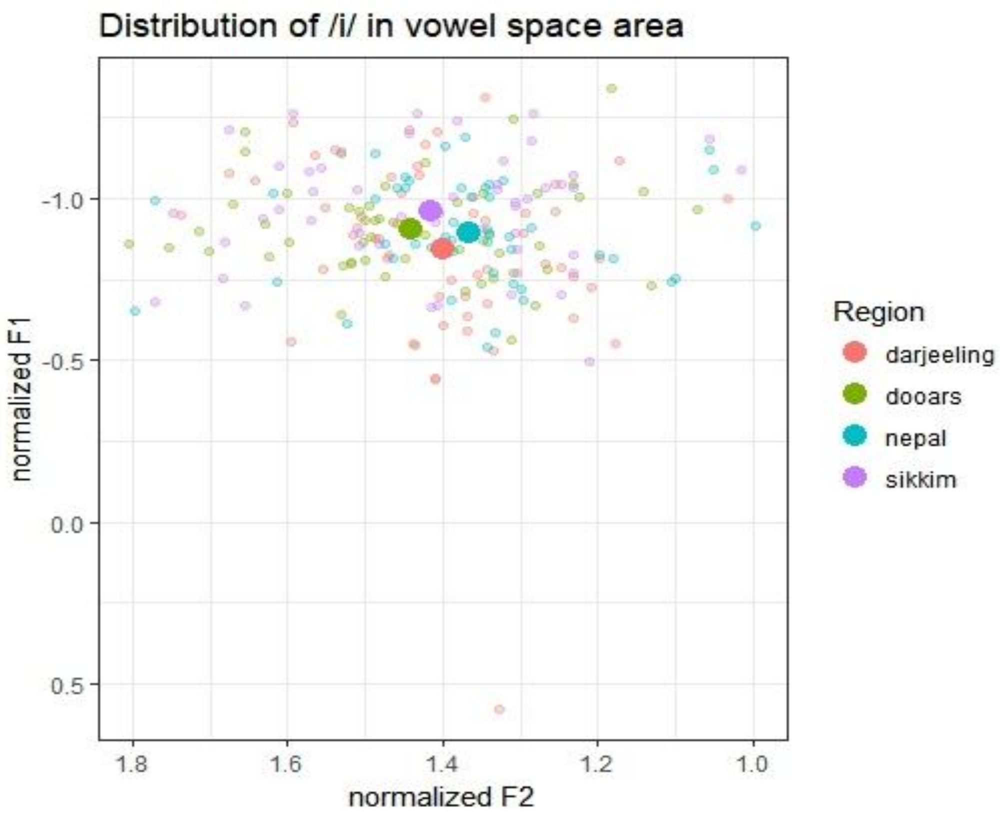


Figure 3. 2 Distribution of [i] in vowel space area

3.4.2 F2 for [i]

The results of a multiple regression model fit with fixed and random effects failed to observe a significant effect of region on F2 for [i]. From the data in this study, region did not affect F2 for [i] ($\chi2$ (3) = 5.6742, p = 0.1286). Figure 3.3 shows the sample density distribution of normalized F2 values for [i]. The data appears to be normally distributed.

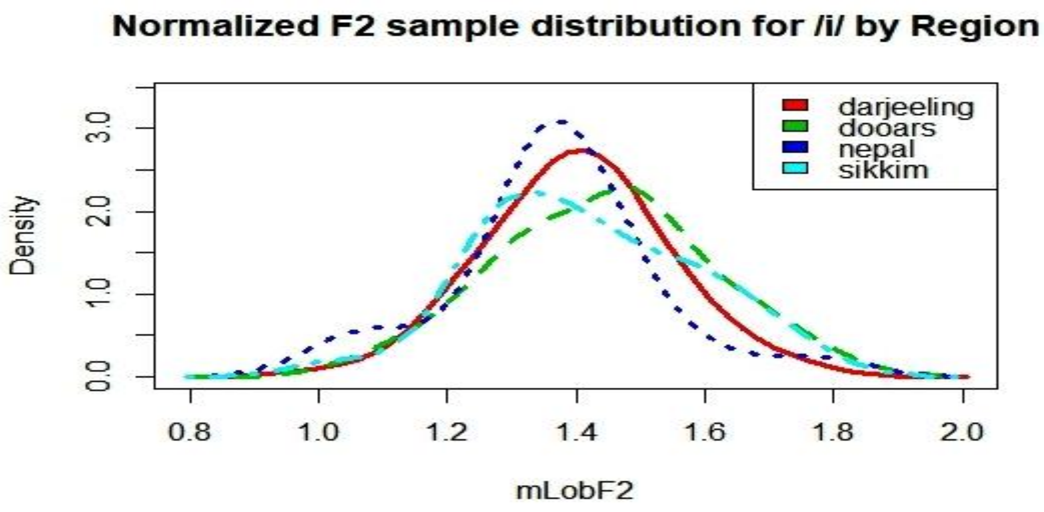


Figure 3. 3 Sample density distribution of F2 for [i] by region

F2 is associated with frontness or backness of a vowel. [i] is classified as a front vowel on the front-back dimension with high readings of F2 values expected for [i]. The estimates provided below in table 3.4 based on normalized values so they may not be intuitive to read as compared to unnormalized Hertz values. However, the estimates do indicate the magnitude and direction of the effect.

Table 3. 4 Fixed effects estimates of normalized F2 for [i]

Fixed effects estimates of normalized F2 for [i]					
	Estimate	Std.Error	df	t-value	Pr(>|t|)
(Intercept)	1.06448	0.24267	151.06	4.387	0.0000215 ***
relevel(Region,nepal)darjeeling	0.04222	0.03501	72.72	1.206	0.2317
relevel(Region,nepal)dooars	0.08142	0.03489	72.19	2.334	**0.0224 ***
relevel(Region,nepal)sikkim	0.06036	0.03538	74.26	1.706	0.0922
relevel(Region, "darjeeling")sikkim	0.01814	0.03361	70.14000	0.540	0.5911
relevel(Region, "darjeeling")dooars	0.03920	0.03361	69.90000	1.166	0.2475
relevel(Region, "dooars")sikkim	-0.02106	0.03397	71.25000	-0.620	0.5373
Repetitionsecond	-0.02631	0.02272	138.84	-1.158	0.2489
Repetitionthird	-0.07158	0.02273	142.42	-3.149	**0.002 ****
log.dur.ms	0.06575	0.05129	151.07	1.282	0.2019
Gendermale	0.04238	0.0247	71.51	1.716	0.0906

Fixed effects estimates show a significant effect of the third repetition on normalized F2 values for [i]. The other statistically significant difference is between the Nepal and the Dooars varieties with the estimate for the Dooars variety indicating that it is further forward in the vowel space area than the standard variant. The Darjeeling and Sikkim varieties are also characterized by increasing slopes for F2 indicating that the all regional variants in the current data are fronted, or further ahead in the vowel space area in relation to the standard variety spoken by speakers from Nepal. No significant results were obtained for the effect of gender or vowel duration on the F2 for [i] from the current data set.

3.4.3 Vowel duration of [i]

A log transform was performed on vowel duration to remove skew from the dataset. Figure 3.4 shows the sample density plot for vowel duration for [i]. The mean and the median of the distribution were 4.616 and 4.605, respectively.

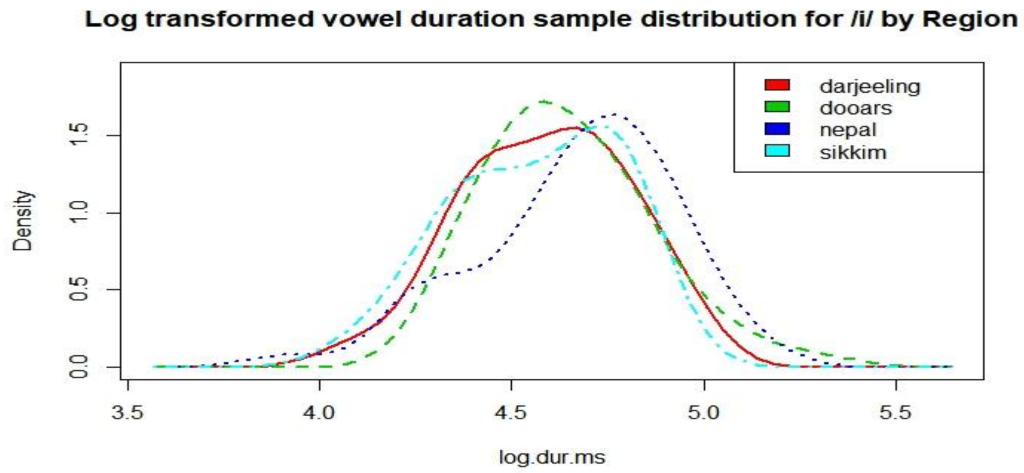


Figure 3. 4 Sample density distribution of vowel duration for [i]

A linear mixed effects model comparison between the models failed to find a significant effect on region on vowel duration for [i] ($\chi 2$ (3) = 3.477, p = 0.3238). Table 3.5 presents fixed effects estimates for all the predictors in the model. Only the third repetition had a significant effect on vowel duration which was expected. Negative estimates for Darjeeling, Dooars and Sikkim in relation to Nepal indicate that [i] in these three regions are shorter in comparison to [i] from speakers from Nepal. The difference however is not statistically significant. The Dooars variant is marginally longer than the Darjeeling variant while the Sikkim variant was found to be shorter than the Darjeeling variant. The Sikkim variant is also shorter than the Dooars variant. In summation it can be observed that in the current data set, the Nepal variant of [i] is the longest followed by the Dooars variant. The Sikkim variant is the shortest.

A negative estimate for F1, coded as mLobF1 in the table, is in line with the assumption that as duration increases, the F1 decreases. F1 for high vowels such as [i] are characteristically low. Estimate for F2 indicates that as the duration increases, the F2 also increases. A front vowel such as [i] is characterized by high F2 values. The fixed effects estimates for F1 and F2 reflect the fact that the inherent length of a vowel strongly affects the articulatory target. This justifies the inclusion of these measures of vowel quality in the model to account for processes such as vowel undershoot.

Table 3. 5 Fixed effects estimates of log transformed vowel duration for [i]

Fixed effects estimates of log transformed vowel duration for [i]					
	Estimate	Std.Error	df	t value	Pr(>|t|)
(Intercept)	4.552325	0.139099	209.840000	32.727	< 2e-16 ***
relevel(Region,nepal)darjeeling	-0.058576	0.063430	76.070000	-0.923	0.358674
relevel(Region,nepal)dooars	-0.008927	0.063678	77.010000	-0.140	0.888871
relevel(Region,nepal)sikkim	-0.103856	0.063615	76.940000	-1.633	0.106645
relevel(Region, "darjeeling")dooars	0.049649	0.061520	75.250000	0.807	0.422186
relevel(Region, "darjeeling")sikkim	-0.045280	0.061716	76.220000	-0.734	0.465397
relevel(Region, "dooars")sikkim	-0.094929	0.061456	75.230000	-1.545	0.126627
Repetitionsecond	-0.007412	0.022440	135.980000	-0.330	0.741686
Repetitionthird	-0.085550	0.023241	142.120000	-3.681	**0.000329 *** **
mLobF1	-0.056701	0.061737	184.490000	-0.918	0.359594
mLobF2	0.057384	0.078463	180.390000	0.731	0.465514
Gendermale	0.006662	0.045009	76.380000	0.148	0.882729

3.4.4 f_0 for [i]

For fundamental frequency analyses, the fundamental frequency of [i] was measured at the midpoint. Since only the second repetitions were subsetted, there were no repeated observations. A linear regression model was fitted to observe pitch as a function of region. No normalization procedure was performed on f_0 values so separate data sets for males and females were created for comparison across regions. It is well attested that males and females differ with respect to fundamental frequency due to differences in the length of the vocal tract. Therefore, the effect of gender on fundamental frequency was not probed.

3.4.5 f_0 for females for vowel [i]

Figure 3.5 displays the sample density distribution for female speakers for the [i] vowel. The mean is 240.3 Hz while the median is 237 Hz. The low difference between the mean and the median and the bell shaped curves of the distributions from all the regions suggest that the data is normally distributed and appropriate for linear regression analysis.

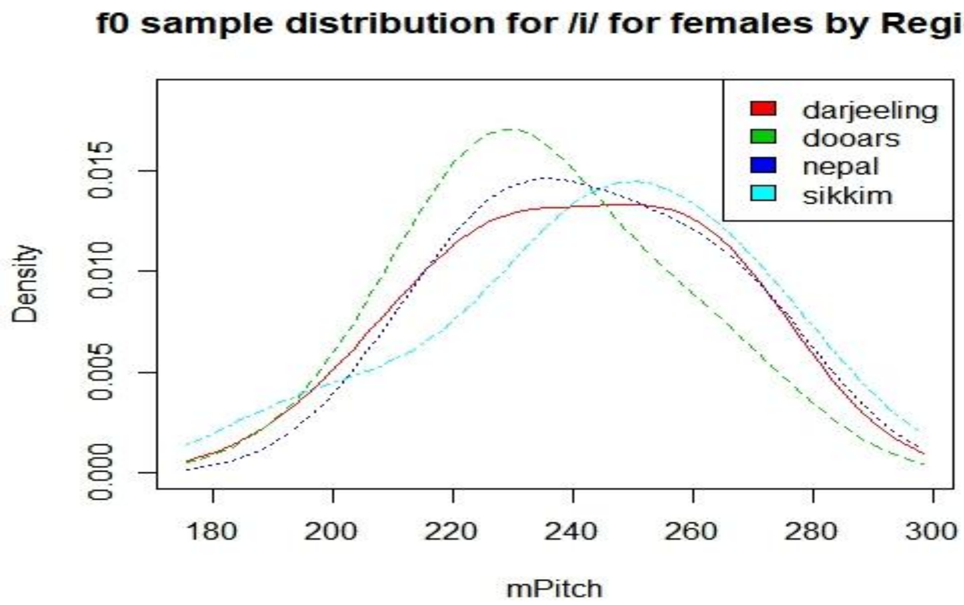


Figure 3. 5 Sample density distribution of f_0 for female speakers for [i]

A linear model of fundamental frequency for females as a function of region and normalized F1 was constructed. This model was not significant ($F(4,24)=0.9987, p=0.4725$). Table 3.6 displays the model estimates. With respect to the standard variety spoken by Nepal speakers, f_0 means for all other regions are lower but the difference is not statistically significant. The pitch for Darjeeling speakers is marginally higher than speakers from Dooars and Sikkim. Females from Dooars have the lowest f_0 value for [i].

Table 3. 6 Fixed effects estimates for fundamental frequency of females for [i] as a function of region

Fixed effects estimates for fundamental frequency of females for [i] as a function of region

	Estimate	Std.Error	t.value	Pr(>|t|)
(Intercept)	211.8773	19.72968	10.73901	**1.20E-10**
relevel(Region, "nepal")darjeeling	-1.14075	12.25067	-0.09312	0.926583
relevel(Region, "nepal")dooars	-7.76351	12.2443	-0.63405	0.532045
relevel(Region, "nepal")sikkim	-2.57399	12.43311	-0.20703	0.837735
relevel(Region, "darjeeling")dooars	-6.62276	10.74456	-0.61638	0.543445
relevel(Region, "darjeeling")sikkim	-1.43324	11.03908	-0.12983	0.89778
relevel(Region, "dooars")sikkim	5.189517	10.96344	0.473348	0.640242
mLobF1	-34.5977	19.40736	-1.78271	**0.087289 .**

3.4.6 f_0 for male speakers for vowel [i]

Comparable to the procedure followed for female speakers, a sample density plot as displayed in figure 3.6 was used to evaluate the normality of distribution of data in the data set. A difference of 7.3 Hz was observed between the mean and the median of the distribution.

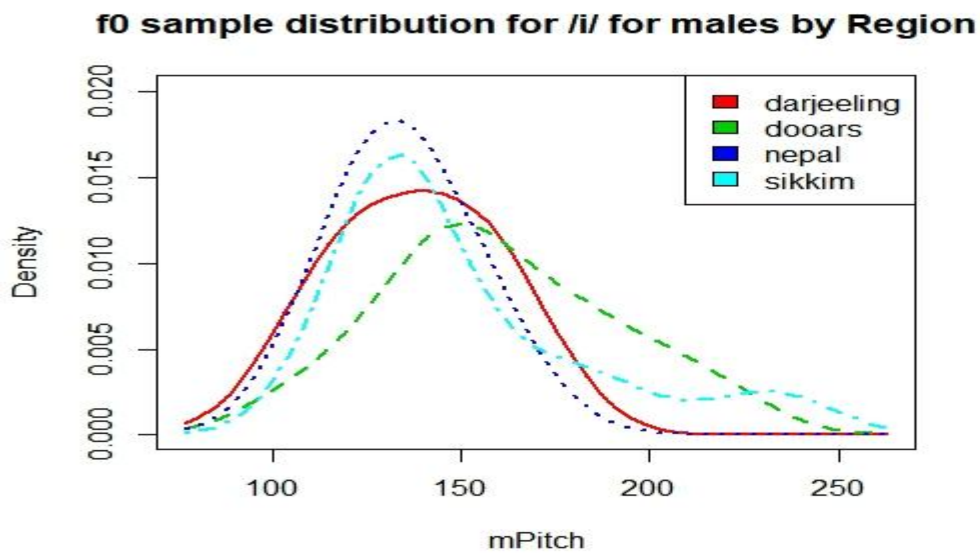


Figure 3. 6 Sample density distribution of f_0 for male speakers for [i]

Subsequently,a linear model of fundamental frequency for males as a function of region and normalized F1 was then constructed. Similar to the model for females for this vowel category, this model was not significant (F(4,34)=1.322,p=0.2816). Table 3.7 below display the model estimates.

Table 3. 7 Fixed effects estimates for fundamental frequency of males for [i] as a function of region

Fixed effects estimates for fundamental frequency of females for [i] as a function of region				
	Estimate	Std.Error	t.value	Pr...t..
(Intercept)	124.7088	30.33494	4.11106	**0.000235**
relevel(Region, "nepal")darjeeling	3.885598	12.89031	0.301436	0.764918
relevel(Region, "nepal")dooars	25.58098	13.0693	1.957334	**0.058558 .**
relevel(Region, "nepal")sikkim	15.85898	12.98285	1.221533	0.230287
relevel(Region, "darjeeling")dooars	21.69539	11.96511	1.813221	**0.078634 .**
relevel(Region, "darjeeling")sikkim	-3.8856	12.89031	-0.30144	0.764918
relevel(Region, "dooars")sikkim	-9.722	12.40161	-0.78393	0.438508
mLobF1	-10.7281	30.34134	-0.35358	0.725836

Even though the model was insignificant for the dataset, the estimates table indicate that compared to the standard variety, the fundamental frequency of male speakers in all the other regions is higher. The p-value (0.058558) for Dooars male speakers is quite close the threshold of 0.05 set for testing statistical significance with a difference of at least 25 Hz between the f_0 means for Nepal and Dooars speakers. These differences can perhaps be probed in subsequent studies. Sikkim speakers have the lowest f_0 compared to other regions in the data set.

3.4.7 Summary of findings for [i]

The following is a summary of findings for [i]:

1. There was no significant effect of region on F1.
2. Statistically significant difference in the height of [i] between the Darjeeling and the Sikkim varieties.
3. Sikkim [i] is positioned significantly higher in the vowel acoustic space compared to Darjeeling [i].
4. Nepal [i] is higher than Darjeeling [i]. Darjeeling [i] is the lowest among all other regions.
5. Dooars variant is positioned lower than the Sikkim variant.

6. No significant effects of either vowel duration or gender on F1.
7. The model for F2 as a function of region was also found to be insignificant for the data.
8. Statistically significant difference was observed between the Nepal and the Dooars varieties. Dooars variant is more fronted in the vowel space area than the standard variant.
9. All regional variants in the current data are fronted, or further ahead in the vowel space area in relation to the standard variety spoken by speakers from Nepal.
10. No significant effect of region on vowel duration.
11. Nepal [i] is longer than all other regional variants. This finding corroborates the findings from the pilot data. Sikkim is shortest. Dooars [i] is longer than Darjeeling [i].
12. Linear models for pitch as a function of region for both male and female speakers proved insignificant for the data. For females, f_0 of [i] in Nepal variety is the highest followed by Darjeeling, Dooars and Sikkim.
13. Fundamental frequency of male speakers for [i] is the lowest for the standard variant. Speakers from Nepal and Dooars may be contrasted on this parameter in subsequent studies as the difference is close to significant.

3.5 Vowel [e]

[e] is classified as a high mid or a half close front vowel in standard literature. Acoustically it is characterized by high F2 values and low F1 values. A total of 209 tokens were included in the data set for analysis of the vowel [e] after removal of outliers from the original data set. 25 tokens had to be excluded because of large deviations from the mean.

3.5.1 F1 for [e]

Figure 3.7 shows the density of normalized F1 for all the four regions under investigation in this study. The mean for the distribution is -0.4603 and the median in -0.4640. The data appears to be normally distributed.

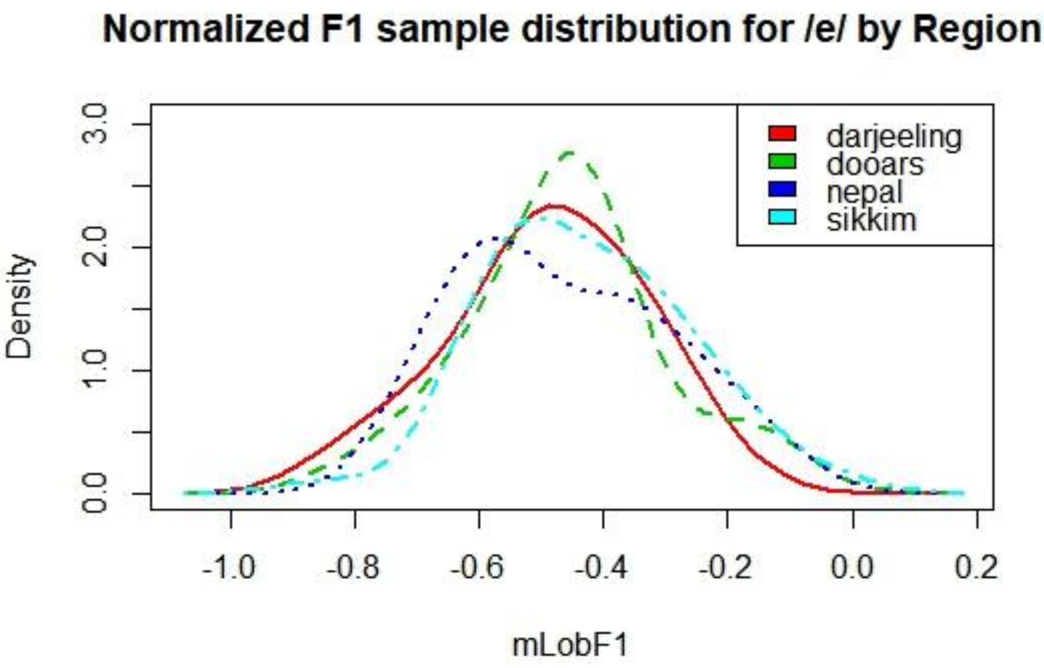


Figure 3. 7 Sample density distribution of F1 for [e] by region

A full linear mixed effects model was constructed with region, repetition, gender and log transformed vowel duration as fixed effects and speaker as a random effect. A second null mixed effects model was created identical to the full model with only region as a fixed effect missing. A one way ANOVA was performed on the full and null models to observe the effect of region on F1. The model for the data set failed to find a significant effect ($\chi 2$ (3) = 1.5557, p = 0.6695).

Table 3. 8 Fixed effects estimates of normalized F1 for [e]

Fixed effects estimates of normalized F1 for [e]					
	Estimate	Std..Error	df	t.value	Pr...t..
(Intercept)	0.040128	0.297021	188.8904	0.1351	0.892676497
relevel(Region, "nepal")darjeeling	-0.03298	0.040122	75.50043	-0.82204	0.413645322
relevel(Region, "nepal")dooars	-0.00787	0.039795	76.34416	-0.19776	0.843759912
relevel(Region, "nepal")sikkim	0.015132	0.039919	76.3271	0.379068	0.705689791

relevel(Region, "darjeeling")dooars	0.025112	0.039065	75.61787	0.642825	0.522282
relevel(Region, "darjeeling")sikkim	0.048114	0.039139	75.6053	1.22929	0.222775
relevel(Region, "dooars")sikkim	0.023002	0.038628	75.84589	0.59546	0.553308
Repetitionsecond	-0.02989	0.018744	138.9086	-1.59474	0.113042107
Repetitionthird	-0.08708	0.01998	160.8425	-4.35824	2.33E-05***
log.dur.ms	-0.0779	0.058919	190.3788	-1.32217	0.187697136
Gendermale	-0.12854	0.029813	83.7173	-4.31154	4.40E-05***

The estimates from the fixed effects estimates indicate that are no significant inter-group differences between any of the four regions included in this study. F1 values for Darjeeling [e] and Dooars [e] are marginally lesser in comparison to the standard variety suggesting that [e] in the Darjeeling and Dooars varieties are slightly higher than the standard variant. Sikkim speakers have a slightly higher F1 among speakers from all four regions indicating that it is positioned lower than the other regional variants. The positive estimate for the Dooars variant with reference to the Darjeeling variant indicates that the Dooars speakers' [e] is lower in comparison to the Darjeeling variant. None of these differences, however, are significant. Figure 3.8 below expresses the relative position of [e] among the four different regions.

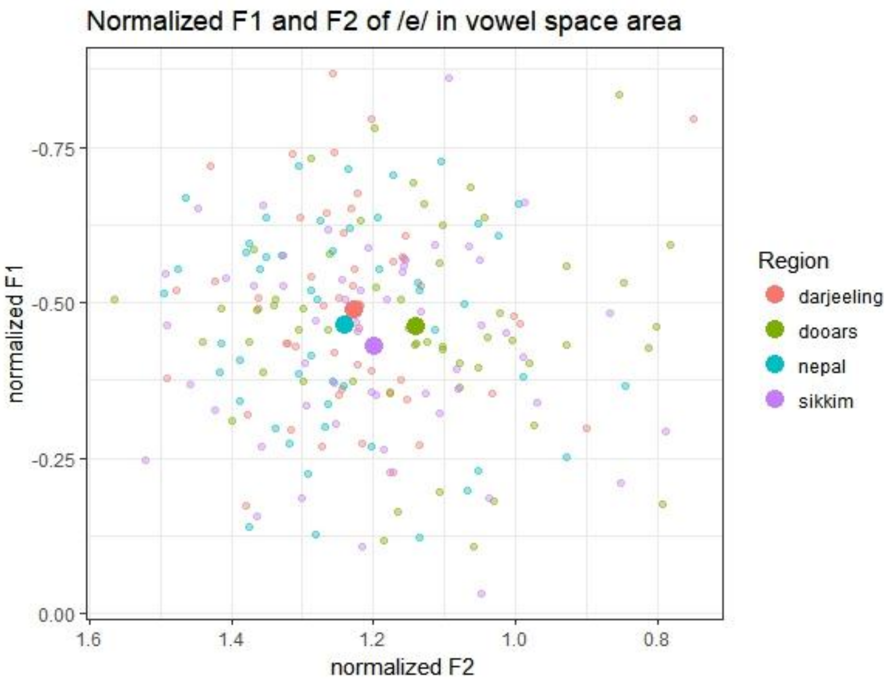


Figure 3. 8 Distribution of [e] in vowel space area

Readings for the third repetition were observed to be significantly different from the first which was in line with the expectation. Repetition was included as a fixed effect in the model for this reason. While no significant effect was found for the relationship of vowel duration as a predictor of F1, the model predicts a negative estimate for the log of vowel duration. This implies that vowel duration decreases with increase in F1.

An interesting effect observed in the coefficient table for fixed effects estimates the significant effect of gender on the height of the vowel [e] considering that the readings were normalized. In relation to females, F1 readings of males speakers were found to be significantly lower suggesting that the [e] for males is higher when compared to female speakers as seen in figure 3.9 below.

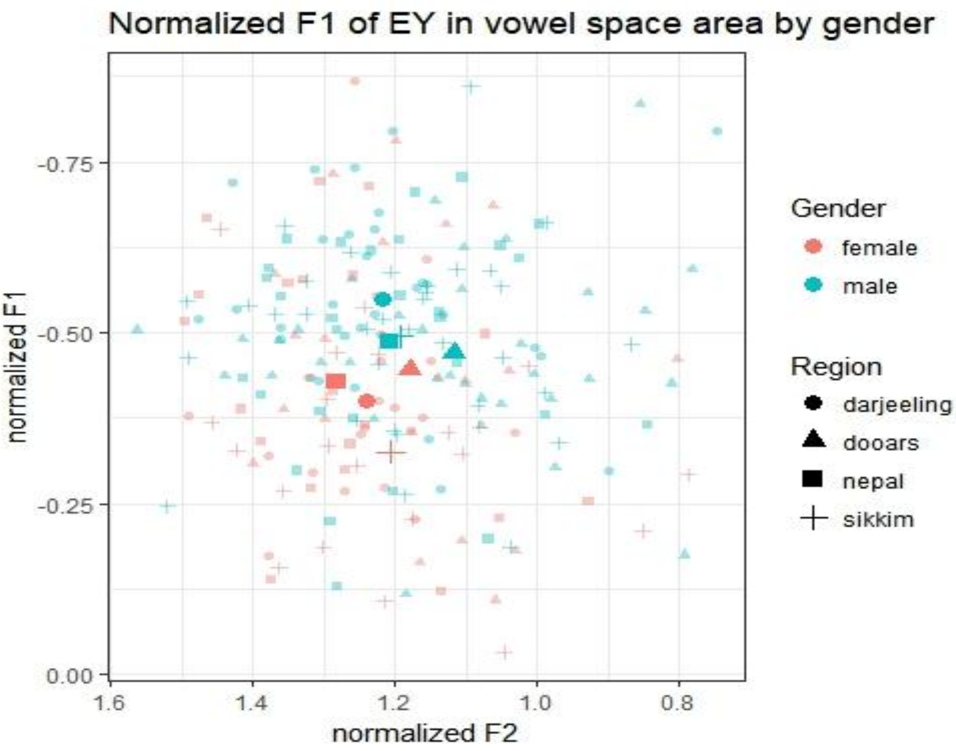


Figure 3. 9 Distribution of [e] between males and females in different regions

This exploratory finding of gender as a predictor of height of [e] provides scope for further studies on this variable. This data set is limited to word-list style data elicitations. An understanding of gender as a sociolinguistic variable in Nepali can be further informed through analysis of data from other elicitation styles such as a reading passage or conversational data.

3.5.2 F2 for [e]

A linear mixed effects model was constructed with region, repetition, gender and log transformed vowel duration as fixed effects and speaker as a random effect and contrasted with a null model without region as fixed effect to check the effect of region on normalized F2 for the vowel [e]. Region affected F2 of [e] ($\chi 2$ (3) = 8.2164, p = 0.04174*. Table 3.9 below provides the estimates for fixed effects estimates from which the magnitude and direction of inter-group differences can be observed.

Table 3. 9 Fixed effect estimates of normalized F1 for [e]

Fixed effect estimates of normalized F1 for [e]					
	Estimate	Std..Error	df	t.value	Pr...t..
(Intercept)	1.27420	0.312206	157.6227	4.08127	**7.10E-**

	1			8	**05*** **
relevel(Region, "nepal")darjeeling	- 0.01422	0.037963	72.27266	- 0.37462	0.709041
relevel(Region, "nepal")dooars	- 0.10119	0.037712	73.35123	- 2.68323	**0.009006****
relevel(Region, "nepal")sikkim	- 0.04258	0.037812	72.77118	- 1.12602	0.263856
relevel(Region, "darjeeling")dooars	- 0.08697	0.036976	72.62086	- 2.35202	**0.021386***
relevel(Region, "darjeeling")nepal	0.01422 2	0.037963	72.27265	0.37461 9	0.709041
relevel(Region, "dooars")sikkim	0.05861 3	0.036573	72.71267	1.60263 1	0.113352
Repetitionsecond	- 0.02474	0.022548	138.8739	- 1.09736	0.274383
Repetitionthird	- 0.05762	0.023586	159.9996	- 2.44291	**0.015657***
log.dur.ms	0.00403 7	0.062014	158.9259	0.06510 3	0.948174
Gendermale	-0.0456	0.028564	80.16993	- 1.59649	0.114314

An analysis of the estimates of fixed effects in the model reveals no significant effect of gender or vowel duration on F2 for vowel [e]. The estimate for vowel duration indicates a directly proportional relationship between F2 and vowel duration. A negative estimate for males indicates that their F2 values are lesser in relation to females.

There is no significant difference between the Darjeeling and the Nepal variants on the axis of fronting. The negative estimate for Darjeeling indicates that [e] in the standard variant is slightly more fronted than its Darjeeling counterpart even though this difference is not statistically significant. In fact [e] in the standard variety is the most fronted of all. The distance between the standard and the Dooars variants, however, is statistically significant. The Dooars variant is comparatively back-er in relation to the standard variant. The Dooars variant also differs significantly from the Darjeeling variety on the front-back dimension. The Darjeeling variant is significantly ahead in the vowel acoustic space than the Dooars variant. Region does not have a significant effect for the Sikkim and Dooars variants. A plot of normalized F2 on the x-axis and normalized F1 on the y-axis as shown in Figure 3.8 visualizes the relative positioning of [e] among the four regional variants.

3.5.3 Vowel duration for [e]

Figure 3.10 display the sample distribution for [e]. The mean for the distribution is 4.797 and the median in 4.787 and the bell shaped curves for the distribution for each of the regions suggest that the data is normally distributed.

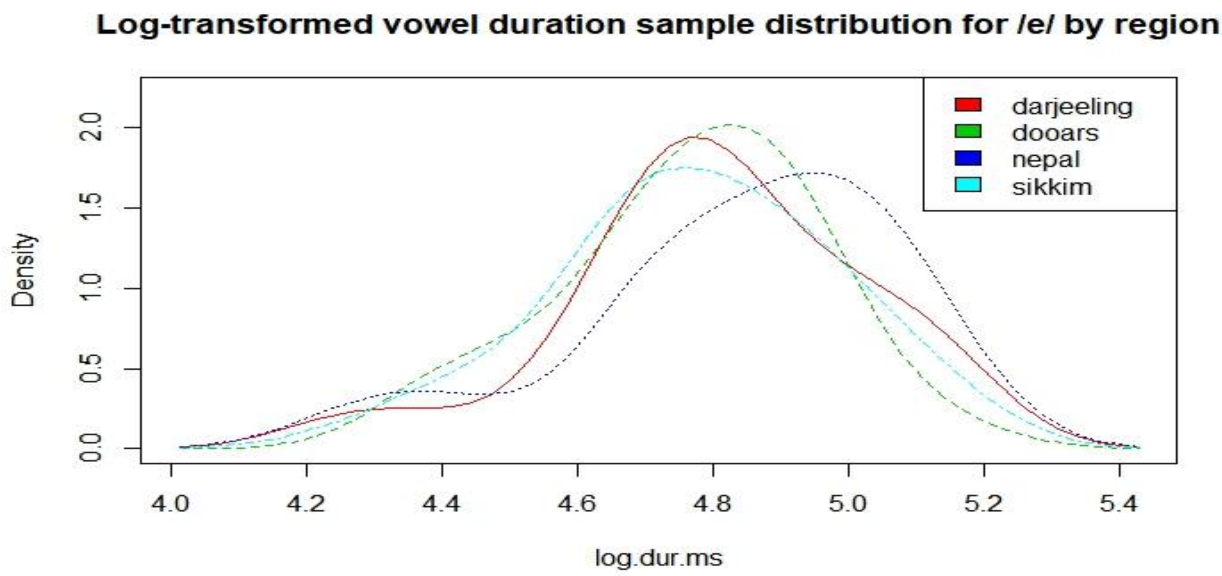


Figure 3. 10 Sample density distribution of F2 for [e] by region

A linear model for vowel duration was constructed with region, gender, normalized F1 and normalized F2 as fixed effects and speaker as a random effect. This model was compared to a reduced or a null model with just region missing as a fixed effect in the null model by performing Likelihood ratio test using the anova function in R. In the current dataset, region did not affect vowel duration ($\chi2$ (3) = 1.9249, p = 0.5881). Table 3.10 provides estimates of the fixed effects.

Table 3. 10 Fixed effects estimates of log transformed vowel duration for [e]

Fixed effects estimates of log transformed vowel duration for [e]					
	Estimate	Std..Erro	df	t.value	Pr...t..

		r			
(Intercept)	4.919835	0.100075	208.0822	49.1616	**< 2e-16 *** **
relevel(Region, "nepal")darjeeling	-0.04141	0.056303	73.06388	-0.7354	0.464453
relevel(Region, "nepal")dooars	-0.05745	0.056067	74.92334	-1.0246	0.308847
relevel(Region, "nepal")sikkim	-0.07489	0.055734	73.57092	-1.34364	0.183195
relevel(Region, "darjeeling")dooars	-0.01604	0.055124	74.46483	-0.29099	0.771864
relevel(Region, "darjeeling")sikkim	-0.03348	0.05498	73.80000	-0.609	0.544400
relevel(Region, "dooars")sikkim	-0.01744	0.054364	74.206	-0.3208	0.749261
Repetitionsecond	-0.05104	0.019039	132.8395	-2.68099	**0.008271** **
Repetitionthird	-0.13841	0.020198	141.6205	-6.85292	**2.05E-10*** **
mLobF1	-0.06888	0.077762	194.6508	-0.88573	0.376856
mLobF2	0.049469	0.067546	175.6177	0.732372	0.464918
Gendermale	-0.16429	0.040979	80.77156	-4.00901	**0.000135*** **

In the data for this current study, no statistically significant inter-region differences can be observed. Conforming to the trend observed in the results of the pilot study, table 3.10 indicates that [e] in the standard variety is longer than all the other regional variants. The Sikkim variant is the shortest. Repetition affected vowel duration as there was significant difference between all the three repetitions. There was no effect of F1 or F2 on vowel duration.

Gender proved to be significant for vowel duration of [e] (p=0.000135). The estimate for males is negative indicating that females produce longer tokens compared to males.

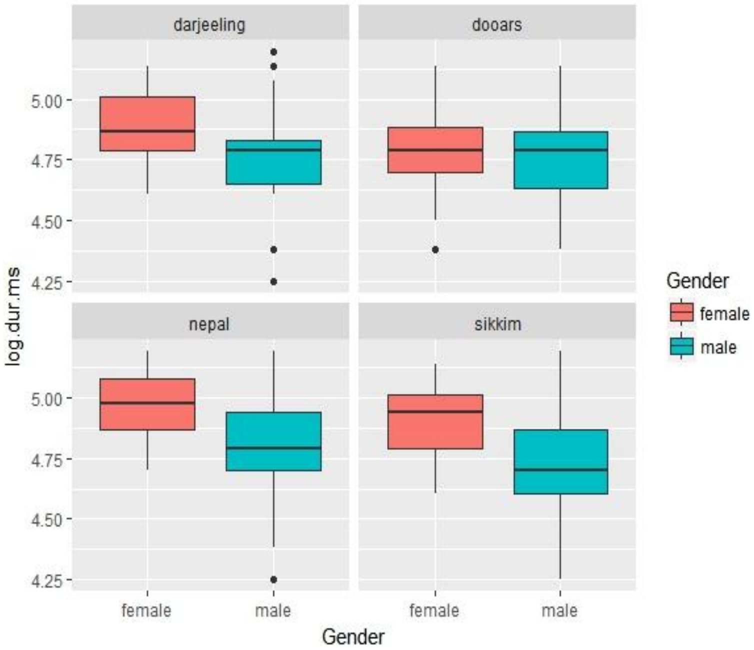


Figure 3. 11 Vowel duration of [e] for males and females across regions

From the plot in figure 3.11 it can be seen that the range and median for females is higher in relation in all the regions except Dooars. In Dooars the medians are similar for males and females. This finding identifies gender as sociolinguistic variable for the vowel duration of [e] and provides scope for further studies with data from other elicitation styles.

3.5.4 f_0 for [e]

Analogous to the procedure followed for the [i] vowel, fundamental frequency data for [e] was divided according to gender groups. The following two sub-sections present the data and analysis for the fundamental frequency for females and males.

3.5.5 f_0 for females for vowel [e]

Figure 3.12 display the density distribution of fundamental frequency for female speakers. The mean of the distribution is 230.6 Hz whereas the median of the distribution is 226.5 Hz. The mean and median of the distribution is not far off and the bell shaped curves for each of the regions indicate that the data is normally distributed.

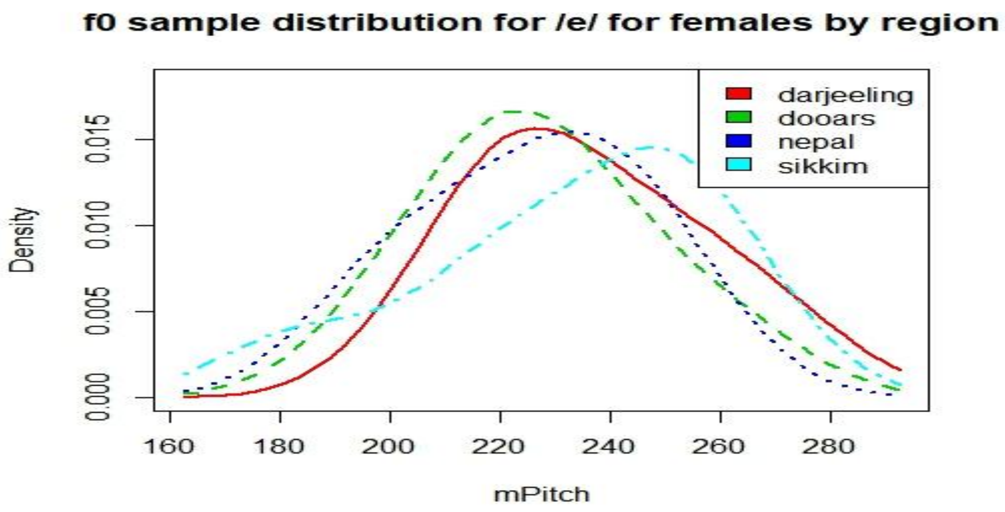


Figure 3. 12 Sample density distribution of f_0 for female speakers for [e]

A linear model of fundamental frequency for females as a function of region and normalized F1 was constructed. The model was not significant (F(4,23)=1.477, p=0.2416). Fixed effects estimates from the model are displayed in table 3.11

Table 3. 11 Fixed effects estimates of f_0 for females for [e]

Fixed effects estimates of f_0 for females for [e]				
	Estimate	Std..Error	t.value	Pr...t..
(Intercept)	250.8383	13.74349	18.25142	**3.52E-15*** **
relevel(Region, "nepal")darjeeling	5.934582	11.13497	0.532968	0.599168
relevel(Region, "nepal")dooars	1.927622	10.98308	0.175508	0.862216
relevel(Region, "nepal")sikkim	3.130721	11.12039	0.28153	0.780822
relevel(Region, "darjeeling")dooars	-4.00696	11.18756	-0.35816	0.723488
relevel(Region, "darjeeling")sikkim	-2.80386	10.97953	-0.25537	0.800706
relevel(Region, "dooars")sikkim	1.203098	11.1707	0.107701	0.915167
mLobF1	62.40773	28.49434	2.19018	**0.038912***

There are no significant inter-group differences between the regional variants. The f_0 for female speakers speaking the standard variant is lower than it is for females from other regions. The second lowest f_0 values are reported for females from the Dooars area. Darjeeling female speakers have the highest values for f_0. The height of [e] denoted by normalized F1 values in the model emerged as a significant predictor of f_0 for female speakers. The estimate indicates that the f_0 increases with an increase in vowel height, a relationship that can be observed in figure 3.13.

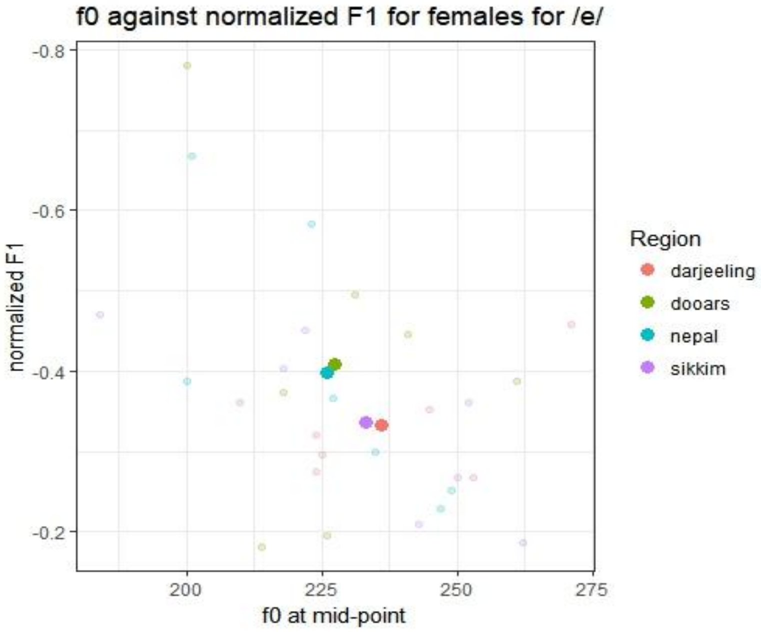


Figure 3. 13 Fundamental frequency against F1 for vowel [e]

3.5.6 f_0 for male speakers for vowel [e]

The mean of f_0 for male speakers is 144.3 Hz whereas the median is 142 Hz. Figure 3.14 below shows the density for the sample distribution of fundamental frequency for male speakers in the four different regions.

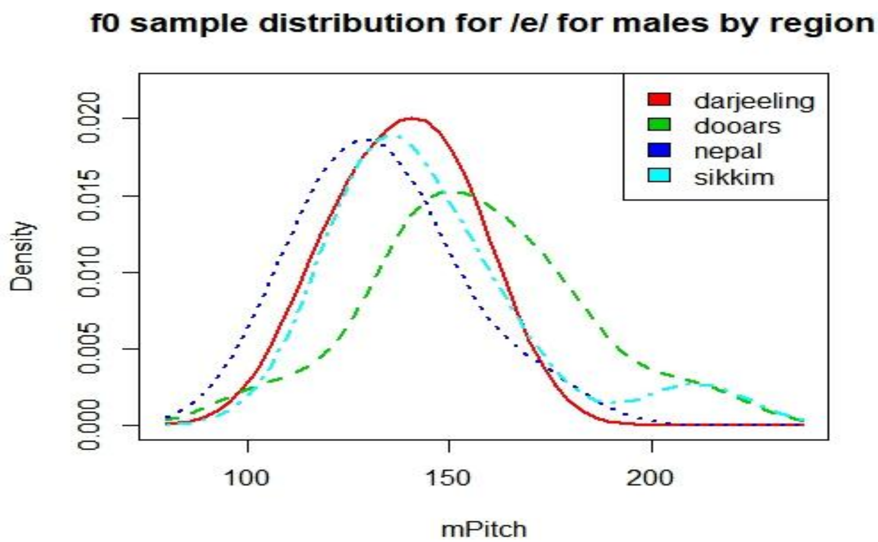


Figure 3. 14 Sample density distribution of f_0 for male speakers for [e]

A linear model of fundamental frequency for male speakers was constructed with region and normalized F1 as the predictors. The model was not significant ($F(4,37)=1.661$, $p=0.1798$). Fixed effects estimates from the model are displayed in table 3.12

Table 3. 12 Fixed effects estimates of f_0 for males for [e]

Fixed effects estimates of f_0 for males for [e]				
	Estimate	Std..Error	t.value	Pr...t..
(Intercept)	128.9932	15.57977	8.279534	**6.03E-10*** **
relevel(Region, "nepal")darjeeling	4.859313	10.53801	0.461122	0.64741
relevel(Region, "nepal")dooars	23.38286	9.87456	2.36799	**0.023222***
relevel(Region, "nepal")sikkim	12.89875	10.09025	1.278338	0.209091
relevel(Region, "darjeeling")dooars	18.52355	9.807358	1.88874	0.066785
relevel(Region, "darjeeling")sikkim	8.039437	9.902932	0.811824	0.422085
relevel(Region, "dooars")sikkim	-10.4841	9.36027	-1.12007	0.269904
mLobF1	-8.18042	28.71804	-0.28485	0.777345

The only statistically significant inter-group difference was observed between male speakers from Dooars and Nepal. Fundamental frequency for male speakers from Dooars is significantly higher in comparison to their male counterparts from Nepal who speak the standard variant as can be seen in figure 3.15 below. It is also higher than male speakers from Darjeeling and Sikkim. Nepal speakers have the lowest values of f_0 for males while readings for the Darjeeling and the Sikkim variant is slightly higher in its comparison.

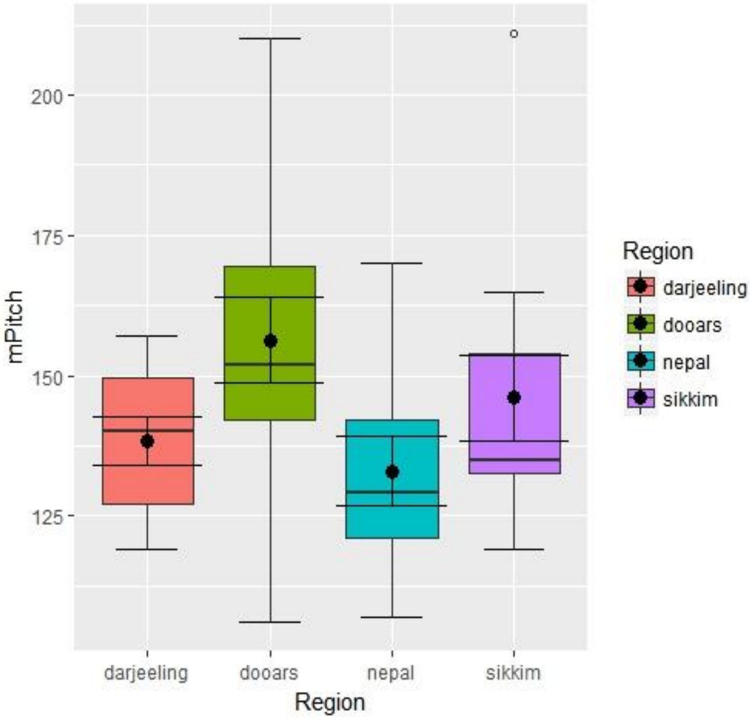


Figure 3. 15 Comparison of fundamental frequency for males speakers for [e]

3.5.7 Summary of findings for [e]

The following is a summary of findings for [e]:

1. Region as an independent variable was observed to have no effect on F1.
2. There were no statistically significant inter-group differences between regions for F1.
3. Darjeeling [e] is highest followed by Dooars, Nepal and Sikkim.
4. Repetition had an expected significant effect on F1.
5. Gender had a significant effect on F1 even though the values were normalized. F1 for males was lower than their female counterparts. This probably is an artifact of the normalization process.
6. Region as an independent variable affected F2 Region affected F2 of [e] ($\chi2$ (3) = 8.2164, p = 0.04174*
7. The Dooars variant is comparatively back-er in relation to the standard variant. In other words, the standard [e] is more fronted in comparison to the Dooars variant. This difference is statistically significant.
8. The Darjeeling variant is significantly ahead in the vowel acoustic space than the Dooars variant.
9. Nepal is the most fronted followed by Darjeeling, Sikkim and Dooars.

10. Mutivariate regression analysis with random effects revealed no significant effect of region on vowel duration.
11. There were no significant inter-group differences between regions.
12. [e] in Nepal variety is longest followed by the Darjeeling, Dooars and Sikkim variants.
13. F1 and F2 did not affect vowel duration for [e].
14. Gender proved to be significant for vowel duration of [e] ($p=0.000135$).
15. Females produce longer tokens compared to males.
16. Linear mixed effects models for fundamental frequency as a function of region for both male and female speakers proved insignificant for the data.
17. Among female speakers, there were no significant between-group differences for fundamental frequency.
18. Vowel height emerged as significant predictor of fundamental frequency for female speakers.
19. Among male speakers, fundamental frequency for male speakers from Dooars is significantly higher in comparison to their male counterparts from Nepal.

3.6. Vowel [ɑ]

The vowel [ɑ] has been classified as a low central vowel by Acharya (1991), Khatiwada (2007) and Pokharel (1979) and the same is observed by Srivastava and Nakkeerar in their reports on Nepali in West Bengal and Sikkim, respectively. For analysis of [ɑ], a total of 211 tokens were analyzed after removing 20 outlier points from the original corpus.

3.6.1 F1 for [ɑ]

The mean for the overall distribution of normalized F1 for [ɑ] in the dataset is 1.836 and the median is also 1.836. The sample density for normalized F1 for [ɑ] indicates that the data is normally distributed for almost all the regions barring Sikkim which displays a bimodal distribution. However, since data for all other factor levels are normally distributed, data for Sikkim has been included for regression modeling.

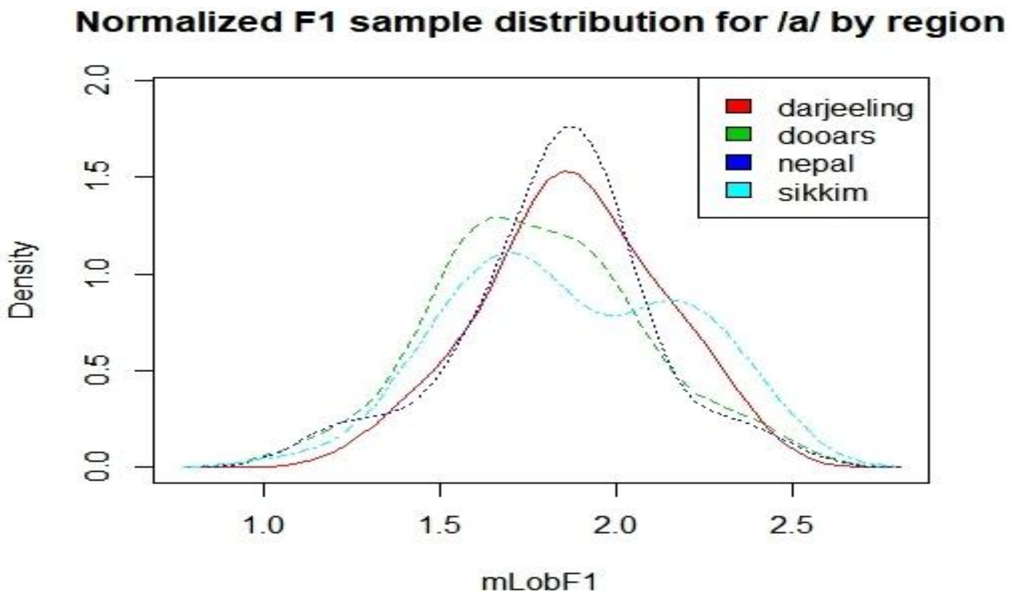


Figure 3. 16 Sample density distribution of F1 for [ɑ] by region

Two linear mixed effects models: full model with normalized F1 as the dependent variable and region, gender, log transformed duration of [ɑ] and gender as fixed effects along with speaker as a random effect. A likelihood ratio test was performed between the full and null models to test the significance of the model. Results indicate that there was not significant with respect to region ($\chi 2$ (3) = 5.4161, p = 0.1437). Table 3.13 below provides estimates of the fixed effects estimates.

Table 3. 13 Fixed effects estimates of normalized F1 for [ɑ]

Fixed effects estimates of normalized F1 for [ɑ]					
	Estimate	Std..Error	df	t.value	Pr...t..
(Intercept)	2.538516	0.452893	211.0001	5.605118	**6.46E-08*** **
relevel(Region, "nepal")darjeeling	0.0391	0.049622	211.0001	0.787967	0.4316
relevel(Region, "nepal")dooars	-0.05856	0.049213	211.0001	-1.18992	0.235413
relevel(Region, "nepal")sikkim	0.031585	0.049281	211.0001	0.640925	0.522267
relevel(Region, "darjeeling")dooars	-0.09766	0.04649	211.0001	-2.10066	**0.036858***

relevel(Region, "darjeeling")sikkim	-0.00751	0.0475	211.0001	-0.15821	0.874441
relevel(Region, "dooars")sikkim	0.090145	0.04698	211.0001	1.918778	0.056362
Repetitionsecond	-0.21423	0.041779	211.0001	-5.12757	**6.63E-07*** **
Repetitionthird	-0.31876	0.043006	211.0001	-7.41185	**2.97E-12*** **
log.dur.ms	-0.11042	0.091217	211.0001	-1.21052	0.227433
Gendermale	0.007494	0.034733	211.0001	0.215751	0.82939

The estimates of the fixed effects estimates indicate that there is a significant difference between the Darjeeling and the Dooars varieties regarding the height of [ɑ]. The negative estimate for Dooars with respect to Darjeeling indicates that the Dooars variant is positioned higher in the vowel acoustic space. Although insignificant, the Dooars variety is also higher than the Sikkim and the Nepal variant. The relative positioning of the four regional variants is demonstrated in figure 3.17 below with the help of a scatter plot with normalized F2 of [ɑ] on the x-axis and normalized F1 of [ɑ] on the y-axis. From the above scatter plot it is evident that the Darjeeling [ɑ] is positioned lowest among other regional variants. Nepal [ɑ] is higher than the Sikkim variant.

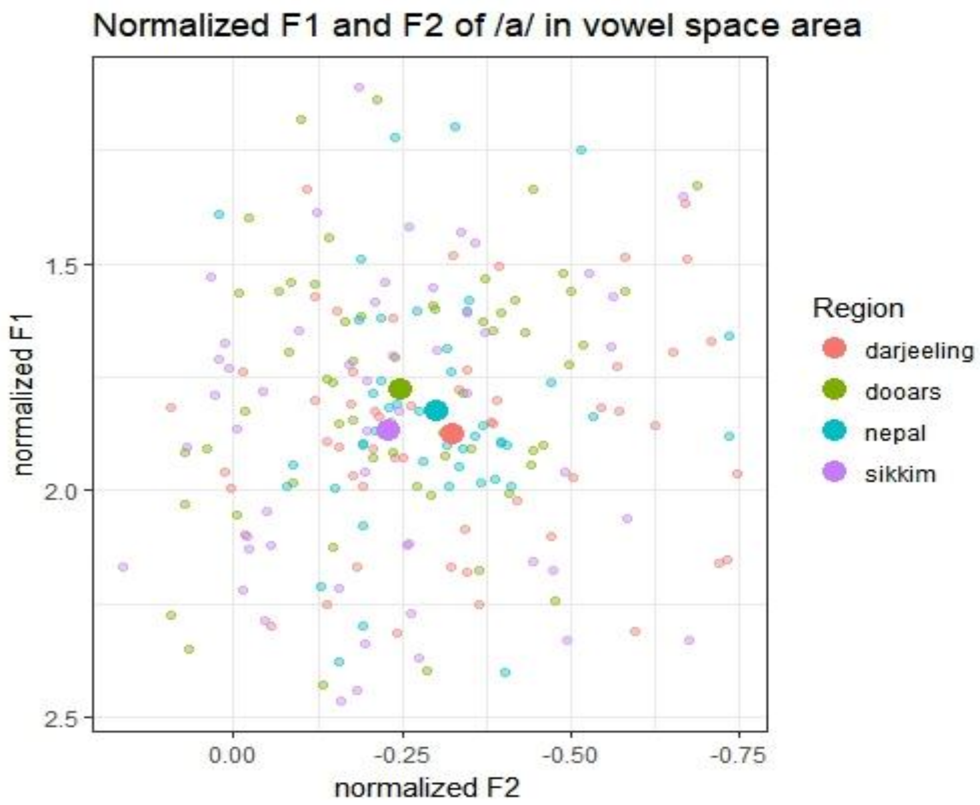


Figure 3. 17 Distribution of [ɑ] in vowel space area

Duration of [ɑ] was observed to have no significant bearing on F1. The negative estimate for duration with respect to F1 signify the process of vowel undershoot where vowels do not reach their intended acoustic target if the duration is short. Gender also had no significant bearing on the height of [ɑ]. The positive estimate indicates that with reference to females, the mean of normalized F1 for male speakers is marginally higher thereby lowering the [ɑ] for males in the vowel acoustic space.

3.6.2 F2 for [ɑ]

An analysis of variance using the likelihood ratio test failed to observe a significant effect of region on normalized F2 measures for the vowel [ɑ] ($\chi2$ (3) = 3.7308, p = 0.292). While speaker was included as a random effect in the model; region, gender, log transformed vowel duration and repetition were included as fixed effects. Table 3.14 below presents the estimates of fixed effects estimates from the model.

Table 3. 14 Fixed effects estimates of normalized F2 for [ɑ]

Fixed effects estimates of normalized F2 for [ɑ]					
	Estimate	Std..Erro r	df	t.value	Pr...t..

(Intercept)	-0.12942	0.351888	209.2202	-0.36779	0.713405
relevel(Region, "nepal")darjeeling	-0.03896	0.057211	74.65044	-0.68105	0.497945
relevel(Region, "nepal")dooars	0.042109	0.057196	74.26038	0.736216	0.463919
relevel(Region, "nepal")sikkim	0.058485	0.057043	73.90535	1.025279	0.308576
relevel(Region, "darjeeling")dooars	0.081073	0.054555	73.52908	1.486077	0.141534
relevel(Region, "darjeeling")sikkim	0.097449	0.054675	74.33108	1.78234	0.078777 .
relevel(Region, "dooars")sikkim	0.016376	0.054543	73.62619	0.300247	0.764835
Repetitionsecond	-0.01127	0.017016	143.0944	-0.66253	0.508698
Repetitionthird	-0.02713	0.019048	164.8559	-1.42428	0.156256
log.dur.ms	-0.04014	0.070495	210.007	-0.56944	0.569667
Gendermale	0.078147	0.04012	74.51668	1.947845	0.0552 .

The fixed effects estimates indicate no significant inter-group differences between the four regions. While the Dooars and Sikkim variants are more fronted than the standard variant, the Darjeeling variant is marginally back in comparison to the standard variant as can be observed in the earlier plot of F2 against F1. No significant effect of repetition, vowel duration or gender was observed on normalized F2 values for [ɑ].

3.6.3 Vowel duration of [ɑ]

Figure 3.18 displays the sample density of the distribution of log transformed vowel duration for the vowel [ɑ]. The mean of the distribution is 4.788 and the median is 4.787 which is almost equal.

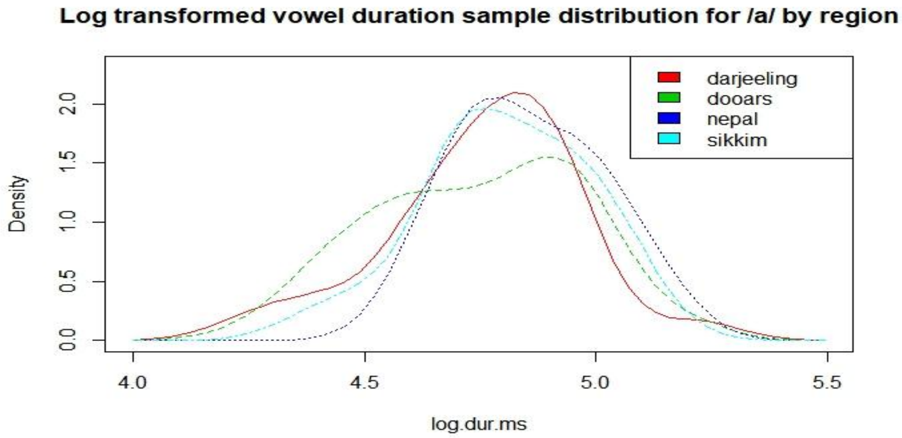


Figure 3. 18 Sample density distribution of vowel duration for [ɑ]

A linear model for vowel duration for [ɑ] was constructed with region, gender, normalized F1 and normalized F2 as fixed effects and speaker as a random effect. This model was compared to a reduced or a null model by excluding region as a fixed effect in the null model through a Likelihood ratio test using the anova function in R. In the current dataset, region did not affect vowel duration ($\chi 2$ (3) = 2.7756, p = 0.4309). Table 3.15 provides estimates of the fixed effects.

Table 3. 15 Fixed effects estimates for vowel duration of [ɑ]

Fixed effects estimates for vowel duration of [ɑ]					
	Estimate	Std..Error	df	t.value	Pr...t..
(Intercept)	5.035396	0.082586	209.0989	60.97188	< 2e-16 ***
relevel(Region, "nepal")darjeeling	-0.06638	0.058486	73.09148	-1.13506	0.26006
relevel(Region, "nepal")dooars	-0.09562	0.058371	72.54162	-1.63813	0.105726
relevel(Region, "nepal")sikkim	-0.04735	0.058482	73.13002	-0.80964	0.420773
relevel(Region, "darjeeling")dooars	-0.02924	0.056291	74.14652	-0.51936	0.605056
relevel(Region, "darjeeling")sikkim	0.019035	0.056385	74.66257	0.337592	0.736618

relevel(Region, "dooars")sikkim	0.04827	0.055874	72.66919	0.863917	0.390476
Repetitionsecond	-0.08197	0.016326	132.1178	-5.02084	1.63E-06***
Repetitionthird	-0.15188	0.017807	134.5719	-8.52915	2.69E-14***
Gendermale	-0.03706	0.041351	74.68202	-0.89624	0.373007
mLobF1	-0.04986	0.029307	143.0944	-1.70135	0.091049
mLobF2	-0.03046	0.067476	209.556	-0.45147	0.652114

For vowel duration, it was not surprising that repetition emerged as a significant predictor with vowel duration decreasing after every subsequent repetition. No significant inter-group differences between regions were observed. The pattern seen for [i] and [e] is seen repeating for this vowel category as well as far as vowel duration is concerned with the standard variant being longer in duration compared to the Darjeeling, Dooars and Sikkim variants. The Dooars speakers have the lowest duration for [ɑ] while the [ɑ] of Sikkim speakers is marginally longer than that of speakers from Darjeeling.

There was no significant effect of gender on vowel duration of [ɑ]. The negative estimate for male speakers indicates that female speakers have marginally higher vowel duration values. There were no significant effects of either F1 or F2 on vowel duration.

3.6.4 f_0 for [ɑ]

Fundamental frequency for [ɑ] has been analyzed separately for males and females as f_0 values were left un-normalized. For both males and females, two linear regression models were created with region and F1 as the independent variable and fundamental frequency at vowel mid-point as the dependent variable.

4.6.5 f_0 for females for vowel [ɑ]

Figure 3.19 displays the sample density distribution of fundamental frequency for female speakers. The mean for the distribution is 209.4 Hz while the median is 214 Hz. The two small humps in the left tail are data points which did not show up as outliers in the data. The bulk of the data however is normally distributed.

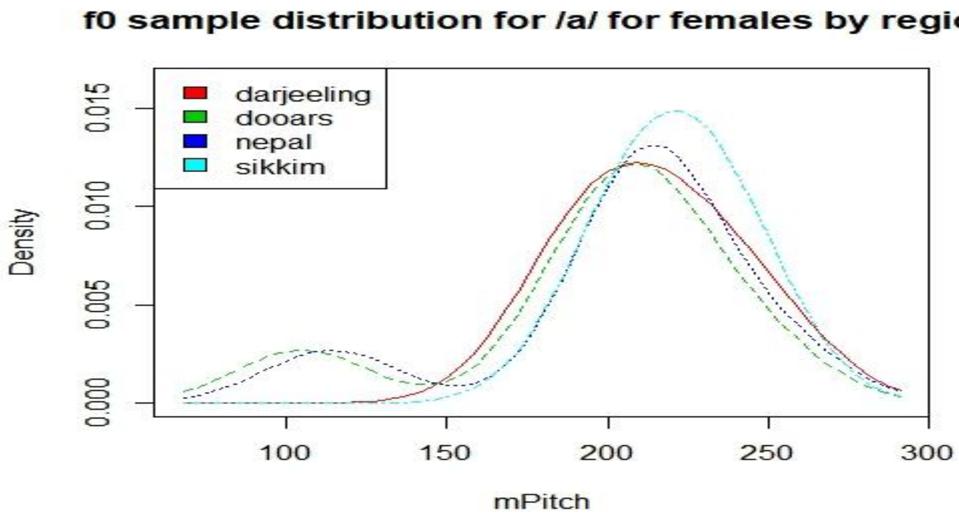


Figure 3. 19 Sample density distribution of f0 for female speakers for [ɑ]

A linear model of fundamental frequency for females as a function of region and normalized F1 was constructed. This model was not significant (F(4,24)=0.9987,p=0.09135). Table 3.16 displays the model estimates.

Table 3. 16 Fixed effects estimates for f_0 for female speakers for [ɑ]

Fixed effects estimates for f_0 for female speakers for [ɑ]				
	Estimate	Std..Error	t.value	Pr...t..
(Intercept)	298.9912	38.80342	7.70528	**6.09E-08*** **
relevel(Region, "nepal")darjeeling	17.55576	16.19292	1.084163	0.289064
relevel(Region, "nepal")dooars	-16.7847	16.9112	-0.99252	0.330846
relevel(Region, "nepal")sikkim	18.10306	16.46703	1.099352	0.28252
relevel(Region, "darjeeling")dooars	-34.3405	17.3023	-1.98473	0.058719 .
relevel(Region, "darjeeling")sikkim	0.547299	16.12251	0.033946	0.973201
relevel(Region, "dooars")sikkim	34.89	17.01	2.050	0.0514.
mLobF1	-52.1718	20.36108	-2.56233	**0.017096***

There are no significant inter-group differences between the four regions for this dataset comprising only females. Female speakers from Sikkim have the highest values for fundamental frequency for [ɑ] followed by females in Darjeeling, Dooars and Nepal. Vowel height is a significant predictor of fundamental frequency for females. The negative estimate for normalized F1 signifies that fundamental frequency increases with an increase in vowel height. Lohagun (2016,pp.318) had demonstrated this directly proportional relationship between vowel height and fundamental frequency for Nepali .

3.6.6 f_0 for males

Figure 3.20 below displays the sample density distribution of fundamental frequency readings for male speakers for the vowel [ɑ]. The mean of the distribution is 133.5 Hz while the median is 130.5 Hz.

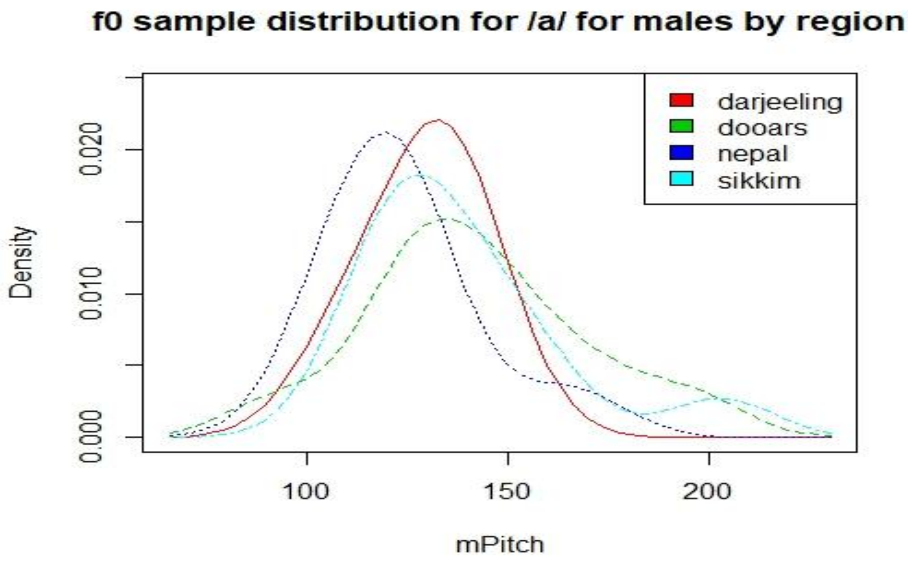


Figure 3. 20 Sample density distribution of f0 for male speakers for [ɑ]

A simple linear regression model of fundamental frequency for males as a function of region and normalized F1 was constructed. This model was not significant ($F(4,37)= 1.118, p=0.3626$). Table 3.17 displays the model estimates.

Table 3. 17 Fixed effects estimates for f_0 for male speakers for [ɑ]

Fixed effects estimates for f_0 for male speakers for [ɑ]				
	Estimate	Std..Error	t.value	Pr...t..
(Intercept)	135.533	28.04207	4.833202	2.36E-05***

relevel(Region, "nepal")darjeeling	5.70803	10.10542	0.564848	0.575586
relevel(Region, "nepal")dooars	17.50078	10.16844	1.721088	0.093588 .
relevel(Region, "nepal")sikkim	14.84621	10.07444	1.473651	0.149033
relevel(Region, "darjeeling")dooars	11.79275	9.834904	1.199071	0.238123
relevel(Region, "darjeeling")sikkim	9.138181	9.643672	0.947583	0.34949
relevel(Region, "dooars")sikkim	-2.65457	9.600636	-0.2765	0.783704
mLobF1	-6.51258	14.97532	-0.43489	0.66617

As in the earlier analysis of female speakers, there are no significant differences in fundamental frequency values for male speakers between the different regions. The estimates indicate that pitch of male speakers from Nepal is the lowest compared to males from other three regions. Dooars speakers have the highest fundamental frequency for [ɑ] whereas it is slightly higher in the Sikkim variant compared to the Darjeeling variant. Vowel height has no significant effect on the fundamental frequency for males but the negative estimate attest the trend observed in the data for females that vowel height is directly proportional to fundamental frequency.

3.6.7 Summary of findings for [ɑ]

The following is a summary of findings for [ɑ]:

1. Region did not affect F1 for [ɑ].
2. With respect to Darjeeling , the Dooars variant is positioned higher in the vowel acoustic space. This difference is statistically significant (p=0.036858*).
3. Darjeeling [ɑ] is positioned lowest among other regional variants while the Sikkim and the Nepal variant is marginally higher.
4. No significant effect of gender of vowel duration on vowel height. Estimates, however, underscore the impact of vowel undershoot on F1.
5. Region did not affect F2 for [ɑ].
6. While the Dooars and Sikkim variants are more fronted than the standard variant, the Darjeeling variant is marginally back in comparison to the standard variant.

7. No significant effect of repetition, vowel duration or gender was observed on normalized F2 values for [ɑ].
8. Region did not affect vowel duration.
9. No significant inter-group differences between regions were observed.
10. Standard variant is longer in duration compared to the Darjeeling, Dooars and Sikkim variants.
11. Dooars speakers have the lowest duration for [ɑ] while the [ɑ] of Sikkim speakers is marginally longer than that of speakers from Darjeeling.
12. No significant effects of either F1 or F2 on vowel duration.
13. Region did not affect fundamental frequency for females.
14. There were no significant inter-group differences with respect to region.
15. Vowel height is a significant predictor of fundamental frequency for females ($p=0.017096^*$).
16. Region did not affect fundamental frequency for males.
17. Pitch of male speakers from Nepal is the lowest compared to males from other three regions.
18. Dooars speakers have the highest fundamental frequency for [ɑ] whereas it is slightly higher in the Sikkim variant compared to the Darjeeling variant.

3.7 Vowel [ʌ]

The vowel [ʌ] in Nepali has been subjected to different interpretations in existing literature. Impressionistic accounts by Bandhu et al. (1971), Dahal (1974) and Acharya (1991) classify it as / ə /. Srivastava has described it as a schwa in his description of Nepali in West Bengal. Nakkeerar in his report on Nepali in Sikkim lists / ə / and / ɔ / as separate phonemes although the contrasts he provides for them indicate that they are allophones rather than phonemes. Results from instrumental accounts such as Pokharel (1989), Khatiwada (2009) and Lohagun (2016) demonstrate that [ʌ]is further back than a schwa in the vowel acoustic space.

Khatiwada (2009) observes that there is considerable variation in the realization of this vowel and proposes that [ʌ] should be adopted as the norm for classifying the vowel. This proposition is strengthened by results of the instrumental studies where the vowel has manifested itself as [ʌ]. This study adopts Khatiwada’s proposition. In the current study, a total of 223 tokens for [ʌ] have been analyzed after removal of 9 outliers from the original corpus.

3.7.1 F1 for [ʌ]

Figure 3.21 displays the sample density distribution of normalized F1 values for [ʌ]. The mean of the distribution is 0.654 whereas the median is 0.664. The density plot indicates that the data is normally distributed.

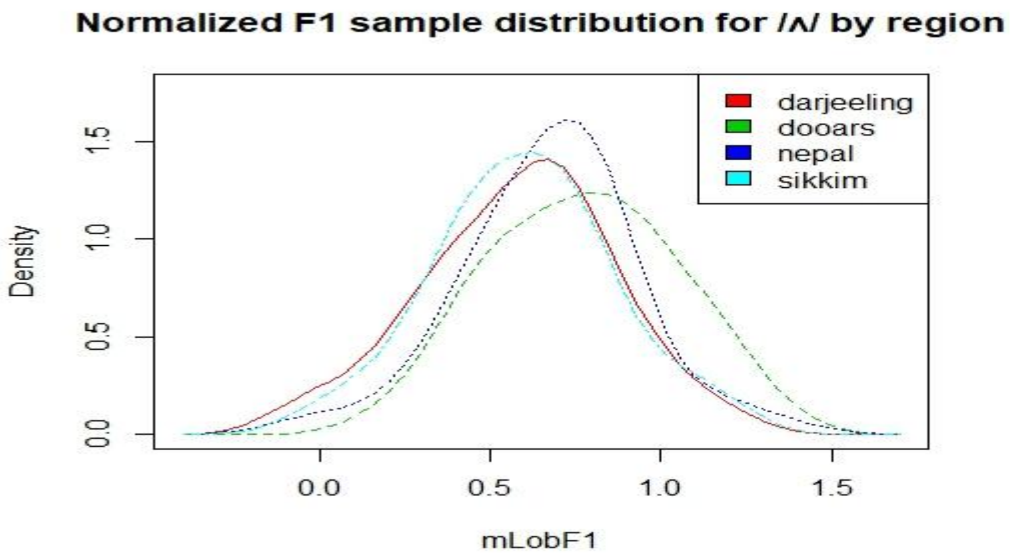


Figure 3. 21 Sample density distribution of F1 for [ʌ] by region

Two linear mixed effects models with random effects were constructed to observe the effect on the height of the vowel [ʌ]. The first was a full model with region, gender, vowel duration and repetition as fixed effects along with speaker as a random effect to account for repeated measurements for the word. The second model was a null identical to the first but with region missing as a fixed effect. The two models were compared using a likelihood ratio test to gauge the significance of region in the model. Region affected the height of a vowel [ʌ]($\chi2$ (3) = 12.941, p = 0.004767**). Table 3.18 below provides the fixed effects estimates for the full model.

Table 3. 18 Fixed effects estimates of normalized F1 for [ʌ]

Fixed effects estimates of normalized F1 for [ʌ]					
	Estimate	Std..Error	df	t.value	Pr...t..
(Intercept)	1.207536	0.440427	183.2819	2.741741	**0.006717****
relevel(Region,	-0.1221	0.062832	76.67911	-1.94327	0.055654 .

"nepal")darjeeling					
relevel(Region, "nepal")dooars	0.082683	0.061891	74.59523	1.335958	0.185625
relevel(Region, "nepal")sikkim	-0.09719	0.063506	78.42171	-1.53039	0.129944
relevel(Region, "darjeeling")dooars	0.204782	0.060336	74.55201	3.394036	**0.001106****
relevel(Region, "darjeeling")sikkim	0.02491	0.059814	73.14892	0.416458	0.678294
relevel(Region, "dooars")sikkim	-0.17987	0.060821	75.78605	-2.9574	**0.004135****
Repetitionsecond	-0.18752	0.030755	142.8304	-6.09727	**9.56E-09*** **
Repetitionthird	-0.23694	0.031074	146.073	-7.625	**2.86E-12*** **
log.dur.ms	-0.09409	0.094376	184.1964	-0.99699	0.320075
Gendermale	0.083946	0.043624	73.51524	1.9243	0.058188 .

From table 3.18 on fixed effects estimates on normalized F1 for /ʌ/, it can be observed that the standard variant does not different significantly with other regional variants as far as F1 is concerned. The estimate for the Darjeeling variant with reference to the standard variant indicates that Is positioned higher in the vowel acoustic space. This difference is close to achieving statistical significance (p=0.1). The positive estimate for the Dooars variant with reference to the standard variant indicates that it is marginally lower than the standard variant. Judging by the estimates, the Sikkim variant is higher than the Nepal variant as can be seen in figure 3.22 below.

There is a significant difference in the height of [ʌ] between speakers from Darjeeling and Dooars (p=0.001106**). The positive estimate for Dooars speakers with reference to Darjeeling speakers suggests that it is lower in the vowel acoustic space than the Darjeeling variant. The Sikkim variant is also lower than the Darjeeling variant but this difference is not statistically significant. The Sikkim variant, however, is higher than the Dooars variant and this difference is significant (p=0.004135**).

Repetition, as expected, had a significant effect on normalized F1 values. The effect of vowel duration on F1 was not significant for this dataset. The estimate indicates that as duration decreases with an increase in vowel height. This is expected because if the vowel lacks the time to reach it intended target, it will be positioned higher in the vowel acoustic space contrary to its inherent character.

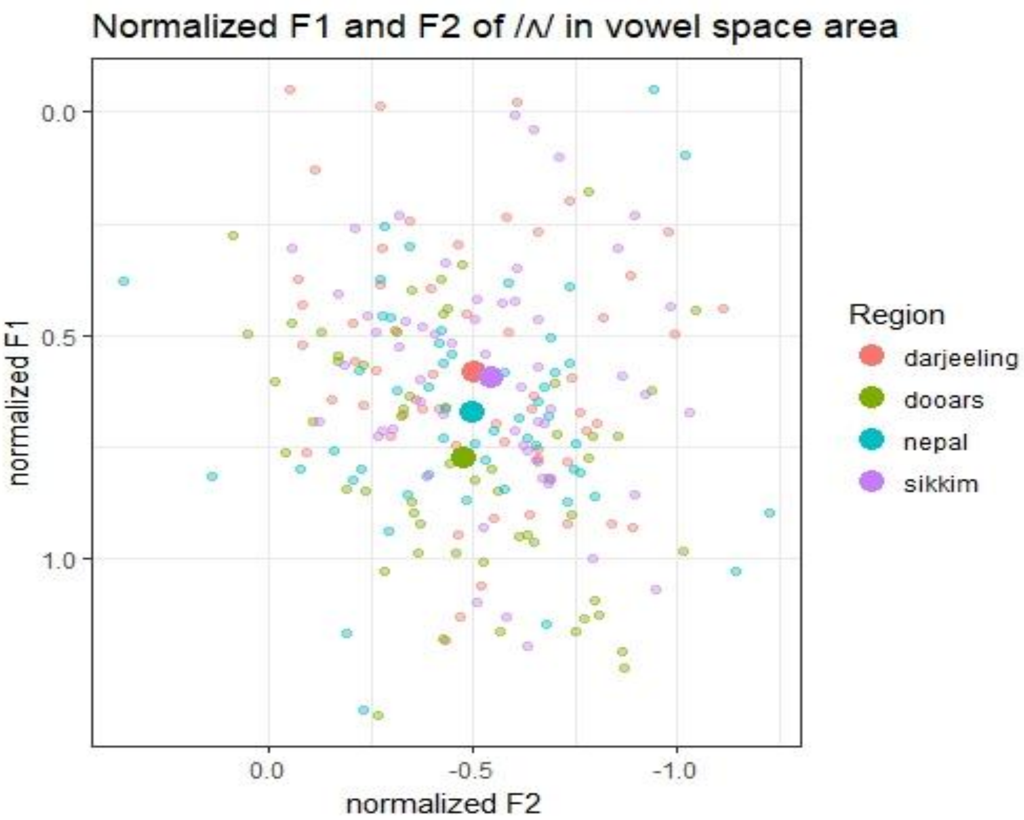


Figure 3. 22 Distribution of [ʌ] in vowel space area

3.7.2 F2 for [ʌ]

F2 in the context of [ʌ]indicates the position of [ʌ]along the front back dimension. At this stage of the analysis, the position of [ʌ] relative to other vowels is not illustrated. This will be taken up in the analysis of vowel acoustic space in the following sections. The current section presents the data and analyses the positioning of [ʌ] in isolation among the four different regions. Figure 3.23 presents the sample density of the distribution of normalized F2 values for [ʌ]. The mean of the distribution is -0.5063 while the median is -0.5080.

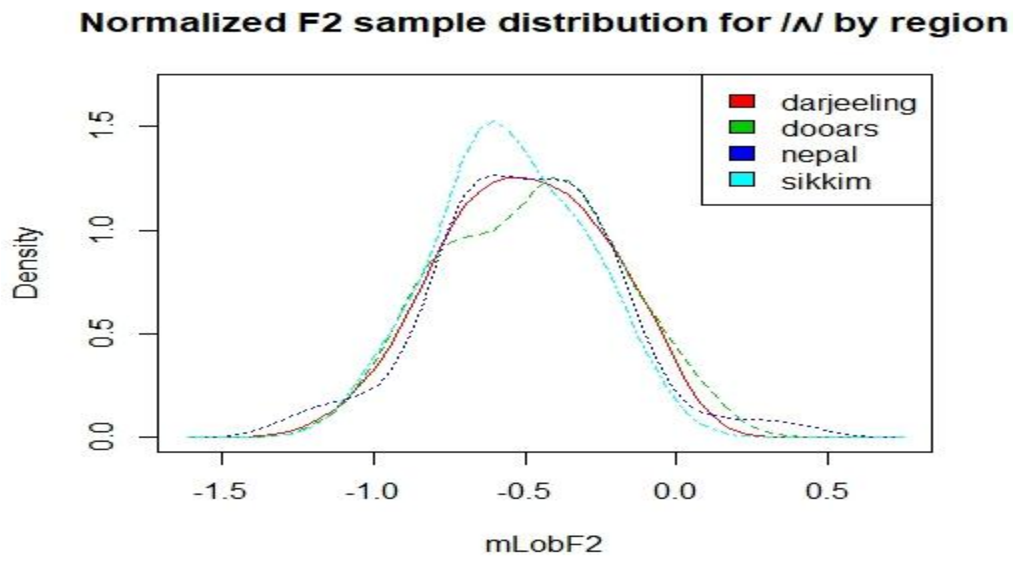


Figure 3. 23 Sample density distribution of F2 for [ʌ] by region

Similar to the two models created for F1 values, two models were created for the analysis of normalized values of F2 as well keeping F2 as the dependent variable. The full and the null models were compared using a likelihood ratio test. Region did not affect the F2 of the vowel [ʌ]($\chi2$ (3) = 2.8037, p = 0.4229). Table 3.19 below present the fixed effects estimates from the model.

Table 3. 19 Fixed effects estimates of normalized F2 for [ʌ]

Fixed effects estimates of normalized F2 for [ʌ]					
	Estimate	Std..Error	df	t.value	Pr...t..
(Intercept)	0.637605	0.393174	222.8709	1.621684	0.106285
relevel(Region, "nepal")darjeeling	-0.03667	0.070871	75.73473	-0.51739	0.606393
relevel(Region, "nepal")dooars	-0.00092	0.07021	73.90309	-0.01313	0.989556
relevel(Region, "nepal")sikkim	-0.1033	0.071312	77.12449	-1.44857	0.151512
relevel(Region, "darjeeling")dooars	0.035746	0.068425	74.12395	0.52241	0.602943
relevel(Region, "darjeeling")sikkim	-0.06663	0.068078	73.07899	-0.97876	0.330925

relevel(Region, "dooars")sikkim	-0.10238	0.068744	75.14403	-1.48927	0.140602
Repetitionsecond	0.0942	0.022396	141.3931	4.206132	4.59E-05***
Repetitionthird	0.138232	0.022721	143.6283	6.083819	1.01E-08***
log.dur.ms	-0.24374	0.083972	222.9966	-2.90263	0.004072**
Gendermale	-0.12758	0.049615	73.23065	-2.57138	0.012159*

The estimates and the plot in figure 3.22 indicate that there are no significant inter-group differences between the four regions. The Dooars variant is marginally fronted compared to all other variants followed by the Nepal variant which is slightly ahead of the Darjeeling and Sikkim variants. Vowel duration has a significant effect on F2 values for [ʌ](p=0.004072**). The negative estimate indicates an increase in F2 with a decrease in vowel duration, again underscoring the process of vowel undershoot. An increase in F2 would mean the vowel being more fronted.

The estimates also indicate a significant effect of gender on F2 of [ʌ] (p=0.012159*). The negative estimate for male speakers signifies that the mean of their F2 is lesser than the mean of female speakers. Despite the process of vowel normalization with the Lobanov formula, the findings suggest that [ʌ] for female speakers is fronted than male speakers.

3.7.3 Vowel duration of [ʌ]

Figure 3.24 below shows the sample density distribution of log transformed duration of the vowel [ʌ]. The mean of the distribution is 4.529 while the median is 4.500. A full and null models with region as the difference in the null model were created for the analysis of vowel duration. Log transformed vowel duration was the dependent variable and region, F1, gender and repetition were added as fixed effects.

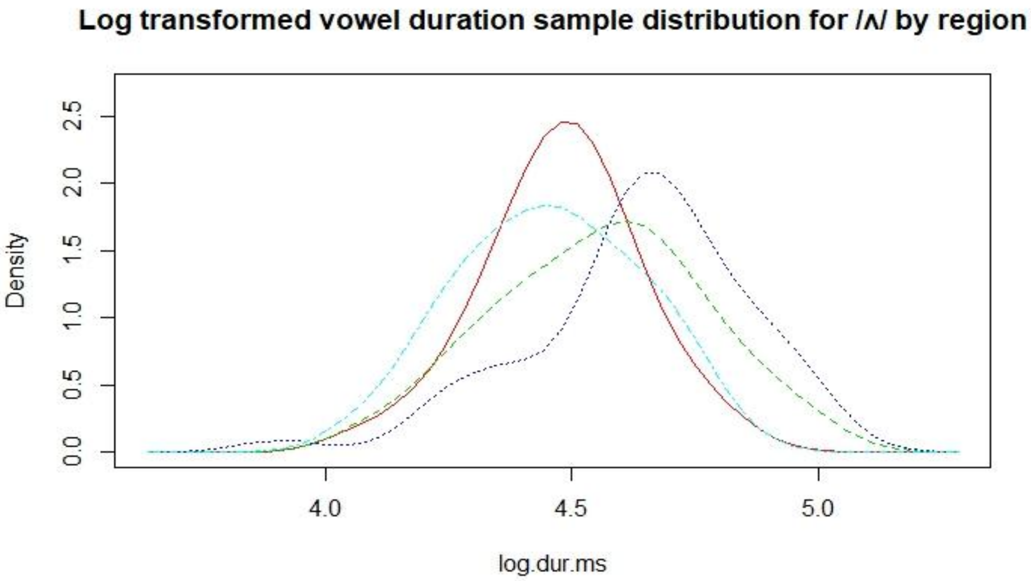


Figure 3. 24 Sample density distribution of vowel duration for [ʌ]

Results from the likelihood ratio test obtained by performing the anova () function in R on the full and null linear mixed effects models indicate that region affects vowel duration ($\chi2$ (3) = 13.79, p = 0.003206**) significantly.

The fixed effects estimates for vowel duration are provided in table 3.20 below.

Table 3. 20 Fixed effects estimates for vowel duration of [ʌ]

Fixed effects estimates for vowel duration of [ʌ]					
	Estimate	Std..Error	df	t.value	Pr...t..
(Intercept)	4.586936	0.059185	156.9227	77.5019	**< 2e-16 *** **
relevel(Region, "nepal")darjeeling	-0.15013	0.051284	73.52367	-2.92753	**0.004545 ** **
relevel(Region, "nepal")dooars	-0.05878	0.051339	73.91094	-1.14494	0.255928
relevel(Region, "nepal")sikkim	-0.17953	0.051403	73.83413	-3.49266	**0.000812*** **
relevel(Region, "darjeeling")dooars	0.091355	0.050564	76.7551	1.806709	0.074727 .

relevel(Region, "darjeeling")sikkim	-0.0294	0.04987	73.22677	-0.58951	0.557337
relevel(Region, "dooars")sikkim	-0.12075	0.050656	76.94165	-2.38379	**0.019601***
Repetitionsecond	0.023746	0.020154	157.0473	1.178204	0.240497
Repetitionthird	-0.01354	0.021923	174.5004	-0.61764	0.537616
Gendermale	-0.02247	0.036978	76.80541	-0.60767	0.545202
mLobF1	-0.05124	0.043351	206.0227	-1.18196	0.238583
mLobF2	-0.16749	0.051114	216.1734	-3.27667	**0.001224****

The fixed effects estimates for vowel duration of [ʌ] indicate that the standard variant is longer than all the other three regional variants. The difference between the Nepal and the Darjeeling variants is significant (p=0.004545 **) and so is the difference between the Nepal and the Sikkim varieties (p = 0.000812***). The Dooars variant is marginally longer than the Darjeeling variant while the Sikkim variant is marginally shorter in comparison to Darjeeling variant. The Sikkim variant is also significantly shorter than the Dooars variant (p=0.019601*). Repetition had no significant effect on vowel duration but the estimates indicate that every subsequent iteration was shorter than the previous one. Similarly gender also had no significant effect on vowel duration of [ʌ] although esitmates indicate that [ʌ]is shorter for males in comparison to female speakers.

Vowel height had no significant effect on the duration of [ʌ]. [ʌ] is a low vowel so the negative estimate for duration indicates that vowel duration decreases with increase in vowel height. Alternatively, it can be said that if vowel duration is short the vowel lacks the time to reach its target in the vowel space.

The effect of F2 on duration of [ʌ]is also significant (p = 0.001224**). The negative estimate again indicates that F2 increases pulling the vowel forward in the acoustic space when duration decreases. Therefore, when vowel duration is short, the vowel tends to be central rather than a back vowel.

3.7.4 f_0 for [ʌ]

The analysis of fundamental frequency was conducted separately for males and females as f_0 measures were not normalized. Out of the three repetitions, only the second repetition was analysed to eliminate effects of list intonation.

3.7.5 f_0 for females for [ʌ]

Figure 3.25 displays the sample density of the distribution of f_0 for females for the vowel [ʌ]. The mean of the distribution is 226 Hz whereas the median is 223.5 Hz.

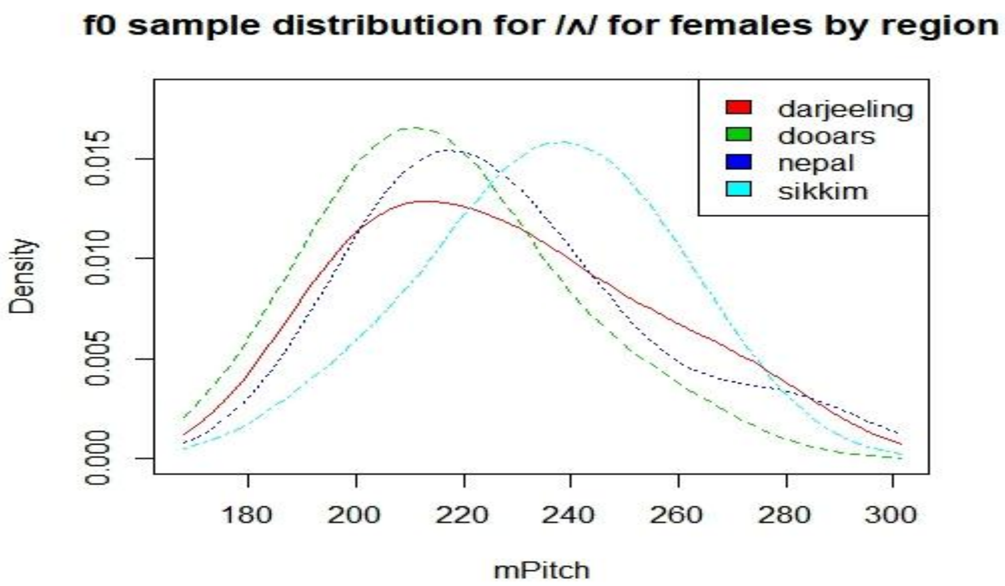


Figure 3. 25 Sample density distribution of f0 for female speakers for [ʌ]

A simple linear regression model of the fundamental frequency of [ʌ]for females was constructed with region and normalized F1 as predictors. This model was not significant. Neither region nor the height of [ʌ] affect f_0 for females ($F(4,25)= 0.998$, $p=0.4271$).

Table 3. 21 Fixed effects estimates for f_0 for female speakers for [ʌ]

Fixed effects estimates for f_0 for female speakers for [ʌ]				
	Estimate	Std..Error	t.value	Pr...t..
(Intercept)	217.8356	11.50811	18.92887	**2.47E-16*** **
relevel(Region, "nepal")darjeeling	0.928764	12.1197	0.076633	0.939526

relevel(Region, "nepal")dooars	-17.8144	12.88705	-1.38235	0.179091
relevel(Region, "nepal")sikkim	5.213164	12.28172	0.424465	0.674857
relevel(Region, "darjeeling")dooars	-18.7431	13.76364	-1.36179	0.185412
relevel(Region, "darjeeling")sikkim	4.284401	12.92177	0.331564	0.742982
relevel(Region, "dooars")sikkim	23.02754	12.486	1.844269	0.077027 .
mLobF1	21.04372	17.94935	1.172395	0.25209

The table of fixed effects estimates for f_0 for females speakers for [ʌ] indicate no significant inter-group differences between the four regions included in this study. Compared to the standard variety, females speaking the Darjeeling and Dooars varieties of Nepali have marginally higher fundamental frequencies for [ʌ] while female speakers from Dooars have lower pitch values than females from Nepal, Darjeeling and Sikkim.The positive estimate for F1 in the model, although insignificant, indicates that f_0 increases with vowel height. Sinc[ʌ]is a low vowel, the estimate is also rather low.

3.7.6 f_0 for male speakers for vowel [ʌ]

Figure 3.26 presents the sample density of the distribution of [ʌ]from male speakers. The mean of the distirbution is 134.1 Hz while the median in 133 Hz.

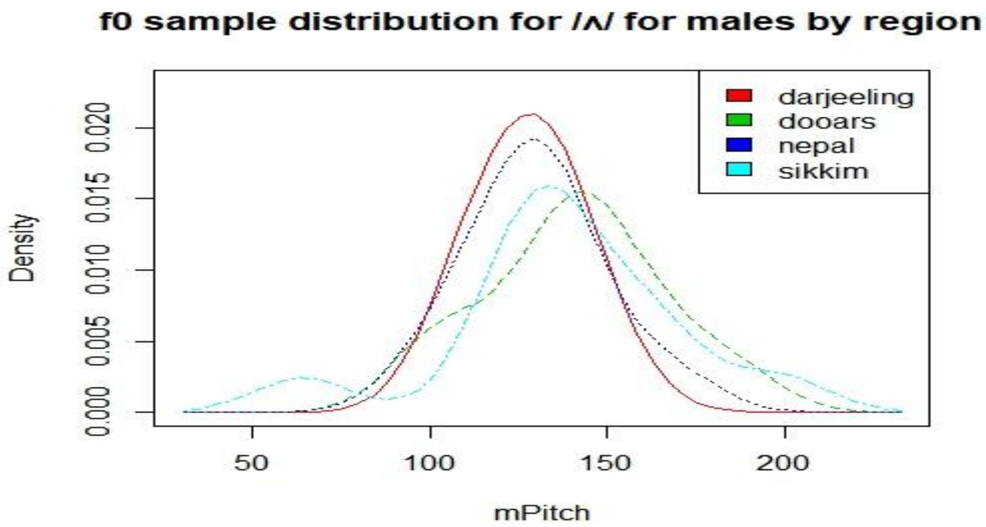


Figure 3. 26 Sample density distribution of f0 for male speakers for [ʌ]

A linear model constructed with f_0 for males as the dependent variable and region along with F1 as predictors was not significant ($F(4,40)=1.02$, $p=0.4089$). The fixed effects cofficients of the predictors in the model is enumerated in table 3.22 below.

Table 3. 22 Fixed effects estimates for f_0 for female speakers for [ʌ]

Fixed effects estimates for f_0 for female speakers for [ʌ]				
	Estimate	Std..Error	t.value	Pr...t..
(Intercept)	217.8356	11.50811	18.92887	**2.47E-16*** **
relevel(Region, "nepal")darjeeling	0.928764	12.1197	0.076633	0.939526
relevel(Region, "nepal")dooars	-17.8144	12.88705	-1.38235	0.179091
relevel(Region, "nepal")sikkim	5.213164	12.28172	0.424465	0.674857
relevel(Region, "darjeeling")dooars	-18.7431	13.76364	-1.36179	0.185412
relevel(Region, "darjeeling")sikkim	4.284401	12.92177	0.331564	0.742982
relevel(Region, "dooars")sikkim	23.02754	12.486	1.844269	0.077027 .
mLobF1	21.04372	17.94935	1.172395	0.25209

There are no significant differences between male speakers in the four different regions. Like females from Dooars, the estimate for male speakers also indicates lower f_0 values in comparison to males speaking the standard dialect. Male speakers from Darjeeling and Sikkim report slightly higher values for f_0 in comparison to male speakers from Nepal. The height of the vowel again is insignificant for f_0 for male speakers and observations from estimates made for female speakers hold true for male speakers as well.

3.7.7 Summary of findings for [ʌ]

The following is a summary of the findings for vowel [ʌ]:

1. Region affected the height of a vowel [ʌ] ($\chi2$ (3) = 12.941, p = 0.004767**).
2. Nepal variant doesn’t differ from other regional variants of [ʌ].
3. The estimate for the Darjeeling variant with reference to the standard variant indicates that is positioned higher.
4. The positive estimate for the Dooars variant with reference to the standard variant indicates that it is marginally lower than the standard variant.
5. The Sikkim variant is higher than the Nepal variant.
6. Significant difference in the height of [ʌ]between speakers from Darjeeling and Dooars (p=0.001106**).
7. Dooars [ʌ]is lower than Darjeeling [ʌ].
8. The Sikkim variant is also lower than the Darjeeling variant but this difference is not statistically significant.

9. The Sikkim variant is higher than the Dooars variant and this difference is significant (p=0.004135**).
10. The effect of vowel duration on F1 was not significant for this dataset.
11. Region did not affect the F2 of the vowel [ʌ] ($\chi 2$ (3) = 2.8037, p = 0.4229).
12. There are no significant inter-group differences between the four regions with regard to F2.
13. The Dooars variant is marginally fronted compared to all other variants followed by the Nepal variant which is slightly ahead of the Darjeeling and Sikkim variants.
14. Vowel duration has a significant effect on F2 values for [ʌ] (p=0.004072**)
15. The estimates also indicate a significant effect of gender on F2 of [ʌ] (p=0.012159*). The negative estimate for male speakers signifies that the mean of their F2 is lesser than the mean of female speakers.
16. Region affects vowel duration significantly ($\chi 2$ (3) = 13.79, p = 0.003206**).
17. Estimates for vowel duration of [ʌ]indicate that the standard variant is longer than all the other three regional variants.
18. The difference in vowel duration between the Nepal and the Darjeeling variants is significant (p=0.004545 **).
19. The difference in vowel duration between the Nepal and the Sikkim varieties is significant (p = 0.000812***).
20. The Sikkim variant is also significantly shorter than the Dooars variant (p=0.019601*).
21. The Dooars variant is marginally longer than the Darjeeling variant while the Sikkim variant is marginally shorter in comparison to Darjeeling variant.
22. Vowel height had no significant effect on the duration of [ʌ]
23. The effect of F2 on duration of [ʌ] is also significant (p = 0.001224**). The negative estimate again indicates that F2 increases fronting the vowel in the vowel acoustic space when duration decreases.
24. Neither region nor the height of [ʌ] affect f_0 for females (F(4,25)= 0.998, p=0.4271).
25. No significant inter-group differences between the four regions included in this study.
26. A linear model constructed with f_0 for males as the dependent variable and region along with F1 as predictors was not significant (F(4,40)=1.02, p=0.4089).
27. There are no significant differences between male speakers in the four different regions.

3.8 Vowel [o]

Traditional accounts of the vowel inventory of Nepali have all classified [o] as half-close or mid back vowel. Acoustically they are characterized by low F2 values and intermediate F1 values. For the analysis of [o] in this study, a total of 217 tokens were included in the data set after removal of 17 outlier points. The following sections present an analysis for each of the acoustic measures delineated in the study.

3.8.1 F1 for [o]

Figure 3.27 presents the sample density of the distribution of normalized values of F1 for the vowel [o]. The mean of the overall distribution is -0.3442 whereas the median for the distribution is -0.3610.

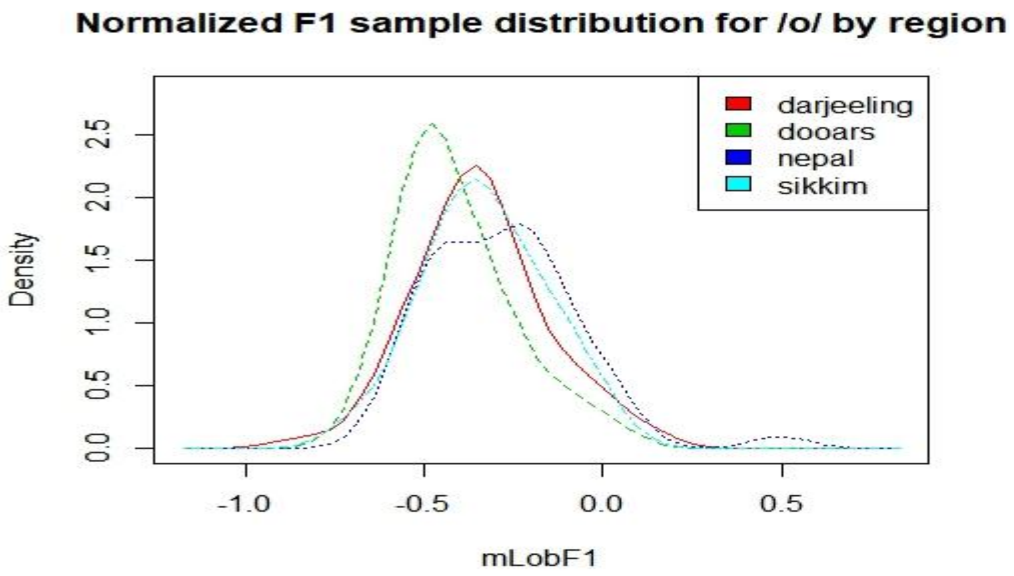


Figure 3. 27 Sample density distribution of F1 for [o] by region

For the analysis of F1, two linear mixed effects models were constructed with fixed and random effects. A full model included region, repetition, vowel duration and gender as fixed effects and speaker as a random effect. The null model was constructed by just excluding region as a fixed effect leaving all other fixed effects intact. Significance of region was determined on the basis of likelihood ratio test by comparing the two models. Region significantly affected F1 of vowel [o] ($\chi2$ (3) = 10.749, p = 0.01316*). Table 3.23 below provides estimates of the fixed effects from the model.

Table 3. 23 Fixed effects estimates of normalized F1 for [o]

Fixed effects estimates of normalized F1 for [o]

	Estimate	Std..Error	df	t.value	Pr...t..
(Intercept)	0.303563	0.368433	197.1105	0.823929	0.410976
relevel(Region, "nepal")darjeeling	-0.06815	0.041712	77.849	-1.63382	0.106334
relevel(Region, "nepal")dooars	-0.14067	0.041936	78.45351	-3.35448	**0.001227****
relevel(Region, "nepal")sikkim	-0.05439	0.041869	79.25749	-1.29897	0.19772
relevel(Region, "darjeeling")dooars	-0.07252	0.03989	77.55838	-1.81808	**0.072912 .**
relevel(Region, "darjeeling")sikkim	0.013763	0.040155	79.88587	0.342746	0.73269
relevel(Region, "dooars")sikkim	0.086286	0.040065	79.55194	2.153648	**0.034295 ***
Repetitionsecond	-0.05996	0.024665	148.445	-2.43092	**0.016252***
Repetitionthird	-0.10401	0.025109	155.7784	-4.14234	**5.62E-05*****
log.dur.ms	-0.10532	0.074046	198.3855	-1.42238	0.156486
Gendermale	-0.02182	0.029325	78.77989	-0.74414	0.459007

The table on fixed effects estimates of normalized F1 for [o] indicates all other regional variants of [o] are raised compared to the standard variant spoken by speakers from Nepal. While differences with Sikkim and Darjeeling are insignificant, Dooars variant is significantly raised in comparison to the Nepal variant (p=0.001227**). Another significant contrast in vowel height for [o] is between Dooars and Sikkim speakers with the Dooars variant significantly higher than the Sikkim [o] (p=0.034295 *). The contrasts are visualized in a plot with normalized F2 values on the x-axis against normalized F2 values on the y-axis in figure 3.28. The Dooars variant is also higher than the Darjeeling variant with the estimate close to achieving significance. Repetition had an expected significant effect on the normalized F1 for [o]. The estimate for vowel duration and gender indicates no significant effect.

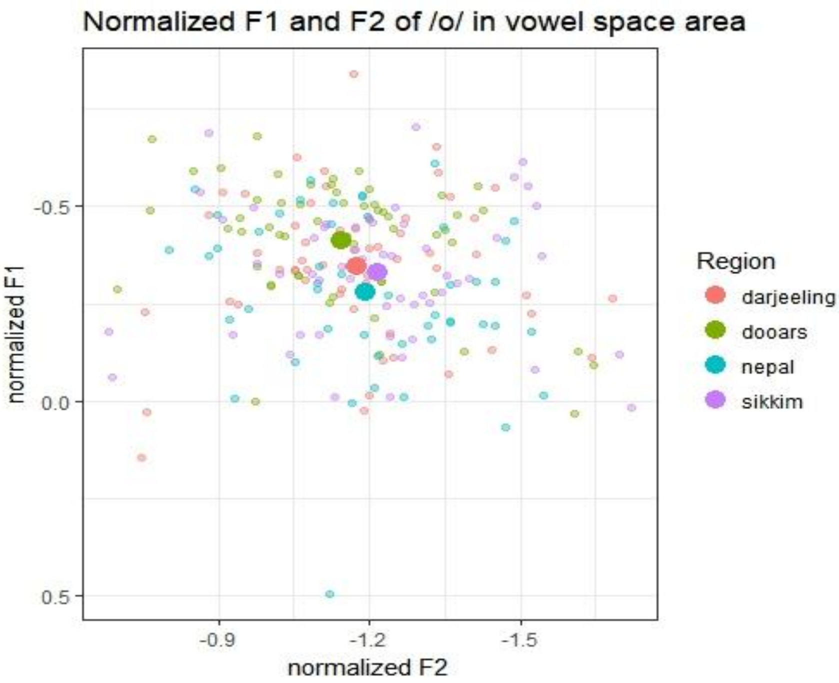


Figure 3. 28 Distribution of [o] in vowel space area

3.8.2 F2 for [o]

Figure 3.29 displays the sample density of the distribution of normalized F2 values for the vowel [o]. The mean of the overall distribution is -1.181 whereas the median is -1.179.

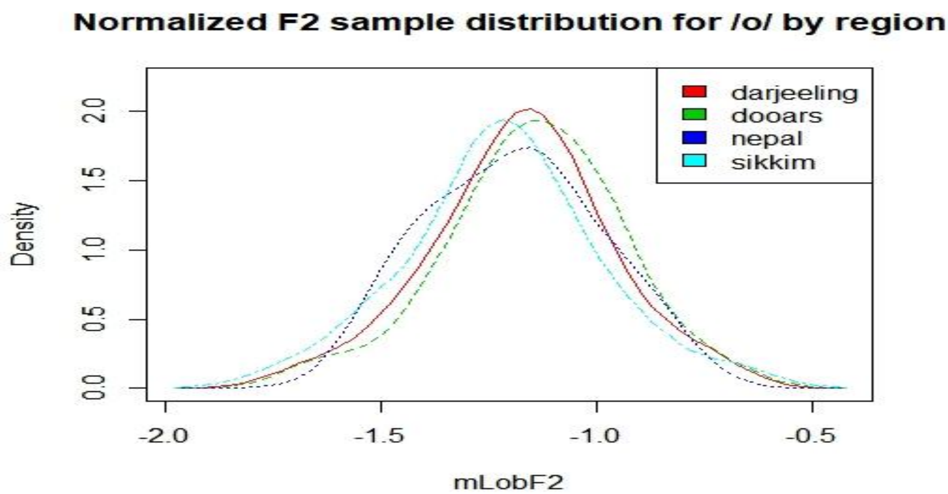


Figure 3. 29 Sample density distribution of F2 for [o] by region

Results of the likelihood ratio test obtained by comparing the full and null linear mixed effects models show no significant effect of region on the normalized F2 for the [o]

($\chi 2$ (3) = 2.8413, p = 0.4167). Table 3.24 below provides estimates of the fixed effect from the model.

Table 3. 24 Fixed effects estimates of normalized F2 for [o]

Fixed effects estimates of normalized F2 for [o]					
	Estimate	Std..Error	df	t.value	Pr...t..
(Intercept)	-1.18109	0.397264	185.219	-2.97305	0.003341**
relevel(Region, "nepal")darjeeling	0.014617	0.043586	67.37792	0.33536	0.738395
relevel(Region, "nepal")dooars	0.041546	0.043831	67.9471	0.947855	0.346562
relevel(Region, "nepal")sikkim	-0.02879	0.043778	68.9146	-0.65769	0.512927
relevel(Region, "darjeeling")dooars	0.026929	0.041677	67.13341	0.646126	0.5204
relevel(Region, "darjeeling")sikkim	-0.04341	0.041997	69.55945	-1.03363	0.304891
relevel(Region, "dooars")sikkim	-0.07034	0.041898	69.26813	-1.67881	0.097695 .
Repetitionsecond	0.134955	0.02747	138.6406	4.912883	2.49E-06***
Repetitionthird	0.189336	0.027918	146.4062	6.781841	2.72E-10***
log.dur.ms	-0.02871	0.079868	186.7465	-0.35952	0.719616
Gendermale	0.033801	0.030655	68.38604	1.102612	0.274061

The fixed effects estimates of normalized F2 for vowel [o] indicate no statistically significant inter-group differences between regions. Compared to the standard variant, the Darjeeling and the Dooars variants are marginally forward whereas the Nepal variant of [o] is marginally fronted in relation to the Sikkim variant. The Dooars variant is the most fronted followed by the Darjeeling, Nepal and the Sikkim variants. The effect of repetition is significant but the effects of vowel duration and gender are insignificant.

3.8.3 Vowel duration for [o]

Figure 3.30 displays the sample density of log transformed vowel duration values for [o]. The mean of the distribution is 4.836 and the median is 4.868.

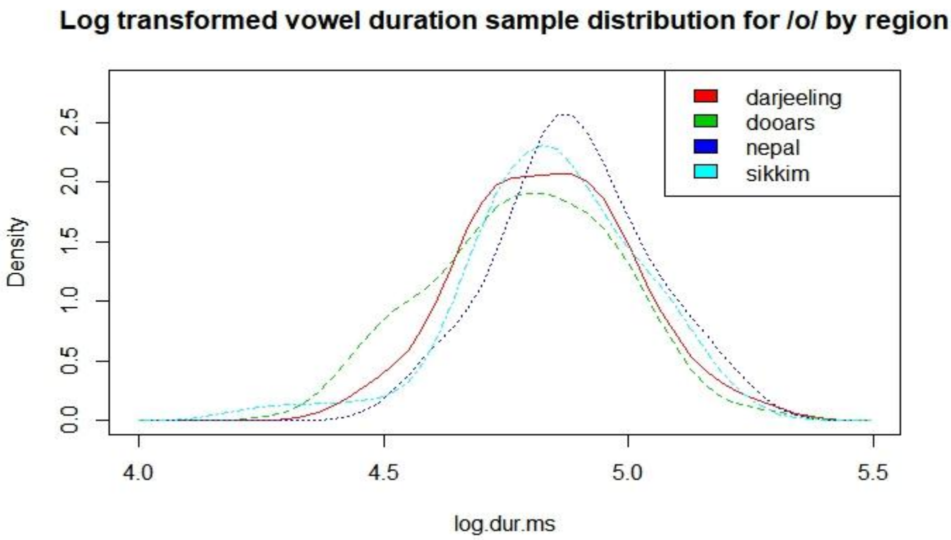


Figure 3. 30 Sample density distribution of vowel duration for [o]

Comparison of the null and full models using a likelihood ratio test indicated that region did not affect vowel duration of [o] ($\chi 2$ (3) = 3.9961, p = 0.2619). Table 3.25 below provides estimates of the fixed effects in the model for vowel duration of [o].

Table 3. 25 Fixed effects estimates for vowel duration of [o]

Fixed effects estimates for vowel duration of [o]					
	Estimate	Std..Error	df	t.value	Pr...t..
(Intercept)	4.810771	0.081092	216.9885	59.32503	< **2e-16 *****
relevel(Region, "nepal")darjeeling	-0.05306	0.047878	67.59674	-1.10828	0.271671
relevel(Region, "nepal")dooars	-0.09836	0.048274	69.43064	-2.03755	**0.045408 ***
relevel(Region, "nepal")sikkim	-0.05592	0.047916	68.04603	-1.16697	0.247294

relevel(Region, "darjeeling")dooars	-0.0453	0.045911	67.84081	-0.98668	0.327308
relevel(Region, "darjeeling")sikkim	-0.00286	0.045982	68.50618	-0.06209	0.950669
relevel(Region, "dooars")sikkim	0.042444	0.046168	69.38132	0.919327	0.36111
Repetitionsecond	-0.03267	0.020434	140.8786	-1.59864	0.112141
Repetitionthird	-0.06246	0.0223	152.4548	-2.80073	**0.00576 ****
Gendermale	-0.03559	0.033558	67.64867	-1.06052	0.292682
mLobF1	-0.04113	0.058041	193.8306	-0.70855	0.479455
mLobF2	-0.10114	0.05266	186.5802	-1.92056	0.056312 .

From the estimates column of the table above, it can be seen that with reference to the standard variety spoken by speakers from Nepal, [o] in all other regional varieties are shorter. Only the contrast with the Dooars variant is statistically significant (p=**0.045408 ***). The Dooars and the Sikkim varieties are shorter than the Darjeeling variant while the Sikkim variant is marginally longer than the Dooars variant. The effect of repetition on vowel duration was on expected lines with every subsequent repetition decreasing in duration. There was no significant contrast between males and females although the estimate for males suggests that vowel duration is marginally shorter for males in comparison to females. F1 and F2 had no significant effect on vowel duration.

3.8.4 f_0 for [o]

The analysis of fundamental frequency was conducted separately for males and females as fundamental frequency values were not normalized. For both male and female speakers, a simple linear regression model was constructed with fundamental frequency as the dependent variable and region along with normalized F1 as the independent variables.

3.8.5 f_0 for females for vowel [o]

Figure 3.31 displays the sample density of the distribution of fundamental frequency values for female speakers for the vowel [o]. The mean and the median of the overall distribution are 230.2 Hz and 229.5 Hz, respectively. The density plot for the distribution indicates that the data is normally distributed and in consonance with the assumptions for fitting linear models.

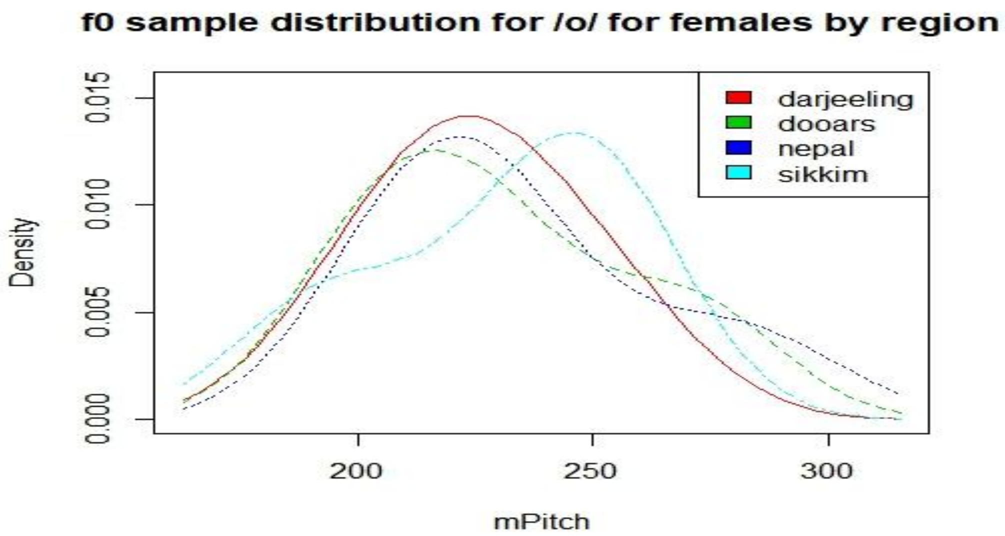


Figure 3. 31 Sample density distribution of f_0 for female speakers for [o]

A linear model constructed with f_0 of females predicted by region and F1 was significant (F(4,25)=5.406, p=0.002812). However, when F1 was dropped as a predictor from the model, region by itself did not significantly affect the fundamental frequency of female speakers for [o] (F(3,26)=0.1279, p=0.9427). Table 3.26 below provides the estimates of the predictors from the model.

Table 3. 26 Fixed effects estimates for fundamental frequency of females for [o]

Fixed effects estimates for fundamental frequency of females for [o]				
	Estimate	Std..Error	t.value	Pr...t..
(Intercept)	267.0103	10.16353	26.27141	**9.99E-20*** **
relevel(Region, "nepal")darjeeling	-1.84296	10.66678	-0.17278	0.864219
relevel(Region, "nepal")dooars	2.612519	10.31716	0.253221	0.802168
relevel(Region, "nepal")sikkim	-8.1832	10.60461	-0.77166	0.447546

relevel(Region, "darjeeling")dooars	4.455474	10.56175	0.42185	0.67674
relevel(Region, "darjeeling")sikkim	-6.34025	11.1791	-0.56715	0.575669
relevel(Region, "dooars")sikkim	-10.7957	10.84892	-0.9951	0.329224
mLobF1	101.2684	22.12762	4.576561	**0.000112*** **

Model estimates indicate that there are no significant differences in fundamental frequency between females from the four regions. Compared to the standard variant, females in Sikkim and Darjeeling have marginally lower f_0 values while female speakers from Dooars have a marginally higher pitch for the [o] vowel than female speakers from Nepal. Females from Dooars have the highest mean for [o] while Sikkim speakers have the lowest. The positive estimate for F1 indicates that fundamental frequency increases with increase in vowel height.

3.8.6 f_0 for male speakers for vowel [o]

Figure 3.32 displays the sample density of fundamental frequency measured at vowel mid-point for male speakers of the vowel [o]. The mean of the overall distribution is 142.9 Hz whereas the median is 140 Hz.

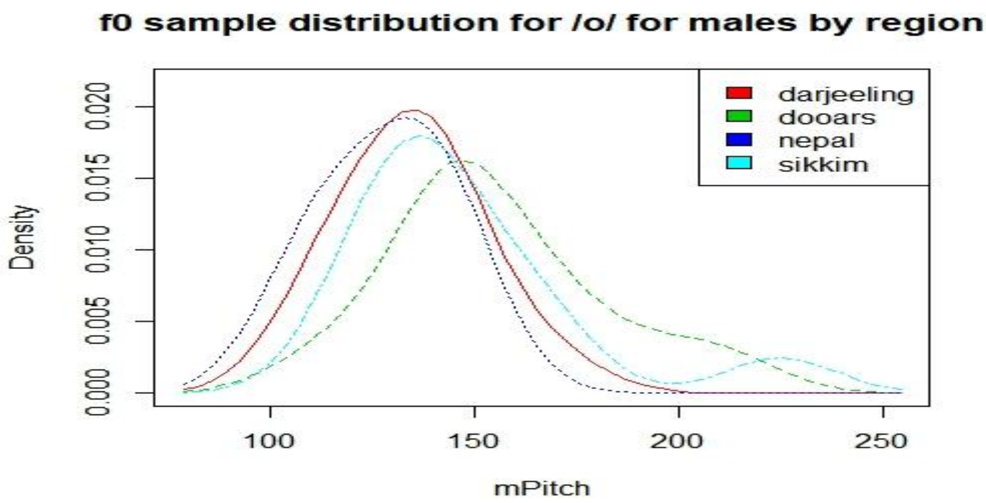


Figure 3. 32 Sample density distribution of f_0 for female speakers for [o]

A simple linear regression model was constructed with fundamental frequency as the dependent variable and region along with F1 as the independent variables. The model was not significant (F(4,38)=2.5, p= 0.0586).

Table 3. 27 Fixed effects estimates for fundamental frequency of males for [o]

Fixed effects estimates for fundamental frequency of males for [o]				
	Estimate	Std..Error	t.value	Pr...t..
(Intercept)	124.2155	8.928256	13.91263	**1.71E-16*** **
relevel(Region, "nepal")darjeeling	2.449017	11.14077	0.219825	0.827184
relevel(Region, "nepal")dooars	21.04107	11.64883	1.806282	0.078799 .
relevel(Region, "nepal")sikkim	15.29696	11.02564	1.387399	0.173404
relevel(Region, "darjeeling")dooars	18.59206	9.817019	1.89386	0.065875 .
relevel(Region, "darjeeling")sikkim	12.84794	9.626675	1.334619	0.189943
relevel(Region, "dooars")sikkim	-5.74412	9.565589	-0.6005	0.551739
mLobF1	-22.7296	19.32503	-1.17617	0.246842

The model estimates indicate no statistically significant inter-group contrasts between regions. According to the estimates, the fundamental frequency of males in all the other three regions is higher than males from Nepal who speak the standard dialect. Dooars males have the highest pitch followed by Sikkim, Darjeeling and Nepal. There was no significant effect of vowel height on fundamental frequency of [o] for male speakers. The negative estimate for normalized F1 indicates that fundamental frequency decreases with increase in vowel height.

3.8.7 Summary of findings for [o]

The following is a summary of findings for vowel [o]:

1. A total of 217 tokens were included in the data set after removal of 17 outlier points.
2. Region significantly affected F1 of vowel [o] ($\chi2$ (3) = 10.749, p = 0.01316*).
3. All other regional variants of [o] are raised compared to the standard variant spoken by speakers from Nepal.
4. Dooars variant is significantly raised in comparison to the Nepal variant (p=0.001227**).
5. The Dooars variant is significantly higher than the Sikkim [o] (p=0.034295 *).
6. The Dooars variant is also higher than the Darjeeling variant with the estimate close to achieving significance.
7. The estimate for vowel duration and gender indicates no significant effect.
8. No significant effect of region on the normalized F2 for the [o] ($\chi2$ (3) = 2.8413, p = 0.4167).

9. No statistically significant inter-group differences between regions along the front-back dimension.
10. Dooars variant is the most fronted followed by the Darjeeling, Nepal and the Sikkim variants.
11. Region did not affect vowel duration of [o] ($\chi2$ (3) = 3.9961, p = 0.2619).
12. Nepal variant is longer than all other regional variants.
13. The contrast with the Dooars variant is statistically significant (p=**0.045408 ***).
14. There was no significant contrast between males and females.
15. F1 and F2 had no significant effect on vowel duration.
16. A linear model constructed with f_0 of females predicted by region and F1 was significant (F(4,25)=5.406, p=0.002812).
17. Region, in isolation, was not significant.
18. No significant differences in fundamental frequency between females from the four regions.
19. Females from Dooars have the highest mean for [o] while Sikkim speakers have the lowest.
20. Compared to the standard variant, females in Sikkim and Darjeeling have marginally lower f_0 values while female speakers from Dooars have a marginally higher pitch for the [o] vowel than female speakers from Nepal.
21. The model for the f_0 of male speakers was not significant.
22. No statistically significant inter-group contrasts between regions.
23. Dooars males have the highest pitch followed by Sikkim, Darjeeling and Nepal.
24. No significant effect of vowel height on fundamental frequency of [o] for male speakers.

3.9 Vowel [u]

[u] has been classified as a high back vowel in existing literature. It is represented acoustically by low values of F1 and F2. The following sections present an analysis of each of the acoustic parameters (F1, F2, vowel duration and f_0 for males and females separately) for [u].

3.9.1 F1 for [u]

Figure 3.33 displays the sample density of the distribution of normalized F1 values for the vowel [u]. The mean of the overall distribution is -0.7623 while the median is -0.7500.

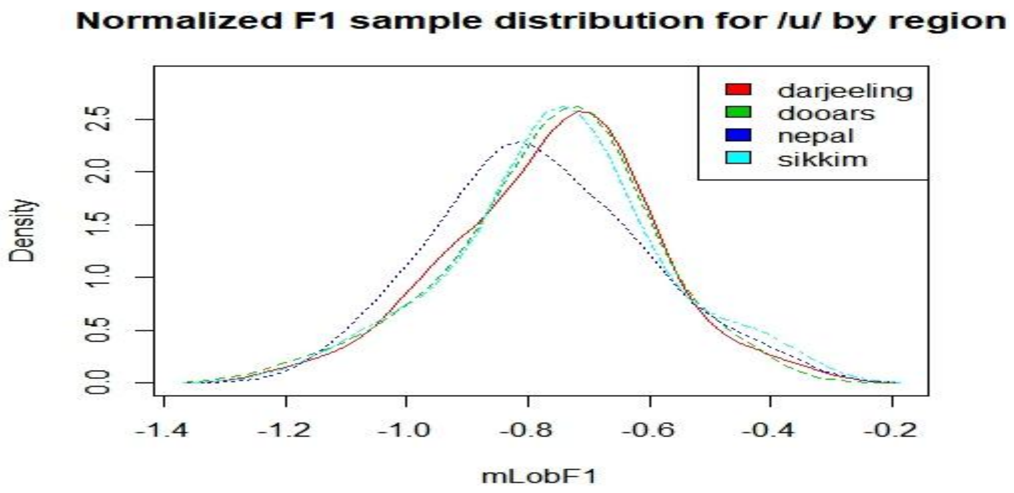


Figure 3. 33 Sample density distribution of F1 for [u] by region

For the analysis of normalized F1 for the vowel [u], two linear mixed effects models were constructed and then compared using a likelihood ratio test. The first model was a full model with normalized F1 as the dependent variable; region, gender, repetition and log transformed values of vowel duration were included in the model as fixed effects along with speaker as a random effect. The null model was different from the full model in the sense that region was excluded as a fixed effect in the model. Results indicated that region had no significant effect on F1 ($\chi2$ (3) = 0.8194, p = 0.8448). Table 3.28 below provides estimates of fixed effects from the full model.

Table 3. 28 Fixed effects estimates of normalized F1 for [u]

Fixed effects estimates of normalized F1 for [u]					
	Estimate	Std..Error	df	t.value	Pr...t..
(Intercept)	-0.92456	0.271555	172.11	-3.40468	0.000824***
relevel(Region, "nepal")darjeeling	0.019894	0.03784	77.38245	0.525738	0.600574
relevel(Region, "nepal")dooars	0.020905	0.037737	76.47395	0.553951	0.581228
relevel(Region, "nepal")sikkim	0.034135	0.037773	76.51691	0.903682	0.369
relevel(Region,	0.001011	0.035811	72.3102	0.02822	0.977561

"darjeeling")dooars			4	5	
relevel(Region, "darjeeling")sikkim	0.014241	0.035816	72.4179 1	0.39762 8	0.692074
relevel(Region, "dooars")sikkim	0.01323	0.03575	71.3545 2	0.37008 3	0.712417
Repetitionsecond	0.029165	0.023637	140.178	1.23389 2	0.219308
Repetitionthird	0.029735	0.023905	149.435 4	1.24387 2	0.215494
log.dur.ms	0.022774	0.055748	174.836 2	0.40852 2	0.68339
Gendermale	0.026138	0.027126	77.0473 8	0.96360 2	0.33826

The estimates of the fixed effects in the model indicate that there are no statistically significant differences between the regions as far as the F1 for [u] is concerned. In comparison to the standard variety spoken by speakers from Nepal, estimates of all the other regions are marginally higher. This suggests that the Nepal [u] is marginally higher than the other variants. Relative to the Darjeeling variant, estimates for Sikkim and Dooars are marginally higher which suggests that the Darjeeling variant is slightly raised in comparison to them. The Sikkim variant of [o] is the lowest but none of the above mentioned contrasts are statistically significant. The differences in their relative positioning can be seen in figure 3.34 below.

There were no significant effects of vowel duration, gender or repetition on F1.

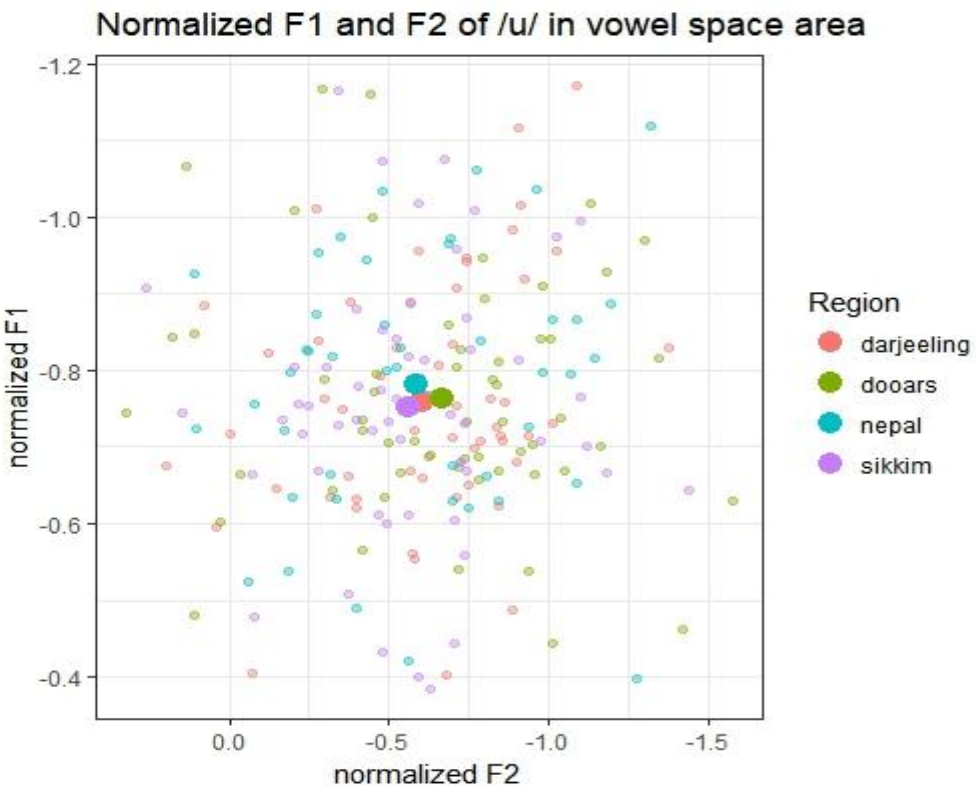


Figure 3. 34 Distribution of [u] in vowel space area

3.9.2 F2 for [u]

Two models similar to the ones created for the analysis of F1 of [u] were created for the analysis of F2. In each of the models normalized values of F2 was the dependent variable. The fixed effects in the full and the null models remained the same as they were for the analysis of F1.

Figure 3.35 displays the sample density of the distribution of F2 for the vowel [u]. The mean of the distribution is -0.6064 whereas the median is -0.6210.

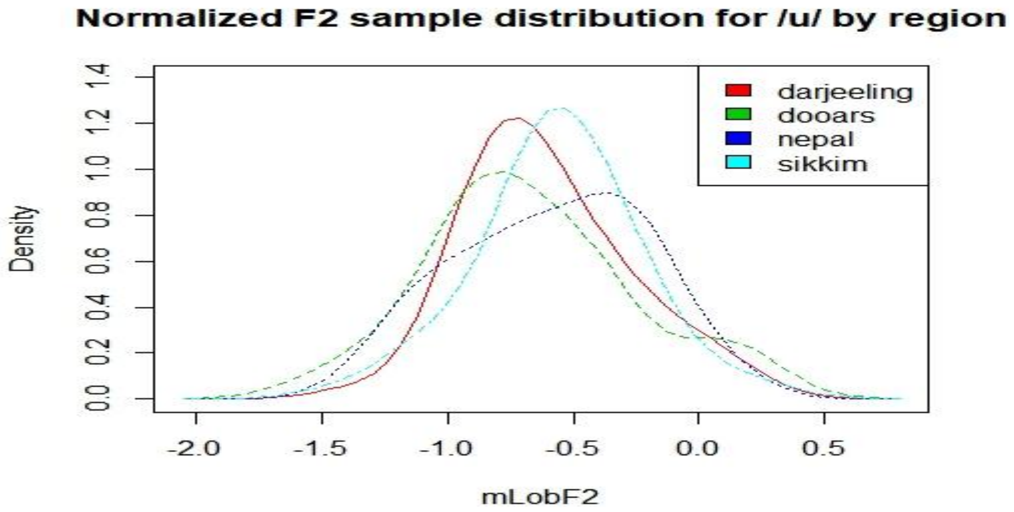


Figure 3. 35 Sample density distribution of F2 for [u] by region

A comparison of the full and null models was performed to check the significance of region on F2 for [u] using the likelihood ratio test. Region did not affect F2 for [u] ($\chi2$ (3) = 1.8954, p = 0.5944). Table 3.29 presents model estimates.

Table 3. 29 Fixed effects estimates of normalized F2 for [u]

Fixed effects estimates of normalized F2 for [u]					
	Estimate	Std..Error	df	t.value	Pr...t..
(Intercept)	0.950044	0.551254	202.2854	1.723425	0.08634 .
relevel(Region, "nepal")darjeeling	-0.0574	0.086188	77.53979	-0.66594	0.507427
relevel(Region, "nepal")dooars	-0.10756	0.086049	77.05321	-1.24999	0.215087
relevel(Region, "nepal")sikkim	-0.01773	0.086123	77.129	-0.20587	0.837435
relevel(Region, "darjeeling")dooars	-0.05016	0.082209	73.31783	-0.61021	0.543612
relevel(Region, "darjeeling")sikkim	0.039666	0.082207	73.38051	0.482514	0.630877
relevel(Region, "dooars")sikkim	0.089831	0.082167	72.87305	1.093267	0.277877

Repetitionsecond	0.135314	0.041974	138.6997	3.22376	0.001577**
Repetitionthird	0.260743	0.042802	147.816	6.091893	9.22E-09***
log.dur.ms	-0.34074	0.112854	204.7178	-3.01932	0.002856**
Gendermale	-0.05857	0.061762	78.25528	-0.94824	0.345927

The estimates of fixed effects summarized in table 3.29 indicate no significant inter-group contrasts between the four regions with regard to F2 for [u]. The negative estimates for the other three regions with respect to Nepal indicate that the Nepal variant of [u] is relatively fronted, marginally. Compared to the Darjeeling variant, a negative estimate for Dooars indicates that the Darjeeling variety is fronted in comparison. Judging by the estimates, the Sikkim variant is fronted in comparison to the Darjeeling and the Dooars variants. Repetition had a significant effect on F2 for [u] with F2 getting more fronted thereby deviating from the target with every subsequent iteration. The effect of vowel duration on F2 was significant and along expected lines. F2 values are seen to be decreasing for every increase in vowel duration. It is worth reiterating that back vowels have low F2 so when the vowel is uttered for the duration of its inherent length, the target is closely approximated. The effect of gender on F2 was not significant however the negative estimate for males with reference to females is along expected lines that females have higher frequencies than males because of differences in the length of the vocal tract.

3.9.3 Vowel duration of [u]

Figure 3.36 below provides the sample density of the distribution of vowel [u]. The mean of the distribution is 4.699 and the median is 4.700.

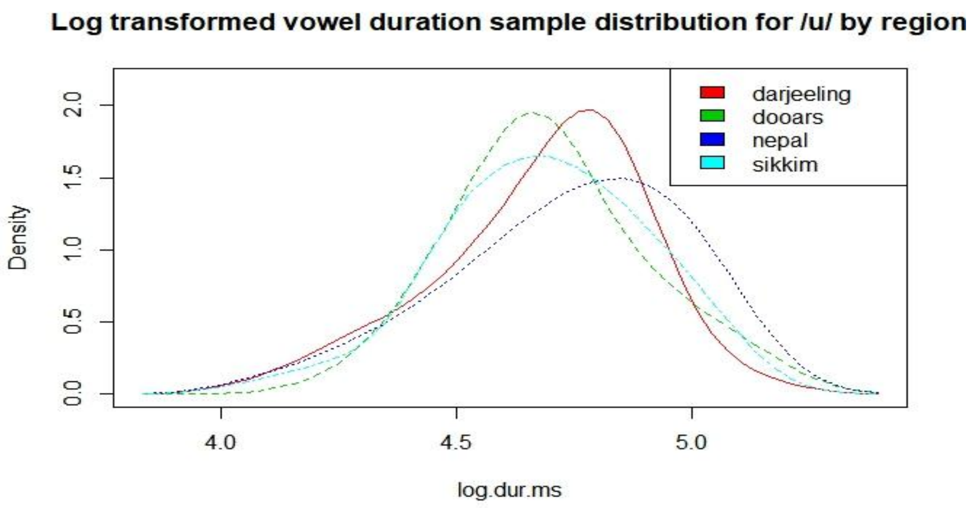


Figure 3. 36 Sample density distribution of vowel duration for [u]

Two models similar to ones constructed for the analysis of F1 and F2, but with minor changes, were constructed for the analysis of vowel duration of [u]. The full model included region, repetition, gender, normalized values of F1 and F2 as fixed effects while speaker was added as a random effect. The null model lacked region as a fixed effect. A likelihood ratio test was performed on the two models to gauge the effect of region on vowel duration. Region did not affect vowel duration of [u] ($\chi2$ (3) = 1.367, p = 0.7133). The coefficient of the fixed effects from the model are presented in table 3.30 below.

Table 3. 30 Fixed effects estimates for vowel duration of [u]

Fixed effects estimates for vowel duration of [u]					
	Estimate	Std..Error	df	t.value	Pr...t..
(Intercept)	4.787033	0.082033	205.5096	58.35487	< 2e-16 ***
relevel(Region, "nepal")darjeeling	-0.06292	0.057177	75.88136	-1.10047	0.274605
relevel(Region, "nepal")dooars	-0.04887	0.057306	76.0841	-0.85283	0.39643
relevel(Region, "nepal")sikkim	-0.05108	0.057228	75.81277	-0.89257	0.37491
relevel(Region, "darjeeling")dooars	0.014049	0.054923	72.77446	0.255801	0.798826

relevel(Region, "darjeeling")sikkim	0.011841	0.054922	72.76815	0.215606	0.829898
relevel(Region, "dooars")sikkim	-0.00221	0.055004	72.97312	-0.04014	0.968093
Repetitionsecond	-0.01637	0.02363	140.0346	-0.6929	0.489522
Repetitionthird	-0.04132	0.025635	153.9674	-1.61187	0.109039
Gendermale	-0.11416	0.040326	74.7157	-2.83098	0.005959**
mLobF1	0.02515	0.074521	189.5797	0.337495	0.736117
mLobF2	-0.10953	0.038988	210.1768	-2.80936	0.005432**

There are no significant contrasts between the regional varieties as the duration of [u] is concerned. The trend observed for all vowel categories so far repeats itself here as well. The standard variant is longer than all other regional variants as indicated by the negative estimates for Darjeeling, Dooars and Sikkim with reference to Nepal. The Dooars variant is longer compared to Sikkim and Darjeeling. The Darjeeling [u] is the shortest among the four regions analyzed in this study. Repetition does not have a statistically significant effect on vowel duration of [u] but estimates for every subsequent reiteration indicate a decrease in vowel duration. The males differ significantly in relation to females. The estimate for males indicates that vowel duration for males is lesser than that of females (p =0.005959**). There was no significant effect of F1 on vowel duration. However, F2 affected the duration of [o] (p=0.005432**). Estimate for normalized F2 indicates a decrease in vowel duration for every unit increase in F2. Figure 3.37 displays a plot of normalized F2 values on the x-axis against vowel duration on the y-axis. The plot indicates that tokens with shorter duration are fronted in comparison to the tokens that are longer in duration.

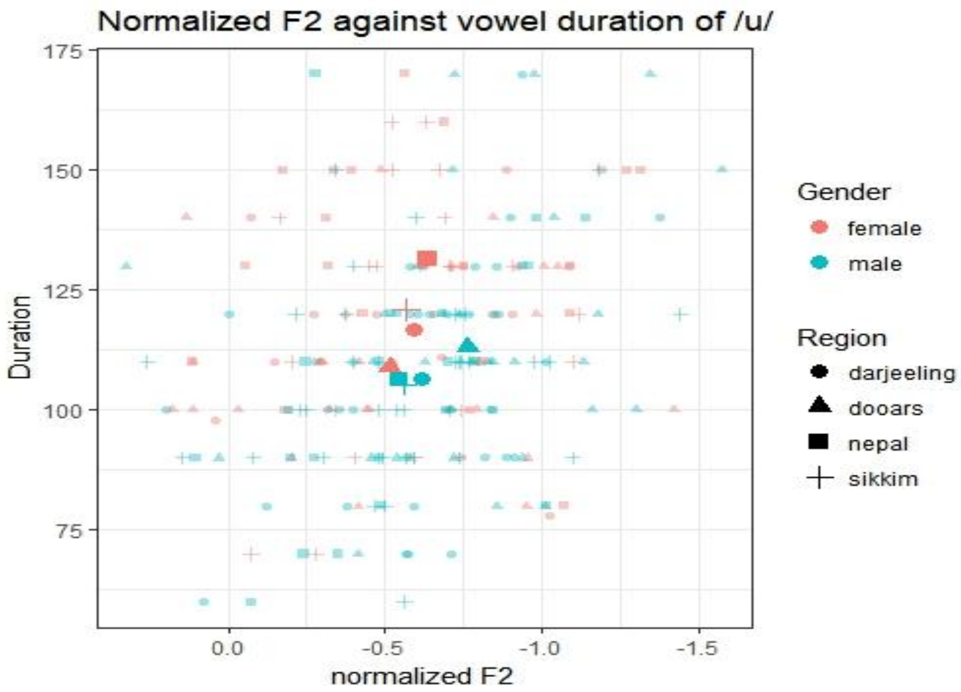


Figure 3. 37 Plot of F2 against vowel duration of [u] for males and females

3.9.4 f_0 for [u]

The procedure for the analysis of fundamental frequency is the same that has been used for other vowel categories in this study. Fundamental frequency was measured at vowel mid-point and only the second repetitions were included for analysis to eliminate effects of list intonation. Since no normalization procedure was carried on fundamental frequency measures, analysis for males and females was conducted separately. A simple linear regression model was fitted to observe the effect of region on fundamental frequency. The model also included normalized measures of F1 as a predictor of fundamental frequency.

3.9.5 f_0 for females for vowel [u]

Figure 3.38 displays the sample density of the distribution of fundamental frequency values for [u]. The mean of the distribution is 235.3 Hz whereas the median is 227.5 Hz.

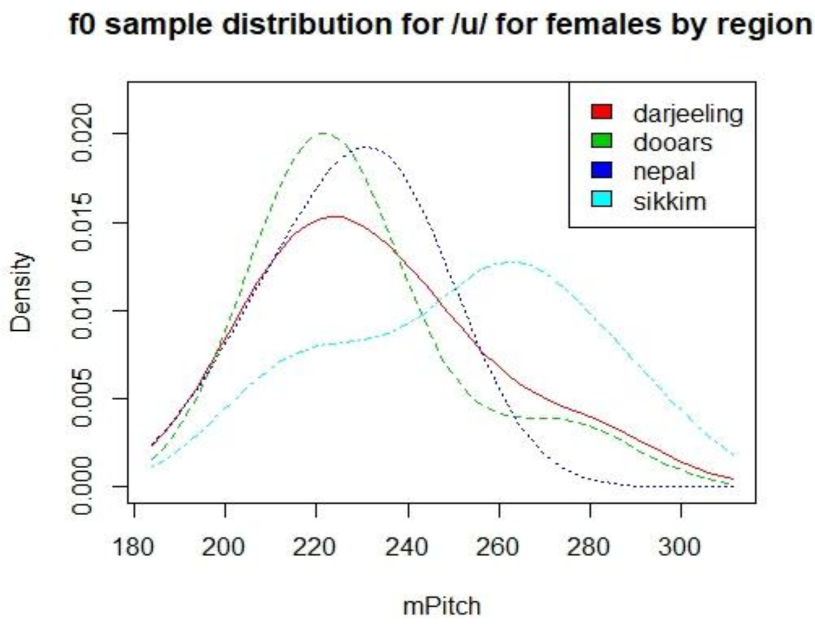


Figure 3. 38 Sample density distribution of f_0 for female speakers for [u]

A simple linear regression model was constructed with fundamental frequency for females as the dependent variable; region and F1 of [u] were the independent variables. The model for f_0 for females as a function of region and F1 was not significant ($F(4,21)=1.011$, $p=0.4241$). The model estimates in table 3.31 below indicate no significant between-group differences between the four regions. The estimates for Darjeeling, Dooars are Sikkim indicate that with reference to female speakers from Nepal, females in the other three regions have higher pitch values. The negative estimate for normalized F1 indicates that pitch increases with a decrease in vowel height.

Table 3. 31 Fixed effects estimates of f_0 for females for [u]

Fixed effects estimates of f_0 for females for [u]				
	Estimate	Std..Error	t.value	Pr...t..
(Intercept)	223.2224	17.63374	12.65883	2.70E-11
relevel(Region, "nepal")darjeeling	4.602422	14.26573	0.322621	0.750172
relevel(Region, "nepal")dooars	0.104462	15.75682	0.00663	0.994773
relevel(Region, "nepal")sikkim	23.16056	15.04102	1.539826	0.138536
relevel(Region, "darjeeling")dooars	-4.49796	12.55639	-0.35822	0.723754
relevel(Region, "darjeeling")sikkim	18.55813	12.76205	1.454166	0.160684
relevel(Region, "dooars")sikkim	23.05609	13.46348	1.712491	0.101535

mLobF1	-7.05751	25.04659	-0.28178	0.780874

3.9.6 f_0 for male speakers for vowel [u]

Figure 3.39 displays the sample density of the distribution of f_0 values for male speakers. The mean of the overall distribution is 150 Hz whereas the median is 143.5 Hz.

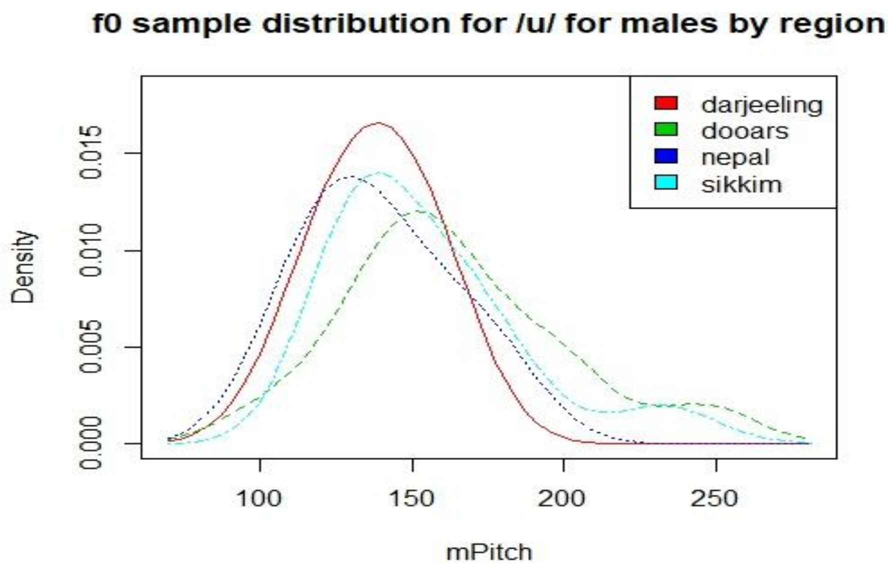


Figure 3. 39 Sample density distribution of f_0 for male speakers for [u]

f_0 for male speakers was modeled as a function of region and F1. The model was not significant (F(4,39)=1.557, p=0.2501). Analysis of estimates of fixed effects from the model indicates the only significant between-group contrast is between Darjeeling and Dooars speakers. Compared to Darjeeling males, the estimate for male speakers from Dooars is more than 25 Hz (p=0.035217*). Male speakers from Dooars report the highest f_0 for [u] followed by Sikkim, Nepal and Darjeeling. There was no significant effect of vowel height on fundamental frequency for male speakers. The negative estimate again indicates that pitch increases with a decrease in vowel height.

Table 3. 32 Fixed effects estimates of f_0 for males for [u]

Fixed effects estimates estimates of f_0 for males for [u]				
	Estimate	Std..Error	t.value	Pr...t..
(Intercept)	136.0083	31.91193	4.26199	0.000124***
relevel(Region, "nepal")darjeeling	-0.8414	13.28995	-0.06331	0.949842
relevel(Region, "nepal")dooars	24.93988	13.32556	1.871582	0.068779 .

relevel(Region, "nepal")sikkim	15.97074	13.91773	1.14751	0.258162
relevel(Region, "darjeeling")dooars	25.78128	11.81667	2.181772	0.035217*
relevel(Region, "darjeeling")sikkim	16.81215	12.16636	1.381855	0.174881
relevel(Region, "dooars")sikkim	-8.96914	12.09989	-0.74126	0.462978
mLobF1	-4.03636	37.63751	-0.10724	0.915146

3.9.7 Summary of findings for [u]

The following is a summary of findings for vowel [u]:

1. Region had no significant effect on F1 (χ^2 (3) = 0.8194, p = 0.8448).
2. No statistically significant differences between the regions as far as the F1 for [u].
3. Nepal [u] is marginally higher than the other variants.
4. The Darjeeling variant is slightly raised in comparison to Sikkim and Dooars.
5. The Sikkim variant of [o] is the lowest.
6. There were no significant effects of vowel duration, gender or repetition on F1.
7. Region did not affect F2 for [u] (χ^2 (3) = 1.8954, p = 0.5944).
8. No significant inter-group contrasts between the four regions with regard to F2 for [u].
9. The Nepal variant of [u] is relatively fronted.
10. The Darjeeling variety is fronted in comparison to Dooars.
11. The Sikkim variant is fronted in comparison to the Darjeeling and the Dooars variants.
12. The effect of vowel duration on F2 was significant.
13. F2 values are seen to be decreasing for every increase in vowel duration.
14. The effect of gender on F2 was not significant.
15. Region did not affect vowel duration of [u] (χ^2 (3) = 1.367, p = 0.7133).
16. No significant contrasts between the regional varieties as for duration of [u].
17. The standard variant is longer than all other regional variants.
18. The Dooars variant is longer compared to Sikkim and Darjeeling.
19. The Darjeeling [u] is the shortest.
20. Vowel duration for males is lesser than that of females (p =0.005959**).

21. There was no significant effect of F1 on vowel duration.
22. F2 affected the duration of [o] (p=0.005432**).
23. Estimate for normalized F2 indicates a decrease in vowel duration for every unit increase in F2.
24. f_0 for females as a function of region and F1 was not significant (F(4,21)=1.011, p=0.4241).
25. No significant between-group differences between the four regions.
26. The model for f_0 for male speakers as a function of region and F1 was not significant (F(4,39)=1.557, p=0.2501).
27. Compared to Darjeeling males, the estimate for male speakers from Dooars is more than 25 Hz (p=0.035217*).
28. Male speakers from Dooars report the highest f_0 for [u] followed by Sikkim, Nepal and Darjeeling.
29. There was no significant effect of vowel height on fundamental frequency for male speakers.

3.10 Acoustic Space/Vowel Space Area

The estimate of vowel space area has been computed as the planar area of the convex hull bound by all the perimeter vowels on a two dimensional plane corresponding to the first and the second formant frequency. Figure 3.40 displays the distribution of vowels of all speakers inside the convex hull area for the four regions.

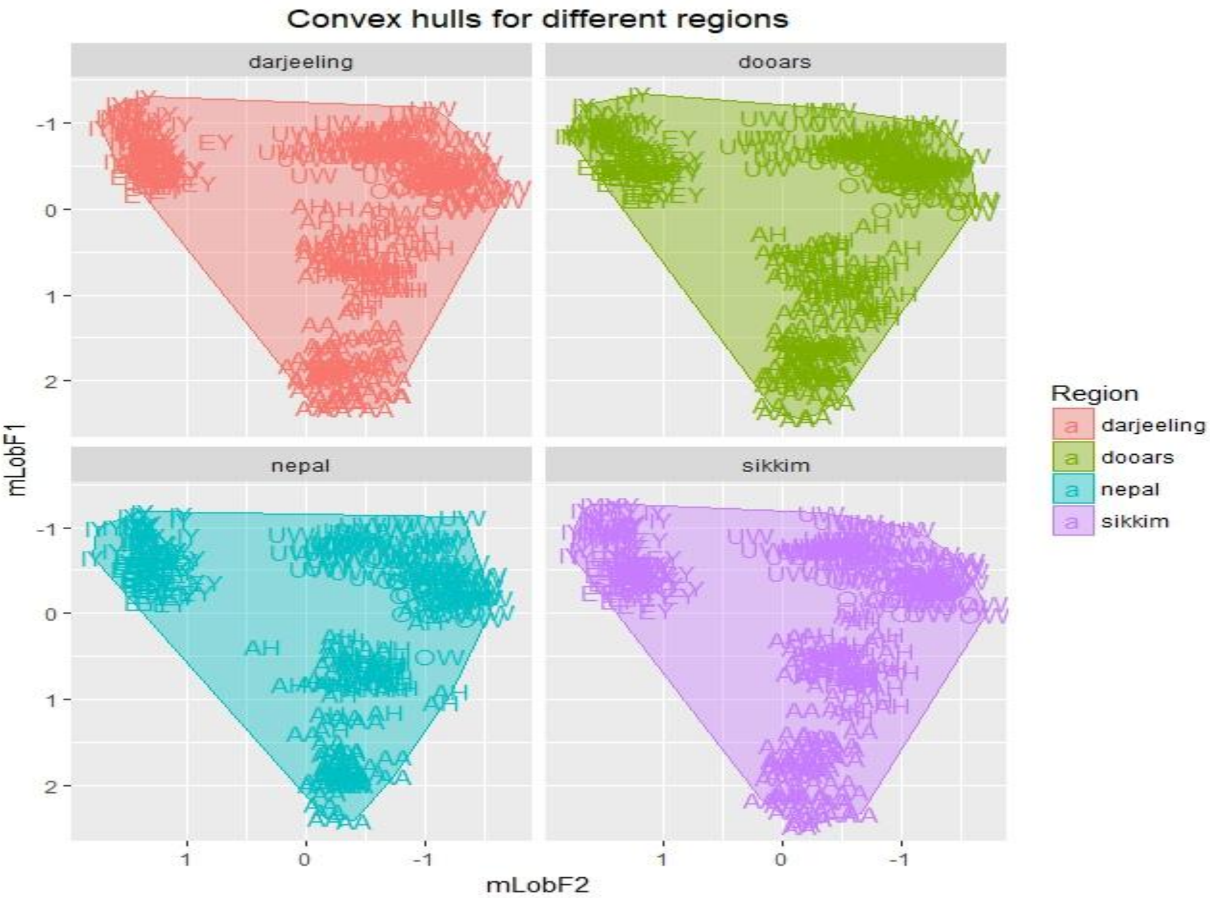


Figure 3. 40 Convex hulls for different regions

Use of convex hull areas as the choice of estimation method for acoustic space was governed by the nature of measurements collected for the study. Measurements for the acoustic parameters were taken at vowel mid-point. Grouped by speaker, hull area for all speakers was therefore calculated using the phonR package (McCloy,2016) in R. Figure 3.41 below provides the sample density of the distribution of the hull area of vowel space in Nepali for all the four regions. The mean of the overall distribution is 4.744 while the median is 4.715.

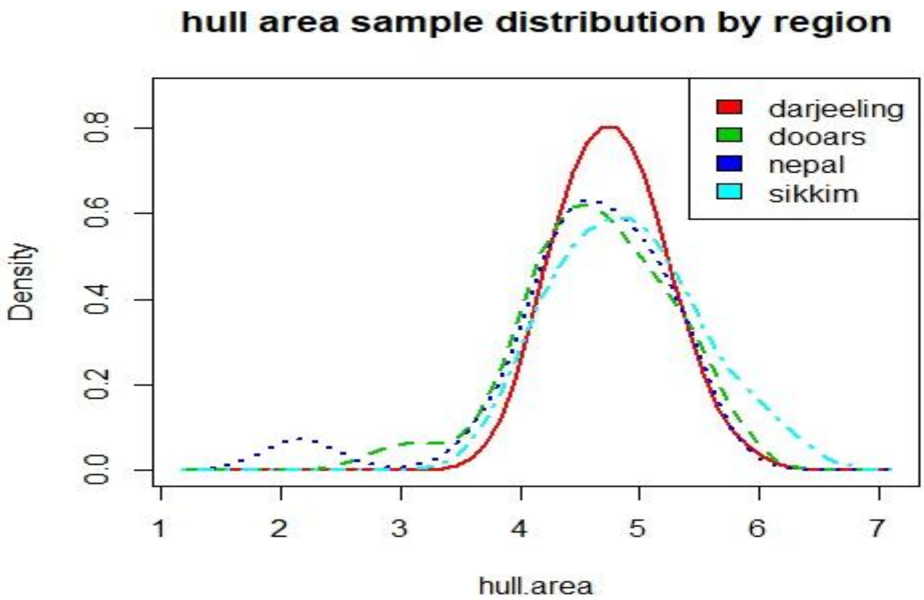


Figure 3. 41 Sample density distribution for convex hull areas by region

A simple linear regression model was fitted with hull area as the dependent variable and region as a predictor using the lme package in R (Bates et al., 2015). Table 3.33 provides the estimates of the fixed effects from the model. The model of acoustic space area as a function of region was not significant (F(3,73)=2.037, p=-0.1161).

Table 3. 33 Model estimates for hull area as a function of region

Model estimates for hull area as a function of region				
	Estimate	Std..Error	t.value	Pr...t..
(Intercept)	4.623293	0.118209	39.11124	**1.03E-50***
relevel(Region, "nepal")darjeeling	0.152059	0.160781	0.945748	0.347397
relevel(Region, "nepal")dooars	-0.03783	0.160781	-0.23526	0.814662
relevel(Region, "nepal")sikkim	0.306679	0.160781	1.90743	**0.060399.**
relevel(Region, "darjeeling")dooars	-0.18988	0.154125	-1.23201	0.221897
relevel(Region, "darjeeling")sikkim	0.154621	0.154125	1.003212	0.319072
relevel(Region, "dooars")sikkim	0.344505	0.154125	2.235227	**0.028462***

The only significant between- group contrast is between the Dooars and the Sikkim varieties (p=0.02). The positive estimate for the hull area of Sikkim speakers with reference to Dooars speakers indicates that the vowel space area in the Sikkim variety is significantly larger than in the Dooars variety. Vowels in the Sikkim variety are dispersed over a larger area than vowels in the Dooars variety. Speakers in both regions are multilingual but the level of multilingualism is higher in Dooars than in Sikkim. Sikkim variety has the largest hull area compared to all the varieties while the Dooars speakers have the smallest. Hull area of Darjeeling speakers is marginally bigger than the hull area for Nepal speakers.

This concludes the chapter on data and analysis. All acoustic parameters have been statistically modeled and the hypothesis that region affects each of these acoustic measures have been tested and findings presented. The next chapter presents a comparative analysis of the vowel dispersion in the acoustic space between the four regional varieties. The chapter also summarizes the research undertaken in this dissertation and presents the scope for further research on regional variation in Nepali.

CHAPTER 4

SUMMARY AND CONCLUSION

This production study was designed to fill the lacuna existing in terms of instrumental studies of regional vowel variation in Nepali. From studies of regional variation in American English, Dutch and Swedish, it emerges that acoustic parameters such as F1, F2 and vowel duration are affected by speakers' geographical location. These studies also show that descriptions of vowel inventories are further enriched with studies on the regional varieties of the language. Considering the fact that Nepali is spoken in pockets throughout the Himalayan belt in different linguistic and social settings, there was a need to supplement existing impressionistic accounts of Nepali vowel inventory with instrumental data. Results from a pilot study conducted in the early stages of the research by taking just two of regions – Nepal and Darjeeling – revealed a significant effect of region on vowel duration across all vowel categories (for all vowels $p<0.05$). All six vowels in the Nepal variety were found to be significantly longer than in the Darjeeling variety. Findings also revealed that [ɑ] in the standard Nepal variety was significantly raised in comparison to the Darjeeling variant whereas [o] was significantly fronted in the Nepal variety relative to the Darjeeling variety.

For the purposes of this study, four geographically contiguous regions where Nepali is widely spoken were chosen as field sites. Speakers of the eastern dialect from Nepal were chosen as participants to provide a frame of reference for comparison with speakers from the other three regions. The eastern dialect of Nepali spoken in Nepal is widely considered to be the standard dialect. Darjeeling district in West Bengal, India was chosen as another field site as there is sizable Nepali-speaking population inhabiting the Darjeeling hills and valley areas. Sikkim is a neighbouring state just across the river Teesta where Nepali is one of the official languages recognized by the state. The final field site was the Dooars region in Alipurduar district of West Bengal. The interviews were conducted with Nepali-speaking residents inhabiting the regions around the Buxa Tiger Reserve. Since a wordlist containing simple bisyllabic words was administered to elicit data, it was necessary for the participants to be literate. The recordings were made using a Zoom H1 Handy recorder in field settings with best efforts to mitigate background noise.

First and second formant frequencies have always been robust acoustic correlates of vowel quality. The first formant (F1) relates to vowel height whereas the second formant (F2) correlates with the tongue advancement along the front-back dimension. The third formant, although not analyzed in the current study, signifies lip rounding in vowel production. The cardinal vowel chart developed by Daniel Jones classifies vowels on the basis of the vowel height and tongue advancement which has its acoustic correlates in formant frequencies. Therefore, it was natural that F1, F2 , fundamental frequency and vowel duration were chosen as acoustic parameters for analysis. The final acoustic parameter which was analyzed was the vowel acoustic space area. Vowel acoustic space area has emerged as an important metric in the analysis of regional variation. Although the traditional method of computing vowel space area as the area of the triangle between the point vowels [i , ɑ, u] or the area of the quadrilateral formed by [i æ, ɑ, u] , in cases of American English , has been used in numerous studies, contemporary methods such as the convex hull area approach (Sandoval et al. 2013) and the "formant space area" approach (Fox and Jacewicz, 2017) have emerged as a better tool in the analysis of regional dialectal variation. A convex hull approach was adopted for the estimation of vowel space area for this study. The hypotheses for this research project based on existing literature and findings from the pilot study were that the region, indicating the geographical locale of the Nepali speakers, affects each of the five acoustic parameters – F1, F2, intrinsic pitch (f_0), vowel duration and vowel acoustic space.

In the previous chapter on data and analysis, the hypotheses formulated for each of the acoustic parameters was statistically tested using multivariate regression models. Linear mixed effects models with fixed and random effects were constructed for statistical modeling of F1, F2 and vowel duration due to non-independencies in the data arising out of repetition of tokens. A simple linear regression model was used for the analysis of fundamental frequency and vowel acoustic space area. Region, repetition, gender and vowel duration were included in the model for F1 and F2 as fixed effects. Speaker was added as a random effect in the model. Vowel duration was modeled with region, repetition, gender, F1 and F2 as fixed effects and speaker as a random effect. Fundamental frequency for males and females were modeled separately using a simple linear regression model with region and normalized values of F1 as the predictors. Finally, vowel space area was a function was region was modeled region as the independent variable.

4. 1 Summary of findings

4.1.1 F1 and F2

The various sections of this chapter 4 have presented an analysis of the six oral monophthongs in Nepali in terms of their first and second formant frequencies, vowel duration and fundamental frequency grouped by gender. The vowel space area was quantified as the area encompassing the convex hull of all peripheral vowels. The hypothesis which was tested was that geographical region of the speaker affects the five acoustic measures outlined earlier. The hypothesis did not hold true for the point vowels [i, ɑ, u]. This is not surprising considering the fact that these vowel categories are preferred crosslingusitically. They lie in regions of acoustic stability and are quantal in nature (Steven 1972, 1989).

The vowels which were affected as a function of region were [ʌ, o, e]. From the analysis of the F1, taking all four regions into consideration, it was observed that region had a significant effect on vowel height mostly for back vowels such as [ʌ] ($\chi2$ (3) =12.941, p = 0.004767**) and [o] ($\chi2$ (3) = 10.749, p = 0.01316*).There was a significant effect of region on F2 for [e] ($\chi2$ (3) = 8.2164, p =0.04174*). The Darjeeling and the Sikkim variants of [ʌ] were significantly raised in comparison to the Dooars variant. Vowel duration of [ʌ] was significantly longer for the standard Nepal variant than Darjeeling and Sikkim variants. [o] was significantly raised in all regional varieties in comparison to the standard Nepal variety (all $p<0.05$). The Dooars variant of [o] was also significantly higher in the vowel space in comparison to the Sikkim variant ($p<0.05$). [e] in Darjeeling and Nepal variants was significantly more fronted compared to the Dooars variant. Although the overall models were not significant for other vowel categories such as [i, ɑ, u], significant contrasts between regions were observed between Darjeeling and Sikkim for [i] for F1 with the Sikkim variant raised in comparison to the Darjeeling variant ($p<0.05$). Findings also revealed that Dooars [ɑ] is raised significantly in relation to the Darjeeling variant ($p<0.05$).

4.1.2 Vowel Duration

The findings for vowel duration in the pilot study had revealed a significant effect of region across all categories. Vowel duration in the standard Nepal variety was significantly longer than vowels in the Darjeeling variety for all the six vowels. The pilot study was based on data from the Nepal and the Darjeeling varieties. With two more regions added, the only significant model was for the vowel [ʌ] ($\chi2$ (3) = 13.79, p = 0.003206**). However, the findings corroborate the results from the pilot study as a similar trend was observed even though the models were not significant. Vowel duration for all the six vowels in the standard variety spoken by speakers from Nepal was found to be longer than the other three regional varieties included in this study. The only significant inter-group contrast was found between Nepal and Dooars speakers for [o] ($p<0.05$).

Figure 4.1 presents a comparison of vowel duration across region and across all vowel categories. All the six vowels are longer in the Nepal. The vowel [ɑ] is longer for Darjeeling speakers compared to Dooars speakers but marginally shorter in comparison to Sikkim speakers. Vowel [ʌ] is significantly longer in the standard variety compared to Darjeeling and Sikkim varities. The Dooars variant of [ʌ] is also longer than the Darjeeling and the Sikkim variants but the contrast isn't statistically significant. [e] is the longest in Nepal followed by Darjeeling, Dooars and Sikkim. None of the between- group contrasts for vowel duration of [e] are statistically significant. Similarly between-group contrasts for vowel duration of [i] is significant. [i] in the Nepal variety is marginally longer than the Dooars variant. The Sikkim variant is the shortest while the Dooars variant of [i] is longer than the Darjeeling variant. Vowel [o] is longest in Nepal followed by the variants in Darjeeling, Sikkim and Dooars. The Dooars variant of [o] is significantly shorter relative to the standard variant. There are no significant between-group contrasts for [u] as far as vowel duration is concerned. [u] in the Nepal variety is the longest followed by Dooars, Sikkim and Darjeeling where it is the shortest.

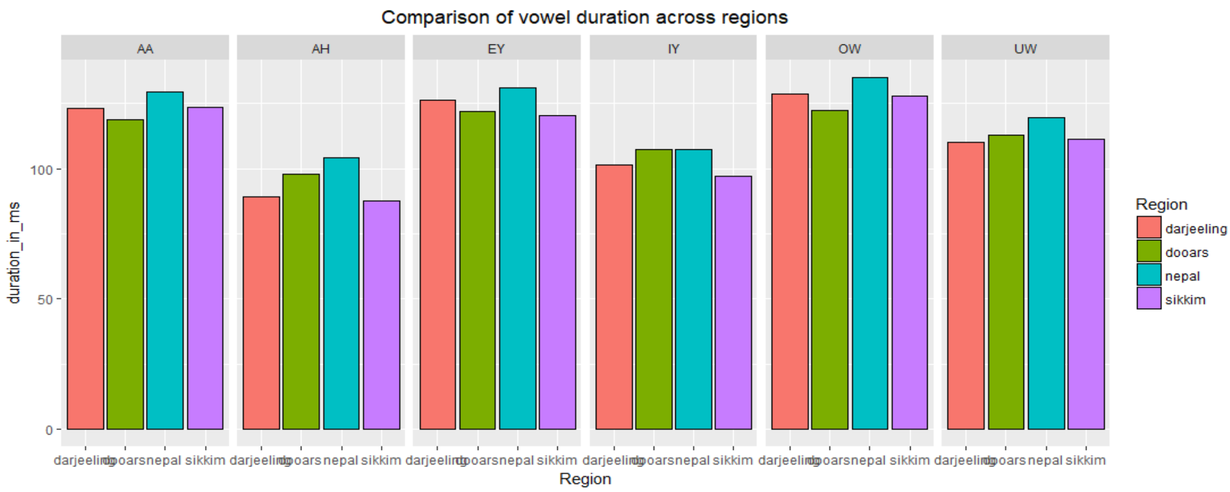


Figure 4. 1 Comparison of vowel duration across regions

4.1.3 Vowel Acoustic Space Area

The model of acoustic space area as a function of region was not significant ($F(3,73)=2.037$, $p=-0.1161$). The comparison of vowel space areas and dispersion of vowels in them are discussed in the next sub-sections.

4.1.3.1 Comparison of vowel acoustic space areas between Nepal and Darjeeling

This section presents a comparative analysis of the vowel dispersion in the acoustic space between the standard Nepal variety and the Darjeeling variety. The model estimates for vowel space area as a function of region presented in table 3.33 indicated a positive estimate for the hull area of the Darjeeling variety with reference to the Nepal variety. This implies that the vowel space area for the Darjeeling variety is bigger compared to the standard dialect. Although this contrast is not significant, it can be observed that the vowels are dispersed in a larger area in the Darjeeling variety when compared to the dispersion pattern in the standard dialect. Figure 4.2 compares the dispersion of vowels of the Nepal and the Darjeeling varieties in the acoustic space with mean values of normalized F1 on the x-axis and normalized values of F2 on the y-axis.

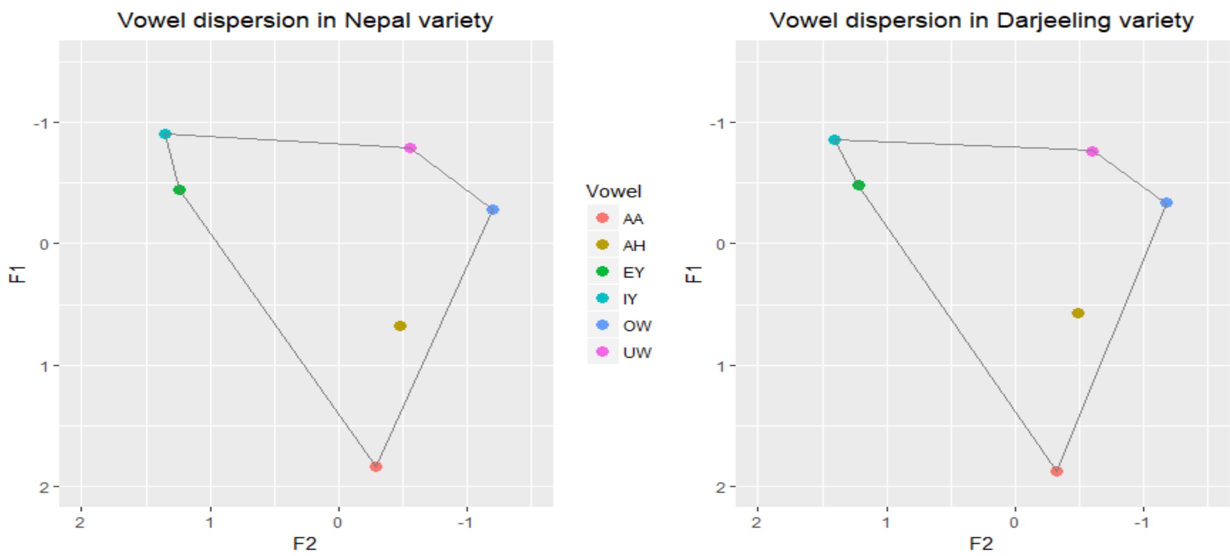


Figure 4. 2 Comparison of vowel acoustic space areas between Nepal and Darjeeling

From the plot in figure 4.2 above, it can be observed that [i] in the Nepal variety is marginally raised in comparison to the Darjeeling variant whereas the Darjeeling variant is more fronted. [e] in the Darjeeling variety in raised in comparison to the standard dialect but the Nepal variant is fronted. The vowel [ɑ] is raised, and fronted in the Nepal variety in contrast to [ɑ] in the Darjeeling variety. The vowel [ʌ] is lower for Nepal speakers but is marginally fronted. The Darjeeling variant of [o] is raised and fronted in comparison to the standard dialect. The vowel [u] is raised and fronted for speakers from Nepal when compared to speakers from Darjeeling.

4.1.3.2 Comparison of vowel acoustic space areas between Nepal and Dooars

This section compares the dispersion of the six oral monophthongs in Nepali between the standard Nepal variety and the Dooars variety in the vowel acoustic space. From the table on model estimates for hull area as function of region presented in table 4.33, a negative estimate for Dooars with reference to Nepal indicates that the vowel space area for Dooars is smaller. This means that vowel dispersion in the Dooars variety takes place in a relatively smaller area compared to the standard variety. Figure 4.3 compares the vowel space areas of the Nepal and the Dooars varieties with mean values of normalized F1 on the x-axis and normalized values of F2 on the y-axis.

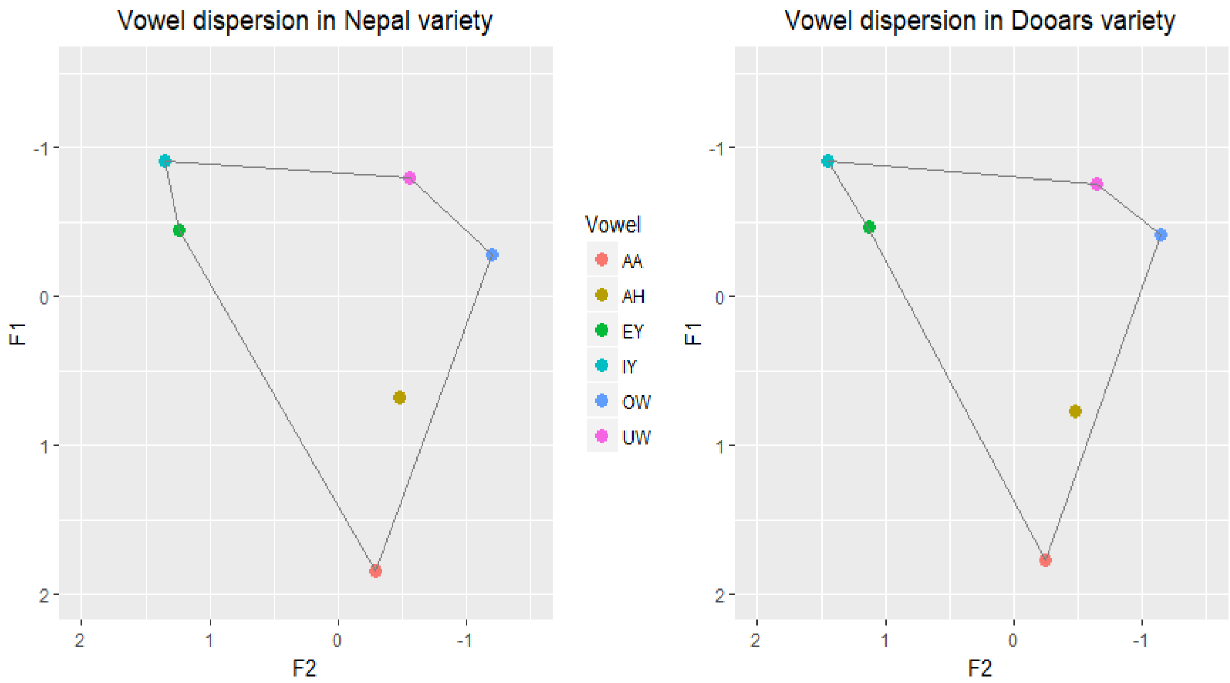


Figure 4. 3 Comparison of vowel acoustic space areas between Nepal and Dooars

It can be observed from the vowel plot that the Dooars variant of [i] is raised and fronted in comparison to the standard variant. [e] in the Nepal variety is lower but it is significantly fronted in comparison to the Dooars variant. The Dooars variant of [ɑ] is raised and fronted in comparison to the standard dialect. The vowel [ʌ] is raised in the Nepal variety when compared with the Dooars variant which is marginally fronted in turn. The Dooars variant of [o] is significantly raised and higher in the vowel acoustic space in comparison to the standard dialect. On the axis of fronting, the Dooars [o] is fronted with higher mean F2 values. The vowel [u] is raised and fronted in the Nepal variety in contrast to the Dooars variety.

4.1.3.3 Comparison of vowel acoustic space areas between Nepal and Sikkim

This section compares the dispersion of the six oral monophthongs in Nepali between the standard Nepal variety and the Sikkim variety in the vowel acoustic space. The model estimates for convex hull area as a function of region presented in table 4.33 indicates a positive estimate for Sikkim with reference to Nepal. This means that the vowel space area for Sikkim is relatively bigger with vowel dispersion taking place in a larger area. Figure 4.4 compares the vowel space areas of the Nepal and the Sikkim varieties with mean values of normalized F1 on the x-axis and normalized values of F2 on the y-axis.

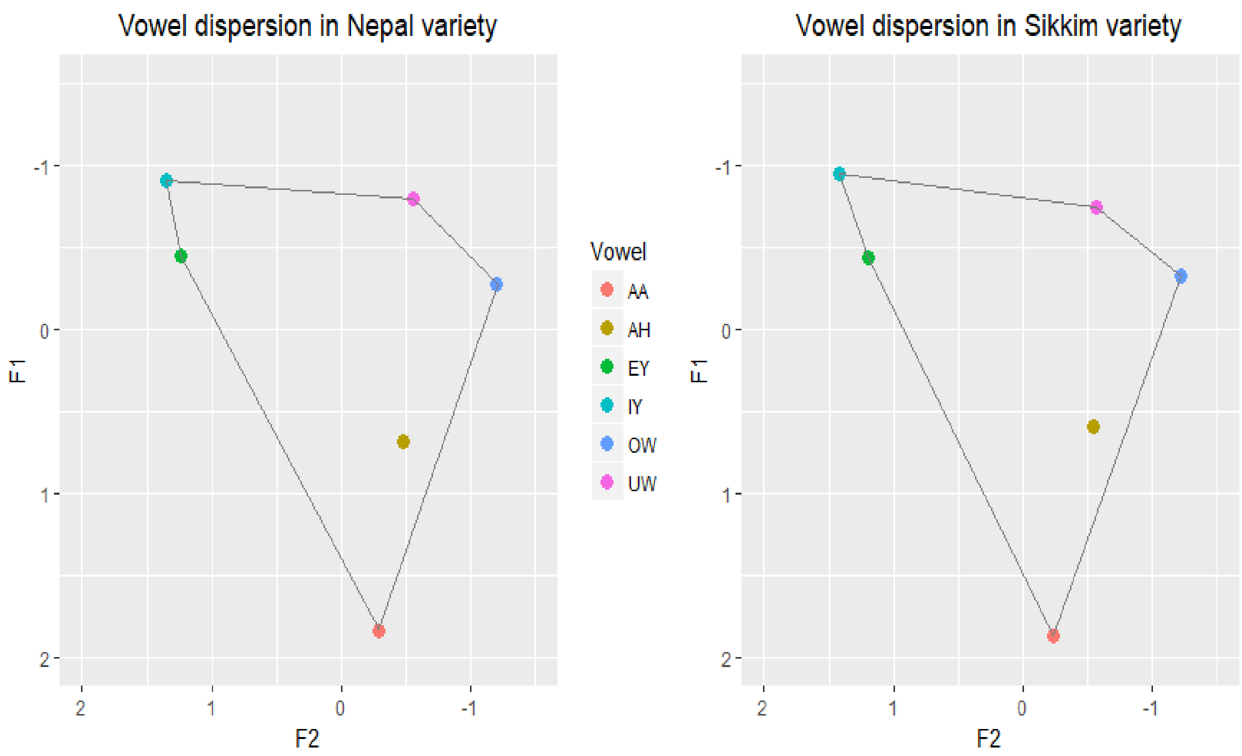


Figure 4. 4 Comparison of vowel acoustic space areas between Nepal and Sikkim

From the plot in figure 4.4 it can be observed that [i] in the Sikkim variety is raised and fronted in comparison to [i] in the standard dialect spoken by speakers from Nepal. The vowel [e] is raised and fronted for speakers of the Nepal variety than for speakers from Sikkim. The vowel [ɑ] is raised for Nepal speakers but the Dooars variant of [ɑ] is fronted in comparison. [ʌ] in the standard dialect is also is lower and fronted in comparison to the Sikkim variant. The vowel [o] in the Nepal variety is fronted with reference to the Sikkim variant of [o] which in turn is raised and higher in the vowel acoustic space in comparison to the standard dialect. The Nepal variant of [u] is raised and fronted in comparison to the Sikkim variant.

4.1.3.4 Comparison of vowel acoustic space areas between Darjeeling and Dooars

This section compares the dispersion of the six oral monophthongs in Nepali between the Darjeeling variety and the Dooars variety. The negative estimate for the Dooars variety with reference to the Darjeeling variety presented in table 4.33 indicates that the vowel space area for the Dooars variety is relatively smaller and vowel dispersion takes place in a smaller area in the Dooars variety *vis-à-vis* Darjeeling variety. Figure 4.5 compares the vowel space areas of the Darjeeling and the Dooars varieties with mean values of normalized F1 on the x-axis and normalized values of F2 on the y-axis.

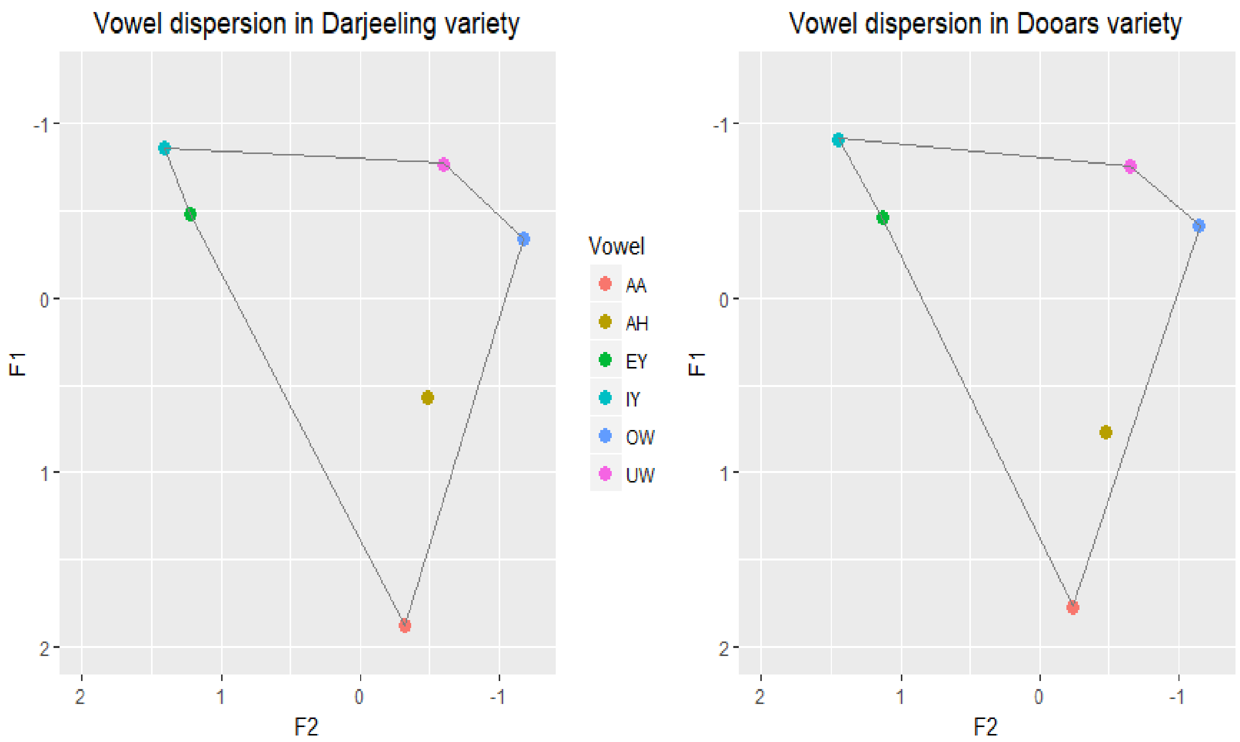


Figure 4. 5 Comparison of vowel acoustic space areas between Darjeeling and Dooars

From the plot in figure 4.5 it is evident that the Dooars variant of [i] is higher and fronted in the vowel acoustic space. The vowel [e] in the Darjeeling variety is significantly fronted compared to the Dooars variant. The Darjeeling [e] is also raised in comparison to the [e] from Dooars speakers. [ɑ] for Dooars speakers is significantly raised in comparison to speakers from Darjeeling. The Dooars variant of [ɑ] is fronted in comparison to the Darjeeling variant. The vowel [ʌ] in the Darjeeling variety is significantly raised with reference to the Dooars variant which is fronted than its Darjeeling counterpart. Compared to the Darjeeling variety, [o] in the Dooars variety is raised, and fronted. With respect to [u], the Darjeeling variant is raised and fronted.

4.1.3.5 Comparison of vowel acoustic space areas between Darjeeling and Sikkim

This section compares the dispersion of the six oral monophthongs in Nepali between the Darjeeling variety and the Sikkim variety in the vowel acoustic space area. From table 4.33 on model estimates for convex hull area as a function of region, it can be observed that the estimate for Sikkim with reference to Darjeeling is positive which indicates that the vowel space for Sikkim is larger and vowel dispersion in the Sikkim variety takes place in a larger area. Figure 4.6 compares the vowel space areas of the Darjeeling and the Sikkim varieties with mean values of normalized F1 on the x-axis and normalized values of F2 on the y-axis. The sizes of the points correspond to mean values of vowel duration for that particular vowel category.

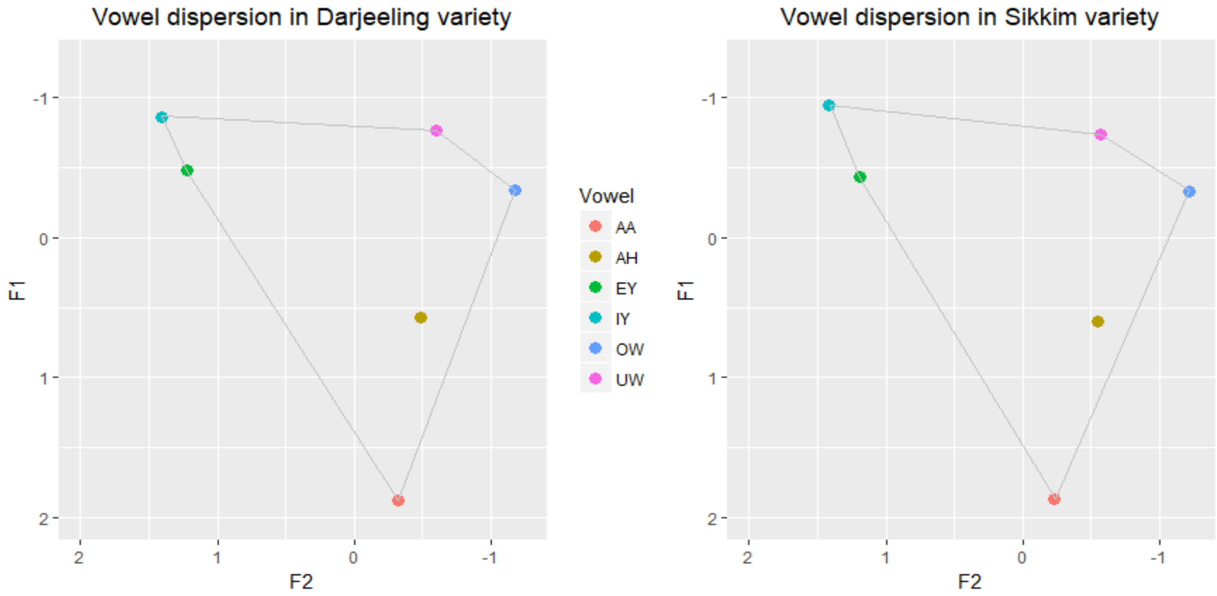


Figure 4. 6 Comparison of vowel acoustic space areas between Darjeeling and Sikkim

Comparison of vowel dispersion patterns between the Darjeeling and the Sikkim varieties indicates that [i] in the Sikkim variety is significantly higher in the vowel space than the Darjeeling variant. The Sikkim [i] is fronted relative to the Darjeeling variant. The vowel [e] is raised and fronted in the Darjeeling variety when compared to the Sikkim variant of [e]. The vowel [ɑ] is higher and fronted in the Sikkim variety than in the Darjeeling variety. Compared to the Sikkim variant of [ʌ], the Darjeeling variant is raised and fronted. The vowel [o] for speakers from Sikkim is lower and back-er in the vowel space in comparison to speakers from Darjeeling. The point vowel [u] is lower and fronted for speakers from Sikkim in contrast to speakers from Darjeeling.

4.1.3.6 Comparison of vowel acoustic space areas between Dooars and Sikkim

This section compares the dispersion of the six oral monophthongs in Nepali between the Dooars variety and the Sikkim variety in the vowel acoustic space. The positive estimate for Sikkim with reference to Dooars in table 4.33 on model estimates for convex hull area as function of region indicates that the vowel space area for the Sikkim variety is significantly larger ($p<.005$) compared to the vowel space area for the Dooars variety. This means that vowel dispersion takes place in a significantly larger area for the Sikkim variety. Figure 4.7

compares the vowel space areas of the Dooars and the Sikkim varieties with mean values of normalized F1 on the x-axis and normalized values of F2 on the y-axis.

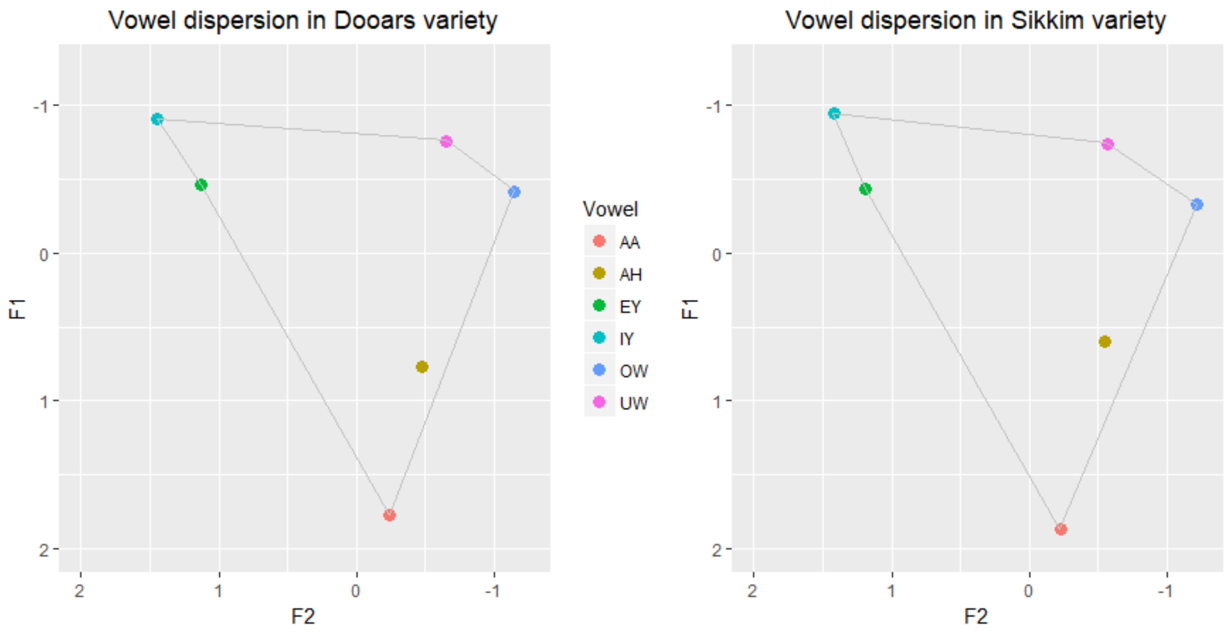


Figure 4. 7 Comparison of vowel acoustic space areas between Dooars and Sikkim

From the plot in figure 4.7 it is evident that vowel [i] in the Sikkim variety is higher than [i] in the Dooars variety. The Dooars [i] is marginally fronted in comparison. Vowel [e] is raised in the Dooars variety but the Sikkim variant of [e] is fronted. The Dooars variant of [ɑ] is raised whereas the Sikkim variant is marginally fronted. The Sikkim variant of [ʌ] is significantly higher in the vowel space compared to the Dooars variant which is fronted relative to the Sikkim variant. [o] in the Dooars variety is significantly higher than [o] from speakers from Sikkim. The Dooars [o] is also fronted in comparison. The vowel [u] is lower but fronted in the Sikkim variety when compared to the Dooars variant.The between-group contrasts for the six vowel categories are presented in table 4.1

Table 4.1 Between-group contrasts for F1, F2 and vowel duration

		Nepal	Darjeeling	Dooars	Sikkim
[i]	Nepal		Raised; back; longer	Lower; **back*;** longer	Lower; back; longer
	Darjeeling	Lower; fronted; shorter		Lower; back; shorter	**Lower***; back; longer
	Dooars	Raised; **fronted*;** shorter	Raised; fronted; longer		Lower; fronted; shorter
	Sikkim	Raised; fronted; shorter	**Raised** *; fronted; shorter	Raised; back; shorter	
	Nepal		Lower; fronted;	Lower;	Raised;fronted;

[e]			longer	**fronted**;** longer	longer
	Darjeeling	Raised;back; shorter		Raised;**fronted**;** longer	Raised;fronted; longer
	Dooars	Raised;**back**;** shorter	Lower;**back**;** shorter		Raised;back;longer
	Sikkim	Lower;back; shorter	Lower; back; shorter	Lower;fronted; shorter	
[a]	Nepal		Raised; fronted; longer	Lower; back; longer	Raised; back ; longer
	Darjeeling	Lower;back; shorter		**Lower***;back; longer	Lower;back; shorter
	Dooars	Raised;fronted; shorter	**Raised***;fronted; shorter		Raised;back;shorter
	Sikkim	Lower;fronted; shorter	Raised;fronted; longer	Lower; fronted; longer	
[ʌ]	Nepal		Lower;bfronted; **longer****	Raised; back; longer	Lower; fronted; **longer*****
	Darjeeling	Raised; back; **shorter ****		**Raised***; back; shorter	Raised; fronted; longer
	Dooars	Lower; fronted, shorter	**Lower***; fronted; longer		**Lower***; fronted, **longer***
	Sikkim	Raised; back; **shorter*****	Lower; back; shorter	**Raised***; back; **shorter***	
[o]	Nepal		Lower; back; longer	**Lower****; back; **longer***	Lower; fronted; longer
	Darjeeling	Raised; fronted; shorter		Lower; fronted; longer	Raised; back; longer
	Dooars	**Raised****; fronted; shorter	Raised; fronted; shorter		**Raised***; fronted; shorter
	Sikkim	Raised; back; shorter	Lower; back; shorter	**Lower***; back; longer	
[u]	Nepal		Raised;Fronted; longer	Raised;fronted; longer	Raised;fronted; longer
	Darjeeling	Lower;back; shorter		Raised;Fronted; shorter	Raised;back; shorter
	Dooars	Lower; back; shorter	Lower;back; longer		Raised; back; longer
	Sikkim	Lower;back; shorter	Lower;Fronted; shorter	Lower; fronted; shorter	

4.1.4 Summary of the analysis of fundamental frequency (f_0)

No significant effect of region on fundamental frequency (f_0) was observed for both gender groups across vowel categories. Analysis of fundamental frequency for females for the vowel [i] reveals that females in Nepal variety have the highest f_0 values followed by Darjeeling, Dooars and Sikkim variants. For male speakers, f_0 is lowest for speakers from Nepal followed by Sikkim, Darjeeling and Dooars. The f_0 for female speakers speaking the standard variant is lower than it is for females from other regions. The second lowest f_0 values are reported for females from the Dooars area. Darjeeling female speakers have the highest values for $f_{0.}$ From the analysis of fundamental frequency for male speakers, the only statistically significant inter-group difference observed was between male speakers from Dooars and Nepal for the vowel [i]. Fundamental frequency for male speakers from Dooars is significantly higher in comparison to their male counterparts from Nepal who speak the standard variant. f_0 of males speakers from Dooars is also higher than male speakers from Darjeeling and Sikkim. Male speakers from Nepal have the lowest values of f_0 for males while readings for the Darjeeling and the Sikkim variant are slightly higher in its comparison.

Analysis of fundamental frequency for female speakers for the vowel [e] revealed no significant inter-group differences between the regional variants for [e] . The f_0 for female speakers speaking the standard variant is lower than it is for females from other regions. The second lowest f_0 values are reported for females from the Dooars area. Darjeeling female speakers have the highest values for $f_{0.}$ The only statistically significant inter-group difference was observed between male speakers from Dooars and Nepal. Fundamental frequency for male speakers from Dooars is significantly higher in comparison to their male counterparts from Nepal who speak the standard dialect. It is also higher than male speakers from Darjeeling and Sikkim. Speakers from Nepal have the lowest values of f_0 for males while readings for speakers from Darjeeling and the Sikkim are slightly higher in its comparison.

Analysis of fundamental frequency for female speakers for vowel [ɑ] revealed no significant between-group differences. Female speakers from Sikkim have the highest values for fundamental frequency for [ɑ] followed by females in Darjeeling, Dooars and Nepal. As in the earlier analysis of female speakers, there are no significant differences in fundamental frequency values for male speakers between the different regions. The estimates indicate that pitch of male speakers from Nepal is the lowest compared to males from other three regions. Dooars speakers have the highest fundamental frequency for [ɑ] whereas it is slightly higher in the Sikkim variant compared to the Darjeeling variant.

Analysis of f_0 for females speakers for [ʌ] indicates no significant between-group differences between the four regions included in this study. Compared to the standard variety, females speaking the Darjeeling and Dooars varieties of Nepali have marginally higher fundamental frequencies for [ʌ] while female speakers from Dooars have lower pitch values than females from Nepal, Darjeeling and Sikkim. There are no significant differences between male speakers in the four different regions as far as f_0 for males speakers for [ʌ] is concerned. Like females from Dooars, the estimate for male speakers also indicates lower f_0 values in comparison to males speaking the standard dialect. Male speakers from Darjeeling and Sikkim report slightly higher values for f_0 in comparison to male speakers from Nepal.

The analysis of fundamental frequency for female speakers for the vowel [o] revealed that compared to the standard variant, females in Sikkim and Darjeeling have marginally lower f_0 values while female speakers from Dooars have a marginally higher pitch for the [o] vowel than female speakers from Nepal. Females from Dooars have the highest mean for [o] while Sikkim speakers have the lowest. According to the estimates, the fundamental frequency of males in all the other three regions is higher than males from Nepal who speak the standard dialect. Dooars males have the highest pitch followed by Sikkim, Darjeeling and Nepal.

Model estimates for the analysis of fundamental frequency for female speakers for the vowel [u] revealed no significant between-group differences between the four regions. The positive estimates for Darjeeling, Dooars are Sikkim indicate that with reference to female speakers from Nepal, females in the other three regions have higher pitch values. Analysis of coefficients of fixed effects from the model constructed for male speakers indicates that the only significant between-group contrast is between Darjeeling and Dooars speakers. Compared to Darjeeling males, the estimate for male speakers from Dooars is more than 25 Hz ($p=0.035217$*). Male speakers from Dooars report the highest f_0 for [u] followed by Sikkim, Nepal and Darjeeling.

4.2 Scope for further research on regional variation in Nepali

Language variation and change is a burgeoning field of research and acoustic methods provide a multitude of ways for conducting dialectological research. Thomas (2018) argues that the use of acoustic methods in regional dialectology should not only be circumscribed to study of vocalic phenomena but expanded to the analysis of non-vocalic elements such as consonants and speech prosody. Even in the studies which have studied vowels, the traditional practice, as has been followed in the current study, has been to take a single measurement, usually at vowel mid-point. Measuring vowel trajectories is relatively new and is shedding newer insights on regional vowel variation. Further dialectal studies on Nepal vowels by measuring vowel inherent spectral change can probably lead to interesting findings. Similarly acoustic studies of consonants in the different regional varieties could also be an interesting topic for further research since there is no such research available for Nepali and processes such as pre-aspiration stop occlusions are difficult to observe auditorily. Factors associated with consonants such as Voice Onset Time (VOT), realizations of /l/ in varieties in English and realizations of /r/ in Spanish have all been studies using acoustic methods which indicates that these methods are robust and could be used in the analysis of Nepali consonants. Although this study has investigated the effect of region on fundamental frequency, there is a large scope for further study on acoustic analysis of prosodic variation in Nepali. Prosodic aspects such as speech rate, loudness, prosodic rhythm and intonation are known to differ on the basis of geographical location of the speakers. Therefore, the ground is fertile for research on prosodic aspects on Nepali and subsequent studies on regional variation on Nepali must analyze these variables.

The development of corpus tools has significantly facilitated major advances in the study of language variation and change. Tools such as forced alignment systems for automatic segmentation of speech has provided an objective, reliable and efficient method which can exponentially increase the number of tokens for analysis thereby increasing the statistical power in studies which employ them. Such resources, however, have mostly been developed for languages spoken in Europe and North America. There is a need to develop language resources for languages spoken in other parts of the world such as India which is characterized by multiplicity of languages and diverse forms of social structure and stratification. Varieties of language thrive and undergo diachronic change but sparse explanations exist for how social factors affect this process. There is a need for developing large spoken corpuses for Indian languages and tools to conduct analysis on variables of interest.

During the course of this study, an attempt was made to develop a forced alignment tool for Nepali (FANEP) which is described in the chapter on method. Developing this tool was a deviation from the research protocol originally envisaged but the utility of the tool cannot be underestimated. Although the tool placed the segment boundaries reasonably well for words in the datasets analyzed in this study, the alignments had to be verified by hand and adjustments had to be made in some cases. There is plenty of scope for its further development beginning with reliability checks, increased training data and regular expansion of the pronunciation dictionary.

Another scope for supplementing the findings of this research is adopting the latest metric known as the ‘formant space’ area (Fox and Jacewicz, 2017) in the estimation of vowel space area. The metric relies on measurements of vowel trajectory at multiple points in the course of a vowel to investigate vowel inherent spectral change. This current study revealed significant effect of region on the F1 of vowel /ʌ/ and /o/. The findings from this study are based on wordlist style data which no doubt has its advantages insofar as it yields longer and stressed tokens in a comparable phonetic context. The tradeoff is that this reduces naturalness in speech which can be in important aspect for analysing variation. Therefore, future studies on regional vowel variation in Nepali must try to incorporate data from other elicitation techniques such as reading passages and conversations.

4.3 Conclusion

Attempt must be made to include a wide array of variables – consonantal, vocalic, prosodic and voice quality – within the same study to gain a holistic understanding of how social identity is indexed by linguistic variation. Linguistic studies should also be conducted by taking more features into consideration. Although it is easier to conduct, single-feature studies may not yield sufficient explanations for phonetic processes such as undershoot which operate across phonetic domains. Similarly perceptual similarity studies across phonetic domains may yield similar principles. Theories of accurate percepetion of speech sounds place varying degrees of importance on motor control or acoustic signal which is centred on segments while excluding prosody. In the contemporary debate on the degree of distinctness between phonological and phonetic encoding, knowledge of how speakers produce and percieve variant forms and their relations to deep structure could inform at what stage of encoding the variant forms are specified. This requires a range of variables to be studied.

Bibliography

Acharya, B. (2053 BS). Allocation of Nepali Dialects Based on Pronunciation [in Nepali]. *M.A. Thesis* . Central Department of Nepali, Tribhuvan University.

Acharya, J. (1991). *A Descriptive Grammar of Nepali and an Analyzed Corpus.* Washington D.C.: Georgetown University Press.

Adank, P., Hout, R. v., & de Velde, H. v. (2007). An acoustic description of the vowels of northern and southern standard Dutch II: Regional varieties. *The Journal of the Acoustic Society of America 121(2)* , 1130-1141.

Adank, P., Hout, R. v., & Smits, R. (2004). A comparison of vowel normalization methods for language variation research. *The Journal of the Acoustical Society of America 116(5)* , 3099-3107.

Adank, P., Hout, R. v., & Smits, R. (2004). An acoustic description of vowels of Northern and Southern Standard Dutch. *The Journal of the Acoustical Society of America 116(3)* , 1729-1738.

Al-Tamimi, J. E., & Ferragne, E. (2005). Does vowel space size depend on language vowel inventories? Evidence from two Arabic dialects and French. *in Proceedings of the 9th European Conference on Speech Communication and Technology INTERSPEECH-EUROSPEECH 2005, Lisbon, Portugal*, (pp. 2465–2468).

Ashby, M., & Maidment, J. (2005). *Introducing Phonetic Science.* Cambridge: Cambridge University Press.

Ayton, J. A. (1820). *A Grammar of Nepali Language.* Calcutta.

Bandhu, C. (1968). *Nepali bhasako utpatti.* Kathmandu: Sajha Prakashan.

Bandhu, C., Dahal, B. M., Holzhausen, A., & Hale, A. (1971). *Nepali Segmental Phonology.* Kathmandu: Summer Institute of Linguistics, Tribhuvan University.

Barney, G. P. (1952). Control methods used in a study of the vowels. *J. Acoust. Soc. Am. 24* , 175-184.

Baskaran, S. G. (2011). *Nepali in Himachal Pradesh: Linguistic Surevey of India-Himachal Pradesh.* Kolkata: Office of the Registrar General.

Baskaran, S. (2011). *Nepali in Himachal Pradesh: Linguistic Surevey of India-Himachal Pradesh.* Kolkata, India: Office of the Registrar General.

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software, 67(1)* , 1-48.

Blevins, J. (2004). *Evolutionary Phonology: The Emergence of Sound Patterns.* Cambridge UK/ New York: Cambridge University Press.

Boersma, P. (2001). Praat, a system for doing phonetics by computer. *Glot International 5:9/10* , 341-345.

Bradley, T. G., & Willis, E. W. (2012). Rhotic variation and contrast in Veracruz Mexican Spanish. *Estudios de Fonética Experimental 21* , 43-74.

Bradlow, A. R., Toretta, G., & Pisoni, D. (1996). Intelligibility of normal speech I: Global and fine-grained acoustic-phonetic talker characteristics. *Speech Commun. 20, https://doi.org/10.1016/S0167-6393(96)00063-5* , 255–272. .

Carver, C. M. (1998). English dialectology and the linguistic atlas. In I. M. (Ed.), *Handbook of dialects and language variation (2nd ed.)* (pp. 5-28). San : Academic Press.

Catford, J. (1988). *A practical introduction to phonetics.* New York: Oxford University Press.

Catford, J. (1977). *Fundamental Problems in Phonetics.* Edinburgh: Edinburgh University Press.

Central Bureau of Statistics, N. (June 2002). National Population Census 2001 - Nepal (NPC 2001) of the public use dataset. National Data Archive, www.cbs.gov.np.

Chiba, T., & Kajiyama, M. (1942). *The Vowel, Its Nature and Structure.* Tokyo: Kaiseikan.

Chomsky, N., & Halle, M. (1968). *The Sound Pattern of English.* New York: Harper & Row.

CIIL Mysore. (n.d.). *Language Information Service (LIS) - India*. Retrieved September 12, 2011, from http://www.lisindia.net/Nepali/Nepali.html

Clopper, C. G. (2009). Computational methods for normalizing acoustic vowel data for talker differences. *Language and Linguistic Compass 3(6)* , 1430-1442.

Clopper, C. G., Pisoni, D. B., & de Jong, K. (2005). Acoustic characteristics of the vowel systems of six regional varieties of American English. *Journal of the Acoustical Society of America; 118* , 1661–1676.

Colantoni, L. (2006). Micro and macro sound variation and change in Argentine Spanish. In Nuria Sagarra and and Almeida Jacqueline Toribio (eds.). *Selected Proceedings of the 9th Hispanic Linguistics Symposium* (pp. 91-102). Somerville, MA: Cascadilla Proceedings Project. www.lingref.com, document #1369.

Cooper, F. L. (1951). The interconversion of audible and visble patterns as a basis for research in perception of speech. *Proceedings of the National Academy of Science, 37,* (pp. 318-325).

Cucchiarini, C., & Strik, H. (2003). Automatic phonetic transcription: An overview. *In Proceedings of the International Congress of Phonetic Sciences (ICPhS), Barcelona,15*, (pp. 347-350).

Dahal, B. M. (1974). *A Description of Nepali: Literary and Colloquial.* Ph.D thesis, University of Poona.

Delattre, P. (1951). The physiological interpretation of sound spectrograms. *Publications of the Modern Language Association, 66* , 864-875.

Disner, S. F. (1980). Evaluation of vowel normalization procedures. *The Journal of the Acoustical Society of America 67(1) DOI:http://dx.doi.org/10.1121/1.383734* , 253-261.

Disner, S. (1983). Vowel quality: The relation between universal and language-specific factors. 58. UCLA Working Papers in Phonetics.

Dixitacharya, B. (1913). *Gorkha Byakarana Bodha.* Kathmandu.

Docherty, G. J., Watt, D., Llamas, C., Hall, D., & Nycz, J. (2011). Variation in voice onset time alongthe Scottish-English border. *In Proceedings of ICPhS XVII, Hong Kong, 17-21 August 2011*, (pp. 591-94).

Eriksson, A. (2004). Swedia-Projektet: Dialektforskning i ett jämförande perspektiv. *Folkmålsstudier 43* , 11–31.

Fant, G. (1960). *Acoustic Theory of Speech Production.* The Hauge: Mouton.

Flynn, N. E., & Foulkes, P. (2011). Comparing vowel formant normalization methods. *Proceedings of the 17th International Congress of Phonetic Sciences* (pp. 683-686). University of Hong Kong.

Fourakis, M. (1991). Tempo, stress, and vowel reduction in American English. *J. Acoust. Soc. Am. 90* , 1816–1827.

Fox, R. A., & Jacewicz, E. (2008). Analysis of total vowel space areas in three regional dialects of American English. *in Proceedings of Acoustics'08 (SFA, Paris, France)*, (pp. 495-500).

Fox, R. A., & Jacewicz, E. (2017). Reconceptualizing the vowel space in analyzing regional dialect variation and sound change in American English. *Journal of Acoustical Society of America 142 (1) July* , 444-459.

Fridland, V. (2000). The Southern Shift in Memphis, Tennessee. *Language Variation and Change 11* , 267-85.

Fry, D. (1979). *The Physics of Speech.* London: Cambridge University Press.

Gorman, K., Howell, J., & Wagner, M. (2011). Prosodylab-Aligner: A Tool for Forced Alignment of Laboratory Speech. *Canadian Acoustics. 39.3* , 192-193.

Grierson, G. (Linguistic Survey of India 9.4). *1916.* Calcutta: Government of India.

Hagiwara, R. (1997). Dialect variation and formant frequency: The American English vowels revisited. *The Journal of the Acoustical Society of America, 102* , 655-658.

Hayward, K. (2000). *Experimental Phonetics.* Harlow, UK: Pearson.

Higgins, C. M., & Hodge, M. M. (2002). Vowel area and intelligibility in children with and without dysarthria. *J. Med. Speech Lang. Pa. 10* , 271–277.

Hillenbrand, J., Getty, L. A., Clark, M. J., & Wheeler, K. (1995). Acoustic characteristics of American English vowels. *The Journal of the Acoustical Society of America 97(5) Pt.1* , 3099-3111.

Hoernle, A. (1880). *Comparative Grammar of the Gaudian Languages with Special Reference to the Eastern Hindi.* London: Rubner.

Jacewicz, E. F. (2007a). Vowel duration in three American English dialects. *American Speech, 82* , 367-385.

Jacewicz, E. F. (2007 b). Vowel space areas across dialects and gender. *Proceedings of the XVIth International. Congress of Phonetic Sciences,* (pp. 1465-1468).

Jacewicz, E., Fox, R. A., & Salmons, J. (2011). Cross-generational vowel change in American English. *Language Variation and Change 23* , 45–86.

Jakobson, R. F. (1952). *Preliminaries to Speech Analysis. The distinctive features and their correlates.* Cambridge, MA: Acoustics Laboratory, Massachusetts Inst. of Technology (Re-published by MIT press, seventh edition, 1967).

Johnson, K. (1997). *Acoustic and auditory phonetics.* London: Blackwell.

Joos, M. (1948). *Acoustic Phonetics.* Baltimore: Linguistic Society of America.

Kendall, T., & Thomas, E. R. (2014). vowels: Vowel Manipulation, Normalization, and Plotting. *R package version 1.2-1* . https://CRAN.R-project.org/package=vowels.

Kent, R. D. (1997). *The Speech Sciences.* San Diego: Singular Publishing Group.

Kewley-Port, D., Akahane-Yamada, R., & Aikawa, K. (1996). Intelligibility and Acoustic Correlates of Japanese Accented English Vowels. *4th ICSLP*, (pp. 450-453). Philadelphia. USA.

Khatiwada, R. (2009). Nepali. *Journal of the International Phonetic Association: Illustrations of the IPA 39/3* , 373-380.

Labov, W. (1963). The Social Motivation of Sound Change. *Word, 19* , 273-309.

Labov, W. (1966). *The Social Stratification of English in New York City.* Washington, D.C.: Center for Applied Linguistics.

Labov, W., Ash, S., & Boberg, C. (2006). *Atlas of North American English: Phonology and Phonetics.* Berlin: Mouton/de Gruyter.

Labov, W., Yaeger, M., & Steiner, R. (1972). *A Quantitative Study of Sound Change in Progress Philadelphia.* U.S. Regional Survey.

Ladefoged, P. (2006 (5th edn.)). *A Course in Phonetics.* New York: Harcourt Brace, Jovanovich, Inc.

Ladefoged, P. (1962). *Elements of Acoustic Phonetics.* Chicago : The University of Chicago Press.

Ladefoged, P. (2001). *Vowels and Consonants: An Introduction to the sounds of Languages.* . Los Angeles: Blackwell Publishing.

Ladefoged, P., & Broadbent, D. E. (1957). Information Conveyed by Vowels. *The Journal of the Acoustical Society of America. Volume 29. Number 1 January* , 98-104.

Lehiste, I. (1970). *Suprasegmentals.* Massachussets: MIT Press, Cambridge.

Leinonen, T. N. (2010). An acoustic analysis of vowel pronunciation in Swedish dialects. *Ph.D. dissertation* . Rijksuniversiteit Groningen.

Lewis, M. P. (2009). Ethnologue: Languages of the World, Sixteenth edition. Dallas.

Liljencrants, J. L. (1972). Numerical simulation of vowel quality systems: The role of perceptual contrast. *Language, 48* , 839-862.

Lindblom, B. (1986). Phonetic Universals in Vowel System. In I. J. (Eds), *Experimental Phonology* (pp. 13-44). New York: Academic Press.

Lisker, L. (1974). On "explaining" vowel duration variation. *Glossa 8* , 233-48.

Lobanov, B. (1971). Classification of Russian vowels spoken by different speakers. *The Journal of the Acoustic Society of America 49(2)* , 606-608.

Lohagun, N. (2016). Formant Patterns and Acoustic Space of Vowels in the Darjeeling variety of Nepali. In D. Gupta, & D. Misra(eds), *Perspectives in General and Applied Linguistics: Papers in honor of Prof.Vaishna Narang.* New Delhi: Vijaya Books.

Mattingly, I. (1999). A Short History of Acoustic Phonetics in the US. *In Proceedings of the XIVth International Congress of Phonetic Sciences, ed. John.J.Ohala et al.* (pp. 1-6). San Francisco: Haskins.

McAuliffe, M., Socolof, M., Mihuc, S., Wagner, M., & Sonderegger, M. (2017). Montreal Forced Aligner: trainable text-speech alignment using Kaldi. *Proceedings of the 18th Conference of the International Speech Communication Association.*

McCloy, D. (2016). phonR: tools for phoneticians and phonologists. R package version 1.0-7.

Misra, D. (2009). *Formant Structures of Hindi and Punjabi Vowels: A study in Comparing Acoustic Space.* Unpublished M.Phil Dissertation. Centre for Linguistics, JNU, New Delhi.

Moosmüller, S., & Granser, T. (2006). The spread of Standard Albanian: An illustration based on an analysis of vowels. *Language Variation and Change, 18(2) doi:10.1017/S0954394506060066* , 121-140.

Nakkeerar, R. (2011). *Nepali in Sikkim: Linguistic Survey of India: LSI-SIKKIM (Part-II).* Kolkata: Office of the Registrar General.

Narang, V. (1989). A Spectrographic Study of Punjabi Vowels. *OPEL,vol.3, CLE/JNU* .

Narang, V. a. (2010). A Contrastive Study of F3 correlation with F2 and F1 in Thai and Hindi. *International Journal on Asian Language Processing 20 (3)* .

Narang, V. a. (2010). A Study in Comparing Acoustic Space: Korean and Hindi Vowels. *International Journal on Asian Language Processing 20 (3)* .

Narang, V. a. (2010). Acoustic Space, Duration and Formant Patterns in vowels of Bangkok Thai. *International Journal on Asian Language Processing 20 (3)* , 123-140.

Office of The Registrar General & Census Commissioner, India, New Delhi. (2010-11). *Abstract of Speakers' Strength of Languages and Mother Tongues - 2001* . Retrieved August 12, 2012, from http://www.censusindia.gov.in/Census_Data_2001/Census_Data_Online/Language/Statement1.htm

Ohala, J. (1997). Phonetics in phonology. *4th Seoul International Conference on Linguistics [SICOL] 11-15 Aug 1997*, (pp. 45-50). Seoul .

Ohala, J. (1979). The Contribution of Acoustic Phonetics to Phonology. In B. a. In Lindblom, *Frontiers of Speech Communication Research* (pp. 355-363). New York: Academic Press.

O'Malley, L. (1907). *Darjeeling District Gazetteer* . New Delhi: Reprint published by Gyan Publishing House.

Peterson, G. a. (1960). Duration of Syllable Nuclei in English. *Journal of the Acoustical Society of America 32* , 693-703.

Peterson, G. E., & Barney, H. L. (1952). Control methods used in a study of the vowels. *The Journal of the Acoustical Society of America 24(175)* , 175-184.

Pickett, J. M. (1996). *The Sounds of Speech Communication - A Primer of Acoustic Phonetics.* Boston: Allyn and Bacon.

Pokharel, M. P. (1989). *Experimental Analysis of Nepali Sound System.* Ph.D. dissertation, University of Pune.

Pokharel, M. P. (2009). Origin and Development of Nepali Language. In A. S. T.B. Subba, *Indian Nepalis - Issues and Perspectives* (pp. 325-338). New Delhi: Concept Publishing Company.

Purnell, T., Salmons, J., & Tepeli, D. (2005a). German substrate effects in Wisconsin English: Evidence for final fortition. *American Speech 80* , 135-64.

Purnell, T., Salmons, J., Tepeli, D., & Mercer, J. (2005b). Structured heterogeneity and change in laryngeal phonetics: Upper Midwestern final obstruents. *Journal of English Linguistics 33* , 307–38.

R Core Team. (2013). R: A language and environment for statistical URL http://www.R-project.org/. Vienna, Austria.

Recasens, D. (2004). Darkness in /l/ as a scalar phonetic property: Implications for phonology and articulatory control. *Clinical Linguistics & Phonetics 18* , 593-603.

Recasens, D., & Espinosa, A. (2005). Articulatory, positional and coarticulatory characteristics for clear /l/ and dark /l/: Evidence from two Catalan dialects. *ournal of the International Phonetic Association 35* , 1-35.

Rosenfelder, I., Fruehwald, J., Evanini, K., & Yuan, J. (n.d.). “FAVE (Forced Alignment and Vowel Extraction) Program Suite". available at http://fave.ling.upenn.edu.

Roy, B. (1968). Darjeeling District Census Handbook 1961. Calcutta.

Sandoval.S, Berisha.V, Utiansk.R.L, Liss.J.M, & Spanias.A. (2013). Automatic assessment of vowel space area. *The Journal of the Acoustical Society of America 134* , EL477-EL483.

Schiel, F. (1999). Automatic Phonetic Transcription of Non-Prompted Speech. *Proc. of the ICPhS 1999. San Francisco, August 1999*, (pp. 607-610).

Schiel, F. (2004). MAUS Goes Iterative. *Proc. of the IV. International Conference on Language Resources and Evaluation, Lisbon, Portugal*, (pp. 1015-1018).

Schiel, F., & Draxler, C. (2003). The production of speech corpora. Bavarian Archive for Speech Signals, Tech. Rep.

Srivastava, S. (2011). *Nepali in West Bengal: Linguistic Survey of India: LSI - West Bengal.* Kolkata, India: Office of the Registrar General.

Srivastava, S. (2011). *Nepali in West Bengal: Linguistic Survey of India: LSI - West Bengal.* Kolkata: Office of the Registrar General.

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

Stevens, K. N. (1989). On the quantal nature of speech. *Journal of Phonetics 17* , 3–46.

Stevens, K. N. (1963). Perturbation of vowel articulations by consonantal context: An acoustical study. *Journal of Speech and Hearing Research, 6* , 112-128.

Stevens, K. (1972). The quantal nature of speech: evidence from articulatory-acoustic data. In I. E. (Eds.), *Human Communication: a Unified View* (pp. 51-66). New York: McGraw Hill.

Strange, W. (1999). Perception of vowels: dynamic constancy. In I. J. (Ed.), *The acoustics of speech communication: fundamentals, speech perception theory, and technology* (pp. 153-165). Surry, Maine: Allyn and Bacon.

Subedi, S. S. (2051 BS). A Survey of Nepali Dialects on Syntactic Features. *M.A. Thesis* . Central Department of Nepali, Tribhuvan University.

Syrdal, A. (1996). Acoustic variability in spontaneous conversational speech of American English talkers. *In ICSLP 96 (Fourth International Conference on Spoken Language Processing), Philadelphia, Oct. 3-–6,* (pp. 3-6).

Thomas, E. R. (2018). Acoustic Phonetic Dialectology. In C. Boberg, J. Nerbonne, & D. Watt, *The Handbook of Dialectology.* Oxford, U.K./Malden, MA: Wiley-Blackwell.

Thomas, E. R. (2001). An Acoustic Analysis of Vowel Variation in New World English. *Publication of the American Dialect Society 85.* Durham, NC: Duke University Press.

Thomas, E. R. (2011). *Sociophonetics: An Introduction.* Basingstoke: Palgrave Macmillan.

Thomas, E. R., & Kendall, T. (2007). NORM: The vowel normalization and plotting suite. [Online Resource: http://ncslaap.lib.ncsu.edu/tools/norm/].

Timilsina, Y. (2050 BS). Allocation of Nepali Dialects on Numerals and Pronominals [in Nepali]. *M.A.Thesis* . Centrral Department of Nepali, Tribhuvan University.

Tronnier, M. (2002). Preaspiration in southern Swedish dialects. *Proceedings of Fonetik, TMH-QPSR 44* (pp. 33-36). http://www.speech.kth.se/qpsr.

Turnbull, A. (1888). *The Nepali Grammar and Vocabulary.* Calcutta: Baptist Mission Press. [3rd edn].

Veatch, T. C. (1991). English vowels: Their surface phonology and phonetic implementation in vernacular dialects. Dissertations available from ProQuest. .

Vorperian, H. K., & Kent, R. D. (2007). Vowel Acoustic Space development in Children: A synthesis of Acoustic and Anatomical data. *Journal of Speech, Language and Hearing Research Vol.50.* , 1510-1545.

Weide, R. (1998). *The CMU pronunciation dictionary, release 0.6.* Carnegie Mellon University.

Weinreich, U., Labov, W., & Herzog, M. (1968). *Empirical foundations for a theory of language change.* Austin: University of Texas Press.

Weismer, G., Jeng, J. Y., Laures, J. S., Kent, R. D., & Kent, J. F. (2001). Acoustic and intelligibility characteristics of sentence production in neurogenic speech disorders,. *Folia Phoniatr. Logop. 53* , 1–18.

Willis, E. W. (2006). Trill variation in Dominican Spanish: An acoustic examination and comparative analysis. In N. Sagarra, & J. A. Toribio, *Selected Proceedings of the 9th Hispanic Linguistics Symposium* (pp. 121-31). Somerville: Cascadilla Proceedings Project.

Wretling, P., Strangert, E., & Schaeffler, F. (2002). Quantity and preaspiration in northern Sweden. *In Proceedings of speech prosody 2002* (p. not paginated). Aix-en-Provence, France: www.ling.gu.se/~anders/SWEDIA/papers/wretling_sp2002.pdf.

Yuan, J., & Liberman, M. (2008). Speaker identification on the SCOTUS corpus. *Proceedings of Acoustics '08.*

Appendix 1

Informed consent proforma

Proforma of Informed consent form for Principal investigator Consent form for patient or the subject. Retrieved from http://www.jnu.ac.in/IERB/Proposals.html

CONSENT FORM (for the subject/ patient)

The advantages and disadvantages of the research in which I am expected to participate, for which I have to donate blood/sputum/hair/voice sample has been explained to me.

I willingly, under no pressure from the researcher-

(i) agree to take part in this research, and agree to participate in all investigations which will help acquire knowledge for the benefit of the mankind,

(ii) agree to provide a voice sample

My consent is explicitly not for disclosing any personal information. For disclosing any such personal information obtained from the investigations conducted on my samples, further consent should be obtained.

I have been informed that JNU and the researchers (PI .... and her/his colleagues) will take my prior consent before they draw benefits from research based on my samples.

Signatures

-------------------- ------------------ -------------------------

Subject/patient Witness Principle Investigator.

Appendix 2

Speaker strength in different Indian states

Census Report (2001) for Nepali published by the Government of India.

India/State/Union Territory#	Persons	Male	Female
India*	2,871,749	1,534,746	1,337,003
West Bengal	1,022,725	514,596	508,129
Assam	564,790	293,122	271,668
Sikkim	338,606	174,068	164,538
Uttar Pradesh	263,982	145,106	118,876
Arunachal Pradesh	94,919	52,276	42,643
Uttaranchal	91,047	54,655	36,392
Himachal Pradesh	70,272	42,346	27,926
Maharashtra	63,480	41,028	22,452
Meghalaya	52,155	28,385	23,770
Manipur *	45,998	24,539	21,459
Delhi#	44,367	27,997	16,370
Nagaland	34,222	19,347	14,875
Haryana	20,362	13,899	6,463
Punjab	19,778	13,328	6,450
Bihar	18,763	9,861	8,902
Jharkhand	17,326	9,558	7,768
Gujarat	17,123	11,336	5,787

Madhya Pradesh	10,923	6,778	4,145
Rajasthan	10,569	7,225	3,344
Karnataka	10,038	6,661	3,377
Orissa	9,927	5,850	4,077
Mizoram	8,948	5,429	3,519
Andhra Pradesh	8,233	5,025	3,208
Jammu & Kashmir	8,199	5,787	2,412
Chandigarh#	5,390	3,516	1,874
Tamil Nadu	4,323	2,719	1,604
Chhattisgarh	3,424	1,995	1,429
Tripura	3,377	2,086	1,291
Kerala	2,715	1,912	803
Goa	2,135	1,478	657
Daman & Diu#	1,407	1,223	184
Dadra & Nagar Haveli#	1,030	840	190
Andaman & Nicobar Islands#	782	479	303
Pondicherry#	411	295	116
Lakshadweep#	3	1	2

Appendix 3

Mean of measurements by ID (normalized and unnormalized)

Region	Vowel	Gender	Mean F1 (Hz)	Mean F2 (Hz)	mean.normF1	mean.normF2	Mean F0 (Hz)	Mean duration (in ms.)
darjeeling	AA	female	860.15	1545.50	1.88	-0.42	216.70	126.00
darjeeling	AA	male	740.50	1425.09	1.87	-0.27	124.41	114.71
darjeeling	AH	female	599.09	1565.59	0.51	-0.36	219.45	93.55
darjeeling	AH	male	546.31	1238.61	0.62	-0.59	128.39	86.39
darjeeling	EY	female	436.10	2575.48	-0.40	1.24	228.86	136.38
darjeeling	EY	male	358.19	2171.48	-0.55	1.22	138.23	117.10
darjeeling	IY	female	362.14	2655.60	-0.78	1.38	240.68	106.41
darjeeling	IY	male	306.65	2276.00	-0.88	1.42	141.03	96.76
darjeeling	OW	female	438.96	1032.21	-0.37	-1.23	220.63	134.50
darjeeling	OW	male	394.63	965.28	-0.32	-1.14	135.69	122.19
darjeeling	UW	female	365.18	1419.91	-0.77	-0.59	231.86	116.68
darjeeling	UW	male	327.79	1218.79	-0.75	-0.62	139.41	106.47

Region	Vowel	Gender	Mean F1 (Hz)	Mean F2 (Hz)	mean.norm F1	mean.norm F2	Mean F0 (Hz)	Mean duration (in ms.)
dooars	AA	female	848.57	1727.38	1.72	-0.27	204.33	110.48
dooars	AA	male	741.81	1407.92	1.81	-0.24	149.19	122.50
dooars	AH	female	689.38	1617.54	0.81	-0.45	220.13	89.17
dooars	AH	male	574.13	1267.44	0.74	-0.50	143.97	103.75
dooars	EY	femal	465.75	2565.5	-0.45	1.18	224.9	123.00

		e		5			0	
dooars	EY	male	388.94	2083.97	-0.47	1.12	153.03	117.65
dooars	IY	female	381.32	2715.59	-0.92	1.42	234.36	102.73
dooars	IY	male	326.73	2241.55	-0.90	1.46	161.18	109.39
dooars	OW	female	480.57	1165.35	-0.36	-1.21	228.52	123.91
dooars	OW	male	395.29	981.26	-0.44	-1.10	155.89	121.14
dooars	UW	female	404.05	1538.82	-0.80	-0.51	228.14	109.09
dooars	UW	male	350.37	1145.80	-0.74	-0.76	162.03	113.14

Region	Vowel	Gender	Mean F1 (Hz)	Mean F2 (Hz)	mean.norm F1	mean.norm F2	Mean F0 (Hz)	Mean duration (in ms.)
nepal	AA	female	879.57	1632.95	1.86	-0.26	206.48	130.48
nepal	AA	male	718.27	1426.73	1.80	-0.33	123.42	129.23
nepal	AH	female	620.96	1476.43	0.55	-0.48	225.43	109.13
nepal	AH	male	549.69	1341.90	0.77	-0.51	128.07	102.41
nepal	EY	female	438.85	2640.40	-0.43	1.28	225.95	144.00
nepal	EY	male	365.21	2121.03	-0.49	1.21	131.48	120.00
nepal	IY	female	353.84	2711.74	-0.88	1.35	240.37	99.47
nepal	IY	male	305.57	2209.39	-0.91	1.38	137.79	117.14
nepal	OW	female	465.92	1039.92	-0.29	-1.17	227.42	134.17
nepal	OW	male	402.60	1014.08	-0.27	-1.21	131.20	134.00
nepal	UW	female	379.60	1374.55	-0.74	-0.64	242.55	131.50
nepal	UW	male	321.04	1300.84	-0.82	-0.55	136.44	106.40

Region	Vowel	Gender	Mean F1 (Hz)	Mean F2 (Hz)	mean.norm F1	mean.norm F2	Mean F0 (Hz)	Mean duration (in ms.)

sikkim	AA	female	876.52	1636.43	1.86	-0.34	221.62	132.86
sikkim	AA	male	747.59	1506.13	1.88	-0.16	136.66	118.13
sikkim	AH	female	625.86	1576.36	0.56	-0.44	230.32	88.18
sikkim	AH	male	534.74	1284.80	0.62	-0.61	143.17	86.86
sikkim	EY	female	452.81	2556.19	-0.33	1.21	226.90	134.29
sikkim	EY	male	350.42	2148.85	-0.49	1.19	148.85	112.73
sikkim	IY	female	330.05	2661.10	-1.02	1.38	240.81	101.90
sikkim	IY	male	281.48	2281.33	-0.92	1.44	151.76	96.06
sikkim	OW	female	461.90	1128.55	-0.28	-1.20	232.70	129.00
sikkim	OW	male	374.94	991.44	-0.36	-1.22	147.65	128.53
sikkim	UW	female	360.57	1500.95	-0.80	-0.57	239.90	120.95
sikkim	UW	male	316.86	1307.50	-0.72	-0.56	158.92	105.28

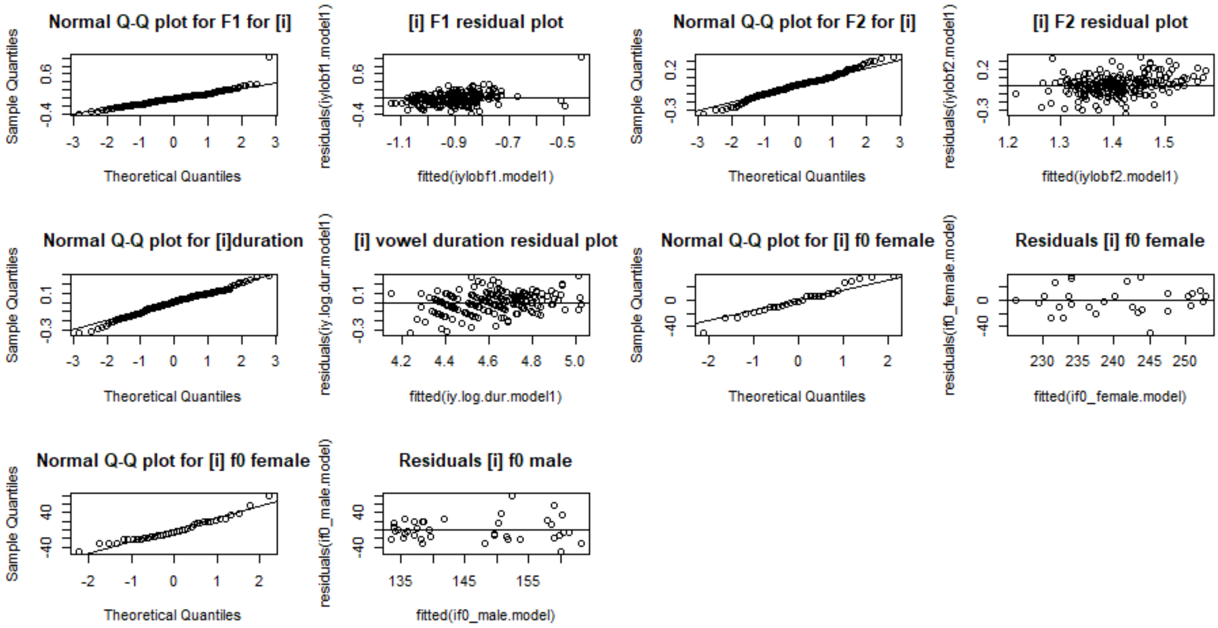
Appendix 4

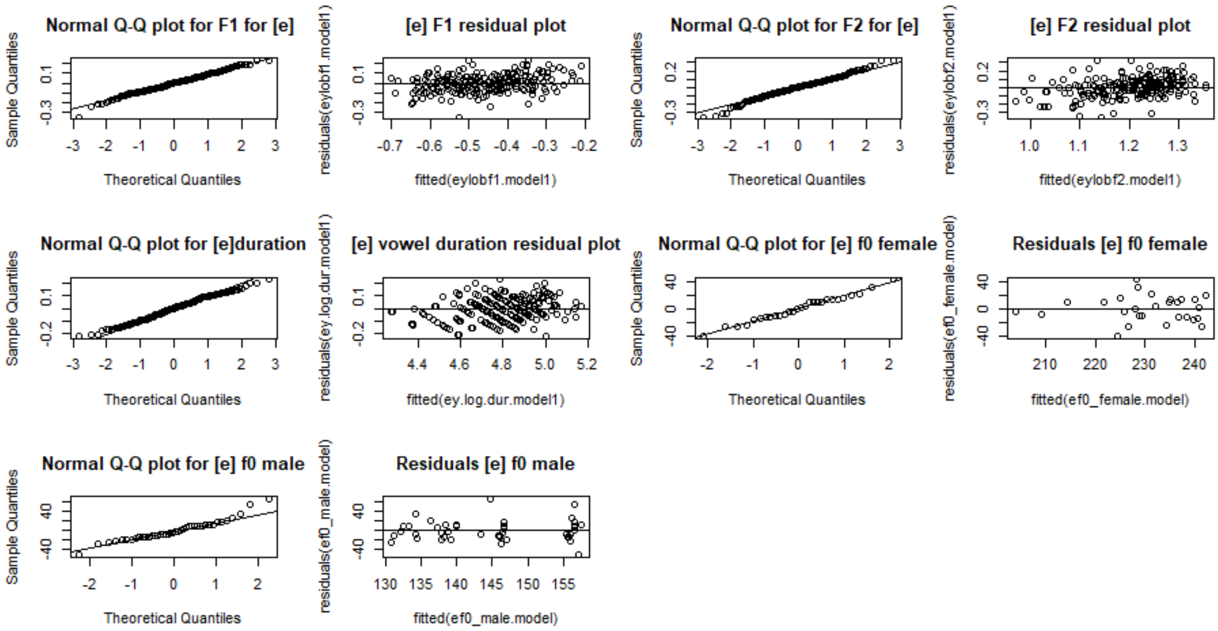
Mean formant values by region

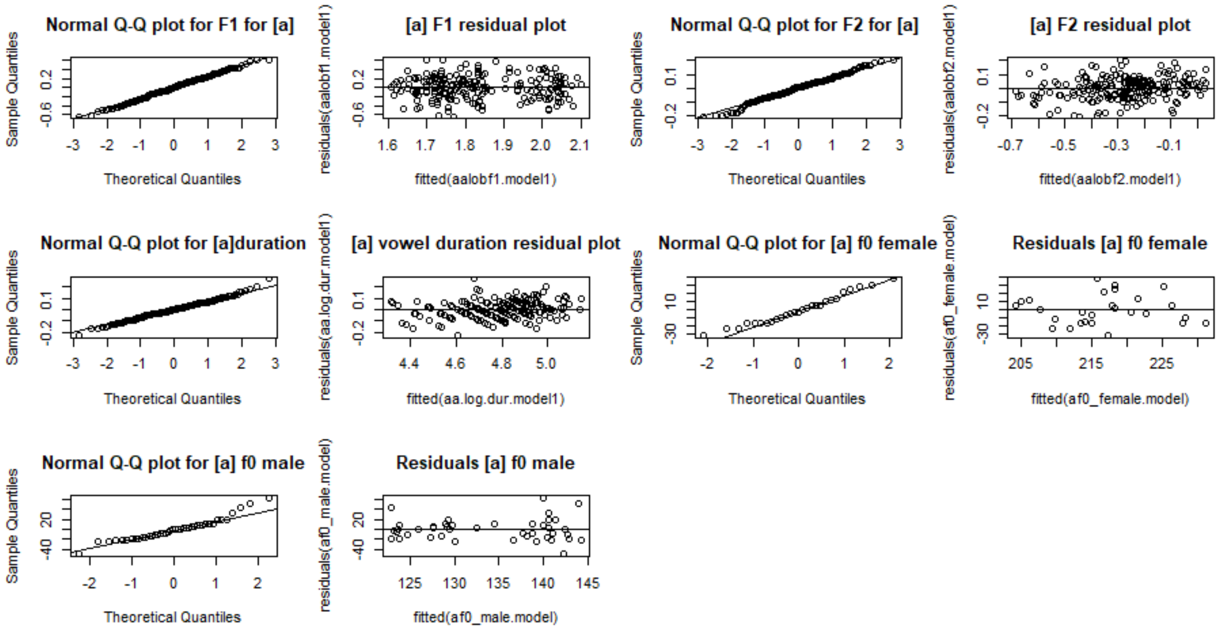
Region	Vowel	meanF1	meanF2	mean.normF1	mean.normF2	meanF0	meanduration
darjeeling	AA	784.81	1469.69	1.88	-0.33	158.59	118.89
darjeeling	AH	566.33	1362.64	0.58	-0.50	162.93	89.10
darjeeling	EY	389.65	2334.63	-0.49	1.23	174.83	124.88
darjeeling	IY	322.02	2416.98	-0.87	1.40	178.89	100.56
darjeeling	OW	413.63	993.96	-0.34	-1.18	172.09	127.46
darjeeling	UW	342.48	1297.80	-0.76	-0.61	175.73	110.48
dooars	AA	781.14	1525.61	1.78	-0.25	169.51	118.07
dooars	AH	623.52	1417.48	0.77	-0.48	176.61	97.50
dooars	EY	417.39	2262.33	-0.46	1.14	179.65	119.63
dooars	IY	348.56	2431.16	-0.91	1.44	190.45	106.73
dooars	OW	429.10	1054.26	-0.41	-1.15	184.69	122.24
dooars	UW	371.09	1297.49	-0.76	-0.67	187.54	111.58
nepal	AA	790.34	1518.87	1.83	-0.30	160.53	129.79
nepal	AH	581.21	1401.40	0.67	-0.50	171.13	105.38
nepal	EY	395.27	2333.02	-0.46	1.24	170.04	129.80
nepal	IY	325.09	2412.47	-0.90	1.37	179.26	110.00
nepal	OW	433.61	1026.73	-0.28	-1.19	178.33	134.08
nepal	UW	347.07	1333.60	-0.78	-0.59	183.60	117.56
sikkim	AA	798.68	1557.75	1.87	-0.23	170.32	123.96
sikkim	AH	569.91	1397.33	0.60	-0.55	176.81	87.37
sikkim	EY	390.24	2307.26	-0.43	1.20	179.20	121.11
sikkim	IY	300.37	2429.02	-0.96	1.42	186.39	98.33
sikkim	OW	407.15	1042.22	-0.33	-1.22	179.15	128.70
sikkim	UW	332.96	1378.77	-0.75	-0.56	188.75	111.05

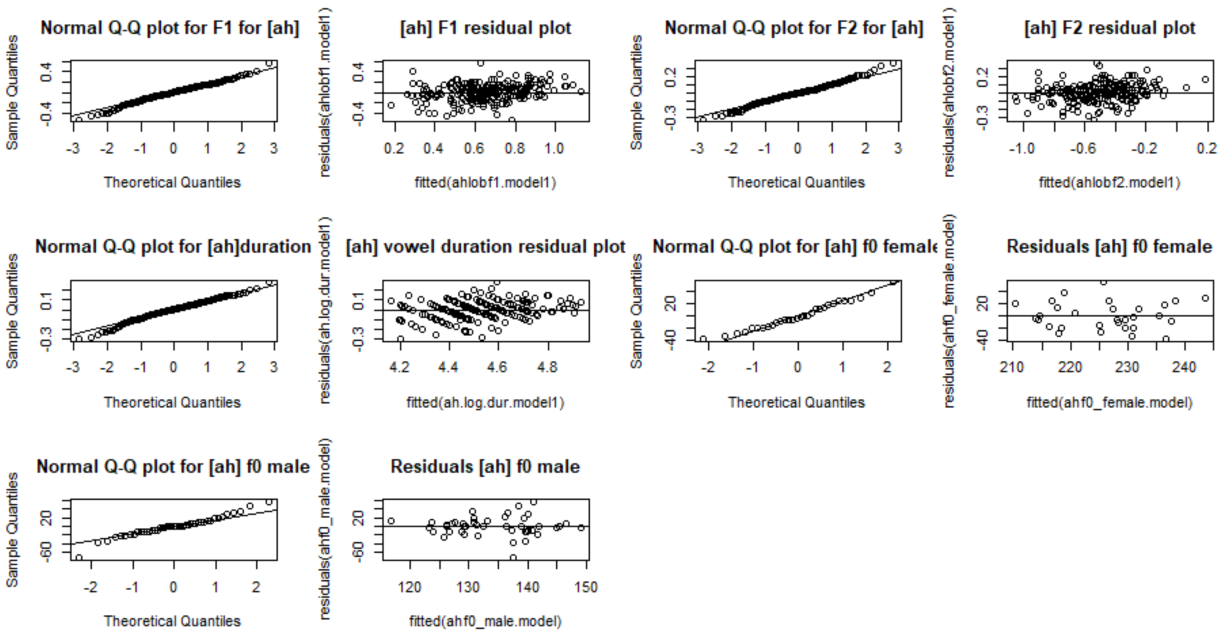
Appendix 5

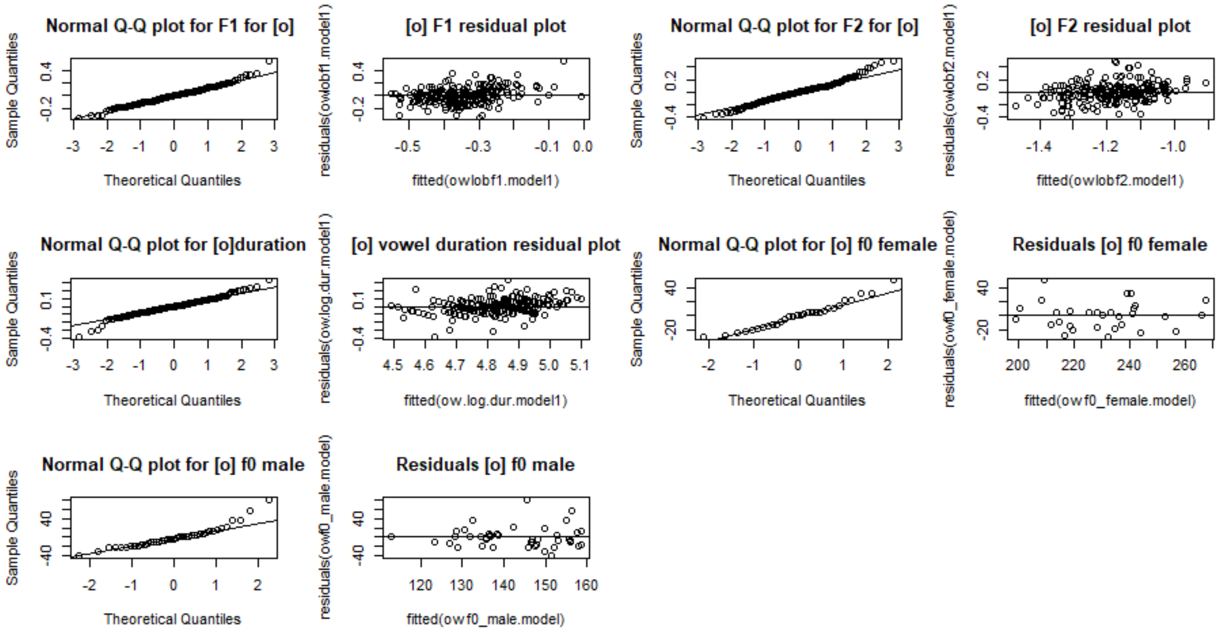
Residual and Normal Plots

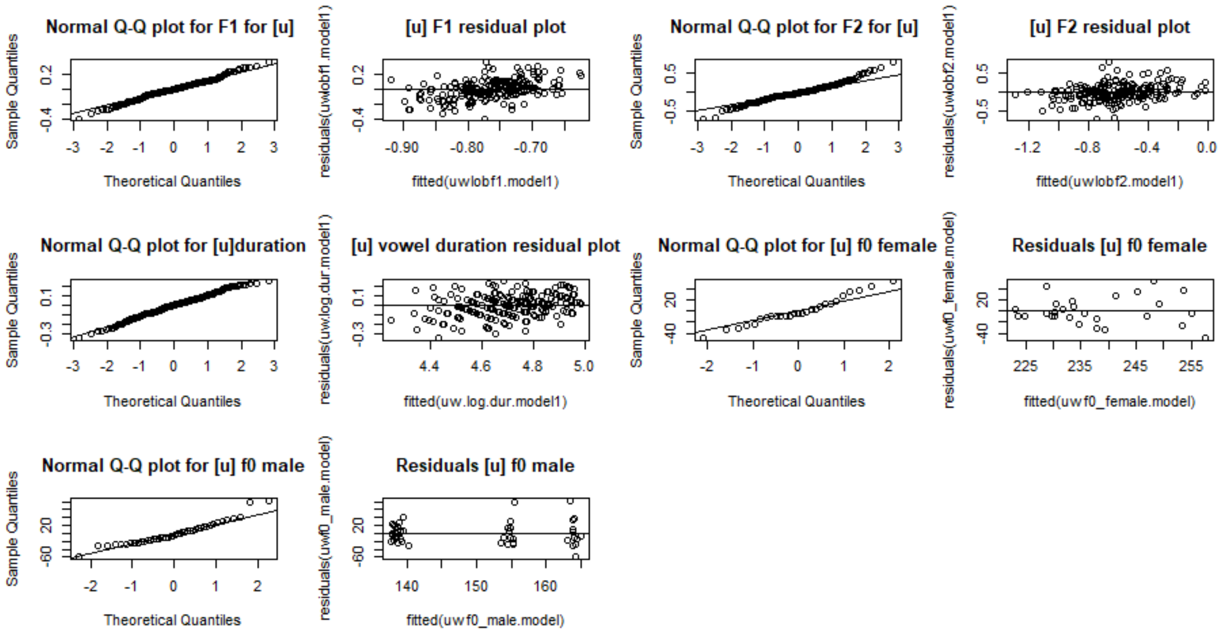





Normal Q-Q plot for F1 for [a]
[a] F1 residual plot
Normal Q-Q plot for F2 for [a]
[a] F2 residual plot
Normal Q-Q plot for [a]duration
[a] vowel duration residual plot
Normal Q-Q plot for [a] f0 female
Residuals [a] f0 female
Normal Q-Q plot for [a] f0 male
Residuals [a] f0 male
Sample Quantiles
Theoretical Quantiles
residuals(aalobf1.model1)
fitted(aalobf1.model1)
residuals(aalobf2.model1)
fitted(aalobf2.model1)
residuals(aa.log.dur.model1)
fitted(aa.log.dur.model1)
residuals(af0_female.model)
fitted(af0_female.model)
residuals(af0_male.model)
fitted(af0_male.model)


Normal Q-Q plot for F1 for [ah]
[ah] F1 residual plot
Normal Q-Q plot for F2 for [ah]
[ah] F2 residual plot
Normal Q-Q plot for [ah]duration
[ah] vowel duration residual plot
Normal Q-Q plot for [ah] f0 female
Residuals [ah] f0 female
Normal Q-Q plot for [ah] f0 male
Residuals [ah] f0 male
Sample Quantiles
Theoretical Quantiles
residuals(ahlobf1.model1)
fitted(ahlobf1.model1)
residuals(ahlobf2.model1)
fitted(ahlobf2.model1)
residuals(ah.log.dur.model1)
fitted(ah.log.dur.model1)
residuals(ahf0_female.model)
fitted(ahf0_female.model)
residuals(ahf0_male.model)
fitted(ahf0_male.model)


Normal Q-Q plot for F1 for [o]
Sample Quantiles
Theoretical Quantiles
[o] F1 residual plot
residuals(owlobf1.model1)
fitted(owlobf1.model1)
Normal Q-Q plot for F2 for [o]
Sample Quantiles
Theoretical Quantiles
[o] F2 residual plot
residuals(owlobf2.model1)
fitted(owlobf2.model1)
Normal Q-Q plot for [o]duration
Sample Quantiles
Theoretical Quantiles
[o] vowel duration residual plot
residuals(ow.log.dur.model1)
fitted(ow.log.dur.model1)
Normal Q-Q plot for [o] f0 female
Sample Quantiles
Theoretical Quantiles
Residuals [o] f0 female
residuals(owf0_female.model)
fitted(owf0_female.model)
Normal Q-Q plot for [o] f0 male
Sample Quantiles
Theoretical Quantiles
Residuals [o] f0 male
residuals(owf0_male.model)
fitted(owf0_male.model)


Normal Q-Q plot for F1 for [u]
Sample Quantiles
Theoretical Quantiles
[u] F1 residual plot
residuals(uwlobf1.model1)
fitted(uwlobf1.model1)
Normal Q-Q plot for F2 for [u]
Sample Quantiles
Theoretical Quantiles
[u] F2 residual plot
residuals(uwlobf2.model1)
fitted(uwlobf2.model1)
Normal Q-Q plot for [u]duration
Sample Quantiles
Theoretical Quantiles
[u] vowel duration residual plot
residuals(uw.log.dur.model1)
fitted(uw.log.dur.model1)
Normal Q-Q plot for [u] f0 female
Sample Quantiles
Theoretical Quantiles
Residuals [u] f0 female
residuals(uwf0_female.model)
fitted(uwf0_female.model)
Normal Q-Q plot for [u] f0 male
Sample Quantiles
Theoretical Quantiles
Residuals [u] f0 male
residuals(uwf0_male.model)
fitted(uwf0_male.model)

Appendix 6

ARPAbet phoneset in CMU Pronunciation dictionary

Phoneme	Example	Translation
AA	odd	AA D
AE	at	AE T
AH	hut	HH AH T
AO	ought	AO T
AW	cow	K AW
AY	hide	HH AY D
B	be	B IY
CH	cheese	CH IY Z
D	dee	D IY
DH	thee	DH IY
EH	Ed	EH D
ER	hurt	HH ER T
EY	ate	EY T
F	fee	F IY
G	green	G R IY N
HH	he	HH IY
IH	it	IH T
IY	eat	IY T
JH	gee	JH IY
K	key	K IY
L	lee	L IY
M	me	M IY
N	knee	N IY
NG	ping	P IH NG
OW	oat	OW T
OY	toy	T OY
P	pee	P IY
R	read	R IY D
S	sea	S IY
SH	she	SH IY
T	tea	T IY
TH	theta	TH EY T AH
UH	hood	HH UH D
UW	two	T UW
V	vee	V IY
W	we	W IY
Y	yield	Y IY L D
Z	zee	Z IY
ZH	seizure	S IY ZH ER

Source: http://www.speech.cs.cmu.edu/cgi-bin/cmudict