Vowel Acoustics of Assamese Spoken in Five Geographical Locations

A Dissertation

Submitted in Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy in Humanities and Social Sciences

Leena Dihingia

Department of Humanities and Social Sciences
Indian Institute of Technology Guwahati
Guwahati, Assam-781039

December 2020

Declaration

I declare that, the dissertation titled, "Vowel Acoustics of Assamese Spoken in Five Geographical Locations", submitted by me to the Indian Institute of Technology Guwahati, for the award of the degree of Doctor of Philosophy in Linguistics, is an original work carried out by me under the supervision of Dr. Priyankoo Sarmah. I have not submitted the dissertation in any form to another university or institute for the award of a diploma or degree.
All external sources used for the completion of this dissertation have been acknowledged and cited according to the rules and regulations given by the Indian Institute of Technology Guwahati.

Leena Dihingia
Department of Humanities and Social Sciences
Indian Institute of Technology Guwahati
Guwahati-781039, Assam, India

December 4, 2020

Certificate

This is to certify that the dissertation titled, "Vowel Acoustics of Assamese Spoken in Five Geographical Locations", submitted by Leena Dihingia (Registration Number: 136141006) is an authentic work carried out by her under my supervision.The dissertation fulfills all the requirements for the award of the degree of Doctor of Philosophy in Linguistics as prescribed by the Indian Institute of Technology Guwahati. The results of this thesis have not been submitted to any university or institute.

Dr. Priyankoo Sarmah
Department of Humanities and Social Sciences
Indian Institute of Technology Guwahati
Guwahati-781039, Assam, India

December 4, 2020

To,
Ma, Deuta, Bubu and Gautam.

Acknowledgements

This thesis would not have been possible without the help and support of several people during these seven years. I extend my sincere gratitude to all them.

I have been extremely fortunate to have worked under a supervisor who has not only guided my research work with his expertise in the field of phonetics, but has also provided continuous support and encouragement throughout my time as his student. Dr. Priyankoo Sarmah has been a great mentor right from the time I joined as a project fellow under him. Working with him has been the most enriching experience. His never ending energy and enthusiasm for quality research kept me on my toes, inspired and motivated me to complete this thesis. All drawbacks and gaps in this work are purely due to my own limitations.

The work presented in this thesis has been critically assessed by an outstanding Doctoral Committee to whom I extend my heartfelt gratitude. Prof. Arupjyoti Saikia's immense knowledge on the history of Assam helped me form the very backbone of this thesis. I thank him for being ever so encouraging throughout the course of this study. I thank Prof. Rohit Sinha for his constructive suggestions during the progress seminars and Dr. Bidisha Som for her continuous support and inputs.

My sincere gratitude also goes to Dr. Shakuntala Mahanta for teaching me the basics of phonology at the start of this program. I thank Prof. S.R.M. Prasanna for his invaluable suggestions and comments during the initial phase of this work.

Prof. K. Samudravijaya, this thesis would have remained incomplete without your expert inputs on language identification, an area that I still dread to venture into. I cannot thank you enough for the many insightful discussions and suggestions, making things a little less intimidating, and easier for me to grasp. Thank you sir, for your advice, motivation and for just being the cheerful person that you are.

A major portion of the data collected for this research is part of a project titled "A broad Sociolinguistic Study of Vowel Variation in Assamese", funded by the Indian Council of Social Science Research (ICSSR), India. I thank the funding agency for letting me use the data for this study. I also thank the Department of Electronics and

Information Technology (DeitY) India for supporting me in the initial years (2013-2016) of my PhD program during which I was a working as project fellow for one of their projects.

This thesis would not have seen the light of day without the help of my informants. I am indebted to each one of them for their co-operation and patience during the data collection procedure. I particularly thank Mr. Satradhar Choudhury and his family for hosting me in their home during my field work in Barpeta.

I have been lucky to have had a working space of my own at the Phonetics and Phonology lab at IITG, which has provided me with all the necessary lab facilities. The seven years spent working in this lab, with the most supportive lab mates, the countless discussions and chats with them over cups of tea and coffee, their constant encouragement and love have made my stay in the IITG campus nothing less than the most cherished part of this journey. Thank you Wendy, for being ever so supportive and nice to me and for constantly pushing me whenever I slow down. Your sincerity towards your work has been worth emulating. Bidisha, for pulling me through my tough days and never saying "no" to any help I ever needed; my go-to person for arranging participants for my tests, you've been a sister, a confidante and a loving friend. Viya, for always being so thoughtful and caring and making sure I do not go hungry during those stressful progress seminars. You three have been my rocks. Prarthana, for literally being by my side these seven years. Luke, for just being there. Malek, you surely made my quarantine days less stressful and kept me in my pink with your generosity and kindness. Dimple, for always believing in me and for being such a great company. Sahiinii for all those discussions and for your keen interest in my work. Pamir, you brighten up the lab whenever you visit us. Phunuma, you will always be missed. Asim da, for providing valuable suggestions and insights about the dialects of Western Assam. Abhas for always being there whenever I needed any technical help. Kanak da, Kalyan da, Amalesh bhaiya, Priti and Krishangi, thank you for being such great lab mates.

Nozomi ma'am, Ixan and Saki, thank you so much for all those wonderful meals and gifts and for being so caring. My sincere thanks also goes to the staff at the Dept. of HSS office. Mala, Parag, Durga da and Rubul, thank you for always helping me out with the departmental and official requirements. Chandana and Devipriya, thank you for your lovely company. Parichita, Sanjib Saikia sir, Saurav, Konku, Trusna, Saswati, Moa, Biswajit, Deepshikha, Sishir, Parismita, Shikha, Bornali ba, Siami, Karter, Emily, thank you all. The security guards at CC and Library deserve special mention for making us feel safe during those late working nights and for always taking

care of our PPLab. My canine buddies in the campus, Chomsky, Chuku, Stew and NB, it would have been a depressing life in the campus had it not been for you four cheering me up every day.

My friends outside the campus, Rebeka, Meenakshi for always checking up on me. Prof. Simon for constantly pushing me to finish my PhD. You have been my foremost well-wisher. Ibu for your care and for keeping me grounded. Junak, thank you for all the "yayy"s in even the tiniest of my happy stories. Pankaj da, for being such a wonderful friend and brother. Prof. Bibha Bharali and Dr. Banani Chakraborty for trusting me and providing me with an opportunity to share what I have learned, with the students of Dept. of Assamese, Gauhati University. Rehena, Ankur, Dhruba and my students at the department for being so understanding and patient. Samim, Jitu da, Bistirna, Chiranjit, Porag and Kalyan, thank you for always wishing me well.

I would not have made it this far without the relentless support of my family. My parents, Mrs. Anita Dihingia and Mr. Surendra Nath Dihingia for their unconditional love, care and encouragement at every stage of life. My younger brother Neelam, for being my greatest support and for always believing in me. I love you all so much. My in-laws, Mrs. Reema Gogoi and Dr. Rukheswar Gogoi for always being so supportive and helping in whatever way they could during this challenging period. Maina, Chayanika and dearest little Zico, thank you for your love. And finally, my husband, Gautam, who has been by my side like a strong pillar throughout this journey, living every single minute of it, cheering for every little success and lifting me up every time I stumble. He has longed to see me achieve this milestone more than anybody else.

Abstract

This dissertation studies the acoustic phonetic properties of vowels in Assamese as spoken in five geographical regions of Assam: Tinsukia, Jorhat, Nagaon, Nalbari and Barpeta. The first part of the dissertation reports results of vowel production experiments conducted on the speakers from the five regions. The results showed that vowel quality across the five varieties is not uniform. Compared to the eight-vowel system proposed for standard colloquial Assamese, the speakers of the five areas produced only six or seven distinct vowels. All varieties showed little evidence of a distinct /ʊ/ in production. Additionally, the Nalbari and the Barpeta varieties are seen to merge the /e/ and /ɛ/ vowels. In all cases the first two formants in vowel production were considered acoustic features. Vowel duration did not show any predictable pattern.

In order to confirm the results obtained from the production experiments, perception tests were also conducted on speakers from the five varieties in this study. In order to see the speakers' ability to distinguish the vowel pairs that merge in production, a series of discrimination and identification tests were conducted. The discrimination tests show low sensitivity of the speakers towards the vowel pairs that merged in production. The identification experiments showed that merged categories elicited more varied responses than distinct categories.

The merger of the vowels in the five varieties of Assamese was explored in detail in order to determine the extent and type of merger. Merger measures such as, Euclidean Distance, Pillai score, SOAM and VOACH were employed to investigate the mergers. The merged vowel pairs showed merger by approximation or expansion. These merger mechanisms differed according to the varieties.

Rhythm characteristics of Assamese read speech were also investigated in order to explore any possible correlation between vowel quality and rhythm characteristics. Attempts to classify the five varieties of Assamese using rhythm measures yielded moderate accuracy. Finally, automatic dialect identification was attempted using a Random Forest classifier with the first two formants as vowel features. The results

showed that vowel features yield high accuracy in dialect identification in Assamese. This confirms that the dialect specific vowel characteristics represented by the first two formants, as shown in the production and perception experiments, are sufficiently robust. Thus, this work confirms that the five varieties of Assamese namely, Tinsukia, Jorhat, Nagaon, Nalbari and Barpeta are distinct in terms of their vowel inventory and vowel quality.

Contents

List of Figures

List of Tables

Chapter 1

Assamese language and the current study

1.1 Introduction

Assamese has been described as having three broad dialectal divisions (G. C. Goswami, 1982; G. C. Goswami & Tamuli, 2003) with several sub varieties within these broad divisions (Kakati, 1941; G. C. Goswami & Tamuli, 2003). Although there have been several studies on Assamese and its varieties, these studies have mainly concentrated on the phonological, morphological and lexical aspects of the language (a discussion on these aspects is provided in Section 1.4). A detailed phonetic study of the sounds of the language has not been conducted yet. This thesis aims at providing a detailed inquiry into the acoustic features of Assamese as spoken in five geographically distinct regions of Assam viz. Tinsukia and Jorhat in Eastern Assam, Nagaon in Central Assam, Nalbari and Barpeta in Western Assam. This study analyzes the language and its spoken varieties using acoustic phonetic and statistical methods. The primary objective is to look into the nature of vowel variation in these varieties of Assamese. A phonetic study of Assamese varieties will throw light on the features of vowels in these varieties and will also enable a proper understanding of the nature of these variations. This work also uses merger measures to quantify the degree of overlaps between vowels and also to determine the type of vowel mergers, if any. Additionally,

a study of the Assamese native speakers' perception of the vowels in merger is also conducted. Further, the study attempts to explore the varieties of the language on individual temporal and rhythm properties.

Dialectal variation in languages have been represented by several phonetic features such as tone (Kristoffersen, 2007; Zhang, 2014), phonation (Pan et al., 2011), voice onset time (VOT) (Jacewicz, Fox, & Lyle, 2009; H. Choi, 2002). Similarly, studies on dialects of American English, British English and Portuguese, among others, have established the importance of variation of vowel features in these languages. A detailed discussion on this is provided in Chapter 2. As reported in previous studies, the difference in vowel sounds is one of the important determinants of dialectal variation in Assamese (G. C. Goswami, 1982; Deka, 2007). Hence, to throw light onto the features of vowels in these varieties and to enable a proper understanding of the nature of these variations, five regions across Assam were considered for the study. These regions represent the three broad dialectal divisions of Assamese as reported in G. C. Goswami (1982) and G. C. Goswami & Tamuli (2003). Sections 1.3 and 1.4 present a brief description of Assamese and an overview of the dialectal variations in the language. While Tinsukia and Jorhat represent the Eastern variety of Assamese, Nagaon represents Central Assamese and Nalbari and Barpeta represent Western Assamese. As discussed in Section 1.4, while these varieties are distinct in several phonological, morphological and lexical aspects, scholars have also reported the presence of several sub varieties within the broad divisions (Kakati, 1941; G. C. Goswami & Tamuli, 2003). Kakati (1941) notes that While Eastern Assamese variety may be regarded as uniform in nature with little to no variation in the language within the variety, Western Assamese displays heterogeneity, having several local dialects. G. C. Goswami & Tamuli (2003) however point out that Central Assamese stands out as an intermediate dialect between Eastern and Western Assamese varieties. Kakati (1941) also notes that the differences amongst the dialects of Kamrup (Western As-

sam) are mostly phonetic and rarely morphological or lexical. The differences between Eastern and Western Assamese however, widely range over phonology, morphology and vocabulary. The present research looks into the acoustic features of the vowels of the language with an attempt to discover the points of phonetic differences among the varieties.

Section 1.2 presents the organization of the chapters of this thesis. Section 1.3 of this chapter presents a detailed background of the Assamese language and the relevant studies done on the language. Section 1.4 briefly discusses dialectal variation in Assamese. Section 1.5 talks about the vowel inventory of Assamese. Section 1.6 puts forward the motivation for the study and the discusses the research questions. Finally, Section 1.7 briefly presents a general overview of the method employed for data collection and analysis during the research.

1.2 Organization of the dissertation

Although there have been several studies on Assamese and its varieties, these studies have mainly concentrated on the phonological, morphological and lexical aspects of the language. A phonetic description of the language is next to none. This study attempts to understand the nature of vowel variation in different varieties of Assamese from an acoustic phonetic perspective and provide a detailed inquiry into the acoustic features of five geographically distributed varieties of Assamese. Chapters 2 and 3 provide an insight into the ongoing vowel changes in the language and establish an asymmetrical pattern of vowel merging in Assamese varieties. Chapter 4 looks into the production-perception dichotomy existing in Assamese varieties. Chapter 5 demonstrates how Assamese is more akin to mora-timed languages and also provides the rhythm metrics for the language. The organization of the thesis is in the following manner:

- Chapter 1 includes a detailed background on the Assamese language and an overview of the current study. Various sections of this chapter discuss the history and development of the language, the trends of dialectal variation in Assamese, previous studies on the language and also briefly discuss the phonemic inventory of Assamese. This chapter also provides the motivation for the current study and gives a brief overview of the general methodology employed in the research.

- Chapter 2 discusses the acoustics of vowel variation and presents the acoustic features of vowels in the five varieties of Assamese. This chapter also gives the statistical analysis of the acoustic results.

- In Chapter 3 discusses the types of vowel mergers and their measurements are provided. Analysis of the Assamese vowels according to the proposed measures have been discussed.

- Chapter 4 discusses the perception of Assamese vowels by native speakers. Tests like discrimination and identification of front and back vowels have been employed to look for perceptual evidence of mergers.

- Chapter 5 discusses the rhythm and speech rate measures in the five varieties. This chapter looks into the probable correlation between vowel quality and rhythm characteristics.

- Chapter 6 provides a final discussion of the analysis carried out in the previous chapters and concludes the thesis. The chapter also makes a preliminary attempt at dialect classification using Random Forest (RF) classifiers to see if vowel features discussed in the previous chapters aid in automatic dialect identification in Assamese.

1.3 The Assamese language: history and development

The Assamese language belongs to the Indo-Aryan language family, however, its prolonged contact with several languages of the Tibeto-Burman (Bodo, Rabha, Karbi, Mising, Deuri etc) and Austro-Asiatic (Khasi) families has influenced the phonology, morphology and vocabulary of the language in many ways. Additionally, several features of the Kra-Tai languages, particularly of the Ahoms, also came to be incorporated in the Assamese language given their long rule in substantial areas of the Brahmaputra valley. It is therefore necessary to understand the origin, history and development of the language before commencing on further study on the language. This section provides a brief overview of the origin of the Assamese language, its history and the various stages of its development.

Assamese is the easternmost Indo-Aryan language of India, primarily spoken in the Brahmaputra valley of Assam. As per the Census Report of 2011, Assamese is spoken by 1,53,11,351 speakers[1] and is one of the 22 languages recognized and listed in the 8th Schedule of the Constitution of India. The word *Assamese* is the anglicized version of /ɔxɔmija/, derived from /ɔxɔm/, the native name of the state Assam. Apart from being a major language spoken in the northeast of India, Assamese also serves as a lingua franca among different speech communities residing in the region (G. C. Goswami & Tamuli, 2003). Scholars are divided regarding the origin of Assamese. Grierson (1903) and later Kakati (1941) were of the view that the language evolved from the Apabhraṁśa dialects from the eastern Magadhi Prakrit group of languages. Of the four Apabhraṁśa dialects, viz. Rāḍha, Vaṅga, Vārendra and Kāmarūpa, the Kāmarūpa Apabhraṁśa spread to Assam eventually emerging as

[1]Information accessed from the website of the Office of the Registrar General, Government of India (2011). Census of India 2011: Language, States and Union Territories. *Table C-16*; accessed on 22nd September, 2018 at http://censusindia.gov.in

Assamese (Chatterji, 1970). According to (Kakati, 1941) Assamese is a dialect that developed as a language together with Bengali from standard Magadhan Apabhraṁśa. U. Goswami (1991) points out that Apabhraṁśa came to be considered as standard only after the tenth century C.E. Hence, if Kakati (1941)'s view is to be accepted, it would mean that Assamese evolved after the tenth century C.E. However, looking at Xuanzang's observation, who visited the country in 643 C.E, that the language of *Ka-ma-lu-p'o* (Kāmarūpa) "differed a little from that of Mid India", and also noting that the copper plate inscriptions found in Assam from the sixth-seventh century C.E bore evidence of phonological, morphological and lexical peculiarities of the language of this region, U. Goswami (1991) opines that an earlier form of Assamese might have emerged as a separate language in the seventh century C.E and calls it Kāmarūpi Prakrit. Based on the several literary materials discovered, (Kakati, 1941) proposes three periods of the history of the Assamese language:

1. Early Assamese: This period begins from the fourteenth century till the end of the sixteenth century and may be divided into Pre-Vaishnavite and Vaishnavite sub-periods. The pre-Vaishnavite period is characterized by the writings of several court poets of the time like Hema Saraswati, Kaviratna Saraswatī and Mādhava Kandali. This period saw the use of an archaic form of the language (Kakati, 1941). Among many other features, the use of conjunctive participles, e.g. *hāni-ere* (does pierce); *kari-era* (do you do); *gucāi-erő* (I do remove) etc, has been found to be distinctive (Kakati, 1941). The Vaishnavite subperiod is marked by the literary works composed by the Vaishnavite reformer Śaṅkaradeva (1449-1567) to propagate his tenets. Kakati (1941) notes the absence of archaisms and conjunctive participles observed in the Pre-Vaishnavaite period and the extensive use of personal endings after participial suffixes during this period.

2. Middle Assamese: The period between the seventeenth to the beginning of the

nineteeth century is characterized by the prose chronicles of the Ahom court. With the decline of the Koch kingdom in western Assam and the consolidation of the Ahom kingdom in eastern Assam, the Assamese of this region came to be used in the standardized literary prose like the diplomatic writings and administrative records. Most importantly, the historical chronicles of the Ahoms called *Burañjīs* began to be written. The first such Buranji was written on the instructions of the first Ahom king Sukaphaa who established the kingdom in 1228. *Burañjīs* were at first written in Tai Ahom, the original language of the Ahoms, but after the adoption of Assamese as the court language around the 16th century, these began to be written in Assamese (Kakati, 1941). The language used during this period gave way to the literary language of the modern period. Kakati (1941) points out that several new features were observed in the language of this period:

a. The use of plural suffixes of nouns like *-bor* and *hāt* were observed for the first time.

b. Conjunct consonants were reduced to single ones, for example, *hrax* became *rax*

c. Conjunctive participles were used as pleonastic suffixes to add emphasis to inflected verbal forms, for example, *dharile gai* (caught him up); *rahil gai* (he stayed there) etc.

d. The transfer of plural suffixes from nouns to verbs, for example, *ami zaõ* (we go) became *ami zaõhɔ̃̃k* (we do go).

3. Modern Assamese: This is the period from the beginning of the nineteenth century till present times. The publication of the Bible in Assamese from Sreerampore Mission Press in Bengal in 1813 AD by a Christian missionary, William Carey, heralded the beginning of the modern Assamese period. Two mission-

aries by the names of Reverend Nathan Brown and Oliver T. Cutter reached Gauhati on January 18, 1836 and Sadiya on March 23, 1836 with the intent of introducing Christianity among the Shans in North-East India. Brown realized that the language of the tribes at Sadiya was different from those of the Burmese and other Shan countries. He discovered Assamese and began to learn the language. However, the missionaries were compelled to leave Sadiya during the Khamti insurrection of 1839. They decided to make Sibsagar the next station for their activities and established the Sibsagar Mission Press in 1843. In the publications that came out of the press, the dialect spoken around Sibsagar was used and several native writers were encouraged to bring out books and periodicals in the language of eastern Assam. Several landmarks were achieved during this period. Firstly, the first Assamese periodical "Arunodoi" started publishing in 1846. Secondly, Nathan Brown published a treatise titled, *Grammatical Notices on Assamese language* in 1848. Following that the first Assamese-English Dictionary compiled by M. Bronson was printed in 1867.

Assam is also home to a number of indigenous languages belonging particularly to the Sino-Tibetan language family, which have influenced the Assamese language to a large extent. Hakacham (2000) maintains that the original inhabitants of this region were speakers of languages belonging to the Mon-Khmer family followed by speakers of Sino-Tibetan languages. Sircar (1990) also notes that Greek travel accounts from the first and second century C.E, "appear to call the land including Assam Kirrhadia after its Kirata population" who were Mongoloid in appearance and spoke a Tibeto-Burmese language. G. C. Goswami & Tamuli (2003) note that during the reign of the Brahmin king Kumara Bhaskara Varma the Aryan language spread among the aboriginal people, who gave up their undeveloped and unwritten language in favour of a much more developed Aryan tongue. Kakati (1941) and U. Goswami (1991) cite several lexical and grammatical features that Assamese language has acquired from

the indigenous Austroasiatic and Sino-Tibetan languages.

Systematic study of the Assamese language started during the colonial rule. The change of guard from the Ahoms to the British East India Company, effected by the Treaty of Yandaboo in 1826, began a new chapter in the history of Assam. The functionaries of the colonial government realized that for an effective administration, they needed a working knowledge of the local language. Initially the British had introduced the Bengali language for administrative conveniences and Assamese was removed from the schools and offices of Assam. According to G. C. Goswami (1982, p.9) "cunning native officials from Bengal" conspired and explained away Assamese as a local variant of the Bengali language. Goswami calls it a "black-hole tragedy for the Assamese language". It was only after vehement opposition from Assamese nationalists like Anandaram Dhekial Phukan, Hemchandra Barua and Gunabhiram Barua among others, that Assamese was reinstated as the official language in 1873. Meanwhile, William Robinson, a British official, wrote *A Grammar of Assamese Language*, published from Serampore Press in 1839. Written in English, it was designed as a manual for the British officials working in Assam. In 1848, a Baptist Missionary by the name Reverend Nathan Brown published *Grammatical Notices on the Assamese Language* and was probably the first grammarian to dismiss the idea that Assamese was a dialect of Bengali. Subsequently, Hemchandra Barua realized the need for a grammar of Assamese in the Assamese language and became the first native scholar from the region to write a grammar of the Assamese language in Assamese. In 1859 he published *Asamiya Bhasar Byakaran*, for the general people. After the reinstatement of Assamese as the official language in 1873, the need for a grammar of the Assamese language became all the more evident particularly in schools, offices and courts. Apart from the 2nd edition of Hemchandra Barua's *Asamiya Bhasar Byakaran* published in 1873, and his *Asamiya Lorar Byakaran*, for school children in 1882, several other writers began to write grammars of the Assamese language. In 1894 G.F. Nicholl,

an Oxford University professor authored a grammar of Bengali titled *Manual of the Bengali Language*. In it he added some 40 odd pages on Assamese grammar as *Appendix V- Sketch of Assamese Grammar*. His attempt was however full of pitfalls. An example from his grammar is the uninformed use of the plural suffix *hot* after words like *para* (pigeon), *hah* (duck) etc. Father O Paviotti in the year 1949 brought out *An Assamese Grammar with vocabulary and exercises*. He made an attempt to explain the probable origin of Assamese language and also mentioned that Assamese had elements of the tribal and sub-tribal languages of the valley.

Debananda Bharali's book *Asamiya Bhasar Moulik Bichar* (1912) threw ample light on the origin, growth and structure of Assamese. He debated that Assamese is not an indigenous growth but a variety developed after blending various speeches of Indo-European origin from different parts of northern India at different periods. Herein, the report of the Linguistic Survey of India, published in 1927 and edited by George Abraham Grierson must be mentioned. The report made analytical study of several languages of India. Volume 5 (pp 393-425) of the report contained a detailed analysis of the Assamese language. Kaliram Medhi's *Asamiya Byakaran aru Bhasatatya* (1936), *Assamese grammar and origin of the Assamese language* (1988) are other works in linguistic research which give a historical analysis of Assamese. According to Medhi, like Hindustani, Assamese is a form of Vedic or pre-Vedic Sanskrit which has greatly assimilated various other elements from the Tibeto-Burman family. In 1941 Banikanta Kakati published his doctoral thesis titled *Assamese: Its formation and development* (AFD). The thesis guided by Dr. Suniti Kumar Chatterji, was inspired by Dr. Chatterji's thesis *The origin and development of Bengali language* and provided a systematic linguistic study of the Assamese language. Kakati took a scientific approach in his description of Assamese phonology and morphology and was able to establish the individuality of the language with adequate illustrations thereby refuting the claim of previous writers that Assamese is an offshoot of Bengali.

After Kakati, mention must be made of his three students who carried forward the study of Assamese language within their own areas of interest. Golokchandra Goswami was interested in phonemics, phonology and grammar. His *Asamiya Golokbyakaran* published in 1972 was a grammar of Assamese written in light of linguistic principles. *An introduction to Assamese phonology* (G. C. Goswami, 1966) and *Structure of Assamese* (G. C. Goswami, 1982) gave a detailed description of the sounds and structure of the language. Dr. Upendra Nath Goswami, another student of Dr. Kakati, shows interest in dialectal studies, with reference to Kamrupi. His work titled *A study of Kamrupi: A dialect of Assamese* (U. Goswami, 1970a) is an important work on dialectal study, and the first of its kind in Assam. The work threw light on the history and development of Assamese and talked about how Kamrupi retained certain important phonological and morphological features of the Magadhan languages, which have gradually become lost in the standard language.

This section looked into the origin and development of the Assamese language and some of the important works done on the language. It is evident that although there are different views regarding the origin of the language, scholars have agreed upon the stages of the development of the language. A look into the history of the Assamese language makes it clear that the development of the language has been largely affected by the political and cultural history of the state. From the Kamrupi dialect being used as the language of literature in the early Assamese period to Eastern Assamese being developed as the standard form due to the patronage of the Ahom kings and the subsequent establishment of the printing press in Sibsagar, to Assamese losing its identity in the mid 1800's and finally being reinstated as the state language in the 1870s, Assamese language has gone through considerable ups and downs. It is also evident that the first works on the language were done by non-natives, primarily to further the administrative and missionary requirements. It was only when the native Assamese realized that these works contained several misrepresentations regarding

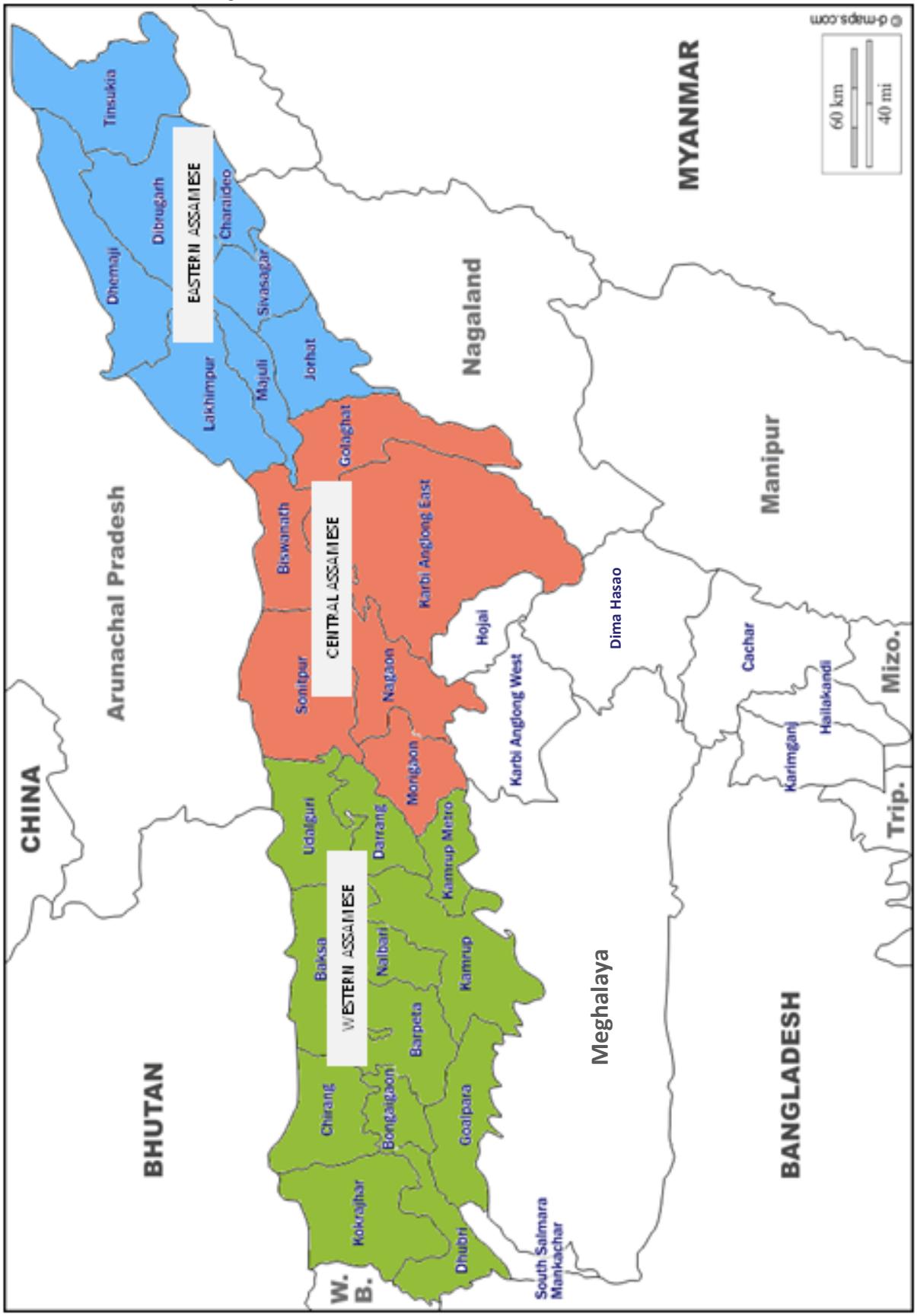
the origin, distinctiveness of Assamese as well as several mistakes, did they bring out grammars in the Assamese language. Later Banikanta Kakati's doctoral thesis came to be regarded as one of the most important and widely studied works on the language which subsequently generated several works on the Assamese language, particularly concentrating on the phonology, morphology and vocabulary of the language and its dialects.

This study focuses on the dialectal variation of Assamese vowel phonemes. Hence, in the following Section 1.4, the distinct phonological features of Assamese dialects are briefly described and in later chapters these dialects are analyzed on the basis of acoustic and statistical parameters.

1.4 Dialectal variation in Assamese

As mentioned in Section 1.3, over time, several changes have occurred in the language and today the Assamese spoken is very different from how it was spoken in the early periods. Here it is also important to note that over the ages several dialects of the language have also evolved, based on the geographical location and the social construct. In this section, the main divisions of the language and the salient phonological features of the varieties are discussed briefly.

Figure 1.1: Broad division of Assamese varieties



As mentioned earlier, the Assamese language is spoken in the Brahmaputra valley of Assam. This valley stretches over an area of about 720 kilometres from east to west. Apart from the Assamese people, many tribes speaking several indigenous languages have been residing in this valley for centuries. The Assamese spoken in this valley has varied enormously giving rise to several dialects and sub dialects. Studies on the Assamese language have maintained that there are noticeable variations in the language. Grierson (1903) and Kakati (1941) both mention that Assamese has two distinct varieties – the Eastern Assamese (EA) and Western Assamese (WA). G. C. Goswami & Tamuli (2003), on the other hand, divide Assamese into three distinct varieties namely, Eastern, Western, and Central varieties. The account lists the phonological, morphological and syntactic features of Assamese in detail and also gives a brief account of the origin and history of the language, dialects, dialectal divergences, and the writing system. Grierson (1903) mentions the form of speech spoken in and around Sibsagar came to be established as the literary language of the entire province and eventually as the standard dialect of Assamese after the establishment of the printing press in 1843. Kakati (1941) adds that this shift of literary importance from Western to Eastern Assam was also a result of the consolidation of Ahom power in that region and the decline of the Koch kingdom in the west. Under the Ahom patronage, as mentioned in Section 1.3, the Assamese of this region came to be used in the standardized literary prose like the diplomatic writings, administrative records and the historical chronicles called *Burañjīs*. This dialect, as Grierson points out, does not change across the upper part of the Assam valley, however, towards the west a distinct variety which he calls Western Assamese is spoken in Kamrup and Eastern Goalpara districts. Grierson however does not mention in what aspects the two varieties he identified are different.

Kakati (1941) also reports two distinct varieties - the Eastern Assamese (EA) and Western Assamese (WA). According to Kakati, while the Eastern variety, spoken

from the easternmost frontier Sadiya down to Guwahati, exhibits homogeneity, with least or no variation, and the Western variety, spoken in the two western districts of Kamrup and Goalpara, comprises many local dialects each showcasing distinctiveness from the others. Kakati points out that a steady commanding central influence gives homogeneity to manners as to speech. Since western Assam was never for a long period of time under any dominant power it lacked homogeneity in speech, thereby allowing a good deal of local variations to crop up in the language of western Assam. Adding to Kakati's comment on the large number of local dialects in Western Assam, U. Goswami (1988) mentions the river Brahmaputra and many of its large tributaries running through the area have acted as a natural barrier among the people. He mentions that in some places forests formed a dividing line. Additionally, the means of communication among these people divided by rivers and forests were also meagre. On the other hand, the Ahom court had a levelling influence in Eastern Assam. Hence, the language of this region shows very few dialectal variations (Kakati, 1941). Although Kakati has referred to dialectal variations in his work, he does not venture much into mentioning more about these local dialects except adding an appendix of dialect specimens, one standard colloquial (eastern Assam) and six western Assam dialects. He points out that the dialects of Kamrup district differ mostly in phonetic terms and hardly morphologically and lexically. However, Eastern and Western Assamese differ in a wide range of phonology, morphology and vocabulary. The most salient points of phonological difference are (Kakati, 1941):

1. Word stress in Kamrupi dialect is initial as opposed to penultimate in the standard colloquial (henceforth, 'St. Coll').

2. As a result, medial vowels are rarely pronounced and mostly slurred over.

3. Kamrupi abounds in epenthetic vowels as opposed to Eastern Assamese. For example, epenthesis comes in whenever disyllabic words are lengthened by affixes.

ghat, a ferry, but *ghaute*, a ferry-man (*ghatuwai*).

4. Absence of diphthongal vowels in word final syllables in Kamrupi. For example, gɒle for gɒlɔi, *galai*, prow of a boat; kabo for kabɔu, *kabau*, supplication.

5. Diphthongs and triphthongs are heard in initial syllables of Kamrupi words. For example, *haula* for *haluwa*, a draught ox; *keuila* for *kewaliya*, a hermit.

6. Predominance of high-vowels as opposed to mid vowels in St. Coll. For example, *kapur* (cloth), *tule* (raises),*indur* (rat), *nimu* (lemon) for St. Coll. *kapor*, *tole*, *endur*, *nemu* respectively.

7. Western Assamese favours aspiration whereas Eastern Assamese favours de-aspiration in the same phonological contexts. For example, Kamrupi shows aspirations after the gluttaral spirant x, the sibilant s, and in the neighbourhood of another aspirate in the same word as against de-aspiration in St. Coll. xit^{h}a (dregs), sɔk^{h}u (eye),$b^{h}uk^{h}$ (hunger) as against St. Coll xita, sɔku,$b^{h}uk$, etc.

While G. C. Goswami & Tamuli (2003) agree on the heterogeneity of Western Assamese dialect, they regroup Eastern Assamese into Eastern and Central dialects based on a close scrutiny of the phonology and morphology. – Eastern Asamiya (EA): Sivasagar and Lakhimpur, some areas of Arunachal in the east, Sonitpur and Nowgong in the west. – Western Asamiya (WA): little east of Guwahati in the south and Darrang district in the north to Goalpara in the west. – Central Asamiya (CA): area between the two regions.

The distinguishing phonological features of the dialects mentioned in G. C. Goswami (1982, p.12-15)are:

1. Eastern and Central Assamese have medial stress, whereas Western Assamese has initial stress. The strong initial stress of Western Assamese generally shortens words. In the other two dialects the nucleus of the syllable immediately

following the stressed one is attenuated in timbre. For example, Eastern *komóra* 'pumpkin'; Central *kumúra*, Western *kúmra*.

2. There appears to be a predominance of high-vowels in Western (and often in Central) Assamese as opposed to higher-mid and low vowels to the lower-mid vowels of Eastern dialects; For example, Eastern *kapór*, *nemú*; Central *kapór*/*kapúr*, *nemú*; Western *kápur nímu*.

   Eastern and Central Assamese have lower-mid front /ɛ/ in place of the higher-low front /æ/ in Western Assamese. Also, there is no phonemic contrast between /ɒ/ and /ɔ/ in Western Assamese, making it a seven-vowel system.

3. Diphthongs in the Eastern and Central dialects are generally falling, whereas Western dialects have both falling and rising ones. For example, Eastern *láthua* 'wicked', *kátia* 'born in the month of *Kati*'; Central *láthua*, *kátia*; Western *láutha*, *káita*.

   The final diphthong /ɔu/ is common and prominently articulated in Eastern and Central Assamese. /ɔi/ has a tendency to become /ɔ/ medially and /e/ or/ej/ finally in Eastern Assamese. In western Assamese, the final /ɔu/ and /ɔi/ either become monophthongs /o/ and /e/ or change the quality and become /ow/ and /ej/ respectively.

4. Triphthongization in common in Western Assamese but absent in Central and Eastern Assamese.

Kakati (1941, p.19) observes that while Eastern Assamese is more or less homogeneous in its linguistic features, Western Assamese is not. U. Goswami (1970b) divides Western Assamese into three further sub varieties, namely, Barpeta, Nalbari and Palashbari-Chaigaon sub-varieties. Based on this observation G. C. Goswami (1982) divides Assamese into three distinct varieties namely, Eastern, Western, and

Central varieties. Agreeing on the divisions of Assamese varieties, Moral (1991) adds that Central Assamese includes the dialect spoken primarily in Nagaon and Morigaon districts and some parts of the adjoining Sonitpur, Darrang, Jorhat and Golaghat districts. Further, Das (2010) claims that both the EA and WA varieties can be divided into two subvarieties each, namely, Sibsagar variety & Central Assam variety and Kamrup variety & Goalpara variety. Apart from that, they also suggest that the Assamese spoken in Guwahati area is a mixture of Eastern and Western Assamese varieties and can also be considered a distinct variety.

To summarize, Assamese can be divided into three broad dialect groups. These are: Eastern Assamese, Central Assamese and Western Assamese. A deeper look into these broad varieties suggest that there are several sub varieties, particularly within Western Assamese, which has been mentioned by Kakati (1941) as a non-homogeneous dialect group. These sub-varieties again have their own peculiarities, phonological variation being one of them. In the following chapters, therefore, an attempt has been made to study these variations based on phonetic analysis of Assamese vowels as spoken in five geographical regions.

1.5 Vowel inventory of Assamese

This section presents an overview of the phonemic inventory in St. Coll. Assamese. However, since the goal of this dissertation is to study the vowels of Assamese, more importance has been given to the discussion of the vowel phonemes. The section first discusses the Assamese vowel inventory as envisaged by different scholars and then provides a brief overview of the consonant phonemes in the language.

1.5.1 Vowels

There have been conflicting views regarding the number and types of vowels in Standard Assamese, particularly the back vowels. Kakati (1941) provided a balanced vowel system for Assamese, as presented in Table 1.1: three front vowels /i, e, ɛ/, three back vowels /u, o, ɔ/ and one low central /a/ vowel.

Table 1.1: Vowels in Assamese (Kakati, 1941)

	Front	Central	Back
Close	i		u
Half-close	e		o
Half-open	ɛ		ɔ
Open		a	

This has been contested in Taid (1988), who points out that there are not three, but four back vowels in the standard colloquial Assamese, viz /ɒ, ɔ, o, u/, the phonemicity of which is evident in the following minimal set in Table 1.2. He, however, also mentions in a footnote that the symbols used for these phonemes do not refer to their exact values in the system of cardinal vowels and have been chosen merely to represent the contrasts.

Table 1.2: Phonemic contrast of back vowels (Taid, 1988)

Assamese word (IPA)	English meaning
/kɒla/	deaf
/kɔla/	black
/kola/	lap
/kula/	winnowing fan

Taid (1988) clarifies that in some dialectal varieties, such as Kamrupi, such a

balanced vowel system does exist, with the difference that instead of three back vowels /ɔ o u/, we come across /ɒ ɔ u/, /o/ being neutralized with either /u/ as in the speech of upper Assam or with /ɔ/ as in the Kamrupi dialect. This was further investigated by Deka (2007) who mentions that compared to the standard variety of Assamese with eight vowels, the Kamrupi variety has only seven vowels, i.e. the vowel /o/ is absent in the variety [2].

Table 1.3: Neutralization of /o/ in Assamese varieties (Taid, 1988)

St. Coll. Assamese	Upper Assam dialect	Kamrupi	English meaning
/bhok/	/bhuk/	/bhɔkh/	hunger
/dokan/	/dukan/	/dɔkan/	shop
/kola/	/kula/	/kɔla/	lap
/gotei/	/gutei/	/gɔtei/	the whole of
/mor/	/mur/	/mɔr/	mine

G. C. Goswami & Tamuli (2003) seem to have noticed this discrepancy and included four back vowels in their description of Assamese vowels, as presented in Tables 1.4 and 1.5.

Table 1.4: Vowels in Assamese (G. C. Goswami & Tamuli, 2003)

	Front	Central	Back
High	i		u
Higher-mid	e		o
Lower-mid	ɛ		ɔ
Low		a	ɒ

[2]Even though Deka (2007) makes this observation, he does not provide any data to support his claim.

Table 1.5: Vowel contrasts in Assamese (G. C. Goswami & Tamuli, 2003).

Vowel	Assamese word	English meaning
/i/	/bil/	small lake
/e/	/bel/	bell (Eng)
/ɛ/	/bɛl/	wood apple
/a/	/bal/	pubic hair
/ɒ/	/bɒl/	strength
/ɔ/	/bɔl/	let's go!
/o/	/bol/	colour
/u/	/bul/	walk

In Mahanta (2008, 2012)'s descriptions of Assamese vowels, we see the inclusion of the vowel /ʊ/ in the inventory. Mahanta (2008) conducts an acoustic study of the Assamese vowels in order to confirm the distinctions among the eight-vowel set in the language. Mahanta (2008, 2012) shows that in terms of the height of the back vowels, /ɔ/ is lower than /o/ and /ʊ/ is a little lower than /u/. About the quality of [ɑ], while in Mahanta (2008) she presents it as a low back vowel, in her later work she mentions that it may vary depending on the context. Nevertheless, it is mostly pronounced in the mid position (Mahanta, 2012). Ladefoged & Maddieson (1990, 1996) mention that this vowel sounds like the British English /ɑ/ as in the word 'father' which is a low back vowel cardinal vowel. However, Kakati (1941); G. C. Goswami (1982); U. Goswami (1991) have called it a low central vowel. Further, looking at the formant frequencies of the Assamese vowels presented in Mahanta (2008), this vowel seems to be a low central vowel rather than a low back vowel. In our analysis in Chapter 2, where we plot the vowels in an x-y plane for the first two formant frequencies, we see that for all the regions this vowel is consistently placed in the low central position. Another vowel that has been the subject of discussion is the vowel /ʊ/.

Ladefoged (2003, p.115) mentions the presence of a back vowel in Assamese that has the tongue position of /ɔ/ but maximum lip rounding as in /u/ and transcribes it as /ʒ/. Mahanta (2012) however represents this vowel as /ʊ/, that changed to /u/ as a result of vowel harmony. Table 1.6 presents the vowel inventory as given by Mahanta (2012).

Table 1.6: Vowels in Assamese (Mahanta, 2012)

	Front	Central	Back
High	i		u
Near-high		ʊ	
Higher-mid	e		o
Lower-mid	ɛ		ɔ
Low		ɑ	

Regarding the front vowels in Assamese, Kakati (1941, p.64) describes /ɛ/ as the "usual sound value of the e-phoneme". Taid (1988) speculates that probably Kakati was unsure of the status of /ɛ/ since, despite the presence of minimal lexical contrasts such as /k^{h}ed/ 'regret', /k^{h}ɛd/ 'drive away'; /bes/ 'fine', /bɛs/ 'sell' etc. that establish /ɛ/ as a distinct phoneme, Kakati seems to suggest that /ɛ/ is only an allophone of /e/.

While there have been differing views regarding the status of Assamese vowels, scholars have agreed upon the phenomenon of vowel harmony as a distinguishing feature of the language, where the vowels /i/ and /u/ trigger a change in the quality of the preceding /ɛ, ɔ, ʊ/ vowels to /e, o, u/, as seen in Table 1.7 from Mahanta (2012):

Table 1.7: Examples of Vowel harmony as in (Mahanta, 2012)

Root	Gloss	Suffix	Derivation	Gloss
/bʰɛkʊlɑ/	'frog'	i	/bʰekuli/	'frog' (dim)
/ʊpɔɹ/	'above'	i	/upoɹi/	'in addition'
/kʰɔɹɔs/	'spend'	i	/kʰoɹosi/	'spend-thrift'
/pɔlɔx/	'silt'	uwɑ	/poloxuwɑ/	'fertile land'
/mɛɹ/	'curl'	uwɑ	/meɹuwɑ/	'curled'
/gʊbɔɹ/	'dung'	uwɑ	/guboɹuwɑ/	'fly living in dung'

While typically vowels are phonetically characterized in terms of the regular dimensions of height and backness, Ladefoged & Maddieson (1996) brings in tongue root features in their discussion Akan and Igbo Tense/Lax vowels in comparison to Germanic languages. Their analysis had shown that in Igbo and Akan the tongue height is not correlated with tongue root position, as opposed to English and German where tongue height and tongue root have shown to correlate, particularly for the back vowels. Mahanta (2008) subsequently, characterizes vowel harmony in Assamese as tongue root harmony and discusses the participation of [-ATR] /ɛ, ɔ/ and /ʊ/ vowels in this phenomenon whereby the [-ATR] change to [+ATR] /e/, /o/ and /u/ in the presence of the two [+ATR] vowels /i/ and /u/ following them, as seen in Table 1.8. Hence, the eight Assamese vowels contrast in terms of the ATR feature, with /i, u, e, o/ being [+ATR and /ɛ, ɔ, ʊ/ being [-ATR].

Table 1.8: [+ATR] and [-ATR] distinctions in Assamese (Mahanta, 2008)

	Front	Back	
High	i	u	[+ATR]
		ʊ	[-ATR]
Mid	e	o	[+ATR]
	ɛ	ɔ	[-ATR]
Low		ɑ	[-ATR]

Archangeli & Yip (2020) uses ultrasonic imaging to explore the physical properties of Assamese tongue root harmony patterns and establishes RTR and ATR vowel distinctions in Assamese vowel harmony between [e, o] and [ɛ, ɔ]. Articulatory analysis showed that not all speakers maintained the distinction. Although the dialects of the speakers is not mentioned, two speakers lost the tongue root contrast in mid vowels. These results hint at a probable vowel overlap among these speakers as is one of the research objectives of this study.

1.5.2 Consonants

There are twenty three consonants in Assamese: twelve stops and eleven continuants as illustrated in Table 1.9. It is to be noted that both Kakati (1941) and G. C. Goswami & Tamuli (2003) consider the fricative /h/ to be a voiced glottal phoneme and the approximant /w/ to be a voiced bilabial. However, as seen in Table 1.9, Mahanta (2012) considers them to be voiceless glottal and voiced velar respectively.

Table 1.9: Consonants in Assamese (Mahanta, 2012)

	Bilabial		Alveolar		Palatal		Velar		Glottal	
	vl.	vd.	vl.	vd.	vl.	vd.	vl.	vd.	vl.	vd.
Plosive	p	b	t	d			k	g		
	p^h	b^h	t^h	d^h			k^h	g^h		
Nasal		m		n				ŋ		
Fricative			s	z			x		h	
Approximant				ɹ		j		w		
Lateral approximant				l						

From the descriptions of vowel phonemes discussed in this section it can thus be said that standard Assamese has eight vowel phonemes: three front, four back and one central. The front vowels are: close /i/, close-mid /e/ and open mid /ɛ/. Although the vowel /e/ mostly occurs in conditions of vowel harmony as described in Mahanta (2008), recalling from (Taid, 1988), there is evidence of minimal contrasts of the vowels /e/ and /ɛ/ confirming their distinctiveness. The back vowels of standard Assamese are: high /u/, a little lower than /u/ is /ʊ/, higher-mid /o/ and lower-mid /ɔ/ (Mahanta, 2012), and the low central vowel /a/ (G. C. Goswami & Tamuli, 2003). Table 1.10 presents the vowel contrasts in Assamese along with their IPA notations that will be used in this study.

Table 1.10: Vowel phonemes in Assamese

Vowel	Assamese word	English meaning
/i/	/bis/	twenty/ to fan
/e/	/bes/	fine
/ɛ/	/bɛs/	sell
/a/	/bas/	to pick
/ɔ/	/bɔl/	strength
/o/	/bol/	let's go!
/ʊ/	/bʊl/	hue
/u/	/bul/	walk/ a person's name

As discussed in the previous sections, there are regional variations in Assamese phonemes. Hence, it is pertinent to sample these phonemes from different regions of Assam to better understand the nature of these variations. This forms the motivation to study the vowels in different geographical locations of Assam. The rest of the chapters in this thesis use various acoustic and statistical measurements to see how these vowels behave in the five regional varieties.

1.6 Motivation and Research Objectives

From Robinson's grammar to Dr. B.K Kakati's work on the formation and development of Assamese, numerous scholars have studied and have written extensively on this language. In recent years, several works have touched upon different aspects of the Assamese language. While Moral (1992) focused on the phonology of Assamese dialects, Hakacham (2000) acknowledges the influence of several features of the Tibeto-Burman languages on Assamese. Mahanta (2001, 2012) tries to present a definitive inventory of standard Assamese vowels using phonetic analysis, and Twaha

(2017) discusses the differences in prosody among the dialects of Assamese. From these studies it is seen that even though a theory of standard Assamese exists, it is however, not a monolithic variety and depending on geographical locations, distinct variations are observed. Scholars have maintained that the Assamese varieties differ on a number of factors including phonetic features. However, these variations lack an in-depth acoustic study. A detailed acoustic phonetic study of the language is pertinent for a proper understanding of the language and its varieties and also to provide insight into the ongoing vowel changes in the language. The lack of a detailed acoustic phonetic study of Assamese varieties as well as the need to understand the general vowel changes in the language has been the motivation for this study.

The research objectives of this thesis are as follows:

- To study the pattern of vowel variation in Assamese varieties.
- To determine the nature of vowel mergers in the Assamese varieties.
- To see if speakers across varieties are able to perceive the differences in each vowel pair.
- To explore any possible correlation between vowel quality and rhythm characteristics.

In order to study the pattern of vowel variation in Assamese varieties, vowel data was collected from the five geographical regions of Assam namely, Tinsukia, Jorhat, Nagaon, Nalbari and Barpeta. In order to characterize the vowels, formant measures F1 and F2 are analyzed and plotted on F1-F2 planes. The analysis also enables us to investigate if there is any vowel merger happening in the varieties. To confirm and determine the nature of vowel mergers in the varieties, a series of measurements are calculated, such as Euclidean Distance, Pillai Bartlett Trace Spectral Overlap Assessment Metric (SOAM) and Vowel Overlap Assessment with Convex Hulls (VOACH).

We expect that the results will show significant acoustic variation among the vowels of the geographically distinct varieties. We also expect to see a reduction in the vowel inventories of the Assamese varieties owing to mergers of vowels, particularly in the Eastern and Western Assamese varieties, as evidenced by impressionistic observations.

To see how the Assamese vowels are perceived, two kinds of perception experiments are conducted. The first is a discrimination test where participants are required to distinguish between two stimuli played to them. The second is an identification test where participants are required to identify the sound played to them. These experiments are described greater detail in Chapter 4. Results of the perception experiments will tell us if speakers across varieties are able to categorically perceive the Assamese vowels or not.

For speaking rate and rhythm measures, data was collected from a read speech. Measures like syllable per second and segments per second are calculated to determine the speaking rate in the varieties. These measures will tell us if a variety is faster or slower than the others. For rhythm measures ΔC, ΔV, %V, nPVI-V and Varco V were calculated. These measures will enable us to explore possible correlations between vowel quality and rhythm characteristics and also position Assamese varieties among languages belonging to three rhythm groups: stress-timed, syllable-timed and mora-timed.

1.7 Methodology

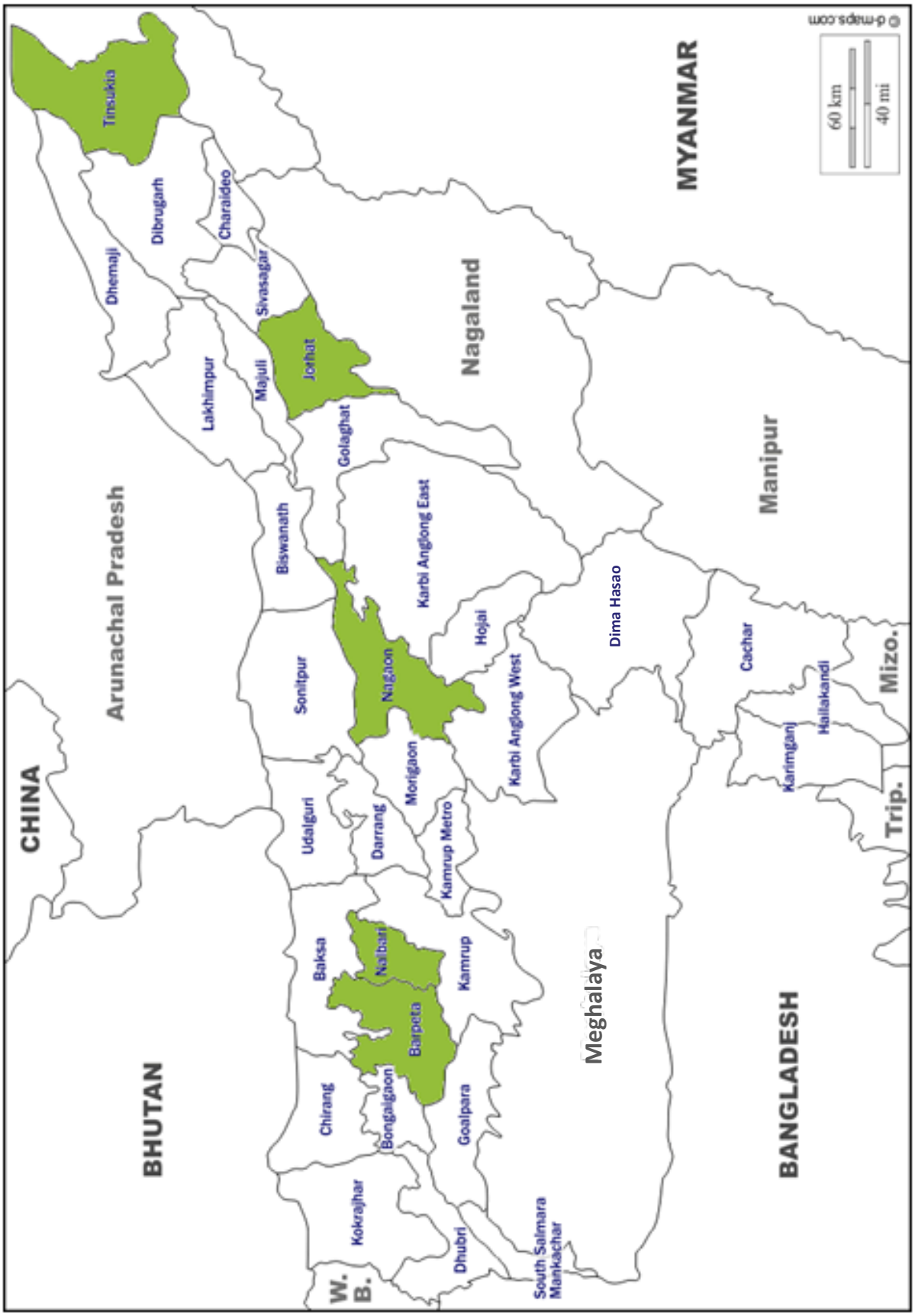
This section gives a general overview of the method employed for data collection and analysis in the research work. Detailed methodologies for each of the chapters covering specific research objectives have been explained in the respective chapters.

1.7.1 Geographical area covered in the current work

G. C. Goswami & Tamuli (2003) divide Assamese into three broad dialectal groups based on geographical location: Eastern Assamese, spoken in the easternmost districts of Assam, such as Tinsukia, Dibrugarh, Lakhimpur, Sivasagar and Jorhat, including the adjoining areas of Arunachal Pradesh to Sonitpur and Nagaon districts on the western boundary. Western Assamese is spoken from a little east of Guwahati in the south and the Darrang district in the north, all the way to the Goalpara district in the western border. Central Assamese or the 'intermediate dialect' is spoken in areas between Eastern and Western Assam. The current study aims to look at the variation in Assamese as spoken in five regions in Assam and representing the three broad dialectal groups. These five regions are Tinsukia in the extreme east of Assam, followed secondly by Jorhat, and belonging to Eastern Assam. The third region is Nagaon, which geographically belongs to Central Assam. Following Nagaon are Nalbari and Barpeta, which belong to Western Assam. Tinsukia and Jorhat are representatives of Eastern Assamese. With a population of about 1.2 lakh people, Tinsukia is also home to several indigenous tribes like Morans, Motoks, Chutias, Sonowals, Deoris and Ahoms. While almost all of these tribes have adopted Assamese as their mother tongue, due to the long contact with the languages spoken by these tribes, subtle affects on the Assamese spoken in this region can not be ruled out. About 183 km west of Tinsukia is Jorhat, which is the second largest city of Assam. Jorhat was the last capital of the Ahom kingdom and the administrative headquarters of the undivided Sibsagar district during the British rule. The language spoken in this region eventually came to be established as the standard form of Assamese. Nagaon belongs to the Central Assamese group, which is considered to be an intermediate variety between Eastern and Western Assamese. Along with Assamese, Tiwas are a dominant tribal group residing in this region. Other indigenous tribes in this region are Karbi, Bodo, Rabha, Koch and Garo. Nalbari and Barpeta are representatives

of Western Assamese. Although both the regions are categorized under the same dialectal region, the Assamese spoken in these two regions are very different from each other and are considered to be two main sub-varieties of Western Assamese. Owing to the dialectal variations observed in these geographically distinct regions, this thesis covers these five regions with an attempt to study the distinct acoustic phonetic features of the Assamese spoken here. The study looks into the geographical progression of vowel change in these Assamese dialects and presents an understanding of the vowel space in each of these varieties.

Figure 1.2: Areas covered in the current work



1.7.2 Participants

Since the study concentrates on the spoken varieties of Assamese, participants taken for the study were native and fluent speakers of their respective varieties. It was also ensured that they did not report any speech or hearing disorders. Once the participants fulfilled these basic criteria, they were asked to fill a form to elicit other sociolinguistic responses. A total of 75 individuals participated in the production data collection. Table 1.11 summarizes the data of the participants. For the perception experiment, data was collected from 51 participants belonging from the five region. Hence, the total number of participants for the entire study is 126. A detailed list of the participants is provided in Appendix A.

Table 1.11: Participants' data

	Region	Avg. age	Avg. education	Female	Male	Total participants
Production	Barpeta	28	Undergraduate	6	4	10
	Jorhat	37	Undergraduate	11	13	24
	Nagaon	36	Undergraduate	5	6	11
	Nalbari	36	Graduate	9	11	20
	Tinsukia	39	Undergraduate	5	5	10
	Total			36	39	75
Perception	Barpeta	26-30	Post graduate	3	8	11
	Jorhat	31-35	Post graduate	10	2	12
	Nagaon	26-30	Post graduate	5	4	9
	Nalbari	36-40	Post graduate	5	4	9
	Tinsukia	26-30	Post graduate	2	8	10
	Total			25	26	51

1.7.3 Data collection

Data collection for the present research involved recordings of the eight vowels of standard Assamese viz. /i, e, ɛ, a, ɔ, o, ʊ, u/ from a list of CVC words. The CVC wordlist consisted of 73 unique words in total presented in a natural sentence, isolation and a sentence frame /mɔi X buli kolʊ̃/ ("I X said") written in the Assamese script. The natural sentence ensured that speakers identified the correct meaning of the target

word. Apart from the worldlist, a short passage was also recorded. The passage was an Assamese translation of the story 'The North Wind and the Sun' and consisted of 103 words, 252 syllables and 447 phonemes. All data was collected in the field and the recordings were done in a noiseless environment. A Shure unidirectional head-worn microphone (Model: SM10A) connected to a Tascam linear PCM recorder (Model: DR100 MKII) via xlr jack was used for the recording. The sampling frequency was kept at 44.1 kHz and 24 bit in .wav format. However, prior to this, a preliminary data collection was done from speakers of Jorhat and Nalbari varieties using the datalist given in Mahanta (2012). These recordings were conducted in a sound attenuated recording booth. A detailed description of the methodology has been provided in Section 2.2.1.

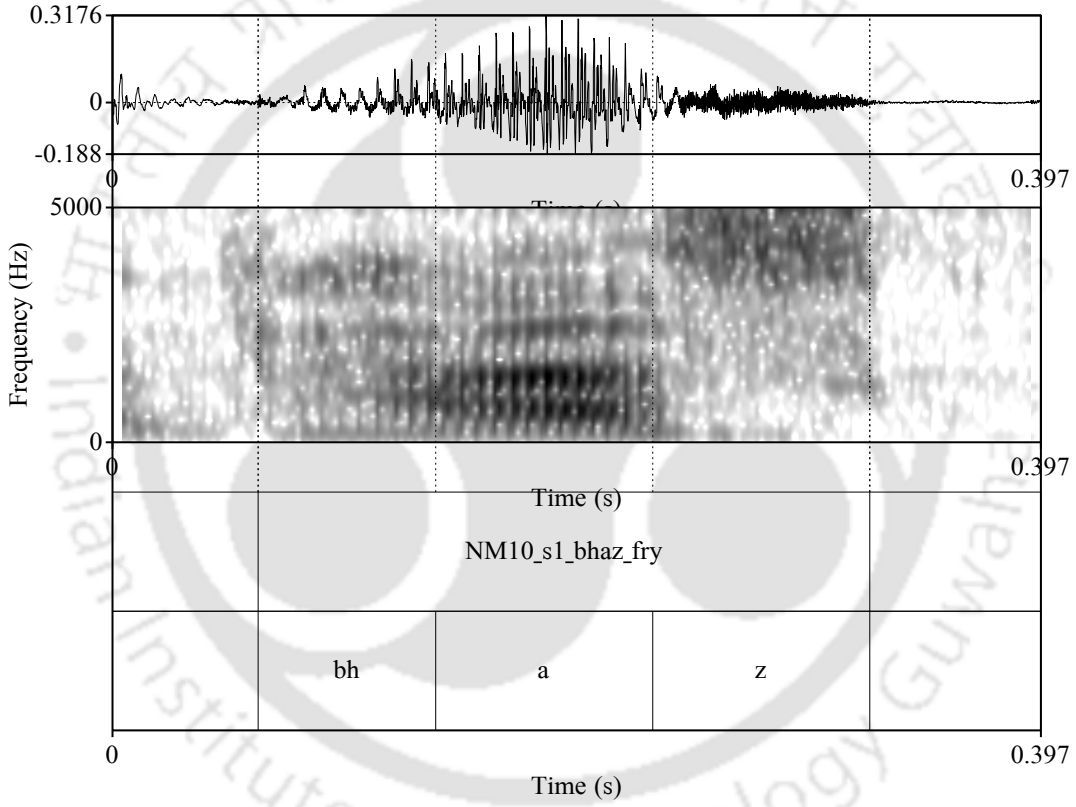
To collect data on the perception of Assamese vowels, two kinds of tests were conducted on 51 speakers from the target areas. The first was a discrimination test where participants were required to distinguish between two stimuli played to them. They were presented with two sounds one after the other and were asked to judge if the sounds they heard were similar or different. The second was an identification test, where participants were required to identify the sound played to them. The participants were presented with a sound and were asked to choose the correct meaning of the word they heard, from a list of options. The methodology for the two tests is explained in detail in Section 4.2.

1.7.4 Annotation

The data recorded was manually segmented in Praat 5.4 (Boersma & Weenink, 2014). For vowel segmentation, onset of a steady state formant was considered vowel beginning and the termination of the steady state was considered vowel end. The annotation was marked in two tiers. In the first tier the details of the sound file was entered as shown in Figure 1.3 which included the speaker number, iteration number,

Assamese word and English meaning. In the second tier, the phonemes were marked. The segmented phonemes along with the time indices, in textgrid form, were saved separately. For annotation for rhythm and speech rate measures, breath groups with at least seven syllables were marked in the first tier and phonemes were marked in the second tier.

Figure 1.3: Praat annotation of speech data



1.7.5 Normalization of data

Recorded speech data may contain anatomical/physiological and talker-related information systematically affecting formant frequencies and creating unwanted variation (Peterson & Barney, 1952). In a study by Adank, Smits, & Van Hout (2004), the z-score method of frequency normalization used in Lobanov (1971) for vowel nor-

malization, has been found to be the best vowel normalization metric. Lobanov's z-score transform is a vowel extrinsic, talker intrinsic normalization procedure reducing anatomical variation of the formant frequencies but maintaining phonological and sociolinguistic variation. Hence, in this study before proceeding with the analysis of the formant frequencies, the data was normalized with Lobanov method for speaker effects using the *normLobanov()* function from the *phonR* package on the open source R platform (R Core Team, 2019).

1.7.6 Acoustic and statistical methods used

Formant frequency has been considered a reliable measure of vowel quality in speech (Deterding, 1997; Ladefoged & Maddieson, 1996; Lindblom, 1986). Peterson & Barney (1952)'s study on American English vowels clearly demonstrated how F1 and F2 are related to vowel height and backness respectively. Hence, in order to investigate the vowel quality of the two varieties of Assamese, the first two formants of the vowels were measured. A Praat script using the Burg algorithm was used to calculate the first two formants (F1 and F2) in Hertz at the mid 20% of the vowel duration. The values of the first two formants were then used to plot the vowels and also to test for statistical validity of acoustical differences.

Exploratory statistical analyses were conducted by building linear mixed effects (LME) models for F1, F2, F3 (Lobanov normalized) for the vowel types. LME models were built using vowel type, region, context, gender, speaker, onset types, coda types and words as factors. The sections describing the results of the statistical tests have more information about the models. All statistical analyses were conducted using the open source R platform (R Core Team, 2019). LME models were built using the *lme4* package on R (Bates et al., 2015). The LME models were subjected to Type II Wald chisquare tests for analysis of deviance using the *car* package and pairwise comparisons were conducted using the *emmeans* package (J. Fox & Weisberg, 2019;

Lenth, 2019).

In order to see the overlap of the distribution of F1 and F2 of vowels in the Assamese varieties, the vowel clusters of interest were subjected to four kinds of tests: Euclidean distance, Pillai score, Spectral Overlap Assessment Metric and Vowel Overlap Assessment with Convex Hulls. Euclidean distance of F1 and F2 values provide an average distance between two vowel categories, ignoring the amount of variability and deviation of the vowels from the average values. Pillai-Bartlett trace or simply, Pillai score, is one of the statistics in the multivariate analysis of variance (MANOVA) and considered robust in measuring the separability of two data categories. Another method of determining and visualizing overlap in vowel clusters is the Spectral Overlap Assessment Metric (SOAM), proposed by Wassink (2006, 1999). In this method the normalized scatter for two vowel distributions is modeled as two best-fit, weighted ellipses. These ellipses are oriented at angles with respect to F1 and F2. The output of the metric is an overlap fraction. This fraction represents the area of the region of overlap between the two ellipses. In Nycz & Hall-Lew (2013), among four methodological approaches to representing and assessing vowel differences, the SOAM method was considered the best. Lastly, the Vowel Overlap Assessment with Convex Hulls (VOACH) method has been used. This method proposed by Haynes & Taylor (2014), is an improved version of the SOAM method which uses using best-fit convex shapes to quantify merger.

For speaking rate, the measures calculated are syllable per second and segments per second. To analyze the rhythm variation, interval measures that are evaluated are: %V, percent-age of vocalic intervals in the complete utterance; ΔC, standard deviation of duration of consonantal intervals; ΔV, standard deviation of the duration of vocalic intervals in a breath group and Varco V (Ramus et al., 1999).

Chapter 2

Acoustics of vowel variation

2.1 Introduction and literature review

From the previous chapter we know that Assamese has been broadly divided into three dialectal divisions. Each of the divisions have several sub varieties within them (G. C. Goswami, 1982; G. C. Goswami & Tamuli, 2003). In this chapter we look into the vowel acoustics of Assamese and show that vowels plays a significant role in dialectal variation in Assamese. Results of the experiments show that speakers of Tinsukia and Jorhat varieties do not distinguish between the vowels /u/ and /ʊ/. Statistical tests confirm that these two varieties of Assamese maintain significant acoustic separation only among seven vowels. The distinction between /u/ and /ʊ/ vowels reported for standard Assamese (Mahanta, 2012) is lost and /ʊ/ is neutralized to /u/. On the other hand, in case of Nalbari and Barpeta varieties, significant overlap of /e-ɛ/ and /ʊ-o/ is observed indicating merger of the pairs and thereby reducing the eight-vowel system proposed for standard Assamese to a six-vowel one. Again, the considerable overlap of the back vowels /u,ʊ,o/ in the Nagaon variety suggests that this variety is at an intermediate position between Eastern and Western Assamese. In order to understand dialectal variation, Section 2.1.1 presents an overview of language variation in general and discusses the factors that may lead to dialectal change. Section 2.1.2 discusses how vowel variation has been a major indication of dialectal

change. Also discussed are the vowel measures that have brought about variation in other languages. Section 2.2.2 then presents the results of the analysis of the acoustic features of vowels in Assamese varieties.

2.1.1 Language variation

Several factors bring about variation in the way a language is spoken. These factors may range from political reasons, geographical distinctiveness, ethnicity, to purely social factors like age, gender, class etc. The English language, for example, has a wide range of varieties. These include American English, British English, Australian English, Canadian English etc, spoken in countries where English is used as the primary language. Then there are varieties of English spoken in countries which have some colonial history. This includes nations such as India, Ghana, Kenya, Malaysia, Nigeria, Singapore, Sri Lanka, Tanzania, Zambia etc. Within the English speaking countries, the language again has several regional varieties. For example, different geographical varieties such as South-Western ('West Country') English, Northern English, Scottish English, and so forth are seen in the British Isles. Similarly, American English has several varieties spoken in geographically distinct areas like North Central, Midland, Southern, Mid-Atlantic, New York etc. Both British and American Englishes again have varieties that have cropped up due to purely social factors. For example, in England, Received Pronunciation (RP) or the 'BBC accent' is seen as being more pleasing, articulate and prestigious than other varieties. Another example of social class playing a role in how a language is spoken is the famous study by William Labov (Labov, 2006) which illustrated that (r) in New York City was stratified by class. The pronunciation of /r/ depended on the social-class membership of the speakers. Those with higher socioeconomic status pronounced post- vocalic /r/ more frequently than those with lower socioeconomic status. Additionally, those aspiring to be in a higher socioeconomic class were also fund to pronounce post-vocalic

/r/ more frequently.

Among sociolinguists, there is no consensus for well-defined criteria regarding the terms 'language', 'dialect', 'variety' and 'accent'. These terms usually have a more social, political and cultural underpinning than a linguistic one. Very often, a dialect is considered to be a *substandard, low-status, often rustic form of a language* generally associated with communities or groups considered to be lacking in prestige, such as the peasantry or the working class (Chambers & Trudgill, 1998, p.3). A 'language' is understood as what Joseph (1987, p.1) defines as, "a [linguistic] system of elements and rules conceived broadly enough to admit variant ways of using it. A dialect is understood as one of these variant ways". In other words, what resolves this confusion whether a language or a dialect is being spoken seems to lie in its perceived importance of usage.

One criterion that is sometimes used by linguists to differentiate between a language and a dialect is that of 'mutual intelligibility', that is to say, 'a language is a collection of mutually intelligible dialects'. The problem with this criterion is illustrated with the example of the Scandinavian languages. While Norwegian, Swedish and Danish are mutually intelligible, they are considered to be different languages[1]. On the other hand, although German is considered to be a single language, certain dialects of German are not mutually intelligible (Chambers & Trudgill, 1998). A similar example is seen among the Indian languages Hindi and Urdu, which are mutually intelligible but are considered different languages. Again, certain dialects of Hindi are not mutually intelligible. In other times, mutual intelligibility can be a question of degree rather than being an all-or-nothing matter. For example, Portuguese speakers understand Spanish better than Spanish understand Portuguese (Penny, 2000). Chambers & Trudgill (1998) consider the term 'language' to be a non-technical one from a linguistic point of view and uses the term 'variety' for any particular kind of

[1]While this differentiation is due to political reasons, but relying on the mutual intelligibility criterion, linguists would still identify these varieties as belonging to the one language.

language that can be considered as a single entity. They also distinguish between the labels 'accent' and 'dialect' which are also often used interchangeably. 'Accent' refers to a variety which is phonetically and/or phonologically different from other varieties. 'Dialect' on the other hand refers to a variety characterized by systematic differences in phonology, grammar and vocabulary from other varieties of the same language.

A variety that in different ways is recognized as more correct and acceptable than other varieties has come to be called a 'standard variety' or 'standard language'. A language thus, is a dialect that has undergone the various processes of standardization. Haugen (1966) proposed four stages of standardization. The first stage is 'selection' in which one of the many forms of language that exist in a society is selected to be the standard. The second stage is 'codification' in which norms and rules of grammar that govern the language as well as the writing system are formulated and codified through authorised documents, media publications, texts etc. In the third stage the functions of the new codified language is extended for use in new domains such as education, government agencies, law and judiciary, bureaucracy, diplomacy, trade and commerce etc. This stage is called 'elaboration'. The fourth stage is that of 'acceptance' where the variety chosen as the standard gains wide recognition and acceptance within the community consisting of speakers belonging to diverse dialect-groups. In most cases the standardized language is also recognized as the 'national' or 'official' language, thus giving its users a distinct national-linguistic identity. A 'language' differs from 'dialect' in the degree to which it has been subjected to each of the processes of standardization (Penny, 2000). However, the term 'dialect' is often used by linguists as a neutral term to refer to a systematic usage of a language variety in a particular region or by a particular social class.

While several factors may bring about dialectal change, such as ethnicity, cultural, religious and racial differences as well as social differences, one of the most important factors that may lead to dialectal change is geography. When dialects are described in

terms of geography, linguists usually find a 'dialect continuum'. Chambers & Trudgill (1998, p.5-6) explains a 'dialect continuum' as follows:

If we travel from village to village, in a particular direction, we notice linguistic differences which distinguish one village from another. Sometimes these differences will be larger, sometimes smaller, but they will be cumulative. The further we get from our starting point, the larger the differences will become. Dialects on the outer edges of the geographical area may not be mutually intelligible, but they will be linked by a chain of mutual intelligibility. At no point is there a complete break such that geographically adjacent dialects are not mutually intelligible, but the cumulative effect of the linguistic differences will be such that the greater the geographical separation, the greater the difficulty of comprehension. This type of situation is known as a geographical dialect continuum.

One of the canonical examples of a dialect continuum is the European dialect continua where although the standard varieties of French, Italian, Catalan, Spanish and Portuguese are not mutually intelligible, the rural dialects of these dialects form part of the West Romance dialect continuum stretching from the coast of Portugal to the centre of Belgium. Similarly, the West Germanic continuum, which includes all dialects of German, Dutch and Flemish; the Scandinavian dialect continuum and others as seen in the map 2.1 (Chambers & Trudgill, 1998, p.6).

Figure 2.1: Map of European dialect continua (Image taken from Chambers & Trudgill (1998, p.6))

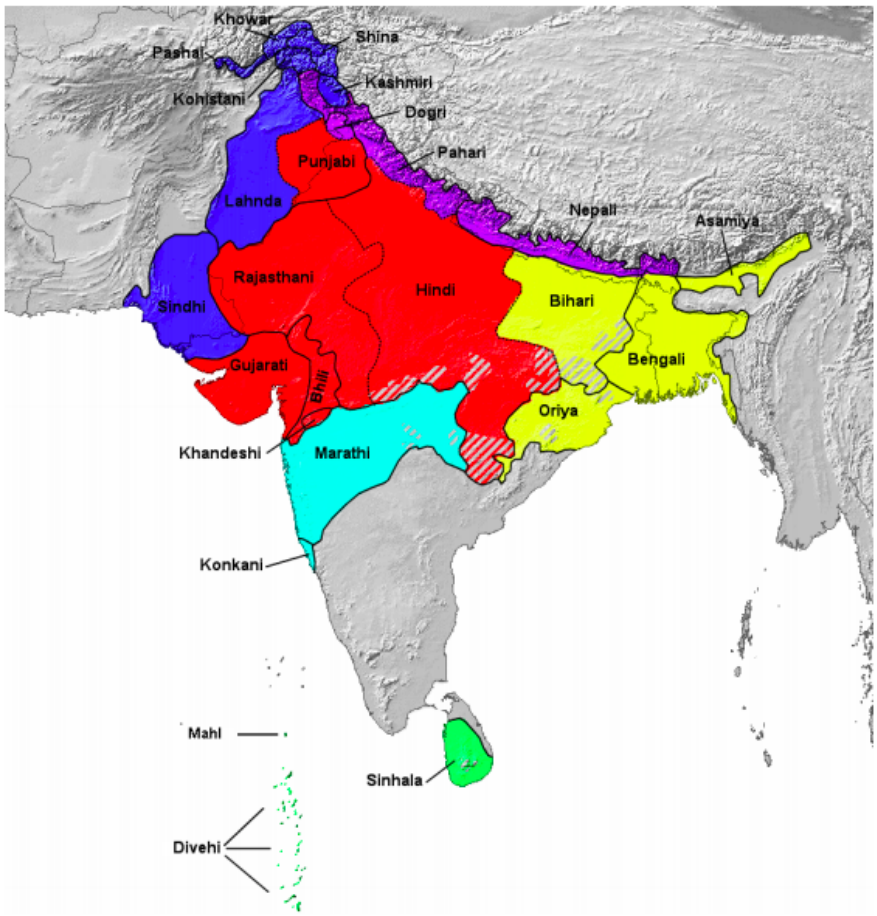
In the current study we are looking at a language that expands over the Brahmaputra valley that stretches over an area of about 720 kilometers from east to west. As mentioned by scholars, Assamese exhibits a wide range of linguistic variations which are geographically distinct. As also mentioned by Patgiri (2011), the Assamese dialects belong to a geographical dialect continuum. Owing to geographical distinctiveness, we expect that the varieties spoken in the far ends of this dialect continua will have more differences than the ones that are closer. Hence, in this study we look

at the vowels spoken five distinct regions across Assam. Section 2.1.1.1 provides a brief note on the geographical varieties of the Assamese language.

2.1.1.1 Variation in Assamese

Similar to the idea of dialect continua discussed previously, among the Indo-Aryan languages of India, instances of language continua has been reported by several scholars. In Grierson (1903-1928)'s synchronic description of the linguistic space, Bengali, Assamese, Eastern Hindi (Magahi), and Oriyā are assigned to the Eastern group, Marāthi and its dialects to the Southern group, Sindhi, Lahnda and the Dardic languages (including Kashmiri) to the North-Western group, while Western Hindi, Punjabi, Rajasthani, Gujarati are assigned to the Central group (as seen in Figure 2.2). Masica (1993) talks about dialect continua among the New Indo-Aryan (NIA) languages of India without sharp boundaries separating mutually unintelligible languages and recognizes a number of *overlapping genetic zones* each defined by specific criteria. These linguistic geographical divisions include the Western region with Hindi and its multiple dialectal variants, North-Central region occupied by Nepali; Northwest includes Punjabi, Sindhi, and Lahnda; Rajasthani and Gujarati to the West; Marathi and Konkani to the South; and the East occupied by Bengali, Oriya, Assamese, and Eastern Hindi (Deo, 2018).

Figure 2.2: The Indo-Aryan language continuum according to SIL, Ethnologue, 1978 (Image taken from Commons (2020))



In the descriptions of the Assamese language, scholars have maintained that dialects of Assamese are geographically distinct. As mentioned in Section 1.4, Grierson (1903) and Kakati (1941) mention two distinct varieties of Assamese – the Eastern Assamese (EA) and Western Assamese (WA). While the Eastern variety, spoken from Sadiya down to Guwahati, exhibits homogeneity, the Western variety, spoken in the two western districts of Kamrup and Goalpara, comprises of many local dialects. U. Goswami (1970b) further divides Western Assamese into three sub varieties namely, Barpeta, Nalbari and Palashbari-Chaigaon subvarieties. G. C. Goswami (1982) divide Assamese into three distinct varieties namely, Eastern, Western, and Central

varieties. Moral (1991) adds that Central Assamese includes the dialect spoken primarily in Nagaon and Morigaon districts and some parts of the adjoining Sonitpur, Darrang, Jorhat and Golaghat districts. Das (2010) divides EA and WA into two subvarieties each, namely, Sibsagar variety, and Central Assam variety, Kamrup variety and Goalpara variety. As can be understood, Assamese dialects show considerable variation with only the neighbouring dialects being mutually intelligible. Many a times, some dialects are not intelligible to people who only know the standard variety. However, all Assamese dialects belong to the geographical dialect continuum where Assamese is spoken in and around the Brahmaputra valley (Patgiri, 2011).

The most significant variations within languages occur lexically, phonologically, grammatically and according to usage. The variations are mostly qualitative, in the sense that one dialect uses a feature more often than another dialect does (Rickford, 2002). In the case of Assamese, as mentioned earlier, scholars divide the language into three different dialects based on the geographical area. These dialects differ from one another in a wide range of features of phonology, morphology and vocabulary (discussed in Section 1.4). This study focuses on one such aspect of dialectal variation of Assamese i.e., vowel variation. G. C. Goswami (1982, p.12) mentions that, "...the Eastern and the Western Assamese are sharply differentiated in almost all areas, viz., in phonology or pronunciation and intonation including speech tempo...". G. C. Goswami & Tamuli (2003) reiterate these findings and consider vowel variation as a major reason for their grouping of Assamese varieties into three groups. Another study on Kamrupi, a variety of WA (Deka, 2007) mentions that compared to the standard variety of Assamese with eight vowels, the Kamrupi variety has only seven vowels, i.e. the vowel /o/ is absent in the variety . Taid (1988, p.74) pointed out that there are not three, but four back vowels in "educated Assamese speech", viz /ɒ, ɔ, o, u/, the phonemicity of which is evident in several minimal sets. However, he also mentions that in some dialectal varieties such a balanced vowel system does exist,

with the difference that instead of three back vowels /ɔ, o, u/, we come across /ɒ, ɔ, u/, /o/ being neutralized with either /u/ as in the speech of upper Assam or with /ɔ/ as in the Kamrupi dialect.

In the following section a brief discussion on vowels as a measure of language variation has been provided. In later sections we try to analyze the vowels in Assamese varieties using several acoustic and statistical measures.

2.1.2 Vowel variation in languages

Cross-linguistically, vowel variation and merger has been considered to be a prominent source of linguistic variation. Dialectal studies on American English (Hagiwara, 1997; Clopper et al., 2005; Fridland et al., 2014), British English (Williams & Escudero, 2014), Dutch (Adank, Smits, & Van Hout, 2004) and Portuguese (Escudero et al., 2009) among others, have established the importance of vowel variation in languages. While some varieties of languages seem to differ in terms of spectral qualities such as formants and vowel space, others have been found to differ in terms of temporal qualities like vowel duration, rhythm and speech rate. One of the pioneering studies that used acoustic analysis in the study of language variation was the quantitative study of sound change in English by Labov et al. (1972). The primary aim of this study was to determine the principles governing vowel shifting. Since then, acoustic enquiry into language variation and change has "grown steadily" (Thomas, 2000, p.368). In the following years a number of acoustic studies of various languages and their dialects were carried out. Such studies have looked into various components of language variation ranging from formant frequencies and duration of vowels, to speech rate, rhythm, intonation, perception, and consonantal variation etc. A major portion of these studies is concentrated on the ways in which production of first and second formants of vowels vary among dialects and languages. The first and second formant values being indicative of vowel height and frontness are important to the

study of how two vowels differ in languages. In the case of Assamese, the accounts of vowel variation and vowel merger in the varieties are largely impressionistic (Kakati, 1941; G. C. Goswami, 1982; G. C. Goswami & Tamuli, 2003). Hence it is important that a detailed acoustic phonetic analysis of the vowels is conducted. Considering that, in this section we present a brief note on vowel acoustics followed by a review of the literature on the theories of vowel systems as well as the various studies of vowel systems and variations across languages.

2.1.2.1 A brief note on vowel acoustics

Speech acoustics finds its roots in a theory of speech production that models speech as a combination of a sound source, and a linear acoustic filter. The relationship between a vowel's acoustic properties and the shape of the vocal tract was famously first shown by Tsutomu Chiba and Masato Kajiyama (Chiba & Kajiyama, 1958). This theory later came to be known as the "source-filter" theory of speech production (Fant, 1970). Speech is viewed as an audible gesture whereby the speaker creates variations in the air pressure and airflow from the lungs using various articulatory configurations. These changes in the air pressure and airflow give rise to acoustic signals that we hear when perceiving sounds. In humans, the vocal folds/larynx serve as the source of sound energy. Vibration in the vocal folds produces a complex periodic wave (represented as the sum of a number of sine waves or pure tone components). The fundamental frequency, in the case of voiced speech sounds is the rate of vibration of the vocal folds. The pure tone components are found at multiples of the fundamental frequency and are called harmonics. The vocal tract above the larynx which is a sort of chamber or tube has certain resonant frequencies at which the contained air naturally tends to vibrate. Resonances in the vocal tract give rise to distinctive frequency components of the acoustic signal carrying a lot of energy. These are known as "formants". The vocal tract thus functions as an acoustic filter thereby amplifying the source frequencies

close to a resonant frequency and attenuating the other frequencies.

In phonetics, vowels are defined as sounds that are produced with a relatively open vocal tract and free passage of the airstream. Vowels are frequently described with reference to their formant structure, which provides an indication of vocal tract resonance and therefore articulatory shape (Fant, 1970). Articulatorily, three parameters are used to describe vowel sounds: tongue height, tongue backness and lip roundedness. Ladefoged & Disner (2012) mention that these three parameters are correlated with the first three formants. Changes in the position of the lips and tongue (and thus the jaw), result in changes to the shape of the oral cavity. Different shapes produce different resonances which result in the production of different formant frequencies. These acoustic resonances appear as dark bands in a spectrogram of the voice signal. Vowel height, which corresponds to the tongue height, refers to the lowest resonance of the voice or the first formant (F1). The first formant (F1) in vowels is inversely related to vowel height i.e., the higher the vowel, the lower the first formant (and vice versa). Likewise, vowel backness is named for the position of the tongue during relative to the back of the mouth. It is defined by the second formant (F2). The more front the vowel, the higher the second formant. Although vowel roundedness does not always have a direct correspondence with lip rounding, acoustically, rounded vowels are defined by the third formant (F3).

Apart from height, backness and roundedness, there are secondary qualities that distinguish vowels. Several languages phonemically distinguish between long and short vowels while many distinguish between oral and nasalized vowels. A nasalized vowel is produced with a lowered soft-palate, allowing airflow through both oral and nasal cavities simultaneously. The resulting vowel has a characteristic nasal quality due to the addition of the resonances of the nasal cavity to those of the rest of the vocal tract. While Assamese does not distinguish between long and short vowels, nasalized vowels do form a part of the phoneme inventory. An overwhelming majority

of vowel sounds in speech are voiced (since vowel formants are modifications of a voiced airstream from the larynx), however, there are some languages, like Japanese, where some vowels become voiceless in some environments (in Japanese, high vowels /i/ and /u/ are devoiced between voiceless consonants (Jaeger, 1978; Vance, 1987)). These vowels are produced without an apparent vibration of the vocal cords. The following section discusses how these acoustic features of vowels account for variation in languages.

2.1.2.2 Vowels as a measure of language variation

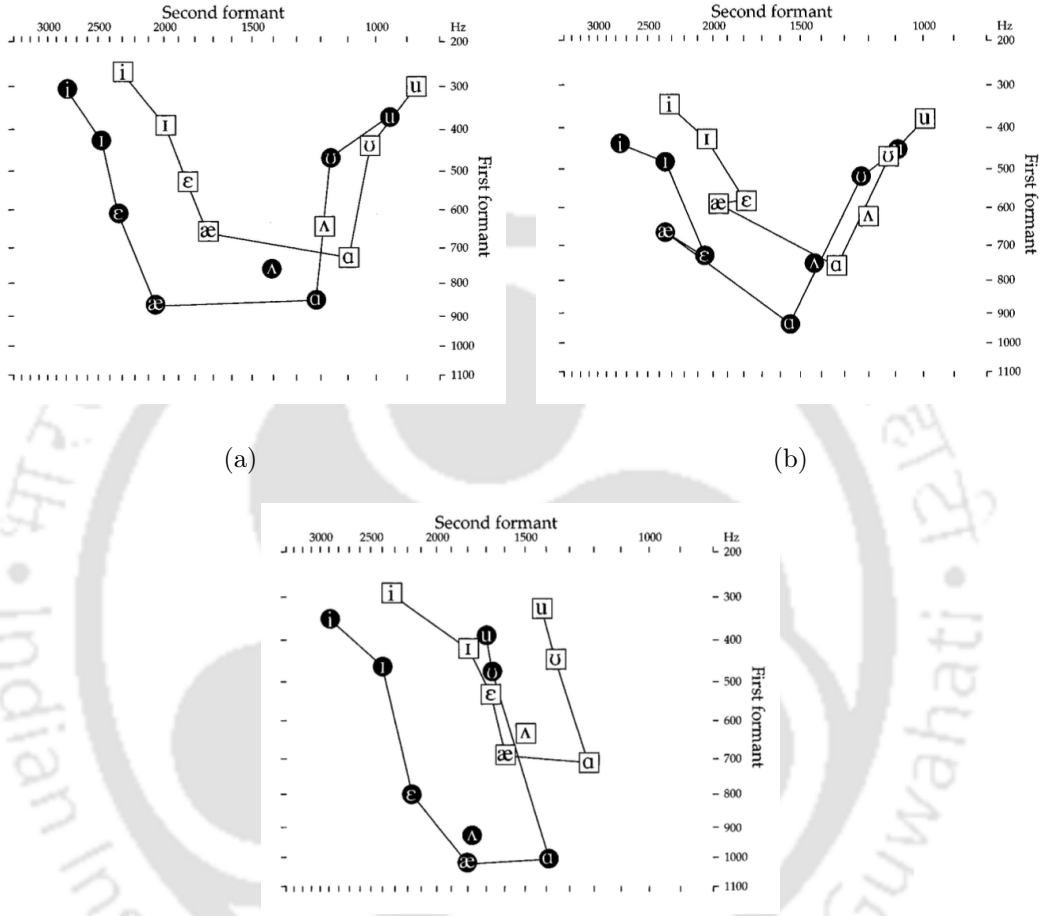
Studies of language variation have shown that vowel quality changes considerably across language varieties. Thomas (2000) in his summary of the phonetic methods applied to language variation mentions that most of the studies on acoustic analysis of vowels have been concentrated on the comparisons of F1 and F2 values and other components of vowels like duration, vowel space etc receiving almost no attention. In recent years however, researchers have delved into these components to understand language variation. Notable mention of the studies on vowel duration may be made of Clopper et al. (2005), Jacewicz et al. (2007), Fridland et al. (2014) on American dialects; Adank, Smits, & Van Hout (2004) on dialects of Dutch etc. In all these studies a significant main effect of dialect on vowel duration was observed. It may however be mentioned that these dialect differences were not always consistent across all vowels. Clopper et al. (2005) in their study of American dialects observed that Southerners produced only two vowels, /ɛ/ and /ʌ/ significantly longer in duration than Northerners. Further, along with the significant main effect of dialect, a significant main effect of gender too was observed on vowel duration in case of Dutch (Adank, Smits, & Van Hout, 2004), whereby females showed longer vowel duration than males. Similarly, studies of dynamic spectral change include works by R. A. Fox & Jacewicz (2009), Williams & Escudero (2014) and Adank, Van Hout, & Smits (2004) among

others. This section primarily focuses on the discussions of steady-state F1 and F2 variations in languages, however at the same time, touches upon some of the analyses on duration and dynamic spectral change as they come.

Some of the earliest studies on the acoustics and perception of vowels were done on varieties of American English. Peterson and Barney (Peterson & Barney, 1952) (PB hereafter) had done a simple study on American vowels at Bell Telephone Laboratories shortly after the introduction of the sound spectrograph, which played a central role in the development and testing of theories of vowel recognition by researchers. Hagiwara (1997) compared the vowels of Southern California (SC) with PB's American vowels and another variety, Northern Midwesterners (NM) previously studied by Hillenbrand et al. (1995). Although little can be said of the dialectal homogeneity of the PB group, the majority of the speakers among 33 men, 28 women and 15 children, were reported to be speakers of "General American" (GA) (Peterson & Barney, 1952). F0, F1, F2 and F3 of the vowels /i, ɪ, ɛ, æ, ɑ, ɔ, ʊ, u, ʌ, ɝ/ were measured during steady state portions of /h_d/ utterances. The /h_d/ signals were also presented to listeners for an identification test. Hillenbrand et al. (1995) had collected the PB vowel data including /e/ and /o/ from 45 men, 48 women and 46 children speaking the northern Midwest variety. Formant frequencies (F1-F4) were extracted from LPC spectra. Listening tests were also performed with the data. Hagiwara (1997) recorded the vowels /i, ɪ, e, ɛ, æ, ɑ, o, ʊ, u, ʌ, ɝ/ in three consonant contexts (/b_t/, /t_k/ and /h_d/) from 9 women and 6 men Southern Californian English-speaking monolinguals. F1-F3 were measured by simultaneous comparison of wideband spectrograms, narrow-band FFT spectral slices and LPC spectra taken at vowel center or steady state where available. The results showed that the three varieties of American English differed greatly from each other with respect to vowel features. Firstly, the shape of the vowel spaces of the three varieties are totally different from each other. While GA vowel space resembles an trapezoid and NM a triangle, SC vowel space

resembles a parallelogram. Between GA and NM difference in the position of the vowels /æ/ and /ɑ/ was observed. As seen in Figure 2.3, /ɑ/in the NM vowel space has moved into a central position relative to the GA space. Additionally the /æ/ vowel has been fronted and raised in NM. The SC /u/ and /ʊ/ are typically unrounded and thus appear with higher F2 values than /ɑ/ and more characteristic of central space in GA. Further, /ʌ/ has a much higher F2 in SC than in GA. In case of NM, extreme raising of /æ/ relative of GA or SC and the centrality of /ɑ/ was observed, which gave the NM vowel space its distinct triangular shape. This clockwise chain shift also indicates a well-established "Northern Cities Chain Shift" (NCCS) as described by Labov (1994).

Figure 2.3: Vowel plots (women's, filled circles and men's, open squares) of (a) General American, (b) Northern Midwestern and (c) Southern Californian varieties of American English (Figures taken from Hagiwara (1997))



(a) (b)

(c)

Sociolinguistic researches, particularly the works of William Labov (Labov et al., 1972; Labov, 1991, 1994) had in the mean time established major regional vowel changes in effect in American dialects. Northern Cities Chain Shift (NCCS) and Southern Vowel Shift (SVS) were the two main shifts apart from several other dialectal variations. Characteristically, the NCCS is identified by the clockwise rotation of the low and low-mid vowels. /æ/ is raised and fronted, /ɛ/ and /ʌ/ are backed, /ɔ/ is lowered and /ɑ/ is lowered and fronted. In SVS the back vowels /u/ and /o/ are

fronted. Additionally, the front lax vowels / ɪ/ and /ɛ/ are raised and fronted and are moved to the periphery of the vowel space. The front tense vowels /i/ and /e/ are centralized through lowering and backing. Another important change occurring in the Midland, Western and New England dialects is the low back vowel merger wherein the vowels /ɑ/ and /ɔ/ merge creating the homophones of words such as the pairs *caught* and *cot* or *dawn* and *don* (Labov, 1991). (A 'merger' is a sound change whereby two or more contrastive sounds are replaced by a single sound. Consequently, phonemic distinctions are lost and previously distinct words become homophones. Detailed explanations on mergers are provided in chapter 3 of this dissertation). Dwelling on these changes Clopper et al. (2005) conducted an acoustic study of vowel systems of six regional varieties of American English. 48 speakers (4 females and 4 males each) from six dialect regions of the United States: New England, Mid-Atlantic, North, Midland, South, and West based on Labov et al. (2008) (then forthcoming) were recorded producing 11 vowels of American English (/i, ɪ, e, ɛ, æ, ɑ, o, ɔ, ʊ, u, ʌ,/) in the /h_d/ context. The measures obtained were vowel duration, F1 and F2 at one-third temporal point in the vowel, F1 and F2 at two-thirds temporal point in the vowel. The formant frequencies were normalized using Lobanov (1971)'s z-score method for each talker and statistically analyzed using a repeated measures ANOVA. Results indicated and further confirmed the presence of NCCS in the Northern talkers and the SVS in the Southern talkers. The Northerners, in particular, produced /ɑ/s that are lowered and fronted and /æ/s that are fronted and raised. Additionally, the Southern speech was found to have several new features: fronting of /æ/ among the male speakers and the raising of /u/ and /ʊ/. Further, the merger of /ɑ/ and /ɔ/ vowels was found to be robust in New England, Mid-Atlantic, Midland, and Western. Some new results were also found: the spreading of the high front vowel shifted from the Southern to the Midland dialect, Southern talkers were found to have raised the high back vowels /æ/, Southern male talkers had fronted /æ/ and the low back vowels

/ɑ/, /ɔ/, and /ʌ/ aligned with respect to F2 in the Mid-Atlantic dialect. Clopper et al. (2005)'s findings were able to indicate that a single set of idealized acoustic-phonetic baselines of "General" American English does not characterize the vowel systems of American English and provided benchmark for six regional varieties. Further, the study established the phenomenon of vowel merging in regional varieties, something that we see in our analysis of Assamese vowels and are discussed in later chapters.

Adank, Van Hout, & Smits (2004) studied 15 vowels of Standard Dutch as spoken in the Netherlands (Northern Standard Dutch(NSD)) and Belgium (Southern Standard Dutch(SSD)). 10 female and 10 male speakers each speaking NSD and SSD and were recorded producing the Dutch vowels (/ɑ, a, ɛ, ɪ, i, ɔ, u, ʏ, y, e, o, ø, ɛɪ, ɔu, œy/) in /sVs/ context. Apart from duration, both steady state formant frequencies and dynamic specifications of the vowels were analyzed. Results indicated several points of difference between the two varieties. Firstly, the duration of the diphthongs /ɛɪ, ɔu, œy/ were longer in case of SSD speakers. Secondly, although the two varieties differed little in the steady state characteristics of the nine monophthongal vowels, differences were seen in case of the three diphthongs and the three long mid vowels /e, o, ø/. For the diphthongal vowels /ɛɪ, ɔu, œy/, the female NSD talkers were found to diphthongize these vowels more as opposed to the female SSD talkers. Second, more diphthongization of the F1 and F2 of the long mid vowels was seen for NSD than for SSD (Adank, Van Hout, & Smits, 2004).

Similarly, Williams & Escudero (2014) in their acoustic study of British English compare the formant frequencies of vowels and find considerable variation in the vowels of Northern and Southern British English. The study analyzes 11 monophthongs /iː, ɪ, ɛ, ɜː, a, ɑː, ʌ, ɒ, ɔː, ʊ, uː/ and five diphthongs /eɪ, əʊ, aʊ, aɪ, ɔɪ/ produced by 17 speakers of Standard Southern British English (SSBE) and 19 speakers of Sheffield English (SE) dialects. Apart from vowel duration, the study analyzed the time varying nature of vowel formants known as 'vowel inherent spectral change'

(VISC) which has been defined by Nearey & Assmann (1986) as the "relatively slowly varying changes in formant frequencies associated with vowels themselves, even in the absence of consonantal contexts". The F1 and F2 trajectories were fitted with parametric curves using the discrete cosine transform (DCT). These were sampled at 30 equally spaced time points within the central 60% of each token. The formant trajectory means was represented by the zeroth DCT coefficient and the magnitude and direction of formant trajectory change to characterize VISC was represented by the first DCT cefficient. Results confirmed the impressionistic descriptions of the major vowel differences between Northern and Southern dialects. /ʌ/ was found to be a phonologically distinct vowel in SSBE but not in SE specifically exhibiting length, lower F1 and F2 and more diphthongization than SSBE /ʊ/. Secondly, SE /ɑː/ was found to be longer with lower F1 and F2 trajectories than SSBE. Further SSBE /ɒ, ɔː/ were phonetically more retracted and raised with lower mean F1 and F2 trajectories, and /ʊ, uː/ were more front with higher F2 trajectories. With respect to VISC, /uː/ was found to be more diphthongized in SSBE however what was most striking was that the F2 trajectory of the diphthong /əʊ/ proceeded in opposite directions in the two dialects demonstrating that the measures of directionality in VISC play an important role in the acoustic characterization of diphthongs, particularly when it comes to dialectal variation (Williams & Escudero, 2014).

These studies show that vowel features can be majorly accountable for variations among dialects. Variations in F1 and F2 measures of vowels affecting their distributions in the vowel space, vowel mergers and shifts leading to dialectal variations have been primary in these discussions and have largely corroborated impressionistic observations of variations in these languages. This prompts us to acoustically look into the impressionistic accounts of vowel variation and vowel merger in Assamese (Kakati, 1941; G. C. Goswami, 1982; G. C. Goswami & Tamuli, 2003) as mentioned in Sections 1.4 and 1.5. How these changes affect their behaviour within the vowel

space is also pertinent to our understanding of vowel systems. The following section briefly discusses the theories of vowel systems which will help us to understand the behaviour of Assamese vowels within the respective dialectal vowel spaces.

2.1.3 Theories of vowel systems

Languages differ not only with respect to vowel features but also with respect to the size of the vowel systems. While the most frequently occurring vowel systems are said to have five to seven vowels (J.-L. Schwartz et al., 1997), the simplest system, such as, Abkhaz (Hewitt, 1979), Kabardian (Wood, 1991) and Marshallese (J. D. Choi, 1992), contain only three vowels – [i, a, u] (also known as "vertical systems"), however, vowel systems with the same number of vowels may not always have the same vowels. Several theories of have been developed regarding the distribution of vowels with the vowel systems. Vaux & Samuels (2015) mentions that these theories seek to find answers to three main queries regarding vowel systems (i) why vowels fall into the particular perceptible/produceable spaces that they do, (ii) why some sounds are more basic than others, (iii) why some vertical vowel systems appear to violate these generalizations. These questions have their answers in two different theories:

(a) Prioritizing perceptual invariance, the Quantal theory (QT) K. N. Stevens (1972, 1989) claims that the vowels produced in the regions of acoustic stability or regions where variability in production has minimal acoustic impact (vowels /i/, /a/ and /u/) are highly preferred cross-linguistically.

(b) The Theory of Adaptive Dispersion (TAD) (Liljencrants & Lindblom, 1972) has its roots in the principles of "minimizing effort" and "maximizing contrast". This theory claims that the vowels /i/, /a/ and /u/ are at the extremes of the physiologically possible vowel space and so are maximally acoustically distinct. TAD further claims that the overall vowel space expands with the increase in the number of vowels.

The Quantal theory of speech is an attempt to explain that some speech sounds are

easier to produce than others and hence, are more common in languages. The basic assumption of the quantal theory of speech is that there are nonlinear relationships between articulation and acoustic output. That is, while certain relatively large changes in the position of the articulator will cause little change in the acoustic signal, other relatively small changes in the placement of the articulator will cause large changes in the acoustic signal. To what extent an acoustic change has happened appears to be related to the particular region of the vocal tract where the articulation is located. In certain critical regions, even a minor adjustment in the of placement of the articulation causes a quantal change in sound.

The theory of Adaptive Dispersion on the other hand offers an alternative explanation for the cross-linguistic preference for the vowels [i a u]. According to this theory, apart from being in potentially stable articulatory regions, these vowels are also the extremes of the physiologically possible vowel space, i.e., vowels are dispersed in the phonetic space in such a way as to maximize auditory differences among the vowels. Therefore, they are maximally acoustically distinct and are unlikely to be confused by a listener. As a corollary to this assumption the theory also predicts that languages that have relatively larger vowel inventories will have larger acoustic vowel spaces, than the ones that have smaller vowel inventories. Section 2.1.3.1 discusses the two theories in detail.

2.1.3.1 The Quantal Theory of Speech

The Quantal Theory of Speech was propounded by K. N. Stevens (1972, 1989). The vocal apparatus is able to articulatorily assume a large number of positions. However, among these large number of positions, preference is given to only a small number of "natural regions" of the articulatory space and these are actually used in phonological systems of languages. Quantal Theory proposes that speech sounds produced in these regions have certain quantal properties that seem to be good candidates for the

phonetic inventory in a language.

The theory examines the relation between:

a. vocal-tract configurations and the properties of the sounds that result from these articulations

b. acoustic parameters that are observed in speech and the auditory responses to the sounds described by these parameters.

Figure 2.4: Schematization of a hypothetical quantal relation between acoustic parameter and articulatory parameter (K. N. Stevens, 1989)

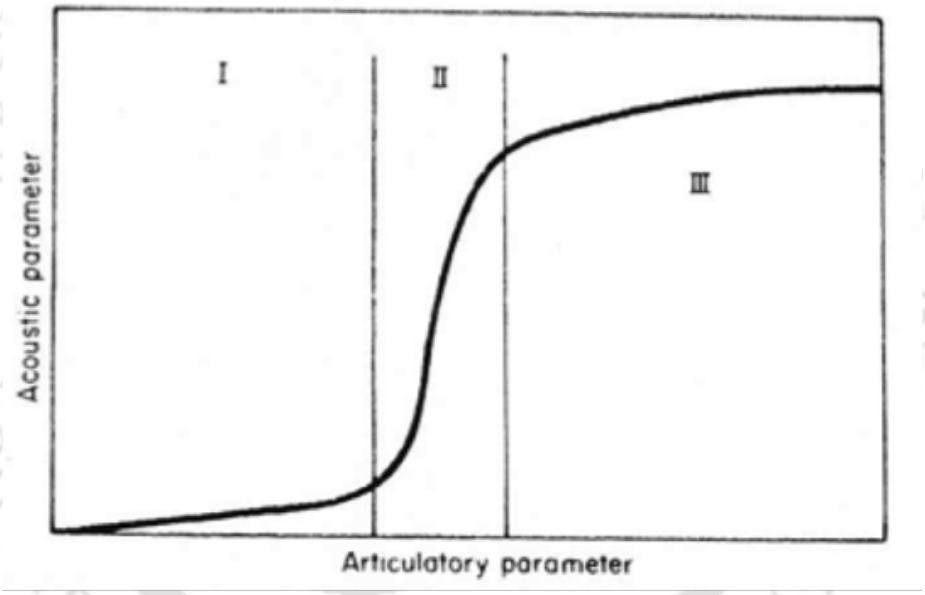


Figure 2.4 shows a hypothetical relation between some acoustic parameter in the sound and some articulatory parameter. Within regions I and III, acoustic parameter is relatively insensitive to changes in the articulatory parameter i.e, if there are perturbations in articulation at one of these natural regions, it would produce only small changes in the acoustic output. As a result the effect on perception will presumably be minimal. However, there are large acoustic differences between regions I and III and as the articulatory parameter moves through region II, the acoustic attribute often undergoes a qualitative change. This region is called the "threshold region". Stevens proposes that sound segments used in languages are selected from

a relatively small range of features. He claims that these configurations occur in precisely those regions which allow variability in articulation without an appreciable effect on the sound output. When applied to vowels, quantal theory has claimed to explain why the vowels [a], [i] and [u] are highly preferred cross-linguistically. There are regions of acoustic stability (quantal regions) where two formants converge.

[a] is produced at the convergence of F1 and F2, where the front and back cavities of the vocal tract are of approximately equal lengths.

[i] is produced at the convergence of F2 and F3, created by a constriction in the palatal region.

[u] is produced at a broad F2 minimum, where F2 is near a stable F1.

Quantal Theory has faced many criticisms. According to Carré (1996) quantal stability is not an intrinsic feature of vowels. Pointing out that the fourth most common vowel /ɛ/ is not as stable as predicted by Quantal theory, Carré claims that it is because the vowels /i a u/ represent the extreme acoustic capabilities of vocal tract production that they are stable. Secondly, Livijn (2000)'s survey of 28 differently-sized vowel inventories selected from the IRIS database (Cunningham-Andersson & Engstrand, 1988) showed no evidence for acoustically favoured hot spots. Euclidean distances (based on F1 and F2) were calculated between the point vowels /i/, /a/ and /u/ as a function of vowel inventory size. Results showed that in 4-8 vowel inventories the Euclidean distances do not to grow with inventory size. The largest systems (11 vowels or more) showed a fairly weak tendency to expand the acoustic vowel space. Second, on plotting the total vowel distribution of the 28 inventories in the F1-F2 space it was seen that the distribution is quite random. The location of point vowels varied to a great extent even in languages which had similar inventory sizes. This did not provide much evidence for 'hot spots' in these data. A third criticism comes from evidences from languages like Abkhaz (Hewitt, 1979), Kabardian (Wood, 1991) and Marshallese (J. D. Choi, 1992) which are vertical vowel systems and do not select

their members from the /i, a, u/ set as predicted by Quantal theory.

2.1.3.2 The Theory of Adaptive Dispersion (TAD)

Dispersion Theory (DT) (Liljencrants & Lindblom, 1972) aimed to predict how the vowel inventories of the world's languages are phonetically structured. Crucially, the DT evaluates the role of perceptual contrast in vowel systems. It posits that the vowels in the phonetic space of a given language are positioned in such a way as to make them highly contrastive. The Dispersion Theory of vowels makes two significant predictions regarding vowel dispersion in vowel systems:

a. the distinctive sounds of a language tend to be positioned in phonetic space so as to maximize perceptual contrast, i.e, maximal acoustic separation of vowels.

b. the overall vowel system expands with the increase in the number of vowels.

Based on these two predictions Liljencrants & Lindblom (1972) reports that the most preferred vowels in languages are the corner vowels /i, u, a/ and provides a comprehensive list of the most preferred and observed sets of vowels in languages starting from three-vowel systems to twelve-vowel systems. The authors used a computer model that maximized the distance between a given number of vowels within a formant space. This model was used to predict F1 and F2/F3 values, which were assigned to corresponding phonetic categories. These categories were then compared to known typology. Table 2.1 exemplifies the sets for an 8 vowel system (like that of Assamese):

Table 2.1: Eight-vowel systems as listed by Liljencrants & Lindblom (1972)

		i		ü/ʉ		ɨ		u						
Predicted			ɛ				ɔ							
				æ		a/ɑ								
	(a)	i		ü		ɨ		u	(d)	i		ɨ		u
		e		ö		a		o		e		ə		o
											a		ɔ	
	(a')	i		ü		ɨ		u	(e)	i		ɨ		u
			e		ö		o			e				o
					a					æ		a		ɔ
	(b)	i			ü			u	(f)	i				u
Observed		e			ö			o		e				o
			æ				a			ɛ				ɔ
										æ				a
	(c)	i		ü		(ɨ)		u	(g)	i				u
		e		ö		ə		o		e				o
						a				ɛ		ʌ		ɔ
												a		

When compared to Liljencrants & Lindblom (1972)'s vowel inventories, the eight-vowel system of Assamese as described in 1.5 looks more like the observed set (g) which was predicted for Portuguese dialects in Hockett (1955), the contrast being the absence of the vowel ʌ and the presence of ʊ in Assamese. J.-L. Schwartz et al. (1997) on the other hand, groups vowels systems into two groups based on the position of vowels within the system. A primary system is a set of vowels within which vowels occupy different positions in the vowel space. If two vowels share the same position but different diacritics, one of them is considered to be a part of the "primary" system

and the other as part of the "secondary" system. They studied 317 vowel inventories reported in UPSID (UCLA Phonological Segment Inventory Database, (Maddieson & Disner, 1984)) and pointed out that 6 vowel and 7 vowel inventories are more common (20% and 14% of the 317 vowel inventories). Of the 317 vowel inventories studied, two-thirds of the primary systems have between 5 and 7 vowels.

Studies of vowel systems across languages have both proven as well as contradicted the predictions of TAD. The first assumption of TAD is that languages that the acoustic spaces of larger vowel inventories will be larger in relation to languages with smaller vowel inventories. Supporting this assumption, Jongman et al. (1989) found that the vowel spaces of English and German, which have 11 and 14 monophthongs, respectively, are more crowded than Greek, which has 5 monophthongs. All the three languages have the vowels /i/, /a/, and /u/, and they occur in similar positions in the languages' vowel spaces. However, the German and English vowels were more peripheral than the Greek vowels i.e, they allowed maximum vocalic acoustic space while minimizing the space contour. Similarly, the vowel spaces of Moroccan Arabic, Jordanian Arabic and French with 5, 8 and 11 vowels respectively were compared using Euclidean distances between point vowels [i, a, u] in an F1 x F2 Bark space (Al-Tamimi & Ferragne, 2005). The comparisons were done in three conditions: isolation, in syllable and in words. It was found that in all conditions French has the largest vowel space followed by Jordanian Arabic. Moroccan Arabic has the smallest vowel space size. In a similar study by Bradlow (1995) it was found that when the vowels were produced in a closed-syllable context, English with 11 vowels had a larger vowel space relative to Spanish with 5 vowels.

The studies discussed above appear to support the hypothesis by TAD that larger vowel inventories have larger vowel spaces. However, there are some studies that seem to reject the hypothesis. Livijn (2000) compared the differently-sized vowel inventories of twenty-eight languages. It was found that only in languages with 11 or more vowels

the distances between them were expanded. Gendrot & Adda-Decker (2007) compared eight languages with differently-sized vowel inventories (English (11), French (10), German (15), Italian (5), Arabic (6), Mandarin Chinese (6), Portuguese (9) and Spanish (5)).The study found that languages with larger vowel inventories did not have respectively expanded vowel spaces.

The second assumption of the TAD is that the vowels in a language's inventory will be maximally dispersed throughout the vowel space. Again, there have been conflicting reports regarding this assumption. Disner (1984) reported that of the 317 vowel inventories reported in UPSID (UCLA Phonological Segment Inventory Database, (Maddieson & Disner, 1984), 96% of the vowel inventories had vowels that were evenly dispersed along the boundaries of the acoustic vowel space. Contradicting this finding is Disner (1983) report that states that the nine vowels of Swedish and ten vowels of Danish are not thoroughly dispersed but are crowded together in their vowel spaces. Similarly, the study of three closely related languages Yoruba, Ghotuo and Edo with seven vowels each (Lindau & Wood, 1977) report that while Ghotua and Edo have evenly dispersed vowels in their vowel spaces, Yoruba does not. Recasens & Espinosa (2006)'s study of three Catalan dialects and a fourth vowel system (Majorcan), all with seven vowels each, also shows comparable dispersion of vowels of the three Catalan dialects across their respective vowel spaces. The vowel space of Majorcan was comparatively larger. However, they found that intervocalic distances varied according to dialect and vowel pair, thus contradicting the TAD prediction that adjacent vowels will be evenly spaced in identical vowel systems.

As mentioned earlier, studies of the acoustic characteristics of vowels in languages have largely dealt with the formant frequencies of vowels. These variations in formant frequencies have also led researchers to studies determining the effects of such variations on the "working" vowel space (in the F1-F2 plane) used by the speakers (Jacewicz et al., 2007). Differences in the size of vowel space area provide insight into

the ongoing vowel chain shifts in languages. Jacewicz et al. (2007) mention two possible outcomes: First, given the significant differences in vowel systems across regional dialects, one might find that this type of variation also affects the size of the vowel space and will differ across dialects. Alternatively, the area of the vowel space may remain the same cross-dialectically and only the relative positions of vowels within the vowel space contribute to the dialectal differences (Jacewicz et al., 2007).

Preliminary analysis of Assamese has shown that there are noticeable differences in vowel formant frequencies, rhythm and speech rate between varieties of Assamese, mainly between the larger Eastern Assamese and Western Assamese varieties. While Eastern Assamese has shown to reduce the eight-vowel set of Assamese to a six-vowel one, Western Assamese varieties have more inconsistent properties with a tendency to reduce it to a six or seven-vowel one (Dihingia & Sarmah, 2017a,b). Such variations in formant frequencies affect the vowel space used by the speakers with two possible outcomes: the size of the vowel space differing cross-dialectically or, the area being same across dialects, only the positions of vowels within the vowel space varying across dialects (Jacewicz et al., 2007). It would be interesting to see how vowels in Assamese dialects behave and what positions they take in their respective vowel spaces.

2.2 Acoustic features of vowels in Assamese

The current study aims to look at the Assamese dialectal variation in five regions from across Assam. This section discusses the results of the production tests done for vowels as spoken in 5 regional varieties of Assam: Tinsukia, Jorhat, Nagaon, Nalbari and Barpeta. In Section 2.2.1, the methodology including the materials and data collection and the acoustic and statistical measures employed for the study have been discussed. Section 2.2.2 discusses the results of the analysis providing a detailed analysis of the acoustic properties of vowels in each geographical region. Section 2.4

concludes the chapter.

2.2.1 Methodology

2.2.1.1 Materials and data collection

The data list for the current study contained the Assamese vowels /i, e, ɛ, a, ɔ, o, ʊ, u/ reported in Mahanta (2012). There were two instances of data collection or this study. A preliminary data collection was done in the beginning from speakers of Jorhat and Nalbari varieties. In this instance, both Jorhat and Nalbari speakers were recorded producing the words in Table 2.2[2]. For this preliminary analysis, the datalist was taken from Mahanta (2012). Eleven speakers of the Jorhat Assamese variety (4 female and 7 male), and ten speakers of the Nalbari variety (3 female and 7 male) were recorded in the first instance of data collection. A detailed list of the participants is provided in Table A.1 of Appendix A. The vowel minimal set was presented to the participants on a computer screen on a slide display application in randomized order. For presenting each word, three different types of slides were used. The first slide contained the English meaning of the target Assamese word, the second slide contained the target Assamese word in Assamese script and finally a third slide where the target word was embedded in a sentence frame- /mɔi tat X buli lik^{h}a dek^{h}ilõ/ ('I saw X written there'). The speakers were instructed to read out the contents of the second and third slides on the microphone, while the contents of the first slide was used for priming the speaker for the meaning of the word. Sets with each word occurred three times in the experiment in random order and the participants were

[2]As Mahanta (2012) clarifies, there are harmonic restrictions on the occurrences of /e/ and /o/ and therefore the word /beli/ has been incorporated in the list. [e] and [o] mostly appear only when /i/ and /u/ occur in the following syllable. However, [e] and [o] appear in borrowed items from English, Hindi, and those which have been directly incorporated from Sanskrit. Therefore, for the second data collection, the loanword /bel/ 'bell' has been used as it is widely used in Assamese and it makes the minimal set more congruent. Although the occurrence of the /e/ vowel in Assamese is somewhat limited, however it is not non-existent. While at least 13 monosyllabic, non-derived, CVC words with the vowel /e/ were found in Assamese, /e/ and /ɛ/ minimal pairs were limited. /bes/ 'fine' ∼ /bɛs/ 'sell' and /khed/ 'regret' ∼ /k^{h}ɛd/ 'drive away' are two such minimal pairs.

allowed to decide the pace of slide change. These recordings were conducted in a sound attenuated recording booth in a lab using an Audio-Technica AT 2020 USB microphone connected to a laptop computer.

Table 2.2: A sample of the words recorded with meanings. (Mahanta, 2012)

Assamese word	Meaning
/bal/	male child/ hair
/beli/	sun
/bɛl/	wood apple/a stupid person
/bil/	swamp
/bol/	let's go
/bɔl/	strength
/bul/	a person's name
/bʊl/	color

The second instance of data collection was conducted in the field and data was collected from speakers of all the five geographical areas. A detailed list of the participants is provided in Table A.2 of Appendix A. For this data collection an additional set of 65 CVC words with different onset and coda consonants was included to the previous list containing the 8 Assamese vowels. The complete list of words is provided in Table B.1 in Appendix B. The entire list of words (73 in total) was presented to the participants in written form (Assamese script) on a sheet of paper in the following manner:

a. Natural sentence: The target word was presented in a sentence in a way that it conveyed the correct meaning of the word and avoided ambiguity in meaning;

b. Isolation: The target word appeared in isolation;

c. Sentence frame: The target word was presented in a sentence frame /mɔi X buli kolõ/ ('I X said') that was kept consistent for all words in the list.

This method was adopted to ensure that the natural sentence that carried the target word would enable the speaker to identify the meaning of the target word and retrieve the correct word from their memory. Each set was repeated two times

randomly. For the analysis of vowels in this study, the target words that appeared only in isolation and in sentence frames were considered. In case of mispronunciation or wrong pronunciation of the target words, the participants were asked to repeat the whole set (i.e, the parts a, b and c for a target word). However, even after additional repetitions there were mispronunciations and wrong pronunciations by some speakers. These tokens were later identified and removed from the final data sheet. The recordings were done in a noiseless environment. The microphone used was a Shure unidirectional head-worn microphone which was connected to a Tascam linear PCM recorder via xlr jack. The sampling frequency was kept at 44.1 kHz and 24 bit in .wav format.

2.2.1.2 Acoustic and statistical measures

For vowel measurements, a Praat script using the Burg algorithm was used to calculate the first three formants (F1, F2 and F3) in Hertz at the mid 20% of the vowel duration. The values were then normalized with Lobanov method for speaker effects using NORM (Thomas, 2000). The z-score method of frequency normalization used in Lobanov (1971) for vowel normalization has been found to be the best vowel normalization metric in a study by Adank, Smits, & Van Hout (2004). Lobanov's z-score transform is a normalization procedure that is vowel extrinsic and talker intrinsic. This reduces anatomical variation of the formant frequencies but maintains phonological and sociolinguistic variation.

$$F_{n[V]}{}^{N} = (F_{n[V]} - MEAN_n)/S_n \tag{2.1}$$

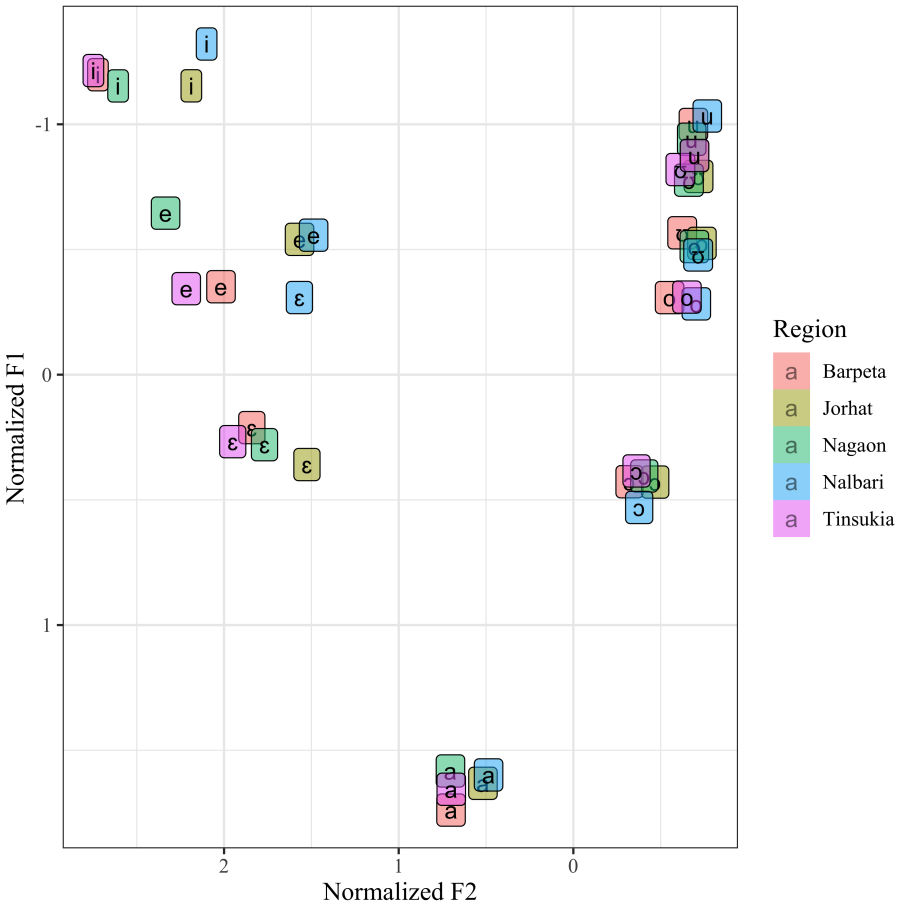
where $F_{n[V]}{}^{N}$ is the normalized value for $F_{n[V]}$ (i.e., for formant n of vowel V). $MEAN_n$ is the mean value for formant n for the speaker in question and S_n is the standard deviation for the speaker's formant n.

Prior to further analysis the data was checked for outliers and recording errors. Plots of vowel group means with 1 standard deviation ellipses were also generated on F1-F2 space separately for male and female speakers. A formant frequency plot gives a visual representation of the vowels in a vowel space. While the positions of the vowels in the vowel space are shown by the formant frequency means, the ellipses show the standard deviation of their distribution within a vowel space. All statistical analyses are conducted using the open source R platform (R Core Team, 2019). The *lme4* package on R (Bates et al., 2015) is used to built the LME models. The models are subjected to Type II Wald chisquare tests for analysis of deviance using the *car* package. Each model is subjected to a post-hoc Bonferroni test using the *emmeans* package on R (J. Fox & Weisberg, 2019; Lenth, 2019) to see the pairwise variability of vowel formants.

2.2.2 Results

In order to see the average vowel characteristics of the five Assamese varieties, a vowel plot showing category means of normalized F1-F2 values is plotted using the *ggplot2* package on the open source R platform (Wickham, 2016; R Core Team, 2019). The plot as seen in Figure 2.5 shows normalized F1 and F2 values for five groups of speakers. This plot is generated from 14289 tokens from the five Assamese varieties. As seen in the figure, the vowels /i, a, ɔ/ in all varieties appear as distinct vowels. However, vowels /e, ɛ, o, ʊ, u/ overlap with each other in varying degrees in the five varieties. In order to see if these mergers are statistically significant or not, various statistical and acoustic experiments have been resorted to.

Figure 2.5: Lobanov normalized formant frequency plots for 8 Assamese vowels by Tinsukia, Jorhat, Nagaon, Nalbari and Barpeta speakers



The data for statistical analysis comes from 14289 tokens produced by the speakers of the five varieties with all the eight vowels of standard Assamese. Separate LME models are constructed with F1, F2 and F3 as dependent variables and vowel type, region and the interaction between vowel and region as fixed effects. Even though Assamese vowels do not contrast in roundedness, apart from F1 and F2 we also considered F3 as a dependent variable. Speaker, gender, coda type, onset type and context are considered random effects in the full model. Once the LME modeling was complete for F1, F2 and F3, the models were subjected to a backwards reduction process using the *step* function of the *lmerTest* package Kuznetsova et al. (2017). In case of all the three models, gender was eliminated. Hence, we built three reduced

models for F1, F2 and 3, with vowel type, region and the interaction between vowel and region as fixed effects and speaker, coda type, onset type and context as random effects. Each reduced model is subjected to a post hoc Bonferroni test to see the pairwise variability of vowel formants using the *emmeans* package on R Lenth (2019). The three models were also subjected to a Wald χ^2 test using the *Anova* function of the *car* package on R. The results obtained are presented in Table 2.3. We notice that the fixed factors, namely vowel type, region and the interaction of vowel and region have a significant effect on F1, F2 and F3 of vowels. The significant effect of vowel type and region indicates that the formant frequencies of vowels are significantly different both across vowels and regions. Further, the significant interaction of vowel and region indicates that the variations in vowel formant frequencies are not uniform across the five regions.

Table 2.3: Wald χ^2 tests on the LME model for F1, F2 and F3

	Fixed Effect	df	χ^2	p-value
F1	Vowel	7	82604.61	< 0.001
	Region	4	17.40	< 0.01
	Vowel*Region	28	1720.25	< 0.001
F2	Vowel	7	99160.59	< 0.001
	Region	4	58.96	< 0.001
	Vowel*Region	28	600.40	< 0.001
F3	Vowel	7	4364.19	< 0.001
	Region	4	24.24	< 0.001
	Vowel*Region	28	335.34	< 0.001

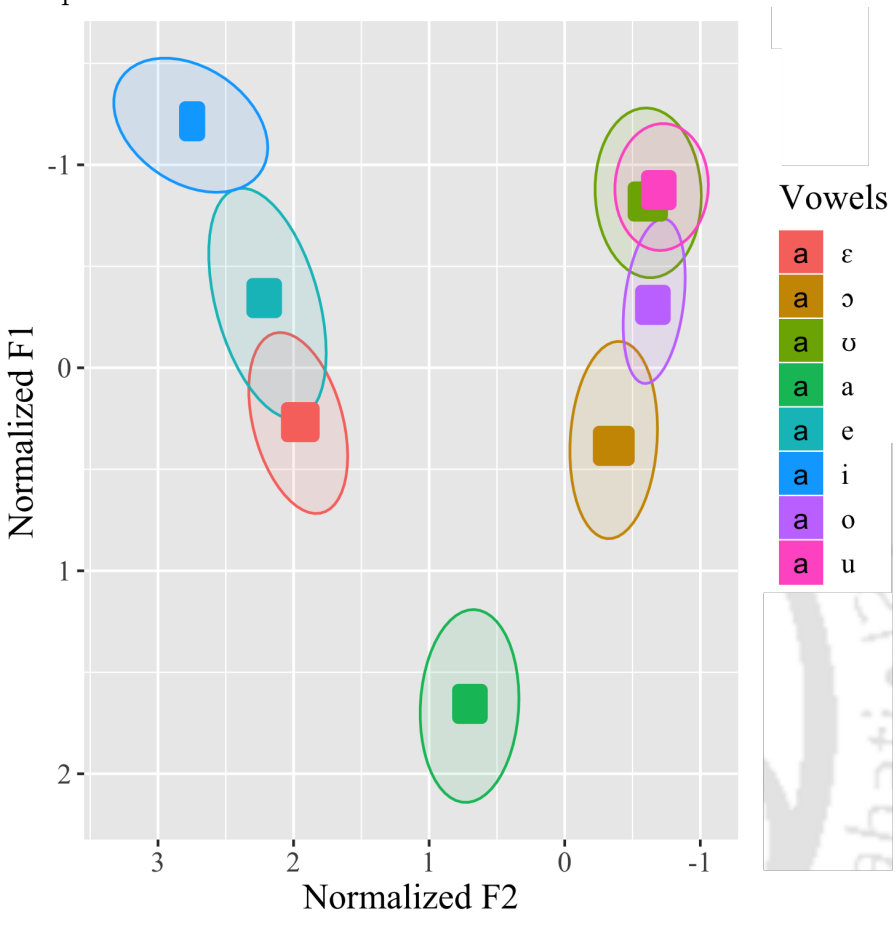
The results obtained from the statistical tests prompt us to look into the formant characteristics for the five regions separately. Hence, in the following sections discuss the formant features of the vowels of each of the five geographical varieties based

on the acoustic and statistical tests conducted. The sections describe the acoustic characteristics of the vowels in the respective regions and present the corresponding plots on F1-F2 planes. Each section first presents an impressionistic overview of the vowels in the regions based on the vowel plots plotted separately for gender and context types. The sections also discuss the average formant frequencies of the vowels for each of the categories. Subsequently, explorative statistics on the vowel measures are conducted for each region, using LME models, chi-square tests and post-hoc tests.

2.2.2.1 Vowel formants in Tinsukia variety

This section describes the acoustic characteristics of the vowel formants in the Tinsukia Assamese variety. Figure 2.6 presents the corresponding vowel plots on an F1-F2 plane. Figures 2.7, 2.8, 2.9 and 2.10, present the vowel plots on F1-F2 plane separately for gender and context types: Tinsukia females in isolation (tfi), Tinsukia males in isolation (tmi), Tinsukia females in sentence frames (tfs) and Tinsukia males in sentence frames (tms).

Figure 2.6: Normalized formant frequency plots for Tinsukia speakers with 1 standard deviation ellipses.



As seen in Figure 2.6, the vowels that are completely distinct with no overlapping ellipses are /i/ and /a/. Some amount of overlap between the vowels /e/ and /ɛ/ has been seen. Similarly, /o/ shows some overlap with /ɔ/, /u/ and /ʊ/. The vowels /ʊ/ and /u/ on the other hand, show complete overlap of the ellipses, indicating a possible merger of these two vowels. Figures 2.7, 2.8, 2.9 and 2.10 present the vowels plots for gender and context types. As seen in the figures, the pattern of overlapping is similar in each category (tfi, tmi, tfs and tms) and corresponds to the vowel plot 2.6. From the visual inspection of the plots we conclude that,

a. there is significant overlap of the vowels /u/ and /ʊ/;

b. while vowel overlaps are seen in all categories, these overlaps are not accidental

since they are uniform across context and gender;

c. even though the overlaps are fairly robust visually, this does not confirm that there is statistical significance of the overlaps.

Hence, in order to analyze these overlaps more conclusively, further statistical tests are conducted.

Figure 2.7: Normalized formant frequency plots for Tinsukia female speakers with 1 standard deviation ellipses in isolation.

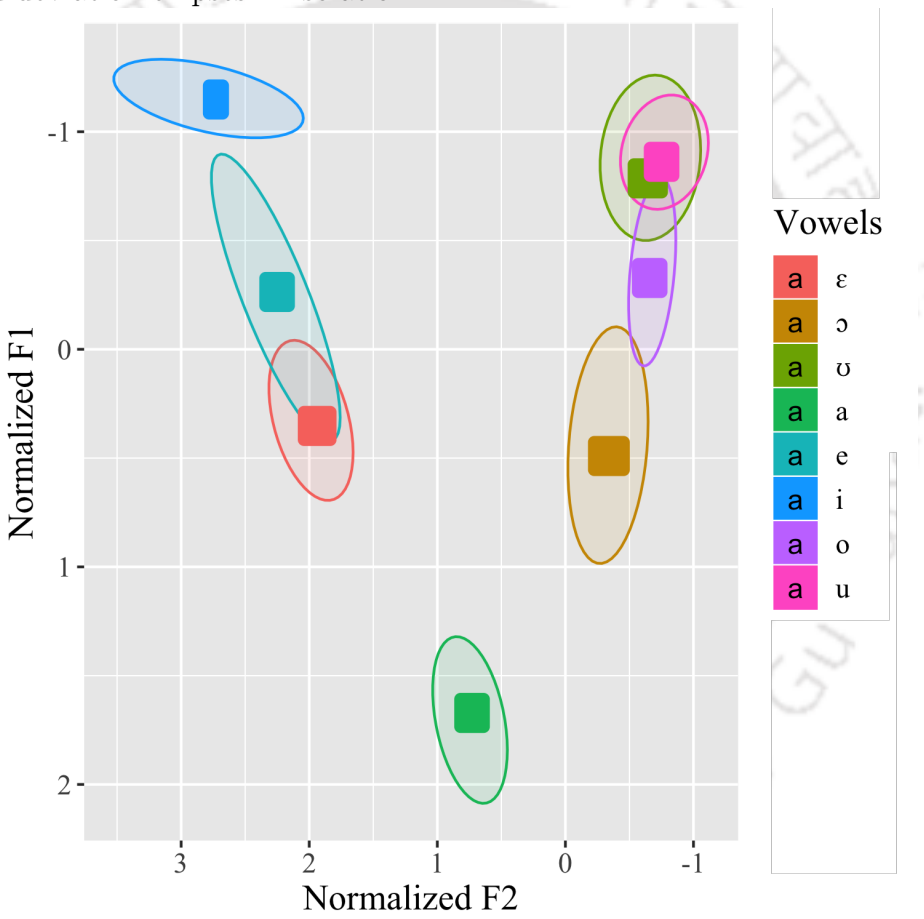


Figure 2.8: Normalized formant frequency plots for Tinsukia male speakers with 1 standard deviation ellipses in isolation.

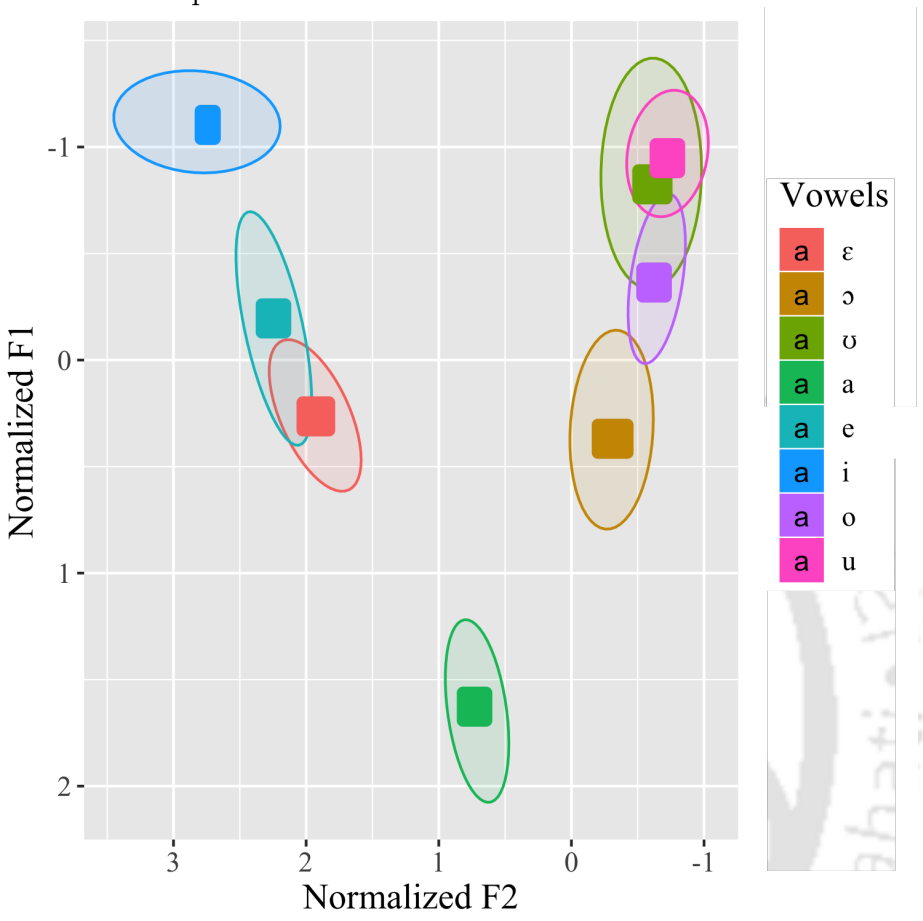


Figure 2.9: Normalized formant frequency plots for Tinsukia female speakers with 1 standard deviation ellipses in sentence frames.

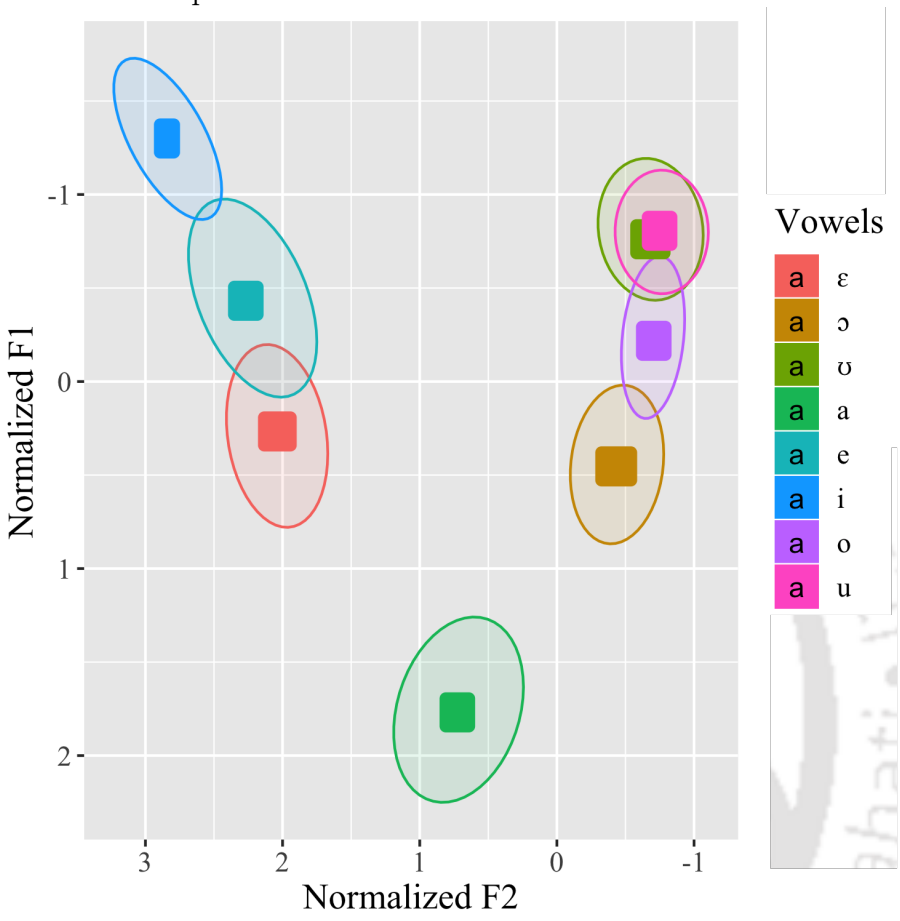


Figure 2.10: Normalized formant frequency plots for Tinsukia male speakers with 1 standard deviation ellipses in sentence frames.

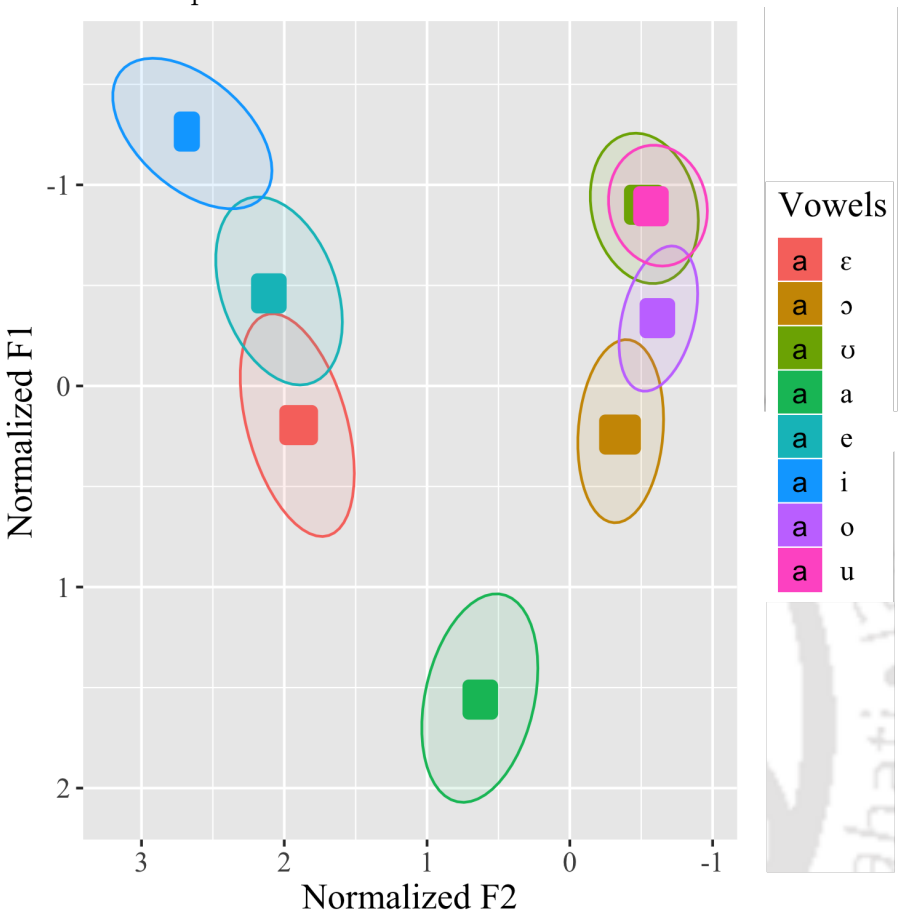


Table 2.4: Lobanov normalized and scaled vowel formant values for Tinsukia speakers in words produced in two contexts

Context	Vowel	Female		Male	
		F1	F2	F1	F2
Isolation	a	521 (35)	1415 (50)	507 (44)	1368 (72)
	e	406 (20)	1675 (152)	404 (25)	1636 (116)
	ɛ	447 (25)	1643 (95)	424 (25)	1560 (127)
	i	364 (16)	1770 (165)	356 (15)	1760 (112)
	o	405 (23)	1160 (43)	419 (29)	1151 (52)
	ɔ	452 (33)	1222 (51)	444 (41)	1208 (57)
	u	378 (24)	1145 (71)	387 (40)	1181 (148)
	ʊ	378 (25)	1161 (76)	392 (51)	1228 (206)
Sentences	a	523 (36)	1405 (51)	502 (52)	1356 (107)
	e	416 (22)	1712 (53)	406 (13)	1623 (109)
	ɛ	446 (23)	1627 (121)	427 (33)	1576 (104)
	i	362 (23)	1805 (112)	368 (15)	1682 (122)
	o	407 (20)	1156 (47)	414 (24)	1154 (51)
	ɔ	450 (24)	1219 (50)	438 (44)	1196 (46)
	u	376 (28)	1148 (61)	391 (42)	1190 (156)
	ʊ	381 (31)	1164 (76)	395 (54)	1250 (226)

The mean Lobanov normalized and scaled F1 and F2 values for all the vowels produced by the Tinsukia speakers are presented in Table 2.4. In order to analyze vowel formant frequencies more conclusively, three separate full LME models were constructed for normalized F1, F2 and F3. In all the three models, vowel type, context and gender were the fixed effects. Speaker, onset type and coda type were the random effects. The significance of the following interactions were also explored:

vowel type * gender, vowel type * context, gender * context and vowel type * gender * context. The full LME models were subjected to backward reduction of effects using the *step* function of the *lmerTest* package on R (Kuznetsova et al., 2017). The reduced models for F1, F2 and F3 were determined as follows.

- F1: vowel + gender + context + (1|speaker) + (1|onset) + (1|coda) + vowel:gender + gender:context
- F2: vowel + gender + context + (1|onset) + (1|coda) + vowel:gender + vowel:context + vowel:gender:context + gender:context
- F3: vowel + gender + (1|onset) + (1|coda) + vowel:gender

Table 2.5: Wald χ^2 tests on the LME model for F1, F2 and F3 for Tinsukia variety.

	F1	F2	F3
Vowel type	15353.6***	23122.5 ***	795.9 ***
Gender	1.3 $^{n.s.}$	0.2 $^{n.s.}$	0.1 $^{n.s.}$
Context	71.3 ***	14.2 ***	NA
Vowel type x Gender	22.3 **	29.2 ***	35.2 ***
Vowel type x Context	NA	45.0***	NA
Gender x Context	8.0 **	3.0 $^{n.s.}$	NA
Vowel type x Gender x Context	NA	25.1 ***	NA

p-values: ***: $p < 0.001$, **: $p < 0.01$, *: $p < 0.05$, $^{n.s.}$: Not significant

The reduced models were subjected to a Wald χ^2 test and the results obtained are presented in Table 2.5. As seen in the table, vowel type has significant effect on F1, F2 and F3 of the vowels. Context effect is significant on F1 and F2, however, vowel type and context interaction is significant only for F2. Similarly, interaction of vowel type, gender and context is also significant for F2. Considering this, we decided to

explore the effects of the variables by context types. Hence, LME models for F1, F2 and F3 were built separately for vowels in isolation and vowels in sentence frames.

For each context namely, isolation and sentence, three full LME models were built with F1 , F2 and F3 as dependent variables, and vowel type and gender as the fixed effects. Speaker, onset type, coda type were the random effects. The interaction between vowel type and gender was also explored. The full LME models were subjected to backward reduction of effects using the *step* function of the *lmerTest* package on R Kuznetsova et al. (2017). The reduced model for F1, F2 and F3 in isolation and in sentence form are as follows.

- Isolation
  - F1: vowel + (1|speaker) + (1|onset) + (1|coda)
  - F2: vowel + (1|onset) + (1|coda)
  - F3: vowel + gender + (1|speaker) + (1|onset) + vowel:gender
- Sentence
  - F1: vowel + gender + (1|speaker) + (1|onset) + (1|coda) + vowel:gender
  - F2: vowel + gender + (1|onset) + (1|coda) + vowel:gender
  - F3: vowel + gender + (1|speaker) + (1|onset) + (1|coda) + vowel:gender

The reduced models were then subjected to a post-hoc Bonferroni test to see the pairwise variability of vowel formants using the *emmeans* function. The results of the post-hoc Bonferroni tests for vowels produced in isolation is presented in Table 2.6 and those produced in sentence frames is presented in Table 2.7. The results of F3 are not included in the two tables as F3 is not a distinguishing feature for Assamese vowels. However, the results for F3 are included in Table C.1.

Table 2.6: Results of Bonferroni post-hoc tests conducted on vowels produced in isolation by Tinsukia speakers.

Contrasts	F1					F2				
	estimate	SE	df	t-ratio	p-value	estimate	SE	df	t-ratio	p-value
a-e	1.89	0.08	1174	22.29	< .0001	-1.64	0.07	1180	-22.52	< .0001
a-ɛ	1.42	0.05	1148	30.96	< .0001	-1.33	0.04	1174	-33.68	< .0001
a-i	2.84	0.07	1175	42.86	< .0001	-2.15	0.06	1185	-37.91	< .0001
a-o	1.91	0.05	1137	37.61	< .0001	1.26	0.04	1145	28.91	< .0001
a-ɔ	1.32	0.03	1030	38.72	< .0001	1.1	0.03	1093	37.42	< .0001
a-u	2.6	0.04	871	71.69	< .0001	1.44	0.03	1026	45.96	< .0001
a-ʊ	2.54	0.03	1077	74.67	< .0001	1.36	0.03	1136	46.46	< .0001
e-ɛ	-0.47	0.09	1168	-5.44	< .0001	0.31	0.07	1177	4.2	0.0008
e-i	0.95	0.09	1161	10.11	< .0001	-0.51	0.08	1171	-6.4	< .0001
e-o	0.02	0.09	1170	0.21	1	2.9	0.08	1177	38.44	< .0001
e-ɔ	-0.57	0.08	1166	-6.92	< .0001	2.74	0.07	1176	38.7	< .0001
e-u	0.71	0.08	1170	8.57	< .0001	3.08	0.07	1179	43.27	< .0001
e-ʊ	0.65	0.08	1166	7.87	< .0001	3	0.07	1176	42.41	< .0001
ɛ-i	1.41	0.07	1168	20.77	< .0001	-0.82	0.06	1177	-14.1	< .0001
ɛ-o	0.49	0.06	1154	8.35	< .0001	2.59	0.05	1141	51.91	< .0001
ɛ-ɔ	-0.1	0.04	1170	-2.35	0.5368	2.43	0.04	1181	64.73	< .0001
ɛ-u	1.18	0.05	1087	26.01	< .0001	2.77	0.04	1149	71.04	< .0001
ɛ-ʊ	1.12	0.04	1168	25.93	< .0001	2.69	0.04	1185	72.76	< .0001
i-o	-0.93	0.07	1175	-12.8	< .0001	3.41	0.06	1177	54.84	< .0001
i-ɔ	-1.52	0.06	1172	-23.81	< .0001	3.25	0.05	1182	59.46	< .0001
i-u	-0.24	0.07	1170	-3.64	0.008	3.59	0.06	1185	64.33	< .0001
i-ʊ	-0.3	0.06	1173	-4.69	0.0001	3.51	0.05	1182	64.34	< .0001
o-ɔ	-0.59	0.05	1168	-12.02	< .0001	-0.16	0.04	1169	-3.85	0.0035
o-u	0.69	0.05	1173	14.69	< .0001	0.18	0.04	1184	4.43	0.0003
o-ʊ	0.63	0.05	1167	12.72	< .0001	0.1	0.04	1156	2.38	0.4959
ɔ-u	1.28	0.03	1087	40.24	< .0001	0.34	0.03	1155	12.44	< .0001
ɔ-ʊ	1.22	0.03	1171	42.18	< .0001	0.26	0.02	1179	10.59	< .0001
u-ʊ	-0.06	0.03	1069	-1.9	1	-0.08	0.03	1151	-2.79	0.1482

Table 2.7: Results of Bonferroni post-hoc tests conducted on vowels produced in sentence frames by Tinsukia speakers.

Contrasts	F1					F2				
	estimate	SE	df	t-ratio	p-value	estimate	SE	df	t-ratio	p-value
a-e	1.76	0.09	1201	20.63	< .0001	-1.63	0.07	1212	-24.43	< .0001
a-ɛ	1.4	0.05	1024	30.72	< .0001	-1.32	0.04	1209	-36.98	< .0001
a-i	2.65	0.07	1161	40.06	< .0001	-2.13	0.05	1215	-40.95	< .0001
a-o	1.89	0.05	1040	37.39	< .0001	1.16	0.04	1188	29.19	< .0001
a-ɔ	1.33	0.03	856	39.32	< .0001	0.97	0.03	1149	36.36	< .0001
a-u	2.54	0.04	588	71.53	< .0001	1.28	0.03	1074	44.8	< .0001
a-ʊ	2.48	0.03	957	73.99	< .0001	1.21	0.03	1182	45.4	< .0001
e-ɛ	-0.36	0.09	1197	-4.2	0.0008	0.31	0.07	1206	4.56	0.0002
e-i	0.89	0.1	1192	9.43	< .0001	-0.5	0.07	1201	-6.68	< .0001
e-o	0.13	0.09	1200	1.51	1	2.8	0.07	1208	40.4	< .0001
e-ɔ	-0.43	0.08	1194	-5.2	< .0001	2.61	0.07	1205	40.1	< .0001
e-u	0.78	0.08	1198	9.32	< .0001	2.91	0.07	1208	44.55	< .0001
e-ʊ	0.72	0.08	1194	8.68	< .0001	2.84	0.07	1205	43.69	< .0001
ɛ-i	1.26	0.07	1198	18.39	< .0001	-0.81	0.05	1206	-15.04	< .0001
ɛ-o	0.5	0.06	1080	8.59	< .0001	2.49	0.05	1190	54.61	< .0001
ɛ-ɔ	-0.07	0.04	1135	-1.56	1	2.3	0.03	1215	67.27	< .0001
ɛ-u	1.14	0.04	889	25.49	< .0001	2.6	0.04	1186	73.2	< .0001
ɛ-ʊ	1.09	0.04	1132	25.41	< .0001	2.53	0.03	1215	75.38	< .0001
i-o	-0.76	0.07	1175	-10.43	< .0001	3.29	0.06	1212	57.51	< .0001
i-ɔ	-1.33	0.06	1193	-20.69	< .0001	3.1	0.05	1211	61.73	< .0001
i-u	-0.12	0.07	1153	-1.81	1	3.41	0.05	1215	66.37	< .0001
i-ʊ	-0.17	0.06	1193	-2.71	0.1902	3.34	0.05	1211	66.52	< .0001
o-ɔ	-0.57	0.05	1119	-11.51	< .0001	-0.19	0.04	1205	-4.91	< .0001
o-u	0.64	0.05	1174	13.63	< .0001	0.11	0.04	1214	3.07	0.0616
o-ʊ	0.59	0.05	1122	11.84	< .0001	0.05	0.04	1197	1.16	1
ɔ-u	1.21	0.03	971	38.17	< .0001	0.3	0.03	1189	12.02	< .0001
ɔ-ʊ	1.15	0.03	1201	39.47	< .0001	0.23	0.02	1208	10.25	< .0001
u-ʊ	-0.06	0.03	873	-1.7	1	-0.07	0.03	1181	-2.65	0.2312

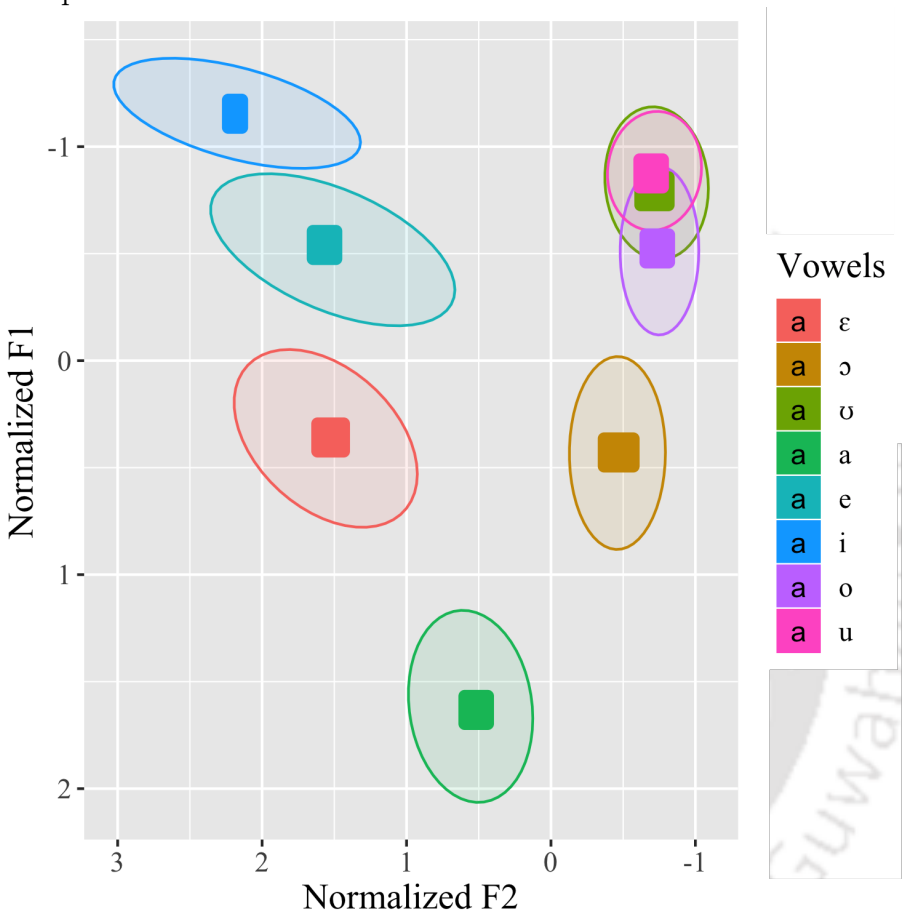
As seen in Tables 2.6 and 2.7, for vowels produced in isolation, post-hoc Bonferroni tests revealed significant F1 differences between all vowels except the following sets: /e-o/ and /u-ʊ/ indicating that the vowels in these pairs did not differ in height. Both /e/ and /o/ are higher-mid vowels in Assamese according to Mahanta (2012) and as reported in Table 1.6. Therefore, the significant similarity in F1 is predictable. However, /u/ and /ʊ/ vowels have been reported as high and near-high respectively. Hence, ideally the two vowels should be significantly different in F1. Significant F2 differences were observed between all vowels except, /o-ʊ/ and /u-ʊ/ indicating that the vowels in the pairs did not differ in backness. Similarly, in case of vowels produced in sentence frames, pots-hoc Bonferroni tests revealed significant F1 differences between all vowels except /e-o/, /ɛ-ɔ/, /i-u/, /i-ʊ/ and /u-ʊ/. Significant F2 differences were observed between all vowels except /o-u/, /o-ʊ/ and /u-ʊ/. The results show that in both contexts, F1 and F2 values of the vowels /u/ and /ʊ/ are not significantly different. Visual inspection of the plot in Figure 2.6 indicated complete overlap of the ellipses of these two vowels. Results of statistical tests show that the two vowels differ neither in height nor in backness, i.e, the positions that the two vowels occupy in the vowel space are not significantly different, thus providing robust evidence of the possibility of a merger between /u/ and /ʊ/ in the Tinsukia variety. However, whether these results actually indicate a merger or not can be confirmed only after testing them for merger measures which we have dealt in detail in Chapter 3 of this dissertation.

2.2.2.2 Vowel formants in Jorhat variety

This section describes the acoustic characteristics of the vowel formants in the Jorhat variety of Assamese. Figure 2.11 presents the corresponding vowel plots on an F1-F2 plane. The vowel plots for gender and context types i.e., Jorhat females in isolation (jfi), Jorhat males in isolation (jmi), Jorhat females in sentence frames (jfs) and Jorhat

males in sentence frames (jms) are presented on F1-F2 planes separately in Figures 2.12, 2.13, 2.14 and 2.15.

Figure 2.11: Normalized formant frequency plots for Jorhat speakers with 1 standard deviation ellipses.



From a visual examination of the plots, vowels /i, e, ɛ, a, ɔ/ appear to be distinct in each category. However, the vowel ellipses of /ʊ, u, o/ show overlap for both genders and contexts. No overlap was observed in case of the front vowels. An impressionistic view of the vowel plots conclude that:

a. there is significant overlap of the vowels /u/ and /ʊ/ with a partial overlap of /o/ with /u/ and /ʊ/;

b. while vowel overlaps are seen in all categories, the uniformity of the overlaps across context and gender confirm that they are not accidental;

c. the robust visual overlaps however, do not confirm their statistical significance.

Hence, further statistical tests are conducted to analyze these overlaps conclusively.

Figure 2.12: Normalized formant frequency plots for Jorhat female speakers with 1 standard deviation ellipses in isolation.

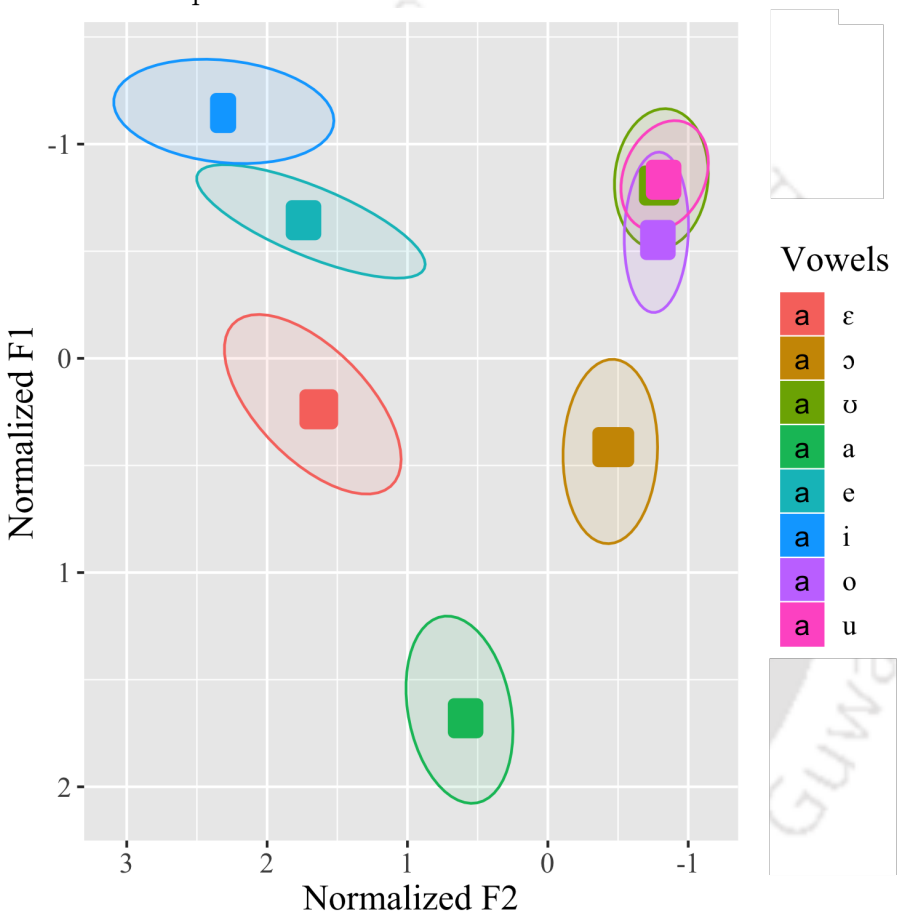


Figure 2.13: Normalized formant frequency plots for Jorhat male speakers with 1 standard deviation ellipses in isolation.

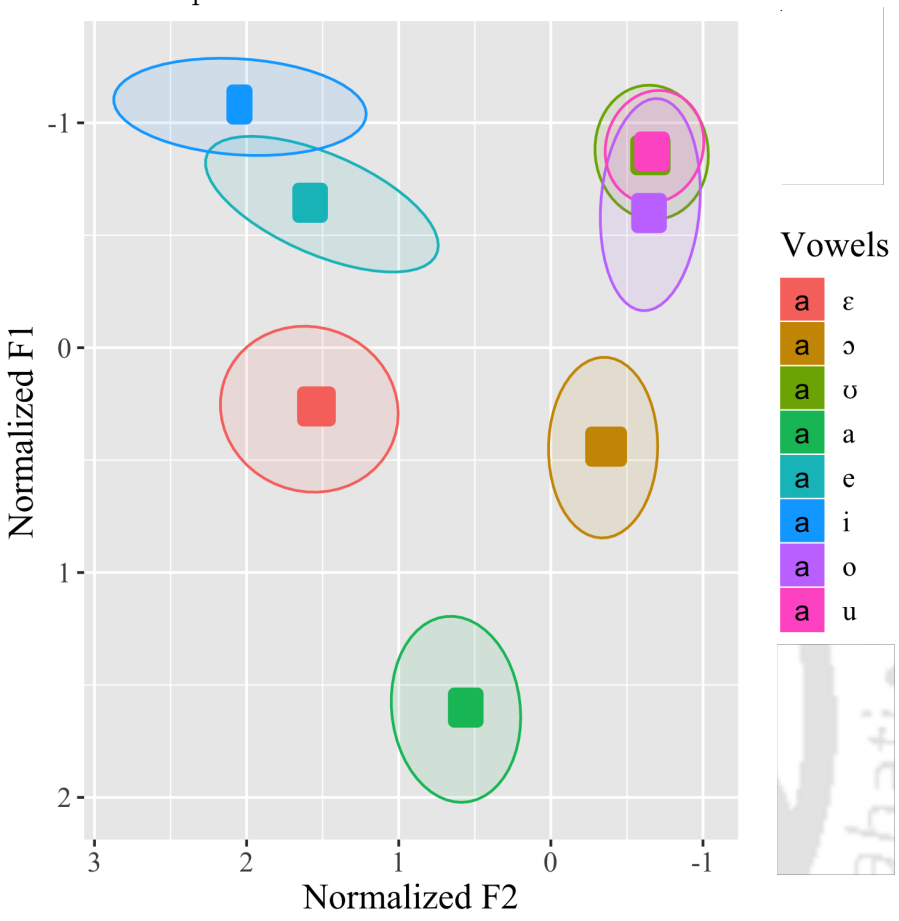


Figure 2.14: Normalized formant frequency plots for Jorhat female speakers with 1 standard deviation ellipses in sentence frames.

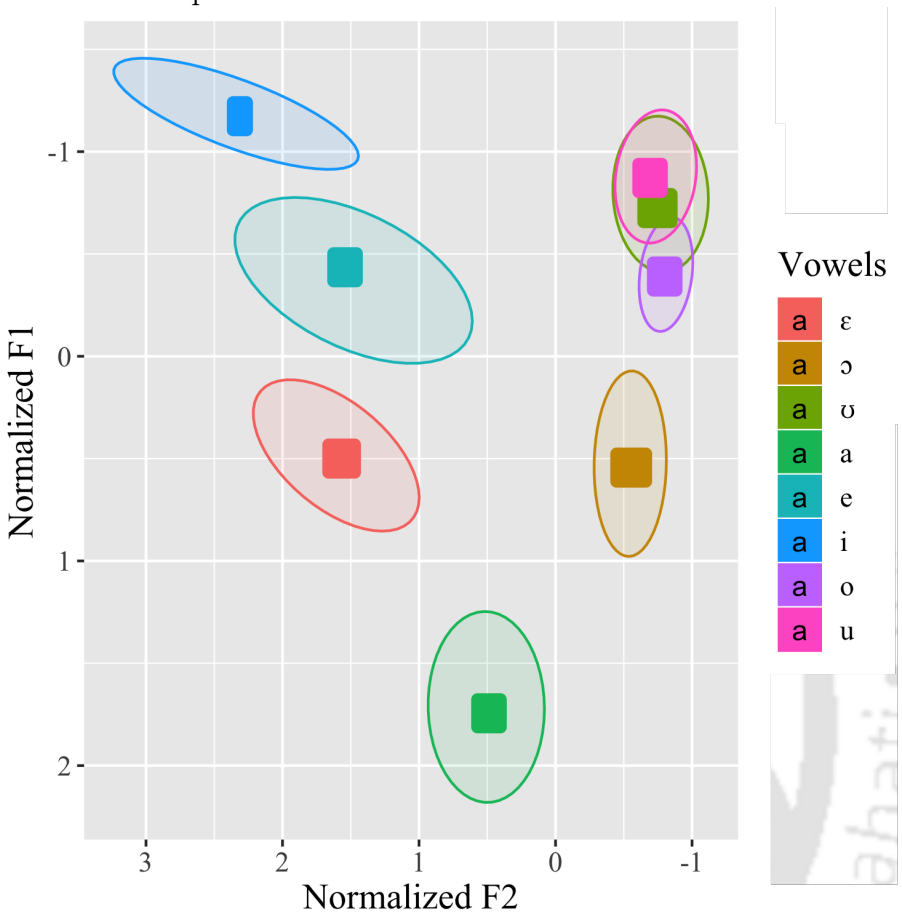


Figure 2.15: Normalized formant frequency plots for Jorhat male speakers with 1 standard deviation ellipses in sentence frames.

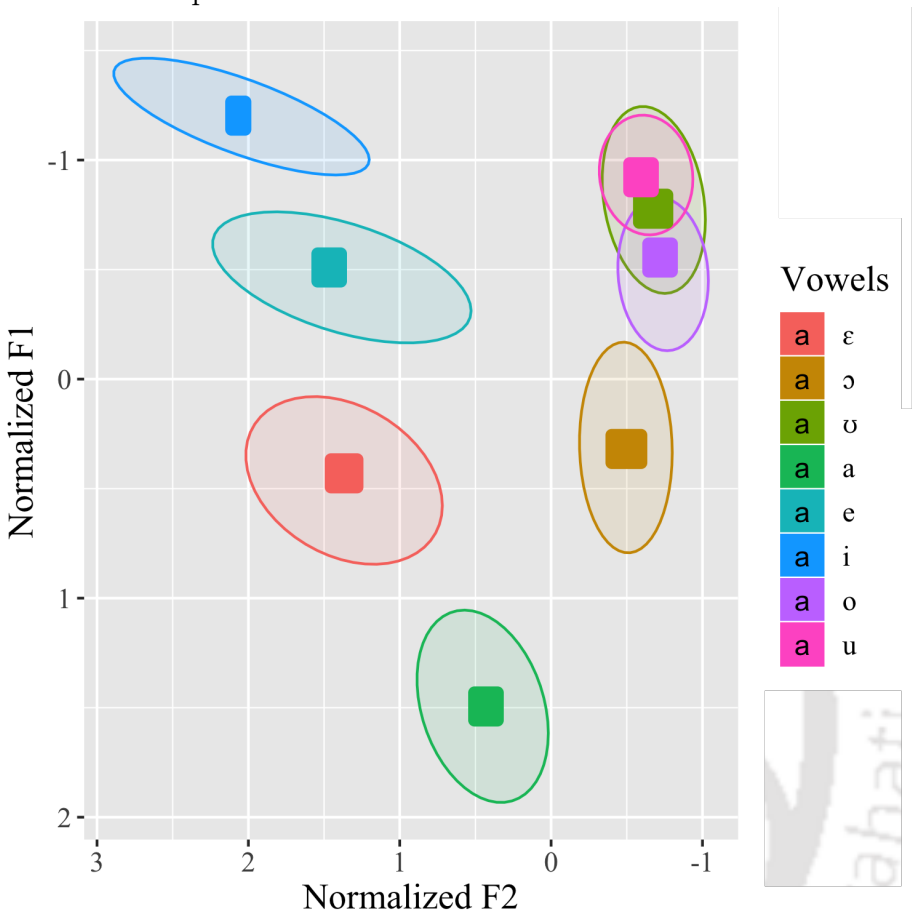


Table 2.8: Lobanov normalized and scaled vowel formant values in Hz for Jorhat speakers in words produced in two contexts.

Context	Vowel	Female		Male	
		F1	F2	F1	F2
Isolation	a	503 (32)	1409 (55)	526 (35)	1358 (53)
	e	351 (13)	1752 (77)	399 (27)	1624 (117)
	ɛ	434 (23)	1600 (90)	451 (26)	1548 (122)
	i	363 (19)	1737 (97)	366 (14)	1733 (85)
	o	395 (21)	1152 (37)	400 (20)	1150 (36)
	ɔ	454 (29)	1221 (55)	448 (36)	1186 (44)
	u	379 (19)	1132 (48)	379 (22)	1163 (77)
	ʊ	382 (23)	1140 (51)	384 (26)	1151 (57)
Sentences	a	524 (32)	1375 (62)	520 (47)	1352 (75)
	e	389 (22)	1710 (110)	397 (27)	1656 (109)
	ɛ	443 (25)	1562 (120)	458 (27)	1526 (125)
	i	367 (20)	1718 (142)	365 (16)	1686 (162)
	o	394 (19)	1146 (53)	400 (18)	1139 (42)
	ɔ	457 (32)	1227 (60)	445 (30)	1191 (67)
	u	378 (18)	1139 (57)	382 (36)	1178 (107)
	ʊ	379 (19)	1148 (68)	388 (30)	1155 (68)

The mean Lobanov normalized and scaled F1 and F2 values for all the vowels produced by the Jorhat speakers are presented in Table 2.8. In order to analyze vowel formant frequencies more conclusively, three separate full LME models were constructed for normalized F1, F2 and F3. In all the three models, vowel type, context and gender were the fixed effects. Speaker, onset type, coda type were the random effects. The significance of the interactions vowel type * gender, vowel type

* context, gender * context and vowel type * gender * context were also explored. The full LME models were subjected to backward reduction of effects using the *step* function of the *lmerTest* package on R (Kuznetsova et al., 2017). The reduced models for F1, F2 and F3 were determined as follows.

- F1: vowel + gender + context + (1|speaker) + (1|onset) + (1|coda) + vowel:gender + vowel:context + gender:context + vowel:gender:context
- F2: vowel + gender + context + (1|speaker) + (1|onset) + (1|coda) + vowel:gender + vowel:context + gender:context
- F3: vowel + gender + context + (1|onset) + vowel:gender + gender:context

The reduced models were subjected to a Wald χ^2 test and the results obtained are presented in Table 2.9.

Table 2.9: Wald χ^2 tests on the LME model for F1, F2 and F3 for Jorhat variety.

	F1	F2	F3
Vowel type	30172.7 ***	28614.4 ***	497.2 ***
Gender	0.7 $^{n.s.}$	0.5 $^{n.s.}$	0.1 $^{n.s.}$
Context	55.9 ***	7.9 **	0.1 $^{n.s.}$
Vowel type x Gender	63.1 ***	73.7 ***	99.0 ***
Vowel type x Context	26.5 ***	123.8 ***	NA
Gender x Context	29.0 ***	6.1 *	6.4 *
Vowel type x Gender x Context	23.9 ***	NA	NA

p-values: ***: $p < 0.001$, **: $p < 0.01$, *: $p < 0.05$, $^{n.s.}$: Not significant

The reduced models were subjected to a Wald χ^2 test and the results obtained are presented in Table 2.9. As seen in the table, vowel type has significant effect on F1, F2 and F3 of the vowels. Context effect is significant on F1 and F2. Vowel type and context interaction is significant for F1 and F2. Interaction of vowel type, gender and

context is significant only for F1. Considering this, we decided to explore the effects of the variables by context types. Hence, LME models for F1, F2 and F3 were built separately for vowels in isolation and vowels in sentence frames.

For each context namely, isolation and sentence, three full LME models were built with F1 , F2 and F3 as dependent variables, and vowel type and gender as the fixed effects. Speaker, onset type, coda type were the random effects. The interaction between vowel type and gender was also explored. The full LME models were subjected to backward reduction of effects using the *step* function of the *lmerTest* package on R Kuznetsova et al. (2017). The reduced model for F1, F2 and F3 in isolation and in sentence form are as follows.

- Isolation
  - F1: vowel + gender + (1|speaker) + (1|onset) + (1|coda) + vowel:gender
  - F2: vowel + gender + (1|speaker) + (1|onset) + (1|coda) + vowel:gender
  - F3: vowel + gender + (1|speaker) + vowel:gender
- Sentence
  - F1: vowel + gender + (1|speaker) + (1|onset) + vowel:gender
  - F2: vowel + (1|speaker) + (1|onset) + (1|coda)
  - F3: vowel + gender + (1|onset) + vowel:gender

The reduced models were then subjected to a post-hoc Bonferroni test to see the pairwise variability of vowel formants using the *emmeans* function. The results of the post-hoc Bonferroni tests for vowels produced in isolation is presented in Table 2.10 and those produced in sentence frames is presented in Table 2.11. The results of F3 are not included in the two tables as F3 is not a distinguishing feature for Assamese vowels. However, the results for F3 are included in Table C.2.

Table 2.10: Results of Bonferroni post-hoc tests conducted on vowels produced in isolation by Jorhat speakers.

Contrasts	F1					F2				
	estimate	SE	df	t-ratio	p-value	estimate	SE	df	t-ratio	p-value
a-e	2.32	0.05	1584	49.67	< .0001	-1.4	0.05	1579	-30.91	< .0001
a-ɛ	1.37	0.03	1541	44.06	< .0001	-1.16	0.03	1570	-38.48	< .0001
a-i	2.89	0.04	1567	79.29	< .0001	-1.9	0.04	1577	-53.9	< .0001
a-o	2.18	0.04	1514	61.96	< .0001	1.12	0.03	1565	32.81	< .0001
a-ɔ	1.28	0.03	1212	48.53	< .0001	1.03	0.03	1400	40.09	< .0001
a-u	2.58	0.03	1258	98.55	< .0001	1.23	0.03	1380	48.18	< .0001
a-ʊ	2.51	0.03	1399	98.59	< .0001	1.26	0.02	1491	50.79	< .0001
e-ɛ	-0.95	0.05	1577	-19.46	< .0001	0.24	0.05	1571	4.99	< .0001
e-i	0.57	0.05	1577	11.01	< .0001	-0.5	0.05	1573	-10.13	< .0001
e-o	-0.14	0.05	1581	-2.79	0.1486	2.52	0.05	1577	51.38	< .0001
e-ɔ	-1.05	0.05	1583	-22.16	< .0001	2.43	0.05	1578	53.2	< .0001
e-u	0.26	0.05	1581	5.52	< .0001	2.63	0.05	1575	58.18	< .0001
e-ʊ	0.19	0.05	1578	4.01	0.0018	2.65	0.04	1571	59.17	< .0001
ɛ-i	1.52	0.04	1582	38.68	< .0001	-0.74	0.04	1578	-19.49	< .0001
ɛ-o	0.81	0.04	1533	20.69	< .0001	2.28	0.04	1580	60.29	< .0001
ɛ-ɔ	-0.1	0.03	1536	-3.03	0.0707	2.19	0.03	1575	70.91	< .0001
ɛ-u	1.21	0.03	1572	38.99	< .0001	2.39	0.03	1581	79.76	< .0001
ɛ-ʊ	1.14	0.03	1575	37.35	< .0001	2.42	0.03	1578	82.29	< .0001
i-o	-0.71	0.04	1561	-16.39	< .0001	3.02	0.04	1580	72.17	< .0001
i-ɔ	-1.61	0.04	1559	-43.46	< .0001	2.93	0.04	1575	81.59	< .0001
i-u	-0.31	0.04	1539	-8.35	< .0001	3.13	0.04	1556	87.37	< .0001
i-ʊ	-0.38	0.04	1575	-10.54	< .0001	3.16	0.03	1576	90.47	< .0001
o-ɔ	-0.9	0.04	1539	-25.21	< .0001	-0.09	0.03	1578	-2.63	0.2425
o-u	0.4	0.03	1583	11.72	< .0001	0.11	0.03	1580	3.31	0.0269
o-ʊ	0.33	0.04	1546	9.3	< .0001	0.14	0.03	1582	3.99	0.0019
ɔ-u	1.3	0.02	1551	52.39	< .0001	0.2	0.02	1574	8.3	< .0001
ɔ-ʊ	1.23	0.02	1576	51.24	< .0001	0.23	0.02	1584	9.78	< .0001
u-ʊ	-0.07	0.03	1564	-2.86	0.1202	0.03	0.02	1576	1.13	1

Table 2.11: Results of Bonferroni post-hoc tests conducted on vowels produced in sentence frames by Jorhat speakers.

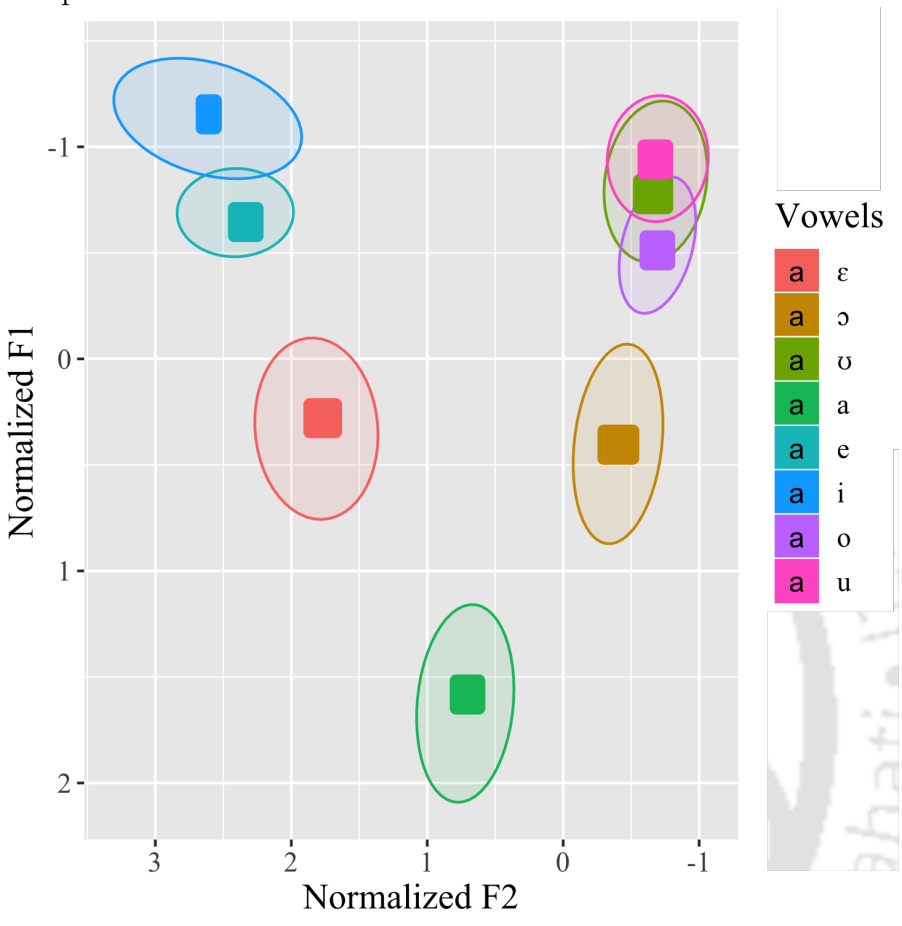
Contrasts	F1					F2				
	estimate	SE	df	t-ratio	p-value	estimate	SE	df	t-ratio	p-value
a-e	2.18	0.05	1590	47.1	< .0001	-1.34	0.05	1593	-27.72	< .0001
a-ɛ	1.26	0.03	1594	42.22	< .0001	-1.05	0.03	1578	-32.83	< .0001
a-i	2.71	0.04	1591	76	< .0001	-1.69	0.04	1591	-44.76	< .0001
a-o	2.12	0.03	1595	65.12	< .0001	1.01	0.04	1549	28.67	< .0001
a-ɔ	1.21	0.02	1590	50.32	< .0001	0.91	0.03	1379	33.78	< .0001
a-u	2.48	0.02	1595	101.97	< .0001	1.09	0.03	1410	40.72	< .0001
a-ʊ	2.4	0.02	1595	99	< .0001	1.12	0.03	1490	43.12	< .0001
e-ɛ	-0.92	0.05	1587	-18.93	< .0001	0.29	0.05	1586	5.79	< .0001
e-i	0.53	0.05	1584	10.35	< .0001	-0.35	0.05	1587	-6.62	< .0001
e-o	-0.06	0.05	1589	-1.25	1	2.34	0.05	1591	45.23	< .0001
e-ɔ	-0.98	0.05	1590	-21.18	< .0001	2.25	0.05	1592	46.23	< .0001
e-u	0.29	0.05	1587	6.37	< .0001	2.43	0.05	1588	50.54	< .0001
e-ʊ	0.22	0.05	1588	4.77	0.0001	2.46	0.05	1585	51.51	< .0001
ɛ-i	1.45	0.04	1587	37.27	< .0001	-0.64	0.04	1592	-15.93	< .0001
ɛ-o	0.86	0.04	1595	22.97	< .0001	2.05	0.04	1549	52.36	< .0001
ɛ-ɔ	-0.06	0.03	1595	-1.96	1	1.96	0.03	1573	60.17	< .0001
ɛ-u	1.21	0.03	1592	40.78	< .0001	2.14	0.03	1588	67.91	< .0001
ɛ-ʊ	1.14	0.03	1591	38.78	< .0001	2.17	0.03	1593	70.14	< .0001
i-o	-0.59	0.04	1593	-14.3	< .0001	2.7	0.04	1578	61.19	< .0001
i-ɔ	-1.51	0.04	1592	-42.46	< .0001	2.6	0.04	1588	67.87	< .0001
i-u	-0.24	0.04	1588	-6.65	< .0001	2.78	0.04	1578	72.94	< .0001
i-ʊ	-0.31	0.04	1588	-8.79	< .0001	2.81	0.04	1592	75.44	< .0001
o-ɔ	-0.91	0.03	1594	-28.37	< .0001	-0.1	0.04	1576	-2.76	0.1662
o-u	0.36	0.03	1591	11.17	< .0001	0.08	0.03	1595	2.4	0.4622
o-ʊ	0.28	0.03	1595	8.6	< .0001	0.12	0.04	1570	3.35	0.0234
ɔ-u	1.27	0.02	1595	53.57	< .0001	0.18	0.03	1581	7.1	< .0001
ɔ-ʊ	1.2	0.02	1593	52.58	< .0001	0.22	0.02	1591	8.81	< .0001
u-ʊ	-0.07	0.02	1593	-3.09	0.0568	0.04	0.03	1591	1.41	1

As seen in Tables 2.10 and 2.11, for vowels produced in isolation, post-hoc Bonferroni tests revealed significant F1 differences between all vowels except the following sets: /e-o/, /ɛ-ɔ/ and /u-ʊ/ indicating that the vowels in these pairs did not differ in height. According to Mahanta (2012) and reported in Table 1.6, /e/ and /o/ are higher-mid vowels and /ɛ/ and /ɔ/ are lower-mid vowels in Assamese. Therefore, the significant similarity in F1 for the two pairs is predictable. However, /u/ and /ʊ/ vowels have been reported as high and near-high respectively and ideally should be significantly different in F1. Significant F2 differences were observed between all vowels except, /o-ɔ/ and /u-ʊ/ indicating that the vowels in the pairs did not differ in backness. Similarly, in case of vowels produced in sentence frames, pots-hoc Bonferroni tests revealed significant F1 differences between all vowels except /e-o/, /ɛ-ɔ/ and /u-ʊ/. Significant F2 differences were observed between all vowels except /o-ɔ/, /o-u/ and /u-ʊ/. The results show that in both contexts, F1 and F2 values of the vowels /u/ and /ʊ/ are not significantly different. Visual inspection of the plot in Figure 2.11 indicated complete overlap of the ellipses of these two vowels. Results of statistical tests show that the two vowels differ neither in height nor in backness, i.e, their positions in the vowel space are not significantly different. This provides robust evidence of the possibility of a merger between /u/ and /ʊ/ in the Jorhat variety. However, whether these results actually indicate a merger or not are dealt with in detail using vowel merger measures in Chapter 3.

2.2.2.3 Vowel formants in Nagaon variety

The acoustic characteristics of the vowel formants in the Nagaon Assamese variety are described in this section. Figure 2.16 presents the corresponding vowel plots on an F1-F2 plane. Figures 2.17, 2.18, 2.19 and 2.20 present the vowel plots on F1-F2 plane separately for gender and context types.

Figure 2.16: Normalized formant frequency plots for Nagaon speakers with 1 standard deviation ellipses.



From a visual examination of the plots, like Nagaon vowels /i, e, ɛ, a, ɔ/ appear to be distinct. However, we see considerable overlap of the /ʊ, u, o/ vowel ellipses for both genders and contexts. An impressionistic view of the vowel plots conclude that:

a. there is significant overlap of the vowels /u/ and /ʊ/ with a partial overlap with /o/ with /u/ and /ʊ/;

b. while vowel overlaps are seen in all categories, the uniformity of the overlaps across context and gender confirm that they are not accidental;

c. the robustness of the overlaps can be confirmed only by using statistical measures.

Figure 2.17: Normalized formant frequency plots for Nagaon female speakers with 1 standard deviation ellipses in isolation.

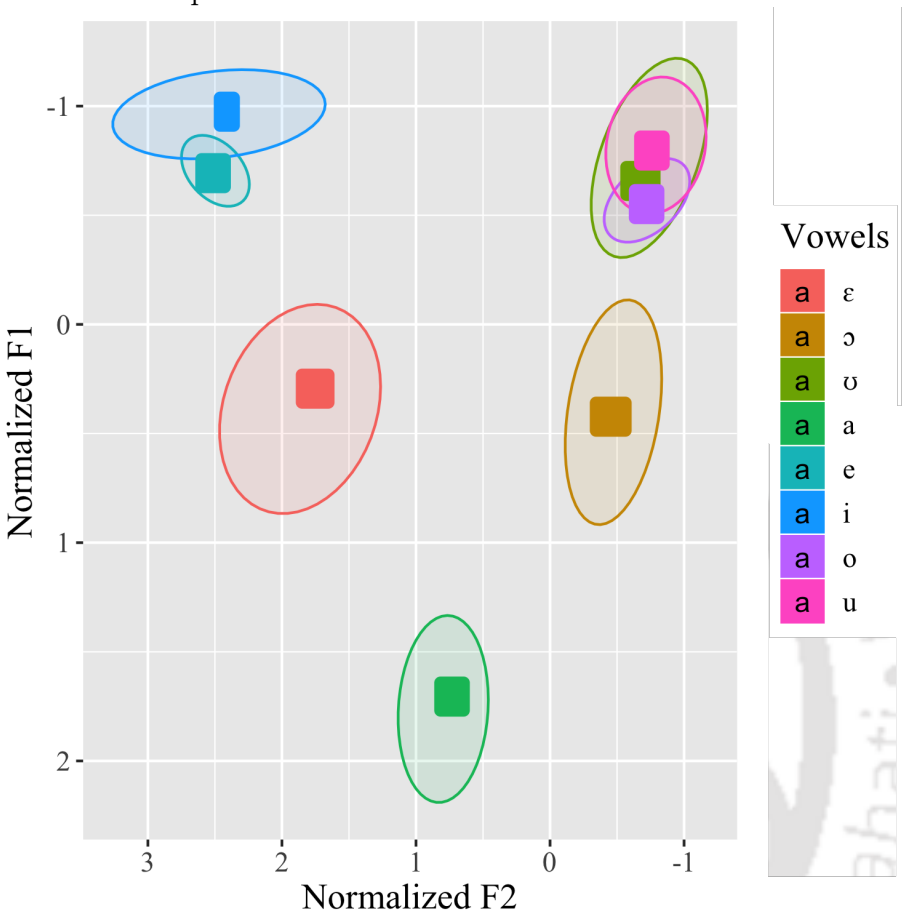


Figure 2.18: Normalized formant frequency plots for Nagaon male speakers with 1 standard deviation ellipses in isolation.

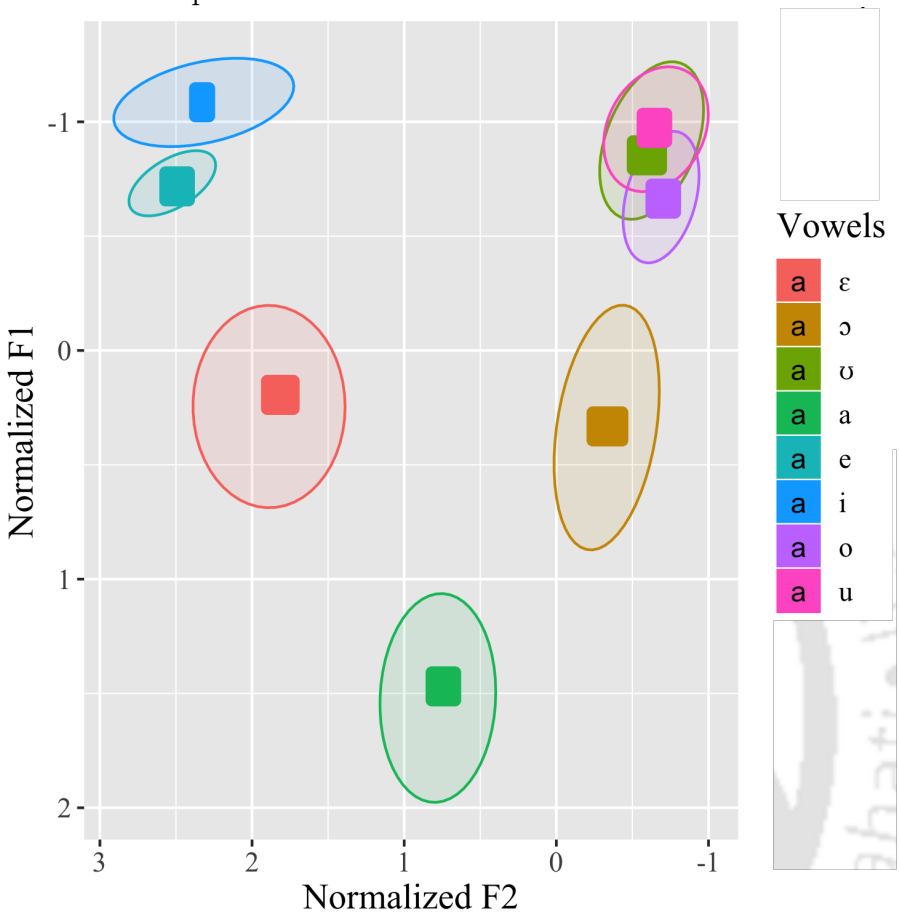


Figure 2.19: Normalized formant frequency plots for Nagaon female speakers with 1 standard deviation ellipses in sentence frames.

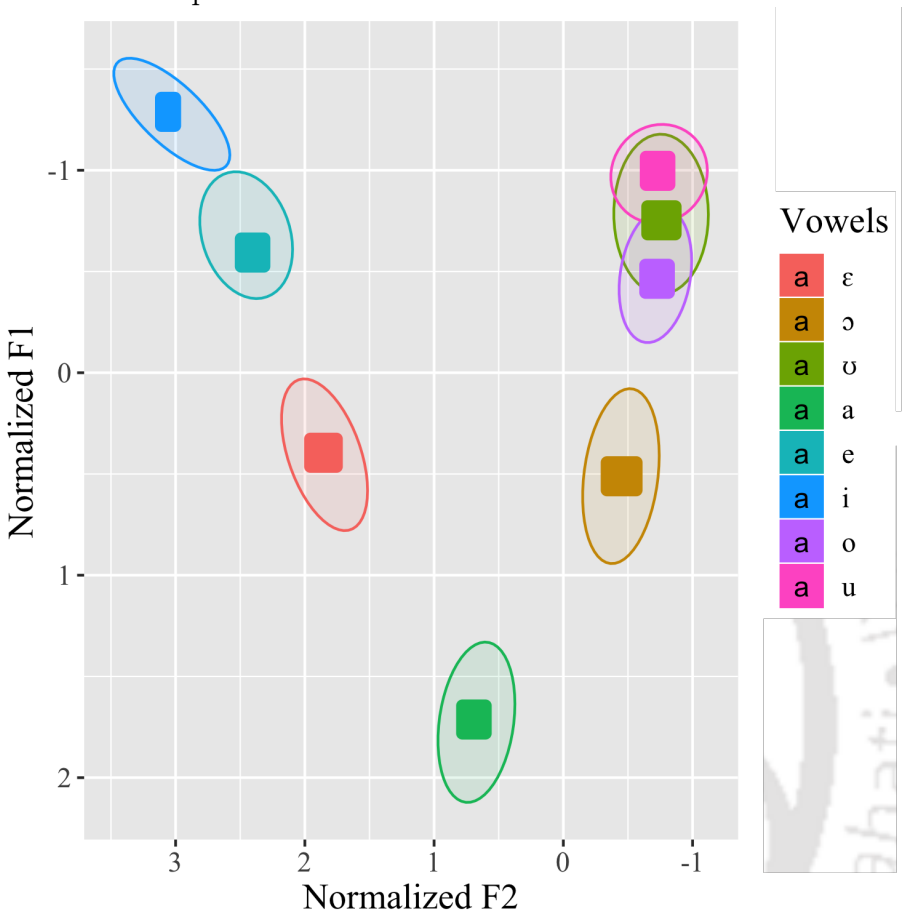


Figure 2.20: Normalized formant frequency plots for Nagaon male speakers with 1 standard deviation ellipses in sentence frames.

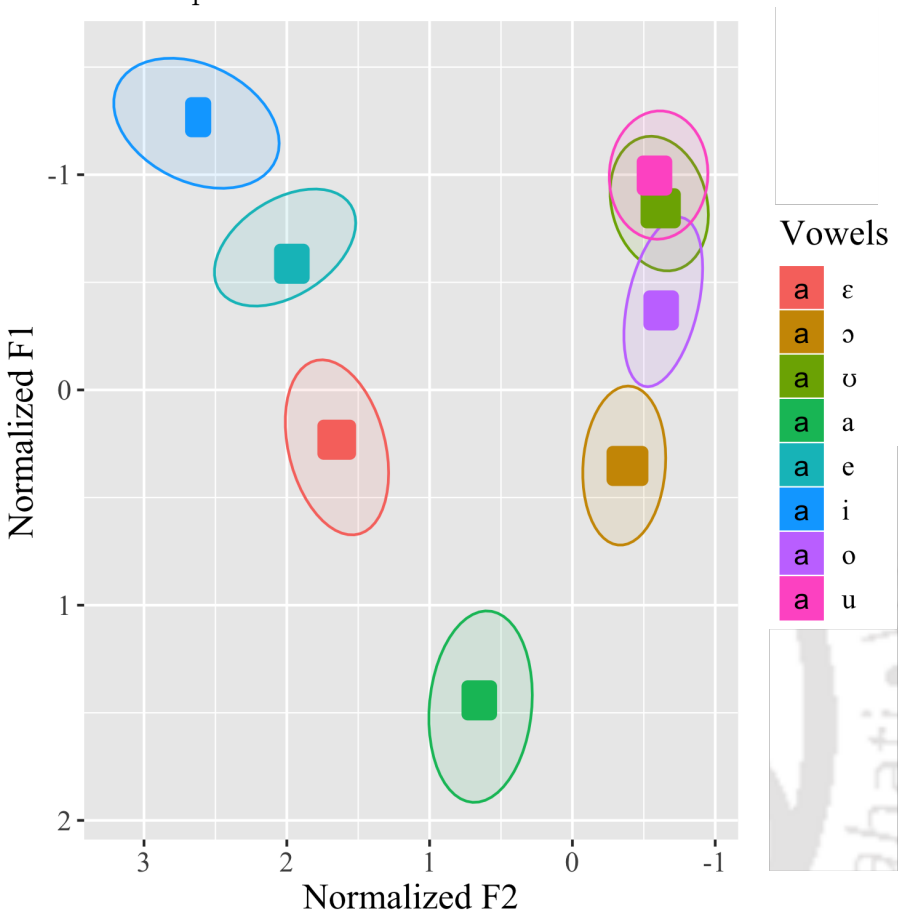


Table 2.12: Lobanov normalized and scaled vowel formant values in Hz for Nagaon speakers in words produced in two contexts

Context	Vowel	Female		Male	
		F1	F2	F1	F2
Isolation	a	526 (25)	1415 (70)	514 (40)	1401 (57)
	e	369 (17)	1753 (61)	395 (21)	1674 (126)
	ɛ	443 (27)	1614 (165)	433 (23)	1615 (67)
	i	369 (19)	1801 (129)	360 (19)	1805 (86)
	o	390 (20)	1154 (39)	401 (30)	1160 (57)
	ɔ	450 (23)	1222 (48)	452 (37)	1223 (75)
	u	377 (24)	1149 (57)	377 (33)	1160 (94)
	ʊ	378 (20)	1162 (69)	387 (30)	1151 (101)
Sentences	a	522 (30)	1406 (81)	511 (37)	1396 (81)
	e	386 (19)	1738 (121)	402 (16)	1653 (142)
	ɛ	442 (26)	1649 (114)	437 (29)	1546 (92)
	i	371 (20)	1713 (195)	359 (16)	1755 (120)
	o	394 (20)	1143 (38)	405 (30)	1152 (54)
	ɔ	454 (28)	1219 (49)	444 (32)	1205 (50)
	u	373 (18)	1151 (64)	384 (51)	1190 (162)
	ʊ	378 (19)	1159 (62)	388 (41)	1194 (187)

Table 2.12 presents the mean Lobanov normalized and scaled F1 and F2 values for all the vowels produced by the Nagaon speakers. In order to analyze vowel formant frequencies more conclusively, three separate full LME models were constructed for normalized F1, F2 and F3. In all the three models, vowel type, context and gender were the fixed effects. Speaker, onset type, coda type were the random effects. The significance of the following interactions were also explored: vowel type * gender,

vowel type * context, gender * context and vowel type * gender * context. The full LME models were subjected to backward reduction of effects using the *step* function of the *lmerTest* package on R (Kuznetsova et al., 2017). The reduced models for F1, F2 and F3 were determined as follows.

- F1: vowel + gender + context + (1|onset) + (1|coda) + vowel:gender + vowel:context
- F2: vowel + gender + context + (1|onset) + (1|coda) + vowel:gender + vowel:context + gender:context + vowel:gender:context
- F3: vowel + gender + context + (1|onset) + vowel:gender

Table 2.13: Wald χ^2 tests on the LME model for F1, F2 and F3 for Nagaon variety.

	F1	F2	F3
Vowel type	14234.2 ***	20435.4 ***	1032.8 ***
Gender	0.3 $^{n.s.}$	0.8 $^{n.s.}$	0.1 $^{n.s.}$
Context	92.9 ***	21.4 ***	5.3 *
Vowel type x Gender	40.6 ***	76.3 ***	67.7 ***
Vowel type x Context	23.1 **	66.4 ***	NA
Gender x Context	NA	2.9 $^{n.s.}$	NA
Vowel type x Gender x Context	NA	25.7 ***	NA

p-values: ***: $p < 0.001$, **: $p < 0.01$, *: $p < 0.05$, $^{n.s.}$: Not significant

The reduced models were subjected to a Wald χ^2 test and the results obtained are presented in Table 2.13. As seen in the table, vowel type has significant effect on F1, F2 and F3 of the vowels. Context effect is significant on F1 and F2. Similarly, interaction of vowel type and context is significant for both F1 and F2. However, vowel type, gender and context interaction is also significant only for F2. Considering this, we decided to explore the effects of the variables by context types. Hence, LME

models for F1, F2 and F3 were built separately for vowels in isolation and vowels in sentence frames.

For each context namely, isolation and sentence, three full LME models were built with F1 , F2 and F3 as dependent variables, and vowel type and gender as the fixed effects. Speaker, onset type, coda type were the random effects. The interaction between vowel type and gender was also explored. The full LME models were subjected to backward reduction of effects using the *step* function of the *lmerTest* package on R Kuznetsova et al. (2017). The reduced model for F1, F2 and F3 in isolation and in sentence form are as follows.

- Isolation
  - F1: vowel + gender + (1|onset) + (1|coda) + vowel:gender
  - F2: vowel + gender + (1|onset) + (1|coda) + vowel:gender
  - F3: vowel + gender + (1|speaker) + vowel:gender
- Sentence
  - F1: vowel + gender + (1|onset) + (1|coda) + vowel:gender
  - F2: vowel + gender + (1|onset) + (1|coda) + vowel:gender
  - F3: vowel + gender + (1|speaker) + (1|onset) + vowel:gender

The reduced models were then subjected to a post-hoc Bonferroni test to see the pairwise variability of vowel formants using the *emmeans* function. The results of the post-hoc Bonferroni tests for vowels produced in isolation is presented in Table 2.14 and those produced in sentence frames is presented in Table 2.15. The results of F3 are not included in the two tables as F3 is not a distinguishing feature for Assamese vowels. However, the results for F3 are included in Table C.3.

Table 2.14: Results of Bonferroni post-hoc tests conducted on vowels produced in isolation by Nagaon speakers.

Contrasts	F1					F2				
	estimate	SE	df	t-ratio	p-value	estimate	SE	df	t-ratio	p-value
a-e	2.21	0.1	1290	22.92	< .0001	-1.86	0.08	1306	-23.55	< .0001
a-ɛ	1.36	0.05	1164	29.77	< .0001	-1.16	0.04	1295	-30.6	< .0001
a-i	2.76	0.07	1270	40.29	< .0001	-2.09	0.06	1310	-37.13	< .0001
a-o	2.1	0.05	993	38.47	< .0001	1.31	0.05	1116	28.79	< .0001
a-ɔ	1.25	0.04	924	35.1	< .0001	1.14	0.03	1153	38.38	< .0001
a-u	2.59	0.04	744	69	< .0001	1.42	0.03	1138	44.75	< .0001
a-ʊ	2.42	0.04	1049	67.82	< .0001	1.4	0.03	1252	47.13	< .0001
e-ɛ	-0.85	0.1	1290	-8.58	< .0001	0.7	0.08	1304	8.69	< .0001
e-i	0.55	0.11	1312	5.13	< .0001	-0.23	0.09	1307	-2.64	0.2333
e-o	-0.11	0.1	1266	-1.05	1	3.16	0.08	1292	37.95	< .0001
e-ɔ	-0.96	0.1	1297	-10.02	< .0001	3	0.08	1306	38.32	< .0001
e-u	0.38	0.1	1295	3.93	0.0025	3.28	0.08	1307	41.66	< .0001
e-ʊ	0.21	0.1	1291	2.15	0.8802	3.26	0.08	1305	41.6	< .0001
ɛ-i	1.4	0.07	1301	19.56	< .0001	-0.93	0.06	1306	-15.96	< .0001
ɛ-o	0.74	0.06	907	12.1	< .0001	2.46	0.05	1040	48.63	< .0001
ɛ-ɔ	-0.11	0.05	1175	-2.53	0.3236	2.3	0.04	1277	62.09	< .0001
ɛ-u	1.23	0.05	1073	26.82	< .0001	2.57	0.04	1239	67.85	< .0001
ɛ-ʊ	1.05	0.04	1271	24	< .0001	2.56	0.04	1314	70.97	< .0001
i-o	-0.66	0.08	1155	-8.48	< .0001	3.4	0.06	1207	52.95	< .0001
i-ɔ	-1.51	0.07	1301	-22.34	< .0001	3.24	0.06	1311	58.28	< .0001
i-u	-0.18	0.07	1258	-2.54	0.313	3.51	0.06	1293	62	< .0001
i-ʊ	-0.35	0.07	1301	-5.12	< .0001	3.49	0.06	1309	63.11	< .0001
o-ɔ	-0.85	0.05	1049	-15.74	< .0001	-0.16	0.04	1178	-3.6	0.0093
o-u	0.49	0.05	1247	9.2	< .0001	0.11	0.04	1296	2.58	0.2832
o-ʊ	0.31	0.05	967	5.75	< .0001	0.1	0.05	1101	2.11	0.9824
ɔ-u	1.34	0.03	1076	38.96	< .0001	0.27	0.03	1276	9.53	< .0001
ɔ-ʊ	1.17	0.03	1307	36.4	< .0001	0.26	0.03	1308	9.76	< .0001
u-ʊ	-0.17	0.04	1037	-4.85	< .0001	-0.02	0.03	1266	-0.56	1

Table 2.15: Results of Bonferroni post-hoc tests conducted on vowels produced in sentence frames by Nagaon speakers.

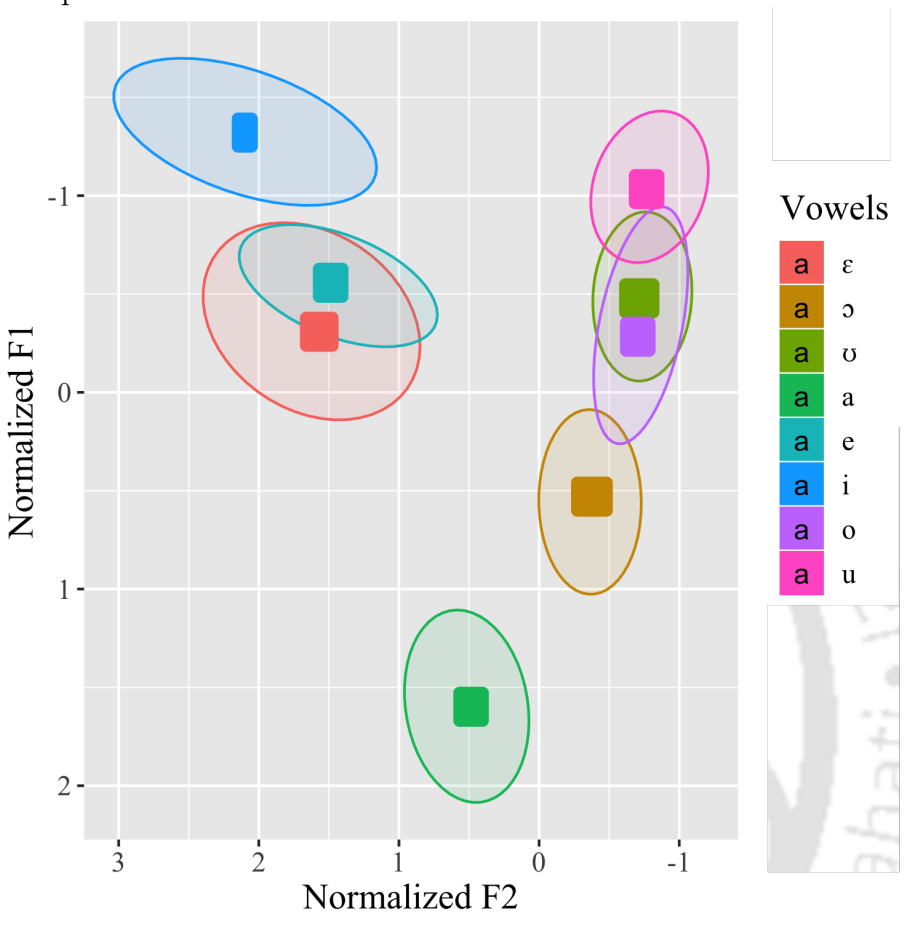
Contrasts	F1					F2				
	estimate	SE	df	t-ratio	p-value	estimate	SE	df	t-ratio	p-value
a-e	2	0.09	1305	23.18	< .0001	-1.66	0.07	1301	-23.07	< .0001
a-ɛ	1.26	0.04	1123	29.13	< .0001	-1.1	0.04	1293	-30.33	< .0001
a-i	2.55	0.07	1234	39.12	< .0001	-1.88	0.05	1306	-34.54	< .0001
a-o	1.89	0.05	1061	36.32	< .0001	1.23	0.04	1221	28.06	< .0001
a-ɔ	1.11	0.03	831	33	< .0001	1.02	0.03	1161	35.83	< .0001
a-u	2.44	0.04	630	69.11	< .0001	1.27	0.03	1089	41.55	< .0001
a-ʊ	2.3	0.03	1003	68.56	< .0001	1.27	0.03	1235	44.44	< .0001
e-ɛ	-0.74	0.09	1303	-8.4	< .0001	0.56	0.07	1298	7.59	< .0001
e-i	0.55	0.1	1301	5.58	< .0001	-0.22	0.08	1296	-2.7	0.1943
e-o	-0.11	0.09	1304	-1.19	1	2.89	0.08	1296	37.77	< .0001
e-ɔ	-0.89	0.09	1304	-10.45	< .0001	2.69	0.07	1298	37.65	< .0001
e-u	0.44	0.09	1306	5.16	< .0001	2.93	0.07	1301	40.84	< .0001
e-ʊ	0.3	0.09	1302	3.45	0.0161	2.93	0.07	1297	41.06	< .0001
ɛ-i	1.29	0.07	1306	19.01	< .0001	-0.78	0.06	1301	-13.79	< .0001
ɛ-o	0.63	0.06	1148	10.88	< .0001	2.33	0.05	1194	47.71	< .0001
ɛ-ɔ	-0.15	0.04	1195	-3.57	0.0105	2.13	0.04	1294	59.84	< .0001
ɛ-u	1.19	0.04	1015	27.54	< .0001	2.37	0.04	1256	64.93	< .0001
ɛ-ʊ	1.04	0.04	1243	25.17	< .0001	2.37	0.03	1306	68.5	< .0001
i-o	-0.65	0.07	1255	-8.82	< .0001	3.11	0.06	1273	50.08	< .0001
i-ɔ	-1.44	0.06	1280	-22.44	< .0001	2.91	0.05	1306	54.23	< .0001
i-u	-0.1	0.07	1234	-1.56	1	3.15	0.05	1300	57.49	< .0001
i-ʊ	-0.25	0.06	1287	-3.91	0.0028	3.15	0.05	1305	58.93	< .0001
o-ɔ	-0.79	0.05	1104	-15.24	< .0001	-0.21	0.04	1258	-4.78	0.0001
o-u	0.55	0.05	1243	11.02	< .0001	0.04	0.04	1305	0.87	1
o-ʊ	0.4	0.05	1137	7.8	< .0001	0.04	0.04	1226	0.84	1
ɔ-u	1.34	0.03	1054	41.26	< .0001	0.24	0.03	1256	8.8	< .0001
ɔ-ʊ	1.19	0.03	1301	39.58	< .0001	0.24	0.03	1301	9.66	< .0001
u-ʊ	-0.15	0.03	973	-4.45	0.0003	0	0.03	1248	-0.01	1

In Tables 2.14 and 2.15 we see that for vowels produced in isolation, post-hoc Bonferroni tests revealed significant F1 differences between all vowels except the following pairs: /e-o/, /e-ʊ/, /ɛ-ɔ/ and /i-u/ indicating that the vowels in these pairs did not differ in height. As mentioned earlier, /e/ and /o/ are higher-mid vowels, /ɛ/ and /ɔ/ are lower-mid vowels and /i/ and /u/ are high vowels in Assamese (Mahanta, 2012). Therefore, the significant similarity in F1 for these pairs is predictable. Significant F2 differences were observed between all vowels except, /e-i/, /o-u/, /o-ʊ/ and /u-ʊ/ indicating that the vowels in the pairs did not differ in backness. Similarly, in case of vowels produced in sentence frames, pots-hoc Bonferroni tests revealed significant F1 differences between all vowels except /e-o/ and /i-u/. Significant F2 differences were observed between all vowels except /e-i/, /o-ʊ/, /o-u/ and /u-ʊ/. Even though visual examination of the Nagaon vowel plots show significant overlap of the three vowels /u, ʊ, o/, results of the statistical tests show their similarity only in backness and not in height. Nevertheless, we check for mergers among vowels in Chapter 3 using vowel merger measures.

2.2.2.4 Vowel formants in Nalbari variety

This section describes the acoustic characteristics of the vowel formants in the Nalbari Assamese variety. Figure 2.21 presents the corresponding vowel plot on an F1-F2 plane. Figures 2.22, 2.23, 2.24 and 2.25 present the vowel plots on F1-F2 plane separately for gender and context types: Nalbari females in isolation (nfi), Nalbari males in isolation (nmi), Nalbari females in sentence frames (nfs) and Nalbari males in sentence frames (nms).

Figure 2.21: Normalized formant frequency plots for Nalbari speakers with 1 standard deviation ellipses.



From a visual examination of the plots, vowels /i/ and /a/ appear to be distinct but the vowel ellipses for /ʊ, o/ and /ɛ, e/ overlap considerably for both genders and contexts. The following can be concluded from a visual inspection of the vowel plots:

a. there is significant overlap of the vowels /e/ and /ɛ/;

b. there is significant overlap of the vowels /ʊ/ and /o/;

c. while vowel overlaps are seen in all categories, the uniformity of the overlaps across context and gender confirm that they are not accidental;

d. the robustness of the overlaps can be confirmed only by using statistical measures.

Figure 2.22: Normalized formant frequency plots for Nalbari female speakers with 1 standard deviation ellipses in isolation.

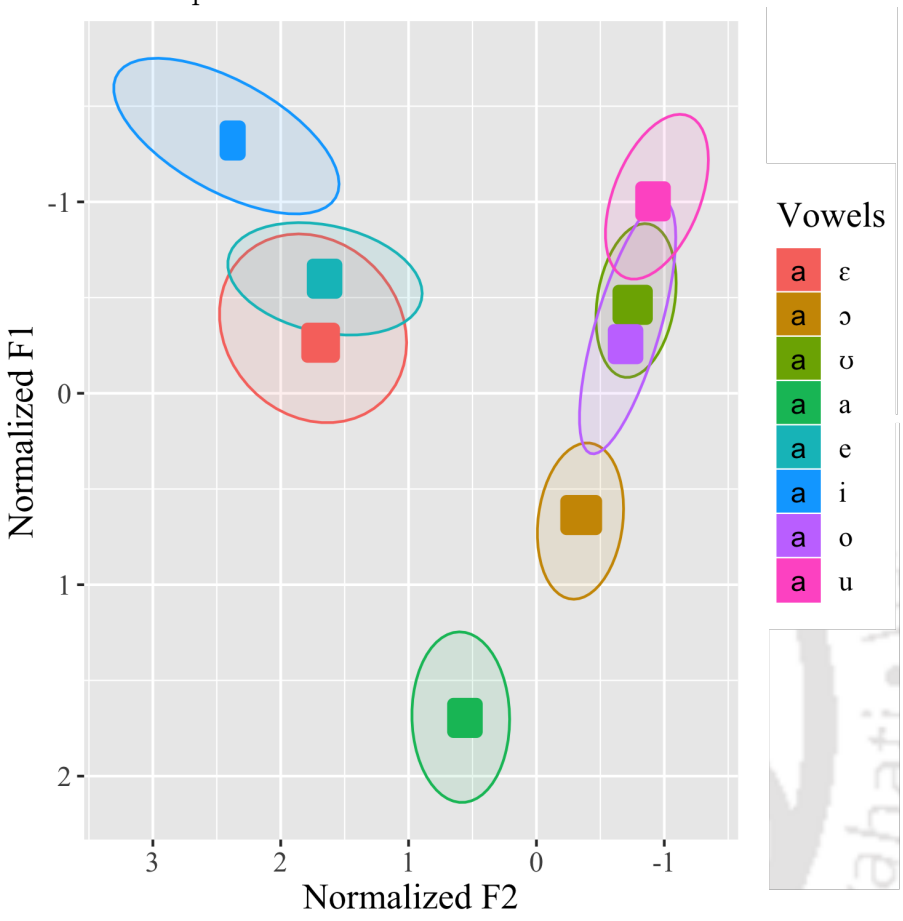


Figure 2.23: Normalized formant frequency plots for Nalbari male speakers with 1 standard deviation ellipses in isolation.

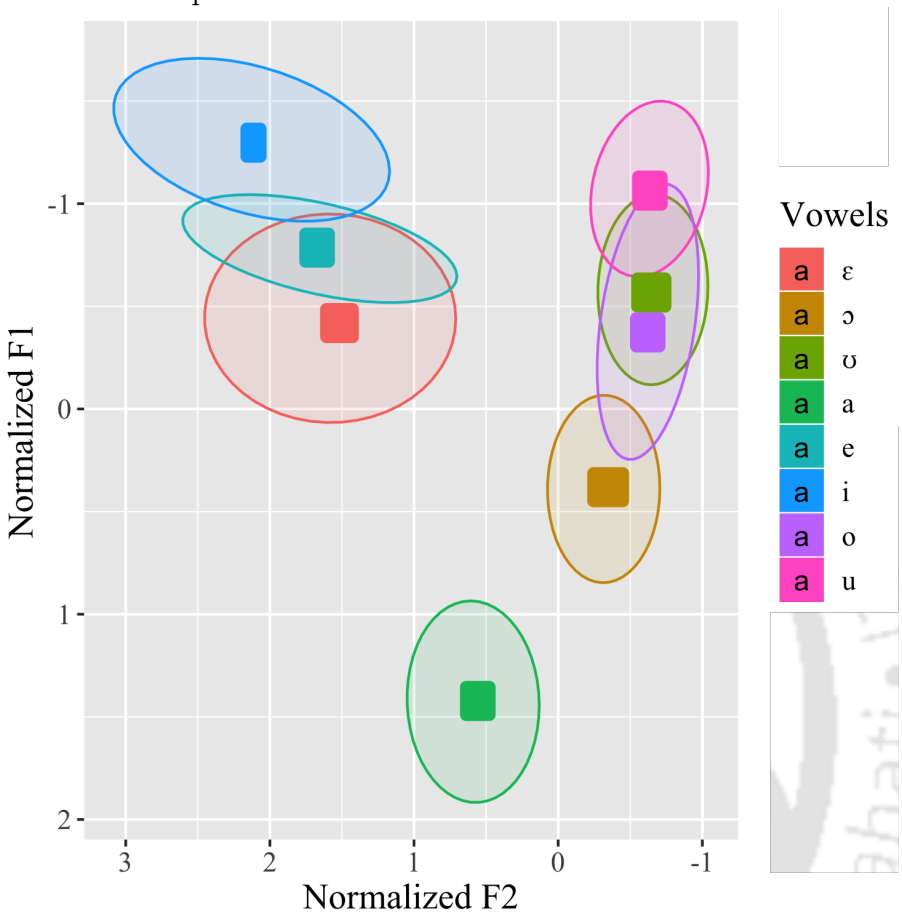


Figure 2.24: Normalized formant frequency plots for Nalbari female speakers with 1 standard deviation ellipses in sentence frames.

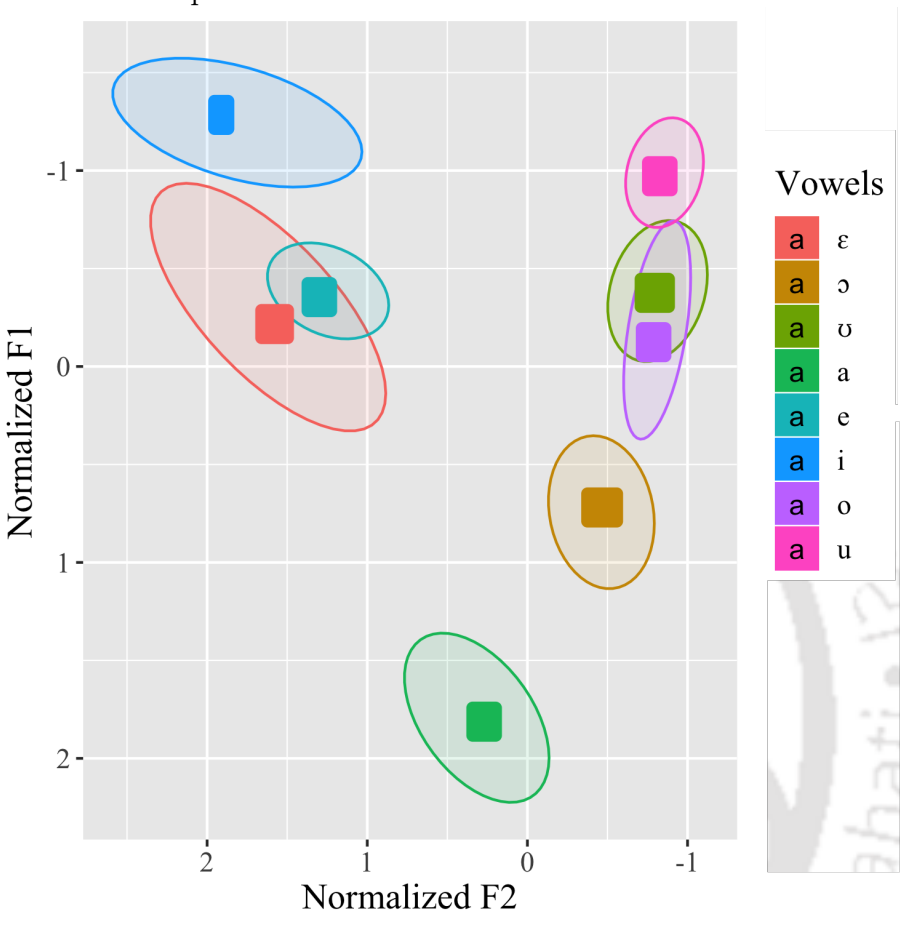


Figure 2.25: Normalized formant frequency plots for Nalbari male speakers with 1 standard deviation ellipses in sentence frames.

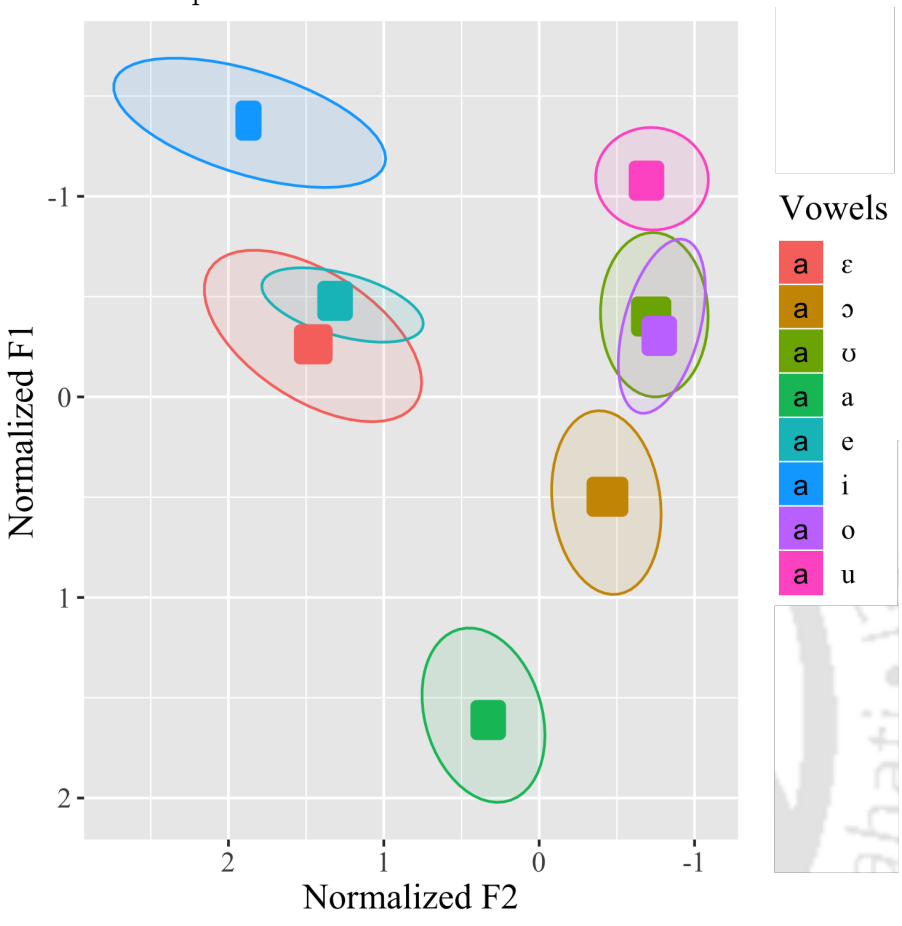


Table 2.16: Lobanov normalized and scaled vowel formant values in Hz for Nalbari speakers in words produced in two contexts.

Context	Vowel	Female		Male	
		F1	F2	F1	F2
Isolation	a	518 (31)	1373 (75)	527 (30)	1316 (48)
	e	374 (27)	1605 (163)	390 (19)	1558 (104)
	ɛ	404 (30)	1576 (185)	418 (33)	1570 (92)
	i	348 (27)	1743 (168)	350 (20)	1653 (110)
	o	408 (41)	1171 (69)	410 (32)	1140 (36)
	ɔ	461 (28)	1245 (55)	466 (28)	1203 (43)
	u	372 (22)	1131 (64)	377 (30)	1141 (87)
	ʊ	399 (29)	1153 (56)	407 (33)	1150 (84)
Sentences	a	516 (37)	1374 (97)	524 (38)	1319 (62)
	e	375 (26)	1567 (222)	389 (21)	1596 (116)
	ɛ	410 (35)	1568 (203)	415 (24)	1581 (128)
	i	349 (27)	1725 (227)	351 (23)	1680 (117)
	o	412 (40)	1158 (75)	406 (37)	1137 (53)
	ɔ	455 (29)	1234 (69)	458 (36)	1204 (53)
	u	367 (24)	1165 (100)	374 (31)	1161 (79)
	ʊ	401 (28)	1167 (75)	413 (33)	1158 (55)

The mean Lobanov normalized and scaled F1 and F2 values for all the vowels produced by the Nalbari speakers are presented in Table 2.16. In order to analyze vowel formant frequencies more conclusively, three separate full LME models were constructed for normalized F1, F2 and F3. In all the three models, vowel type, context and gender were the fixed effects. Speaker, onset type, coda type were the random effects. The significance of the following interactions were also explored:

vowel type * gender, vowel type * context, gender * context and vowel type * gender * context. The full LME models were subjected to backward reduction of effects using the *step* function of the *lmerTest* package on R Kuznetsova et al. (2017). The reduced models for F1, F2 and F3 were determined as follows.

- F1: vowel + gender + context + (1|speaker) + (1|onset) + (1|coda) + vowel:gender + vowel:context
- F2: vowel + gender + context + (1|speaker) + (1|onset) + (1|coda) + vowel:gender + vowel:context
- F3: vowel + gender + context + (1|onset) + vowel:gender + vowel:context + gender:context + vowel:gender:context

Table 2.17: Wald χ^2 tests on the LME model for F1, F2 and F3 for Nalbari variety.

	F1	F2	F3
Vowel type	25241.1 ***	23876.9 ***	1964.4 ***
Gender	5.8 *	5.5 *	3.9 *
Context	199.7 ***	12.6 ***	0.1 $^{n.s.}$
Vowel type x Gender	36.5 ***	53.3 ***	190.8 ***
Vowel type x Context	60.3 ***	124.2 ***	20.7 **
Gender x Context	NA	NA	1.1 $^{n.s.}$
Vowel type x Gender x Context	NA	NA	21.0 **

p-values: ***: $p < 0.001$, **: $p < 0.01$, *: $p < 0.05$, $^{n.s.}$: Not significant

The reduced models were subjected to a Wald χ^2 test and the results obtained are presented in Table 2.17. As seen in the table, vowel type has significant effect on F1, F2 and F3 of the vowels. Context effect is significant on F1 and F2. Vowel type and context interaction is significant for F1, F2 and F3. However, interaction of vowel type, gender and context is significant only for F3. Considering this, we decided to

explore the effects of the variables by context types. Hence, LME models for F1, F2 and F3 were built separately for vowels in isolation and vowels in sentence frames.

For each context namely, isolation and sentence, three full LME models were built with F1 , F2 and F3 as dependent variables, and vowel type and gender as the fixed effects. Speaker, onset type, coda type were the random effects. The interaction between vowel type and gender was also explored. The full LME models were subjected to backward reduction of effects using the *step* function of the *lmerTest* package on R Kuznetsova et al. (2017). The reduced model for F1, F2 and F3 in isolation and in sentence form are as follows.

- Isolation
  - F1: vowel + (1|speaker) + (1|onset) + (1|coda)
  - F2: vowel + gender + (1|speaker) + (1|onset) + (1|coda) + vowel:gender
  - F3: vowel + gender + (1|onset) + vowel:gender
- Sentence
  - F1: vowel + gender + vowel:gender + (1|speaker) + (1|onset) + (1|coda)
  - F2: vowel + gender + (1|speaker) + (1|onset) + (1|coda)
  - F3: vowel + gender + (1|onset) + vowel:gender

The reduced models were then subjected to a post-hoc Bonferroni test to see the pairwise variability of vowel formants using the *emmeans* function. The results of the post-hoc Bonferroni tests for vowels produced in isolation is presented in Table 2.18 and those produced in sentence frames is presented in Table 2.19. The results of F3 are not included in the two tables as F3 is not a distinguishing feature for Assamese vowels. However, the results for F3 are included in Table C.4.

Table 2.18: Results of Bonferroni post-hoc tests conducted on vowels produced in isolation by Nalbari speakers.

Contrasts	F1					F2				
	estimate	SE	df	t-ratio	p-value	estimate	SE	df	t-ratio	p-value
a-e	2.31	0.05	1801	43.01	< .0001	-1.26	0.05	1817	-24.53	< .0001
a-ɛ	2.01	0.04	1630	57.12	< .0001	-1.31	0.04	1798	-36.43	< .0001
a-i	3.08	0.04	1727	74.62	< .0001	-1.83	0.04	1819	-44.84	< .0001
a-o	1.93	0.04	1497	48.38	< .0001	1.02	0.04	1720	25.54	< .0001
a-ɔ	1.09	0.03	1036	37.74	< .0001	0.83	0.03	1543	26.79	< .0001
a-u	2.72	0.03	1110	91.91	< .0001	1.24	0.03	1628	39.69	< .0001
a-ʊ	2.17	0.03	1347	76.84	< .0001	1.18	0.03	1709	39.28	< .0001
e-ɛ	-0.3	0.06	1818	-5.38	< .0001	-0.05	0.05	1823	-0.92	1
e-i	0.77	0.06	1823	13.29	< .0001	-0.57	0.06	1828	-10.39	< .0001
e-o	-0.38	0.06	1806	-6.71	< .0001	2.28	0.05	1813	41.76	< .0001
e-ɔ	-1.21	0.05	1814	-22.84	< .0001	2.08	0.05	1830	40.82	< .0001
e-u	0.42	0.05	1814	7.74	< .0001	2.5	0.05	1821	48.64	< .0001
e-ʊ	-0.13	0.05	1821	-2.53	0.3218	2.43	0.05	1826	48.48	< .0001
ɛ-i	1.07	0.04	1823	24.2	< .0001	-0.52	0.04	1830	-12.08	< .0001
ɛ-o	-0.09	0.04	1621	-1.97	1	2.32	0.04	1713	53.84	< .0001
ɛ-ɔ	-0.92	0.04	1611	-26.07	< .0001	2.13	0.04	1783	59.09	< .0001
ɛ-u	0.71	0.04	1743	20.24	< .0001	2.55	0.04	1813	70.95	< .0001
ɛ-ʊ	0.16	0.03	1771	4.83	< .0001	2.48	0.03	1830	71.67	< .0001
i-o	-1.15	0.05	1690	-23.83	< .0001	2.85	0.05	1775	60.4	< .0001
i-ɔ	-1.98	0.04	1707	-48.15	< .0001	2.66	0.04	1820	64.86	< .0001
i-u	-0.35	0.04	1658	-8.39	< .0001	3.07	0.04	1797	73.74	< .0001
i-ʊ	-0.9	0.04	1782	-22.38	< .0001	3	0.04	1830	75.24	< .0001
o-ɔ	-0.83	0.04	1562	-21.06	< .0001	-0.19	0.04	1785	-4.81	< .0001
o-u	0.8	0.04	1791	20.67	< .0001	0.22	0.04	1830	5.76	< .0001
o-ʊ	0.25	0.04	1616	6.37	< .0001	0.16	0.04	1747	4.01	0.0018
ɔ-u	1.63	0.03	1680	58.36	< .0001	0.42	0.03	1800	13.88	< .0001
ɔ-ʊ	1.08	0.03	1767	41.3	< .0001	0.35	0.03	1821	12.31	< .0001
u-ʊ	-0.55	0.03	1714	-19.4	< .0001	-0.07	0.03	1820	-2.24	0.701

Table 2.19: Results of Bonferroni post-hoc tests conducted on vowels produced in sentence frames by Nalbari speakers.

Contrasts	F1					F2				
	estimate	SE	df	t-ratio	p-value	estimate	SE	df	t-ratio	p-value
a-e	2.2	0.05	2034	44.16	< .0001	-1.26	0.05	2014	-24.17	< .0001
a-ɛ	1.91	0.03	1989	58.46	< .0001	-1.13	0.03	1986	-33.73	< .0001
a-i	2.84	0.04	2016	75.4	< .0001	-1.62	0.04	2011	-41.23	< .0001
a-o	1.83	0.04	1952	49.66	< .0001	1.01	0.04	1867	26.52	< .0001
a-ɔ	1.1	0.03	1733	39.47	< .0001	0.82	0.03	1688	29.46	< .0001
a-u	2.59	0.03	1688	90.13	< .0001	1.07	0.03	1782	36.69	< .0001
a-ʊ	2.01	0.03	1873	74.28	< .0001	1.11	0.03	1878	40.36	< .0001
e-ɛ	-0.29	0.05	2027	-5.69	< .0001	0.13	0.05	2019	2.5	0.3522
e-i	0.65	0.05	2024	12.19	< .0001	-0.36	0.06	2020	-6.39	< .0001
e-o	-0.36	0.05	2031	-6.91	< .0001	2.28	0.06	2007	41.11	< .0001
e-ɔ	-1.1	0.05	2029	-22.43	< .0001	2.09	0.05	2022	40.59	< .0001
e-u	0.39	0.05	2033	7.89	< .0001	2.33	0.05	2017	44.48	< .0001
e-ʊ	-0.18	0.05	2027	-3.72	0.0058	2.37	0.05	2021	46.48	< .0001
ɛ-i	0.94	0.04	2028	23.42	< .0001	-0.49	0.04	2019	-11.78	< .0001
ɛ-o	-0.08	0.04	2010	-1.88	1	2.14	0.04	1852	51.67	< .0001
ɛ-ɔ	-0.81	0.03	2006	-24.99	< .0001	1.95	0.03	1985	58.91	< .0001
ɛ-u	0.68	0.03	2024	20.83	< .0001	2.2	0.03	1990	65.63	< .0001
ɛ-ʊ	0.11	0.03	2019	3.45	0.0158	2.24	0.03	2022	69.82	< .0001
i-o	-1.01	0.04	2001	-22.94	< .0001	2.64	0.05	1940	56.96	< .0001
i-ɔ	-1.75	0.04	2013	-46.75	< .0001	2.45	0.04	2019	62.8	< .0001
i-u	-0.25	0.04	1962	-6.55	< .0001	2.69	0.04	1977	66.45	< .0001
i-ʊ	-0.83	0.04	2020	-22.53	< .0001	2.73	0.04	2022	71.2	< .0001
o-ɔ	-0.74	0.04	2001	-20.27	< .0001	-0.19	0.04	1972	-5.06	< .0001
o-u	0.76	0.04	2031	21.26	< .0001	0.05	0.04	2016	1.43	1
o-ʊ	0.18	0.04	2011	5.08	< .0001	0.09	0.04	1901	2.47	0.3809
ɔ-u	1.49	0.03	2005	55.28	< .0001	0.24	0.03	1991	9.01	< .0001
ɔ-ʊ	0.92	0.03	2030	36.71	< .0001	0.28	0.02	2009	11.37	< .0001
u-ʊ	-0.57	0.03	2004	-21.02	< .0001	0.04	0.03	2006	1.46	1

As seen in Tables 2.18 and 2.19, for vowels produced in isolation, post-hoc Bonferroni tests revealed significant F1 differences between all vowels except the pairs /e-ʊ/ and /ɛ-o/ indicating that the vowels in these pairs did not differ in height. Significant F2 differences were observed between all vowels except, /e-ɛ/ and /u-ʊ/ indicating that the vowels in the pairs did not differ in backness. While /e/ and /o/ are higher-mid vowels, /ʊ/ is a near-high vowel and /ɛ/ is a lower-mid vowel in Assamese as reported in Table 1.6 (Mahanta, 2012). Therefore, significant similarity in F1 for the vowel pairs /e-ʊ/ and /ɛ-o/ indicate a probable shift in the positions of the vowels. Similarly, in case of vowels produced in sentence frames, pots-hoc Bonferroni tests revealed significant F1 differences between all vowels except /ɛ-o/. Significant F2 differences were observed between all vowels except /e-ɛ/, /o-u/, /o-ʊ/ and /u-ʊ/. Even though visual examination of the Nalbari vowel plots show significant overlap of the vowels /e-ɛ/ and /o-ʊ/, results of the statistical tests do not reveal any robust evidence of the overlaps. Nevertheless, we check for mergers among vowels in Chapter 3 using vowel merger measures.

2.2.2.5 Vowel formants in Barpeta variety

The acoustic characteristics of the vowel formants in the Barpeta Assamese variety are described in this section. Figure 2.26 presents the corresponding vowel plots on an F1-F2 plane. Figures 2.27, 2.28, 2.29 and 2.30 present the vowel plots on F1-F2 plane separately for gender and context types. Vowels /i/ and /a/ appear to be distinct but the vowel ellipses for /ʊ, o/ show considerable overlap for both genders and contexts. Both the vowel ellipses also show some overlap with /u/. Vowel /o/ shows some overlap with /ɔ/. Among the front vowels, ellipses of /ɛ, e/ show partial overlap for both genders and contexts. The following points are concluded from the visual examination of the plots:

a. there is significant overlap of the vowels /e/ and /ɛ/;

b. there is significant overlap of the vowels /ʊ/ and /o/;

c. while vowel overlaps are seen in all categories, these overlaps are not accidental since they are uniform across context and gender;

d. even though the overlaps are fairly robust visually, this does not confirm that there is statistical significance of the overlaps.

Figure 2.26: Normalized formant frequency plots for Barpeta speakers with 1 standard deviation ellipses.

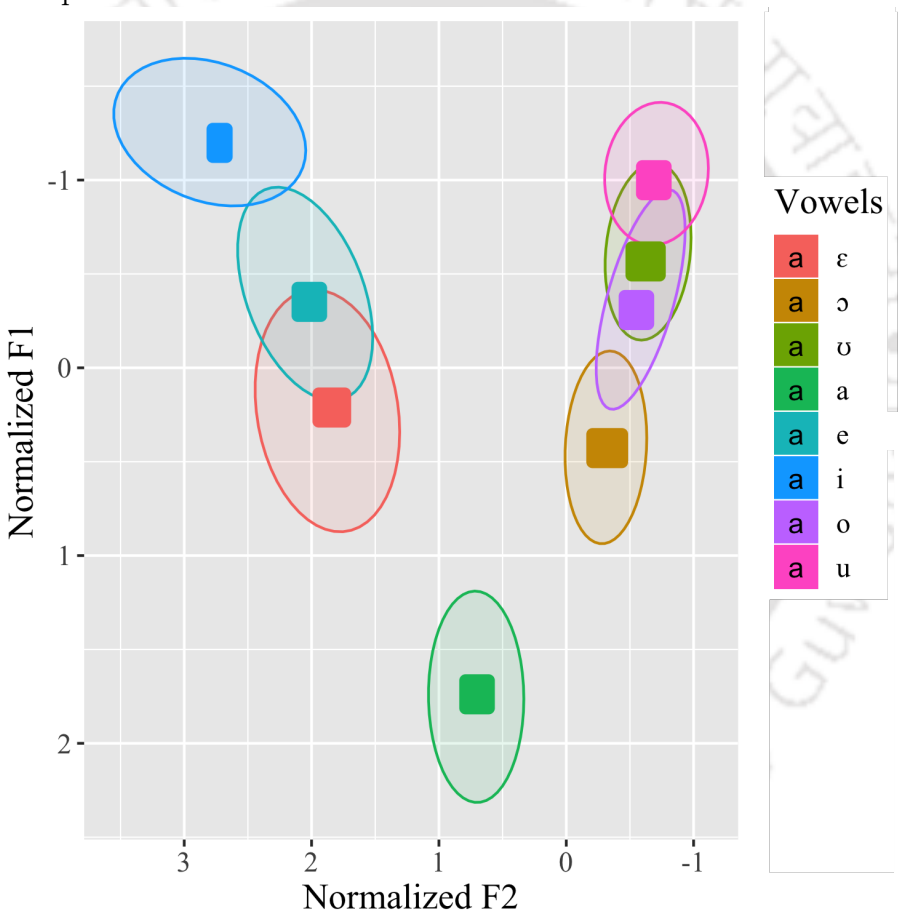


Figure 2.27: Normalized formant frequency plots for Barpeta female speakers with 1 standard deviation ellipses in isolation.

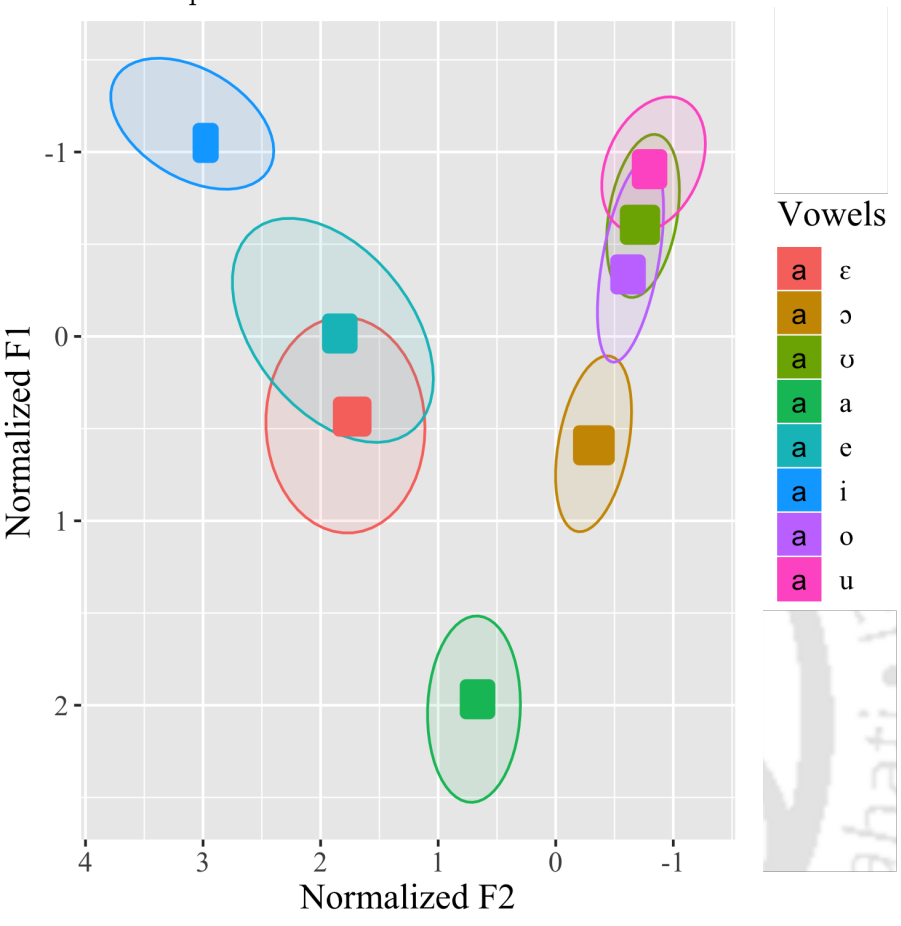


Figure 2.28: Normalized formant frequency plots for Barpeta male speakers with 1 standard deviation ellipses in isolation.

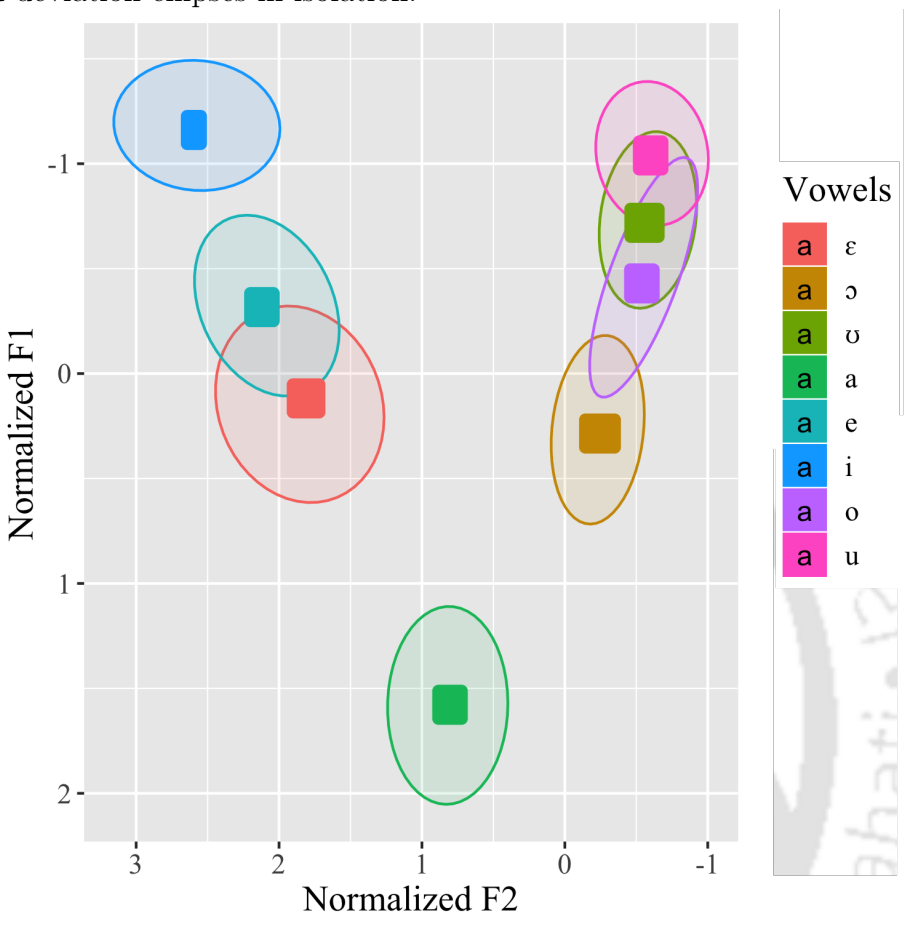


Figure 2.29: Normalized formant frequency plots for Barpeta female speakers with 1 standard deviation ellipses in sentence frames.

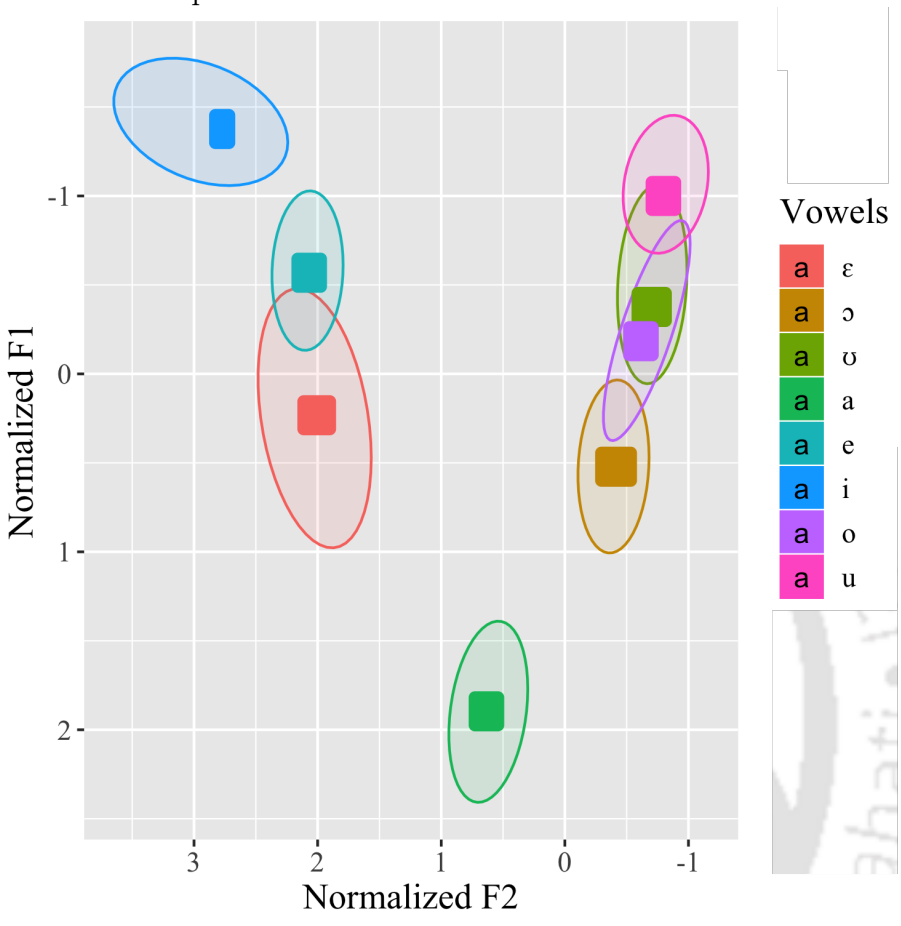


Figure 2.30: Normalized formant frequency plots for Barpeta male speakers with 1 standard deviation ellipses in sentence frames.

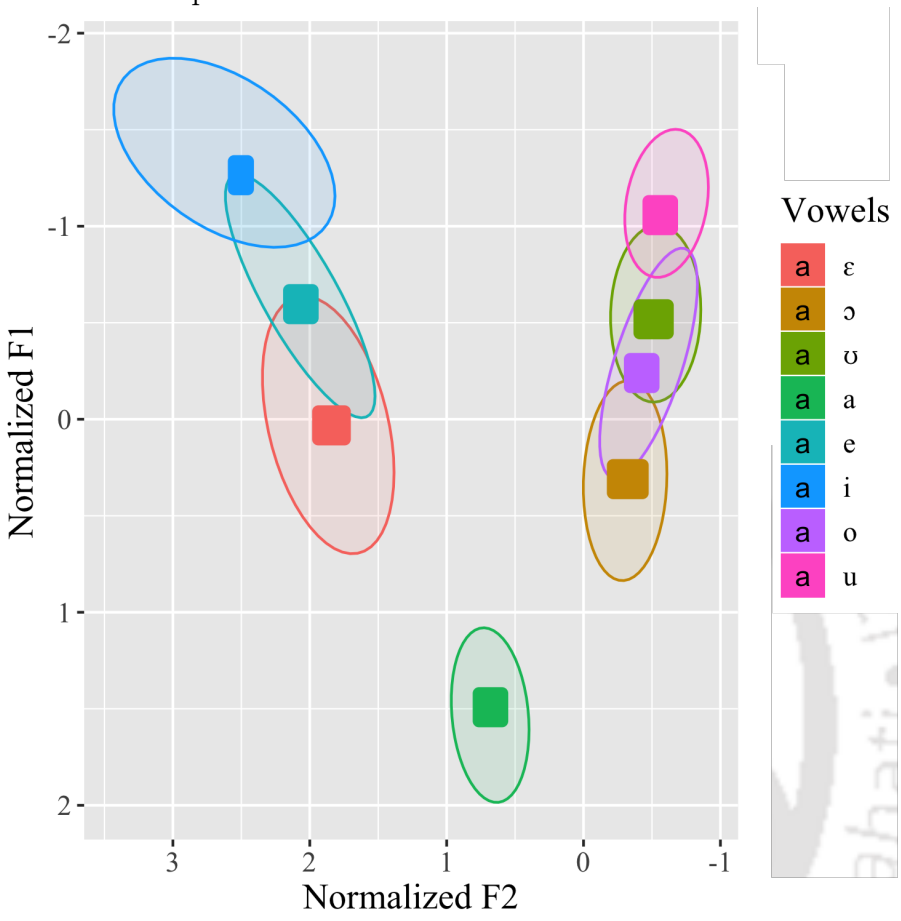


Table 2.20: Lobanov normalized and scaled vowel formant values in Hz for Barpeta speakers in words produced in two contexts.

Context	Vowel	Female		Male	
		F1	F2	F1	F2
Isolation	a	503 (43)	1441 (76)	501 (38)	1415 (66)
	e	386 (33)	1752 (150)	353 (26)	1756 (98)
	ɛ	412 (29)	1689 (170)	400 (40)	1730 (163)
	i	319 (16)	2016 (96)	303 (22)	1983 (101)
	o	365 (29)	1148 (68)	381 (37)	1141 (81)
	ɔ	424 (28)	1230 (73)	422 (35)	1203 (49)
	u	337 (26)	1096 (96)	333 (30)	1108 (83)
	ʊ	350 (26)	1118 (71)	368 (35)	1139 (68)
Sentences	a	503 (39)	1440 (93)	491 (44)	1424 (62)
	e	382 (29)	1844 (93)	357 (29)	1813 (75)
	ɛ	410 (35)	1718 (158)	397 (42)	1725 (163)
	i	329 (24)	1938 (114)	298 (16)	1985 (83)
	o	376 (28)	1123 (71)	384 (38)	1144 (99)
	ɔ	426 (38)	1230 (93)	428 (39)	1195 (69)
	u	333 (22)	1116 (94)	334 (32)	1135 (124)
	ʊ	356 (26)	1126 (87)	370 (39)	1141 (94)

The mean Lobanov normalized and scaled F1 and F2 values for all the vowels produced by the Barpeta speakers are presented in Table 2.20. Two separate LME models were constructed for F1 and F2 in order to statistically analyze vowel formant frequencies. In both cases, gender, vowel type and context were considered fixed effects and speaker, onset type and coda type were considered random effects. The models were further subjected to a Wald χ^2 test and the results obtained are presented

in Table 2.21.

- F1: vowel + gender + context + (1|onset) + (1|coda) + vowel:gender + vowel:context
- F2: vowel + gender + context + (1|onset) + (1|coda) + vowel:gender + vowel:context
- F3: vowel + gender + context + (1|onset) + vowel:gender + vowel:context + gender:context + vowel:gender:context

Table 2.21: Wald χ^2 tests on the LME model for F1, F2 and F3 for Barpeta variety.

	F1	F2	F3
Vowel type	10460.7 ***	15441.8 ***	1072.5 ***
Gender	0.04 $^{n.s.}$	0.3 $^{n.s.}$	0.02 $^{n.s.}$
Context	157.0 ***	74.0 ***	0.2 $^{n.s.}$
Vowel type x Gender	74.5 ***	35.3 ***	24.2 **
Vowel type x Context	49.5 ***	63.3 ***	15.7 *
Gender x Context	NA	NA	NA
Vowel type x Gender x Context	NA	NA	NA

p-values: ***: $p < 0.001$, **: $p < 0.01$, *: $p < 0.05$, $^{n.s.}$: Not significant

The reduced models were subjected to a Wald χ^2 test and the results obtained are presented in Table 2.21. As seen in the table, vowel type has significant effect on F1, F2 and F3 of the vowels. Context effect is significant on F1 and F2. Vowel type and context interaction is significant for F1 and F2. Considering this, we decided to explore the effects of the variables by context types. Hence, LME models for F1, F2 and F3 were built separately for vowels in isolation and vowels in sentence frames.

For each context namely, isolation and sentence, three full LME models were built with F1 , F2 and F3 as dependent variables, and vowel type and gender as

the fixed effects. Speaker, onset type, coda type were the random effects. The interaction between vowel type and gender was also explored. The full LME models were subjected to backward reduction of effects using the *step* function of the *lmerTest* package on R Kuznetsova et al. (2017). The reduced model for F1, F2 and F3 in isolation and in sentence form are as follows.

- Isolation
  - F1: vowel + gender + (1|speaker) + (1|onset) + (1|coda) + vowel:gender
  - F2: vowel + gender + (1|coda) + vowel:gender
  - F3: vowel + (1|speaker)
- Sentence
  - F1: vowel + gender + (1|speaker) + (1|onset) + (1|coda) + vowel:gender)
  - F2: vowel + gender + (1|coda) + vowel:gender
  - F3: vowel + (1|onset)

The reduced models were then subjected to a post-hoc Bonferroni test to see the pairwise variability of vowel formants using the *emmeans* function. The results of the post-hoc Bonferroni tests for vowels produced in isolation is presented in Table 2.22 and those produced in sentence frames is presented in Table 2.23. The results of F3 are not included in the two tables as F3 is not a distinguishing feature for Assamese vowels. However, the results for F3 are included in Table C.5.

Table 2.22: Results of Bonferroni post-hoc tests conducted on vowels produced in isolation by Barpeta speakers.

Contrasts	F1					F2				
	estimate	SE	df	t-ratio	p-value	estimate	SE	df	t-ratio	p-value
a-e	2.05	0.11	1119	18.96	< .0001	-1.32	0.08	1127	-16.42	< .0001
a-ɛ	1.6	0.05	1116	29.52	< .0001	-1.25	0.04	1133	-29.86	< .0001
a-i	3.07	0.08	1123	37.21	< .0001	-2.25	0.06	1129	-35.67	< .0001
a-o	2.08	0.07	1064	31.67	< .0001	1.25	0.05	1131	25.69	< .0001
a-ɔ	1.42	0.04	1070	32.97	< .0001	1	0.03	1132	30.38	< .0001
a-u	2.91	0.05	934	63.39	< .0001	1.47	0.04	1130	40.86	< .0001
a-ʊ	2.44	0.04	1099	56.59	< .0001	1.33	0.03	1135	39.46	< .0001
e-ɛ	-0.45	0.11	1117	-4.17	0.0009	0.07	0.08	1125	0.88	1
e-i	1.02	0.12	1118	8.53	< .0001	-0.92	0.09	1124	-9.87	< .0001
e-o	0.03	0.11	1115	0.28	1	2.57	0.08	1124	30.42	< .0001
e-ɔ	-0.63	0.1	1117	-5.96	< .0001	2.32	0.08	1126	29.22	< .0001
e-u	0.86	0.11	1120	8.14	< .0001	2.79	0.08	1128	35.02	< .0001
e-ʊ	0.39	0.1	1117	3.73	0.0056	2.66	0.08	1126	33.55	< .0001
ɛ-i	1.47	0.08	1123	17.63	< .0001	-0.99	0.06	1126	-15.39	< .0001
ɛ-o	0.48	0.07	1034	6.8	< .0001	2.5	0.05	1127	49.34	< .0001
ɛ-ɔ	-0.17	0.05	1119	-3.4	0.0194	2.25	0.04	1132	57.69	< .0001
ɛ-u	1.31	0.05	1054	24.67	< .0001	2.72	0.04	1135	67.8	< .0001
ɛ-ʊ	0.84	0.05	1123	16.81	< .0001	2.59	0.04	1131	66.36	< .0001
i-o	-0.99	0.09	1099	-10.77	< .0001	3.49	0.07	1124	51.47	< .0001
i-ɔ	-1.64	0.08	1125	-20.66	< .0001	3.25	0.06	1128	52.68	< .0001
i-u	-0.16	0.08	1112	-1.93	1	3.71	0.06	1130	60.05	< .0001
i-ʊ	-0.63	0.08	1126	-7.91	< .0001	3.58	0.06	1127	58.52	< .0001
o-ɔ	-0.66	0.06	1078	-10.46	< .0001	-0.24	0.05	1130	-5.22	< .0001
o-u	0.83	0.06	1122	13.67	< .0001	0.22	0.05	1133	4.74	0.0001
o-ʊ	0.36	0.06	1054	5.68	< .0001	0.09	0.05	1129	1.92	1
ɔ-u	1.49	0.04	1035	37.29	< .0001	0.47	0.03	1132	15.15	< .0001
ɔ-ʊ	1.02	0.04	1124	27.91	< .0001	0.33	0.03	1129	11.57	< .0001
u-ʊ	-0.47	0.04	1025	-11.47	< .0001	-0.13	0.03	1134	-4.24	0.0007

Table 2.23: Results of Bonferroni post-hoc tests conducted on vowels produced in sentence frames by Barpeta speakers.

Contrasts	F1					F2				
	estimate	SE	df	t-ratio	p-value	estimate	SE	df	t-ratio	p-value
a-e	1.83	0.1	1142	18.64	< .0001	-1.39	0.08	1147	-17.58	< .0001
a-ɛ	1.52	0.05	1121	30.62	< .0001	-1.13	0.04	1154	-27.32	< .0001
a-i	2.69	0.07	1137	35.95	< .0001	-1.85	0.06	1149	-30.07	< .0001
a-o	1.78	0.06	1090	30.11	< .0001	1.18	0.05	1152	25.12	< .0001
a-ɔ	1.28	0.04	1070	32.4	< .0001	1	0.03	1153	30.83	< .0001
a-u	2.59	0.04	880	62.54	< .0001	1.31	0.03	1138	37.84	< .0001
a-ʊ	2.18	0.04	1093	55.33	< .0001	1.24	0.03	1154	37.42	< .0001
e-ɛ	-0.31	0.1	1138	-3.13	0.0502	0.26	0.08	1144	3.27	0.0307
e-i	0.86	0.11	1136	7.94	< .0001	-0.46	0.09	1143	-4.99	< .0001
e-o	-0.04	0.1	1140	-0.41	1	2.57	0.08	1143	31.14	< .0001
e-ɔ	-0.55	0.1	1137	-5.79	< .0001	2.39	0.08	1147	30.57	< .0001
e-u	0.77	0.1	1140	8.04	< .0001	2.71	0.08	1147	34.7	< .0001
e-ʊ	0.35	0.1	1137	3.69	0.0066	2.63	0.08	1146	33.86	< .0001
ɛ-i	1.17	0.08	1142	15.39	< .0001	-0.72	0.06	1145	-11.4	< .0001
ɛ-o	0.27	0.06	1093	4.12	0.0011	2.31	0.05	1147	46.76	< .0001
ɛ-ɔ	-0.24	0.05	1134	-5.15	< .0001	2.12	0.04	1153	55.28	< .0001
ɛ-u	1.07	0.05	1047	22.3	< .0001	2.44	0.04	1154	62.63	< .0001
ɛ-ʊ	0.66	0.05	1137	14.34	< .0001	2.37	0.04	1152	61.86	< .0001
i-o	-0.9	0.08	1129	-10.89	< .0001	3.03	0.07	1143	45.96	< .0001
i-ɔ	-1.41	0.07	1142	-19.55	< .0001	2.84	0.06	1149	47.31	< .0001
i-u	-0.1	0.07	1125	-1.3	1	3.16	0.06	1150	52.74	< .0001
i-ʊ	-0.51	0.07	1142	-7.08	< .0001	3.09	0.06	1147	51.78	< .0001
o-ɔ	-0.51	0.06	1114	-9	< .0001	-0.19	0.05	1152	-4.15	0.001
o-u	0.81	0.05	1142	14.9	< .0001	0.13	0.05	1153	2.89	0.109
o-ʊ	0.39	0.06	1105	6.88	< .0001	0.06	0.04	1150	1.26	1
ɔ-u	1.32	0.04	1033	36.8	< .0001	0.32	0.03	1143	10.71	< .0001
ɔ-ʊ	0.9	0.03	1142	27.29	< .0001	0.24	0.03	1149	8.73	< .0001
u-ʊ	-0.41	0.04	1009	-11.28	< .0001	-0.07	0.03	1148	-2.43	0.4226

As seen in Tables 2.22 and 2.23, for vowels produced in isolation, post-hoc Bonferroni tests revealed significant F1 differences between all vowels except the pairs /e-o/ and /i-u/ indicating that the vowels in these pairs did not differ in height. Significant F2 differences were observed between all vowels except, /e-ɛ/ and /o-ʊ/ indicating that the vowels in the pairs did not differ in backness. Similarly, in case of vowels produced in sentence frames, pots-hoc Bonferroni tests revealed significant F1 differences between all vowels except /e-ɛ/ and /i-u/. Significant F2 differences were observed between all vowels except /o-u/, /o-ʊ/ and /u-ʊ/. Both /e/ and /o/ are higher-mid vowels and /i/ and /u/ are high vowels in Assamese Therefore, the significant similarity in F1 is predictable. /ɛ/ being a lower-mid vowel, it should ideally maintain a difference in height from /e/. However, the same is not seen in case of vowels produced in sentence frame. Visual inspection of the plot in Figure 2.26 indicated complete overlap of the ellipses of the vowels /e-ɛ/ and /o-ʊ/. However, results of the statistical tests do not reveal any robust evidence of the overlaps. Nevertheless, we check for mergers among vowels in Chapter 3 using vowel merger measures.

2.3 Vowel duration

In order to explore the interaction between vowel duration and vowel quality, we conducted a set of statistical tests. For each of the five varieties in this study we built five separate LME models with vowel type and context as fixed effects. Speaker, gender, onset type and coda type were the random effects. The interaction between vowel type and context was also explored. The full LME models were subjected to backward reduction of effects using the *step* function of the *lmerTest* package on R Kuznetsova et al. (2017). For all five varieties the reduced model was as shown below.

duration – vowel + context + vowel:context + (1|speaker) + (1|onset) + (1|coda)

In order to see the significance of contrast between vowel pairs of each variety, models for each region were subjected to a Bonferroni post-hoc test using the *emmeans* function on R (Lenth, 2019). While almost all vowel pairs across the varieties showed significant differences within, the pairs listed in Table 2.24 showed no significant difference between them. However, there was no consistent pattern that justified the difference between the duration of the vowel pairs of Assamese spoken in the five regions. Hence, we decided to visually examine only the pairs of vowels that overlapped, as shown in Section 2.2.2. Violin plots were generated for vowel duration for four vowel contrasts, namely, /e-ɛ/, /u-ʊ/, /ʊ-o/ and /o-ɔ/. The violin plots are provided as Figure 2.31 to Figure 2.34. Visual examination of the plots do not indicate any significant difference in duration between the vowels of the overlapping pairs. Hence, in the following parts of this dissertation we do not consider any further exploration of the duration feature of vowels.

Table 2.24: Vowel pairs not significantly different in duration.

Tinsukia	Jorhat	Nagaon	Nalbari	Barpeta
ɛ-e	ɛ-a	ɛ-e	ɛ-ɔ	ɛ-i
ɛ-i	ɔ-o	ɛ-i	ɛ-ʊ	ɔ-ʊ
ɔ-o	ʊ-e	ɔ-a	ɛ-a	ɔ-o
ɔ-u	e-u	ɔ-o	ɛ-o	ɔ-u
ʊ-o		ɔ-u	ɔ-a	ʊ-o
a-e		ʊ-u	ɔ-o	ʊ-u
a-i		e-i	ʊ-i	e-i
e-i		o-u	a-o	o-u
o-u			e-u	
			i-u	

Figure 2.31: Violin plots showing duration of vowel pair /e-ɛ/ by context.

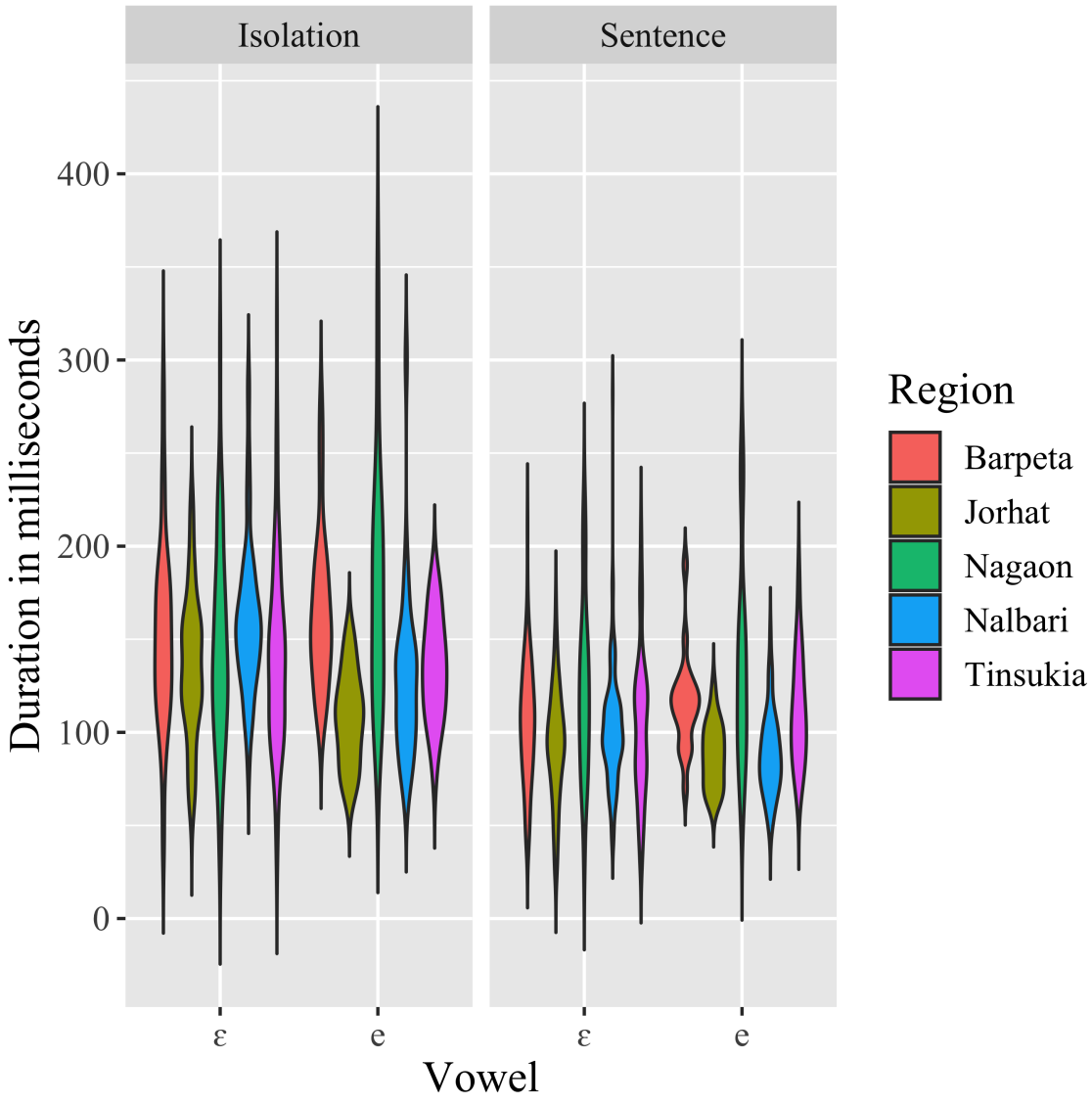


Figure 2.32: Violin plots showing duration of vowel pair /u-ʊ/ by context.

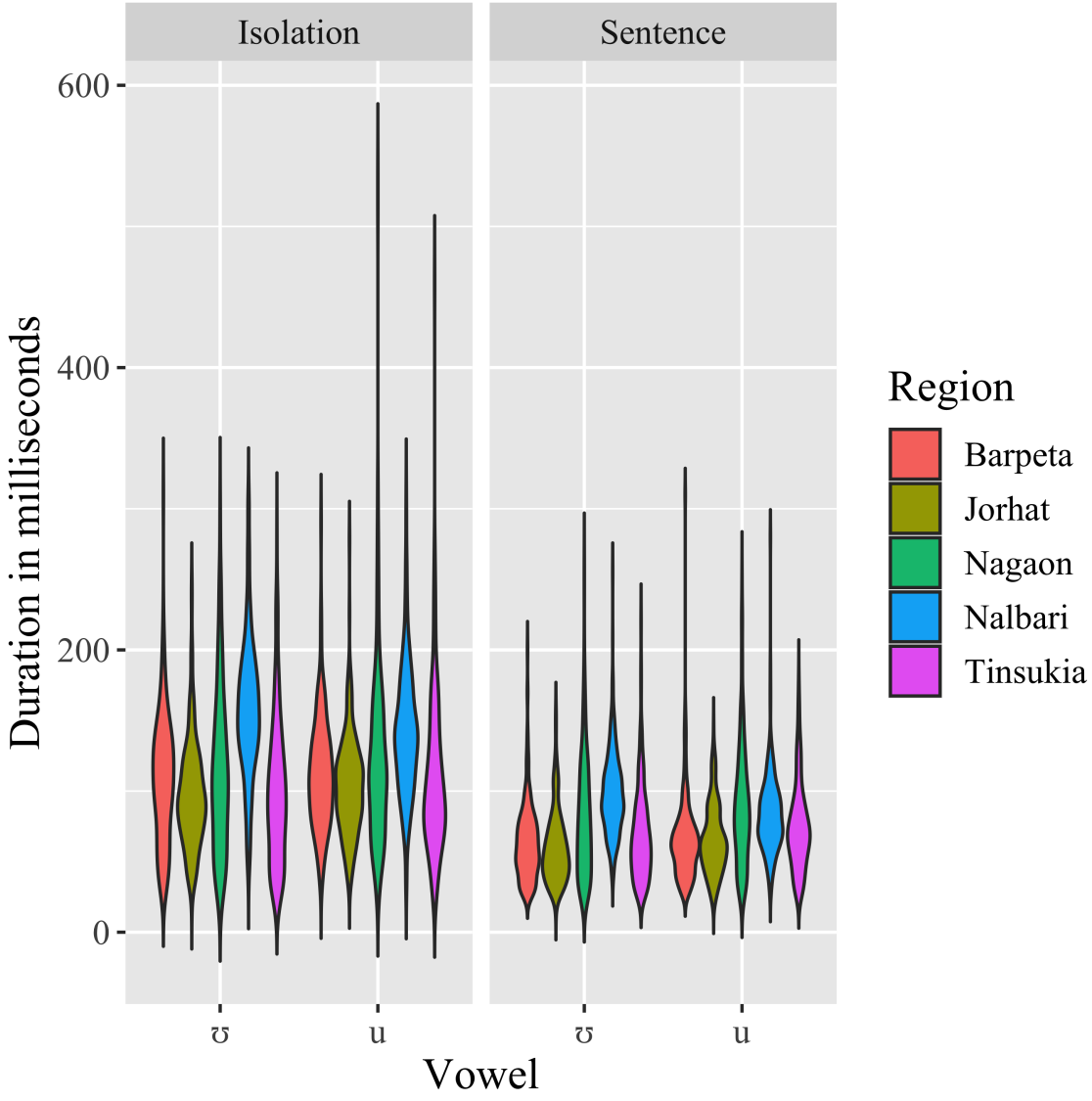


Figure 2.33: Violin plots showing duration of vowel pair /ʊ-o/ by context.

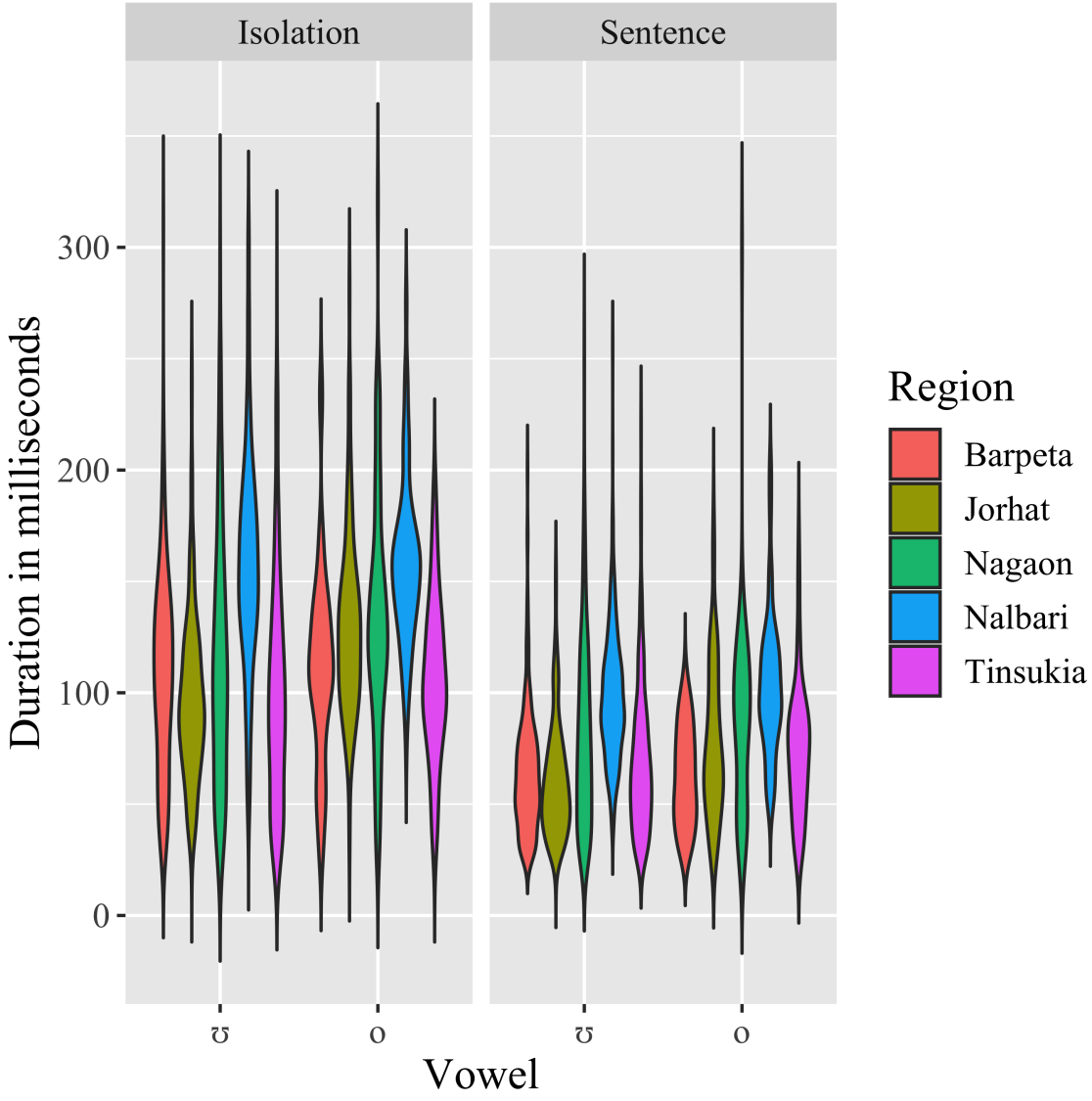
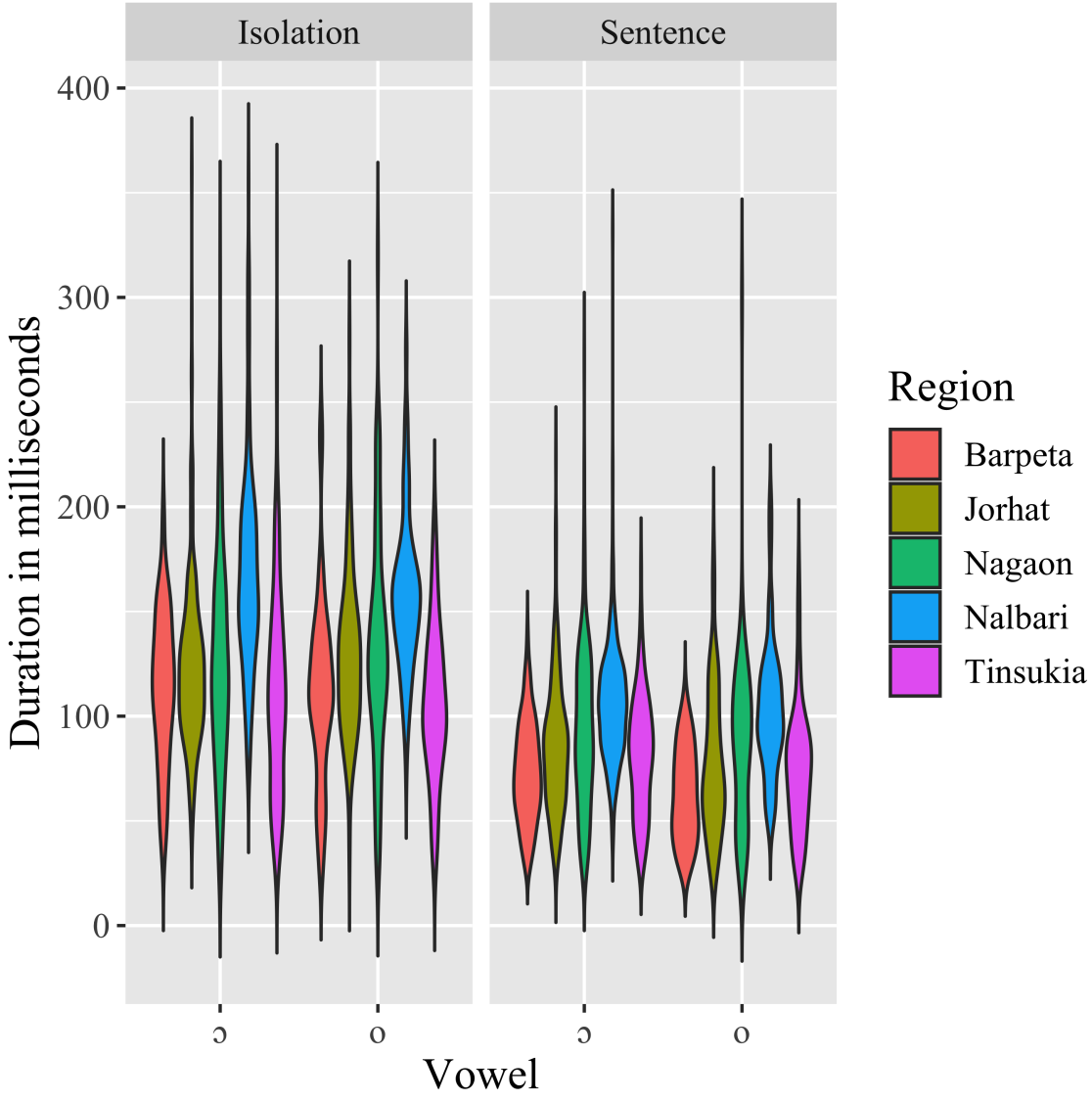


Figure 2.34: Violin plots showing duration of vowel pair /o-ɔ/ by context.



2.4 Conclusion

This chapter looked into the formant frequencies of vowels in Assamese dialects. The vowels were analyzed in three ways: (a) using a visual examination of formant frequency plots with 1 standard deviation ellipses, (b) comparing formant frequency values and (c) using exploratory statistical methods to gauge the significance of dis-

tinctiveness in F1 and F2 among the vowels. The formant frequency plots gave us a visual representation of the position of vowels for each variety in their respective vowel spaces. While the positions are shown by the formant frequency means, the ellipses show the standard deviation of their distribution within a vowel space. It was also observed that in all the five varieties, vowel ellipses overlap with each other to varying degrees. Depending on the degree of these overlaps, vowel ellipses indicate the possibility of mergers between the vowels. From the visual examination of the plots, the overlaps observed are categorized into three different kinds with respect to the degree in which the vowel ellipses overlapped. These are:

a. Complete overlap (Overlap 1), where two or more vowel ellipses are superimposed on each other;

b. Partial overlap (Overlap 2), where the overlap seems to cover at least 40% of the ellipses, in other words, where the outer bound of a vowel ellipse reaches the mean of the other vowel;

c. Minor overlap (Overlap 3), where the overlap covers only a small area of the ellipses.

It is to be noted that these categories of overlap are purely subjective based on visual examination and can not be used as a confirmed measure of overlap. As such, the overlaps observed for the vowels require statistical validity as well as other tested overlap measures to confirm their status. While the statistical validity has been explored in this chapter, the overlap measures have been dealt with in Chapter 3. The overlaps observed in the vowel plots of the five varieties of Assamese are summarized in Table 2.25.

Table 2.25: Summary of vowel ellipse overlap in five varieties of Assamese.

	Tinsukia	Jorhat	Nagaon	Nalbari	Barpeta
Complete overlap	/u-ʊ/	/u-ʊ/	/u-ʊ/	/ʊ-o/ /e-ɛ/	/ʊ-o/
Partial overlap	/ʊ-o/ /e-ɛ/	/u-o/ /ʊ-o/	/u-o/ /ʊ-o/	/ʊ-u/	/e-ɛ/ /ʊ-u/
Minor overlap	/o-ɔ/ /u-o/		/i-e/	/o-ɔ/ /u-o/	/o-ɔ/ /u-o/

As mentioned earlier, the degree of overlap of vowel ellipses may indicate the possibility of mergers between the vowels. Considering this, the Overlap 1 (complete overlap) may be an indication of mergers of the concerned vowels where one vowel is completely replaced by or neutralized to another. Overlap 2 (partial overlap) on the other hand, may be an indication of ongoing or near-mergers where mergers have not been completed yet. While the vowel ellipses provided us an impressionistic view of possible mergers in the varieties, the vowel means showed us their respective positions in the vowel space. As observed in the plots, as well as from the formant frequencies of vowels for each variety presented in respective tables, it is evident that the vowels completely overlapped as the ellipses also did not differ much in formant frequencies and were positioned either very close to or superimposed on each other in the vowel space. The differences in formant frequency means of vowels with Overlap 2 were comparatively higher and the respective vowels were positioned at some distance from each other in the vowel space.

To confirm these visual observations, exploratory statistical tests were conducted on the vowel frequencies. The results showed that in Tinsukia and Jorhat varieties the F1 and F2 values for the vowels /u-ʊ/ are not statistically significantly different. In case of Nalbari and Barpeta varieties, although complete overlap of ellipses was

observed for the vowels /ʊ-o/ and /e-ɛ/ results of statistical tests did not confirm the same. Post-hoc Bonferroni tests for the concerned vowels indicated significant separation in either of the two parameters, F1 or F2. In case of Nagaon, we also see overlap of the three back vowels /u, ʊ, o/. However, again like Nalbari and Barpeta, post-hoc Bonferroni tests revealed no separation only in backness (F2) of these vowels. The vowels were distinct in height (F1). We had also seen some partial or minor overlap of /u-o/ vowels in all varieties. Results of statistical tests showed that the two vowels showed significant similarity only for F2 in all variety. The two vowels being back vowels, this result is predictable. A complete loss or absence of phonetic distinctiveness can be indicated if both F1 and F2 values of the concerned vowels show no statistically significant difference. The results of post-hoc Bonferroni tests for 4 vowel pairs that showed complete or partial overlap are presented in Table 2.26 where $*$ represents 'statistically significant distinctiveness' and ! represents 'no significant distinction'.

Table 2.26: Summary post-hoc Bonferroni test results for 4 vowel pairs.

Region	Context	/e-ɛ/		/u-ʊ/		/u-o/		/ʊ-o/	
		F1	F2	F1	F2	F1	F2	F1	F2
Tinsukia	Iso	*	*	!	!	*	*	*	!
	Sen	*	*	!	!	*	!	*	!
Jorhat	Iso	*	*	!	!	*	*	*	*
	Sen	*	*	!	!	*	!	*	!
Nagaon	Iso	*	*	*	!	*	!	*	!
	Sen	*	*	*	!	*	!	*	!
Nalbari	Iso	*	!	*	!	*	*	*	*
	Sen	*	!	*	!	*	!	*	!
Barpeta	Iso	*	!	*	*	*	*	*	!
	Sen	!	*	*	*	*	!	*	!

*: distinct; !: not distinct

Overall, a few observations are evident from the analysis of the production experiment. Firstly, in terms of both F1 and F2, only vowels /i/ and /a/ are statistically significantly different in all varieties. Secondly, Tinsukia and Jorhat varieties show complete overlap of vowel ellipses of the vowels /u/ and /ʊ/. Formant frequencies as well as post-hoc Bonferroni tests also reveal robust evidence of overlap. This may be a case of possible merger of the two vowels. Third, Nalbari and Barpeta varieties show complete vowel ellipse overlap of two pairs /e-ɛ/ and /ʊ-o/. However, statistical significance of no separation is seen only for backness and not for height, and fourthly, Nagaon shows significant vowel ellipse overlap of /u, ʊ, o/. However like Nalbari and Barpeta varieties the overlaps are not found to be statistically significant.

The above points lead us to the following observed trends among the Assamese vowels.

a. The case of vowel /ʊ/: The vowel /ʊ/ reported for standard Assamese (Mahanta, 2012) seems to be distinct in none of the varieties of Assamese. The vowel showed neutralization with /u/ as in Tinsukia, Jorhat and Nagaon varieties, or with /o/ as in Nalbari and Barpeta varieties.

b. EA varieties vs. WA varieties: While the study focused on individual varieties, within the larger dialectal divisions that they belonged, similar trends have emerged. The two EA varieties Tinsukia and Jorhat showed evidences of /u-ʊ/ merger and WA varieties Nalbari and Barpeta show evidences of /ʊ-o/ merger. Additionally, unlike the EA varieties, both the WA varieties show evidences of merger of another vowel pair, i.e /e-ɛ/. The CA variety, represented by Nagaon, on the other hand, although shows more evidences of a /u-ʊ/ merger, the considerable overlap of /u, ʊ, o/ vowel ellipses suggests that the variety shows an intermediate trend of vowel merger

c. Asymmetrical vowel movement: The different trends of vowel overlap in EA and WA varieties show that the direction of vowel movement is different for the two varieties. In the EA variety, /ʊ/ raises to merge with the /u/ vowel, but in the WA variety the same vowel lowers towards the direction of the /o/ vowel. Again in the same variety /ɛ/ raises to merge with /e/. Thus, there is asymmetrical movement of vowels both between and within the varieties.

It may be noted that the Assamese writing system makes a distinction among /ʊ/, /o/ and /u/ in the orthography. Hence, the participants of this study, proficient readers of Assamese, could see the difference among the vowels in the orthography while reading out the words from the datalist. They nevertheless, made no distinction between the vowels in production. On the other hand, /e/ and /ɛ/ contrast is not distinctly marked in the Assamese orthography, even though there are at least five minimal pairs showing this contrast. While /ɛ/ is more commonly occurring in Assamese, /e/ occurs mostly as a result of tongue root harmony in the language (Mahanta,

2008). As discussed in Section 1.5 of the previous chapter, scholars have reported the reduced vowel inventories for Assamese varieties (Taid, 1988; Deka, 2007). The analysis presented in this chapter provides strong acoustic and statistical evidences that support these reports. However, in order to draw more conclusive evidence for vowel merger in the varieties, a few more measures for vowel merger to determine their nature are calculated and discussed in the following chapter.

Chapter 3

Vowel merger

3.1 Introduction and literature review

In the previous chapter it was observed that Assamese varieties displayed considerable overlap between vowel ellipses indicating vowel mergers in the varieties. While the Tinsukia and Jorhat varieties have a complete overlap of /u-ʊ/ vowels, Nalbari and Barpeta displayed considerable overlap of /ʊ-o/ ellipses. The two WA varieties also showed overlap of front vowels /e-ɛ/ ellipses. On the other hand, Nagaon showed overlap of the three back vowels /u, ʊ, o/. However, to determine whether these overlaps are mergers or not, some theoretical considerations and finer acoustic measurements required to be taken. Hence, in this chapter we discuss the vowel overlaps in light of four measures for vowel mergers. These are: Euclidean Distance, Pillai Bartlett Trace, Spectral Overlap Assessment Metric (SOAM) and Vowel Overlap Assessment with Convex Hulls (VOACH). In Section 3.1.1 the theoretical considerations in identifying the mechanisms of mergers observed in languages have first been discussed. Section 3.1.2 discusses the four measures of vowel mergers employed and then Section 3.2 discusses the results of the merger measures.

3.1.1 Mechanisms of mergers

A situation of language change can have two alternative outcomes: chain shifts and mergers, Both involve the encroachment of one phoneme into the phonological space of another. A 'merger' occurs when two or more contrastive phonemes are replaced by a single phoneme. If the second phoneme changes without bringing about a change in the distinction between the two phonemes, then the result is a chain shift. Mergers and chain shifts both affect two sounds from the same phonological neighbourhood. However, while the phonemic contrast between the two sounds is lost in merger, chain shifts maintain the distinction between the two sounds. Chain shifts have been distinguished into two main types: 'push' chain and 'drag' chain. In a 'push' chain, one sound moves into the space occupied by another sound. This in turn, causes the second sound to move so that the distance between the two is maintained. The 'drag' chain occurs when the movement of one sound creates an opening and another sound moves to occupy the opening (Martinet, 1952). The most well-known example of chain shift is the Northern Cities Chain Shift (NCCS) seen in American dialects. As mentioned in Section 2.1.2.2, in these dialects the low and low-mid vowels rotate clockwise causing /æ/ to be raised and fronted, /ɛ/ and /ʌ/ to be backed, /ɔ/ to be lowered and /ɑ/ to be lowered and fronted (Labov et al., 1972).

In contrast, a merger occurs when the the distinction between two phonemes is lost. Traditionally, mergers have been treated as a structural phenomenon. When the phonological structures of a language are compared at two points of time, it is observed that where in the earlier structure there were two sounds, the distinction is lost in the later structure. W. Maguire et al. (2013, p.230) explains: *if we have two phonemes, A and B, A and B merge if A moves into the phonetic space of B, if B moves into the phonetic space of A, or if A and B move into a single new phonetic space, C.* As Labov (1994, p.229) mentions, "the most spectacular example" of multiple vowel merging is the eventual merging of nine phonemes of Ancient Greek (/i, iː, e, ɛː, a, oi,

yi, y, y:/) as the Modern Greek /i/, through a history of fronting, raising, unrounding and loss of glides and length distinctions. Another example where multiple mergers occurred as a process is that of Classical Latin, where when the system of distinctive vowel length collapsed in transition to Vulgar Latin, regular mergers took place in all varieties (Johnson, 2007, p.3) as shown in Table 3.1.

Table 3.1: Vulgur Latin mergers following loss of Classical Latin vowel length (Johnson, 2007)

<table>
<tr><th>Classical Latin</th><th>Western Romance</th><th>Romanian</th><th>Sicilian</th><th>Sardinian</th></tr>
<tr><td>i:</td><td>i</td><td>i</td><td rowspan="3">i</td><td rowspan="2">i</td></tr>
<tr><td>i̯</td><td rowspan="2">e</td><td rowspan="2">e</td></tr>
<tr><td>e:</td><td rowspan="2">e</td></tr>
<tr><td>e̯</td><td>ɛ</td><td>ɛ</td><td>e</td></tr>
<tr><td>a</td><td rowspan="2">a</td><td rowspan="2">a</td><td rowspan="2">a</td><td rowspan="2">a</td></tr>
<tr><td>a:</td></tr>
<tr><td>o̯</td><td>ɔ</td><td rowspan="2">o</td><td>o</td><td rowspan="2">o</td></tr>
<tr><td>o:</td><td rowspan="2">o</td><td rowspan="2">u</td></tr>
<tr><td>u̯</td><td rowspan="2">u</td><td rowspan="2">u</td></tr>
<tr><td>u:</td></tr>
<tr><td>10 distinct vowels</td><td>7 d.v</td><td>6 d.v</td><td>5 d.v</td><td>5 d.v</td></tr>
</table>

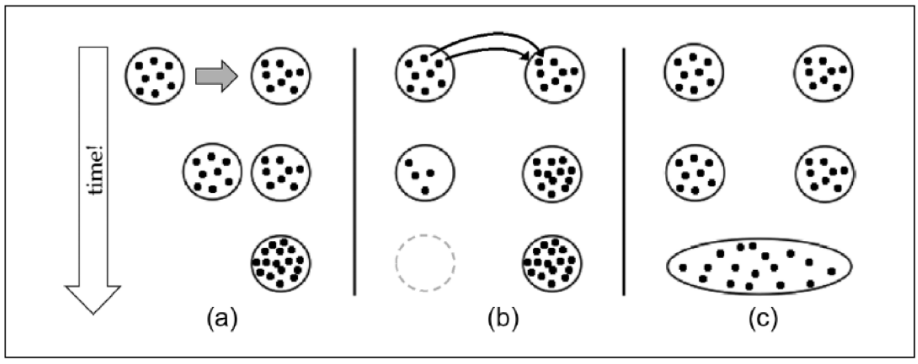
Martinet (1957), describing the functions of a phoneme, is of the view that two phonemes that frequently distinguish between two words are unlikely to undergo a change since they have high 'functional load'. A homonymy created by merger may make comprehension difficult. Martinet says that most vowel shifts are a way of avoiding merger and those that do occur are exceptional cases. This hypothesis that sound changes do not operate independently of communicative needs was challenged by King (1967) who investigated four cases of sound change from historical German:

Old Icelandic to Modern Icelandic, Old Saxon to Middle Low German, Middle High German to Modern German, and Middle High German to certain modern Yiddish dialect. He came to the conclusion that functional load is the least important factor in sound change. Another example challenging the functional load hypothesis is provided by Spa (2015) in his review of Martinet. Present-day French has the phonemes /e/ and /ɛ/ which display a high functional load. However, native speakers tend to coalesce the two phonemes into one phoneme having two allophones: [e] occurs in an open syllable and [ɛ] occurs in a closed one. Similar behaviour is observed for the phonemes /œ/ and /ø/ which are characterized by the same tongue height. In both cases, the functional load of the pairs is fairly low as compared to a third pair, /ɔ/ and /o/, which have a well preserved contrast. Martinet (1957) however shows that the phonemes of a language have a tendency to form regular configurations of the vowel space, such as, rectangular, triangular, or square. To adopt these configurations, the principle of 'economie' or 'least effort' is employed. According to this principle, vowel spaces are related to articulatory space in a way that phonological systems avoid potential crowding in relation to the physical space available for the tongue's movements in the mouth. Johnson (2007) believes that since phonological subsystems have a pressure to achieve symmetry, vowel mergers can create a more symmetrical subsystem and also relieve articulatory crowding at the same time.

Hoenigswald (1965) reports on several patterns of change that involve merger. In 'unconditioned merger', phonemic contrast between phonemes is lost in all phonological environments and only a single phoneme remains. In 'conditioned merger', separate phonemes are retained in the inventory and the merger appears only in certain contexts. American English illustrates the two types of mergers. In case of *cot* and *caught*, the distinction between /ɒ/ and /ɔ/ is completely lost. This is a case of 'unconditioned merger'. Alternatively, in case of *pin* and *pen* merger, the distinction is maintained, except in preceding nasal consonants. This is a case of 'conditioned

merger'. Merger may also refer to a synchronic state where a phonemic distinction which is found in some varieties of a language is lost or does not appear in one or more varieties of the same language. The phonemic distinction may be either historical or contemporary and was, in one form or another, characteristic of earlier forms of the language (Hoenigswald, 1965). Examples can be given of the NURSE–NORTH merger in some varieties of Tyneside English, characterized by a merging of the vowels /ɜː/ and /ɔː/. This merger came about through a combination of internal, gradual phonetic drift and externally motivated transfer from one phonetic target to another (W. N. Maguire, 2008). Another example is the FOOT–GOOSE merger of Scottish Standard English, which is characterized by a merger of /ʊ/ and /uː/ vowels. This merger did not come about because of a historical merger of the two but rather as a result of the Scottish speakers not being able to make a contrast between the two vowels when they learned English in the seventeenth century (W. Maguire et al., 2013). As such, at least three types of segment merging mechanisms have been proposed in the literature. As illustrated in Figure 3.1, Trudgill & Foxcroft (1978) mention (a)merger by approximation, and (b) merger by transfer, while Herold (1990) proposes a third, (c) merger by expansion. The three processes are described in greater detail below.

Figure 3.1: Schematic representations of three mechanisms of merger (Diagram taken from Dinkin (2016, p.164))



.

Circles represent the phonetic value of a phoneme and dots represent lexical items containing the phoneme.

1. Merger by approximation: Trudgill & Foxcroft (1978) in their study of vocalic mergers in East Anglia, describe how the vowel /ɔː/ in words that descended from Middle English (ME) (like *moan, road, nose* etc) had "gradually" merged with the ME diphthong /ʌu/ (words like *mown, rowed, knows* etc) and became /ɔu/. The authors pointed out that while some speakers retain the distinction, most were able to *gradually approximate the two vowels by bringing them closer together phonetically until, finally, they become identical* (Trudgill & Foxcroft, 1978, p.72) and called this tactic "merger by approximation". Several other cases of this merger had been documented. For example, the American English *cot* and *caught* merger characterized by a backing of /ɒ/ and lowering of /ɔ/. While in American English one vowel stays put while the other vowel shifts to join it, in East Anglia the final vowel can end up somewhere between the two historic vowel classes. In other words, the two phonemes move closer to each other in phonetic space and finally merge somewhere between their original positions. This merger results when the phonetic space between the two phonemes become small enough to no longer maintain a contrast between them. In either case, the overall distribution of the final vowel is approximately the same size as the original ones, leaving a gap in the F1-F2 space. According to Guy (1990), these are "spontaneous" and "internally induced" changes that stem from language-internal pressures. Since this merger is gradual, intermediate stages should be observable and may reach a stage towards the end of the change, or constitute 'near merger', where the merger may never complete. For example, in case of the NEAR-SQUARE merger in New Zealand English, different members of the speech community are at different stages of the gradually happening merger. As a result, speakers were exposed to patterns which were different than their own. This leads to a potential mismatch between their production and perception of the merger or the distinction (Hay et al., 2010).

2. Merger by transfer: Merger by transfer is a phenomenon when words containing one category of phonemes are gradually transferred to another category. According to Guy (1990), these mergers are often a result of external factors like dialect or language contact and can be considered as a type of borrowing. The end result is that all the words containing one category are completely replaced by another so that only one phoneme remains. Consequently, the merger is targeted and gradual and phonetically intermediate stages do not exist. Hence, 'near merger' is not possible. However, there could be cases of 'partial mergers' where only some members of a lexical set are merged with another. This may be part of an ongoing 'merger by transfer' or the result of a partial (but no longer ongoing) lexical diffusion. Trudgill & Foxcroft (1978) mention the Long Mid mergers of English which were absent in the traditional dialects of East Anglia. While the vowel /æi/ in these dialects occurred only in items that descended from Middle English (ME) /ai/, items that descended from ME /a:/ had the vowel /e:/ (word pairs like *days-daze, maid-made* etc). These distinctions were gradually being lost by a "transfer" of lexical items from /e:/ to /æi/. Another example is the *merry-marry* merger in Southern Quebec, whereby the merger of mid and low front vowels preceding /r/ did not begin as a conditioned change but rather as a lexically-specific one. This merger gradually spread to all phonologically similar environments (Baxter, 2010, p.15-18). W. Maguire et al. (2013, p.232-233) point out that since merger by transfer is lexically gradual, speakers will be at different stages in the change. Although individuals may transfer words from one phoneme category to the other, it is not the case that they would instantly forget, or even stop using the old value. Since speakers will be in contact with others who have different patterns and may have phonological representations which have changed during their own lifetimes, such mergers may lead to a production and perception anomaly.

3. Merger by expansion: While studying the American low back merger /ɒ-ɔ/ (*cot-caught, lot-thought*), Herold (1990) had noticed that while the older generation in Tamaqua retained the distinction, the younger generation had lost it. The phonetic range of the younger speakers' merged phoneme was found to be very wide, acoustically combining ranges of both original phonemes. Herold coined the term 'merger by expansion' to describe the change in Tamaqua. The categorical boundaries between two phonemes disappear and they constitute a larger categorical space, as illustrated in panel (c) of Figure 3.1. This merger is lexically and phonetically abrupt and there are no intermediate stages in the change. Since the merger "completed within a single generation" (Herold, 1997, p.185), speakers who use the whole phonetic space for a single phoneme co-existed with speakers who have two phonemes in the same phonetic space. As with the other kinds of merger, this co-existence of different systems may lead to a mismatch between production and perception for some speakers (Hay et al., 2010).

Table 3.2: Summary of consequences of three mergers types.

	Approximation	Transfer	Expansion
1.	below the level of consciousness	above the level of consciousness	below the level of consciousness
2.	lexically abrupt	lexically gradual	lexically abrupt
3.	phonetically gradual	phonetically abrupt	phonetically abrupt
4.	intermediate forms observed	no intermediate forms	no intermediate forms

Based on these mechanisms, the nature of vowel merger in the varieties of Assamese is investigated in this chapter. Also determined is the merger mechanism that fits the best for the Assamese varieties. In the analysis of vowels, the five Assamese varieties showed overlapping of vowel ellipses in the F1-F2 space. While visual examination and statistical analyses have largely corroborated the claims of potential mergers, in order to draw more conclusive evidence for vowel merger in the varieties,

a few more measurements for vowel merger have been conducted. These measures are discussed in the following section.

3.1.2 Measures of mergers

Linguists have used several methods to access and quantify vowel mergers. Nycz & Hall-Lew (2013) reviewed and compared four measures of overlap based on the criteria of how well the measures capture distance, how well the measures capture overlap, how well they account for unbalanced data, and whether they enable a comparison between speakers. These measures are, Euclidean Distance (ED), Linear Mixed-Effects Regression modeling, SOAM (Wassink, 2006), and Pillai score. Similar comparative study of overlap measures is done by Kelley & Tucker (2020). They compare *a posteriori* probability (APP)-based metric (Morrison, 2008) and Vowel Overlap Assessment with Convex Hulls (VOACH) (Haynes & Taylor, 2014) along with SOAM and Pillai score. Nycz & Hall-Lew (2013) concluded that to capture distance, ED proved to be better than the other measures, although ED calculated with median may be preferable to ED calculated with mean. In order to capture overlap, SOAM was considered to be the best among the four measures compared. LMER adjusted ED (ED-adjusted) fared the best while comparing unbalanced data and also enabling speaker comparisons. On the other hand, Kelley & Tucker (2020) conclude that the APP-based metric and Pillai score are theoretically preferable to SOAM and VOACH. In our analysis of vowel ellipse overlap, we primarily focus on calculating the distance between vowel categories in an acoustic space and measuring the degree of overlap between them. We use four methods to quantify the vowel overlaps observed in the previous chapter. These are: ED, Pillai score, SOAM and VOACH. The following subsections briefly describe the methods that are used in this study.

3.1.2.1 Euclidean distance

Euclidean distance (ED) (also known as Cartesian distance) of F1 and F2 values provide an average distance between two vowel categories, ignoring the amount of variability and deviation of the vowels from the average values. Two prominent works that used ED to quantify vowel distance are the studies of American English dialects, Baranowski (2006) used it to study vowel mergers in Charleston. Dinkin (2009) used it in his study of vowel change in Upstate New York. The distance between vowels is modeled as the hypotenuse of a right triangle, with the other two sides of the triangle defined by distances along the F1 and F2 dimensions. EDs are calculated using the following equation, where, D(i,j), is the ED between the vowels i and j calculated from the first and the second formants of the two vowels (Yang, 1996). The smaller the value between the vowels, the closer they are in the two-dimensional vowel space.

$$D(\mathbf{i},\mathbf{j}) = \sqrt{(F1_i - F1_j)^2 + (F2_i - F2_j)^2} \tag{3.1}$$

While this ED is effective to measure the distance between two vowels that are phonemically distinct, and fulfils the first requirement listed previously, it may not be the best measure to quantify vowels in a system with vowel mergers. ED is only a measure to quantify the distance between two vowel means and does not capture the amount of overlap between the categories. It also does not provide any information about the distribution of the tokens within each category. Hence, two additional measures are used in this study to quantifying merger.

3.1.2.2 Pillai-Bartlett trace

Hall-Lew (2010) suggested the Pillai-Bartlett trace as a reliable measure to assess vowel mergers. The Pillai-Bartlett trace or simply, Pillai score, is one of the statis-

tics in the multivariate analysis of variance (MANOVA) and is considered robust in measuring the separability of two data categories. MANOVA models variation with respect to multiple dependent variables at a time, such as both F1 and F2. The higher the value of the Pillai score, the greater the difference between the two categories with respect to the dependent variables. The Pillai score values range from 0 to 1: a score of 0 indicates a merger and 1 indicates complete distinctiveness. Depending on the distribution analyzed, the scores can be anywhere between 0 and 1. This measure was introduced by Hay et al. (2006) in their analysis of NEAR and SQUARE mergers in New Zealand English. Later Kennedy (2006) used it to examine CAUGHT-FOOT merger in New Zealand English. Hall-Lew (2009) and Wong & Hall-Lew (2014) used the Pillai score in their analysis of COT-CAUGHT merger in San Francisco and New York City.

3.1.2.3 Spectral Overlap Assessment Metric (SOAM)

Wassink (2006) proposes the Spectral Overlap Assessment Metric (SOAM) to capture the overlap between two vowels. In the current work, using 2-D SOAM has been considered, where F1 and F2 ellipses are generated for a vowel pair being investigated. The ellipses are best fit to account for the variance using least-squares fitting. Finally, vowel overlap is calculated by the fraction of uniformly distributed data points in the region of overlap, compared to the number of total data points in each vowel category. Wassink (2006) also provides a tentative interpretation of the SOAM measures, where 0%-20% is considered 'no overlap', more than equal to 40% as complete overlap and anything between 20%-40% is considered an indication of partial overlap. However, Haynes & Taylor (2014) points out that these estimates of overlap cutoff ranges are arbitrary.

Although SOAM is considered to be one of the best tools to classify vowel systems according to both their spectral and temporal features, Wassink (2006) notes that the

use of best-fit ellipses is perhaps not ideal as the convention has no basis in auditory or acoustic reality and appears to exist out of convenience. She is of the view that the metric may not be ideal in cases where data are not widely available and where the issue of "empty data" or distribution gaps may be more pronounced, as in under-described and endangered languages. In an attempt to address this shortcoming of SOAM, Haynes & Taylor (2014) propose the Vowel Overlap Assessment with Convex Hulls (VOACH) method which takes into account the distribution gaps in data.

3.1.2.4 Vowel Overlap Assessment with Convex Hulls (VOACH)

Haynes & Taylor (2014) provide an improved version of the SOAM method, using best-fit convex shapes, calling it the Vowel Overlap Assessment with Convex Hulls (VOACH) method. This method was originally proposed by Brubaker & Altshuler (1959), creating and analyzing the overlap shapes in F1-F2 space by hand. It was difficult to apply the method to two-dimensional datasets and impossible to extend it to three dimensions. Haynes & Taylor (2014) adopted the method and improved on its applicability in both two and three dimensions using standardized and efficient convex hull algorithms now developed in computational geometry and included in numerical packages such as MATLAB. The use of best-fit convex shapes in this method minimally accounts for space that is not occupied by data, thereby reducing the danger of incorporating "empty" data into calculations of vowel space and allowing for more precision in calculating overlap. In this method convex hulls (H1 and H2) are first fit around the data points that represent the vowel being assessed. Next, each vowel token in the data set is checked to determine whether it is contained in one or both hulls. In the third step, a third convex hull (H3) is fitted around the points contained in the first two hulls. Lastly, ratio of the area between H3 and H1 & H2 are calculated. The larger of the ratios is taken as the overlap.

The formant values of the vowel pairs /u-ʊ/, /ʊ-o/, /o-ɔ/ and /ɛ-e/ of the As-

samese varieties were subjected to four overlap measuring tests mentioned here. The following sections discuss the results of the three tests performed on the Assamese vowels.

3.2 Results

3.2.1 Euclidean Distance

Two-dimensional Euclidean distances (ED) between the vowel pairs were calculated to see the extent of separation between the vowels in the five varieties of Assamese. It has been argued that the perceptual realities of humans are better reflected by the Mel scale (S. S. Stevens et al., 1937; S. S. Stevens & Volkmann, 1940). Therefore, to calculate the Euclidean Distances formant frequencies in Hz were converted to Mel values. The distances were then calculated using the equation in 3.1. The results of the Euclidean Distances are shown in Table 3.3.

Table 3.3: Euclidean distance in Mel for vowels in five varieties of Assamese

Vowel pairs	Tinsukia	Jorhat	Nagaon	Nalbari	Barpeta
a-e	463	346	486	336	414
a-ɛ	378	309	321	348	351
a-i	618	516	568	511	606
a-o	496	480	518	450	467
a-ɔ	375	356	388	309	362
a-u	534	482	530	505	543
a-ʊ	504	488	517	463	502
e-ɛ	89	82	168	36	70
e-i	157	173	83	176	192
e-o	898	735	933	702	809
e-ɔ	792	640	829	586	724
e-u	918	723	929	732	865
e-ʊ	887	732	922	709	837
ɛ-i	246	216	251	165	258
ɛ-o	830	739	791	729	763
ɛ-u	720	633	678	609	674
ɛ-ʊ	854	730	792	761	823
i-o	1035	899	1001	866	986
i-ɔ	935	809	902	757	907
i-u	1051	884	996	890	1037
i-ʊ	1020	894	989	872	1012
o-ɔ	124	133	137	144	110
o-u	57	36	42	77	85
o-ʊ	50	26	28	19	37
ɔ-u	172	149	163	210	194
ɔ-ʊ	146	149	147	159	147
u-ʊ	31	12	17	58	49

In Chapter 2 it was observed that the vowels /u-ʊ/ showed complete overlapping of ellipses for Tinsukia, Jorhat and Nagaon. Statistical tests conducted on the formant values indicated no significant difference between the two vowels for both F1 and F2 values for Tinsukia and Jorhat and only F2 for Nagaon. The ED measures in Table 3.3 show that these three varieties have the smallest distances between the two vowels, 31, 12 and 17 Mels respectively for Tinsukia, Jorhat and Nagaon indicating merging of the pair in the varieties. Similarly, it was observed that the vowels /o-ʊ/ had considerable overlap of vowel ellipses in the Nalbari and Barpeta varieties. Statistical tests conducted on the formant values indicated no significant difference between the two vowels in backness. As seen in the Table 3.3, the two regions have an ED of 19 and 37 Mels respectively for the two vowels indicating that /ʊ/ is closer to /o/ than to /u/ in these two varieties (/u-ʊ/ values 58 and 49 Mels respectively).

In case of the front vowels /e-ɛ/ no significant overlap was observed in Tinsukia, Jorhat and Nagaon varieties. Although Nalbari and Barpeta had showed considerable overlap between the two vowels, statistical results showed the two vowels did not differ only in terms of backness in all categories (except for sentence frames in Barpeta variety in which case the two vowels did not significantly differ in height). The ED measures in Table 3.3 show that in both Nalbari and Barpeta, the distance between /e-ɛ/ is comparatively much smaller than /e-i/: 36 and 70 Mels for /e-ɛ/ respectively for Nalbari and Barpeta as opposed to 176 and 192 Mels for /e-i/. This indicates that /e/ is closer to /ɛ/ than it is to /i/ in both the varieties.

The results of the Euclidean distance measures between the vowel pairs do indicate that there is potential vowel merging in the varieties of Assamese. As stated before, from the evidence drawn from the vowel plots and statistical measurements in Chapter 2, the Nalbari and Barpeta varieties seem to have merged /e-ɛ/ and /o-ʊ/, while the Tinsukia, Jorhat and Nagaon varieties are seen to have merged the /ʊ-u/ vowel pair. It is noteworthy that the same vowel /ʊ/ is merging with two different vowels in

different varieties of Assamese.

3.2.2 Pillai-Bartlett Trace

Pillai score is considered a robust measurement of separability of two data categories. The score can be between 0 and 1 where 0 indicates a complete merger and 1 indicates complete distinctiveness. In case of vowel categories, if two vowels are overlapping, indicating merger, the corresponding Pillai scores will be near 0, on the other hand, if the two vowel categories have sufficient acoustic distance between them, the scores will be near 1. In case of vowels, F1 and F2 values will be the dependent variables for the test. As Pillai scores are statistic derived from MANOVA tests, one MANOVA test was conducted with F1 and F2 as dependent variables and vowel category as factor. Twenty separate MANOVA tests were conducted, that is, for four vowel pairs, namely, /e-ɛ/, /o-ʊ/, /o-ɔ/ and /ʊ-u/, in each of the varieties of Assamese.

Table 3.4: Pillai scores for four vowel pairs by Assamese varieties

Vowel pair	Tinsukia	Jorhat	Nagaon	Nalbari	Barpeta
/e/-/ɛ/	0.29	0.64	0.51	0.08	0.15
/u/-/ʊ/	0.02	0.02	0.05	0.42	0.21
/ʊ/-/o/	0.24	0.12	0.09	0.05	0.06
/o/-/ɔ/	0.41	0.60	0.45	0.46	0.33

Table 3.4, presents the Pillai scores of the vowel pairs /e-ɛ/, /o-ʊ/, /o-ɔ/ and /ʊ-u/ for each of the five Assamese varieties. The Pillai scores for /u/-/ʊ/ vowel pair in Tinsukia and Jorhat is 0.02, whereas for Nagaon the score is 0.05. The Pillai score for the same vowel pair in the Nalbari variety is 0.42. and in Barpeta it is 0.21. Hence, as far as the /u/-/ʊ/ vowel pair is concerned, it is seen that the vowels are completely merged in the Tinsukia, Jorhat and Nagaon varieties and no complete merging in the Nalbari and Barpeta varieties. The /ʊ/-/o/ vowel pair shows Pillai

scores of 0.05 and 0.06 in the Nalbari and Barpeta varieties respectively that can be an indication of complete merger. On the other hand, the same pair has a Pillai score of 0.24 for Tinsukia, indicating no complete overlap. However, Jorhat and Nagaon has scores of 0.12 and 0.09 showing partial overlap of the two vowels. With Pillai scores above 0.20 in each of the five varieties, the /ɔ/-/o/ vowel pair does not show complete overlap in any of the varieties. Lastly, the /ɛ/-/e/ pair produces Pillai scores of 0.29, 0.64 and 0.51 in the Tinsukia, Jorhat and Nagaon varieties indicating that the two vowels are distinct in these two varieties. For the Nalbari and Barpeta speakers the same pair yields Pillai scores of 0.08 and 0.15 indicating strong evidence of merger. To summarize, the complete merger of /u/-/ʊ/ in the Tinsukia, Jorhat and Nagaon varieties is evident from the Pillai scores. On the other hand, the Nalbari and Barpeta varieties show complete merger of /ʊ/-/o/ vowel pairs and strong evidence of merger for /ɛ/-/e/ vowels.

3.2.3 Spectral Overlap Assessment Metric (SOAM)

Following Wassink (2006), an online freeware was used to generate 2-D SOAM plots for the four vowel pairs, /ɛ/-/e/ and /ʊ/-/o/, /o/-/ɔ/ and /u/-/ʊ/. Table 3.5 presents the SOAM measurements for the vowel pairs in each of the varieties. Figures 3.2 to 3.21 show the 2-D SOAM plots for vowel overlap for each of the pairs in the varieties. In each plot the distribution of vowel categories is shown using blue and pink squares or circles. The darker area between them (purple) shows the amount of overlap between the two vowel categories. In the following subsections the SOAM values of each of the concerned vowel pair are discussed.

3.2.3.1 SOAM overlap of /ɛ/-/e/

Figure 3.2: SOAM plot for /ɛ/-/e/ overlap for Tinsukia.

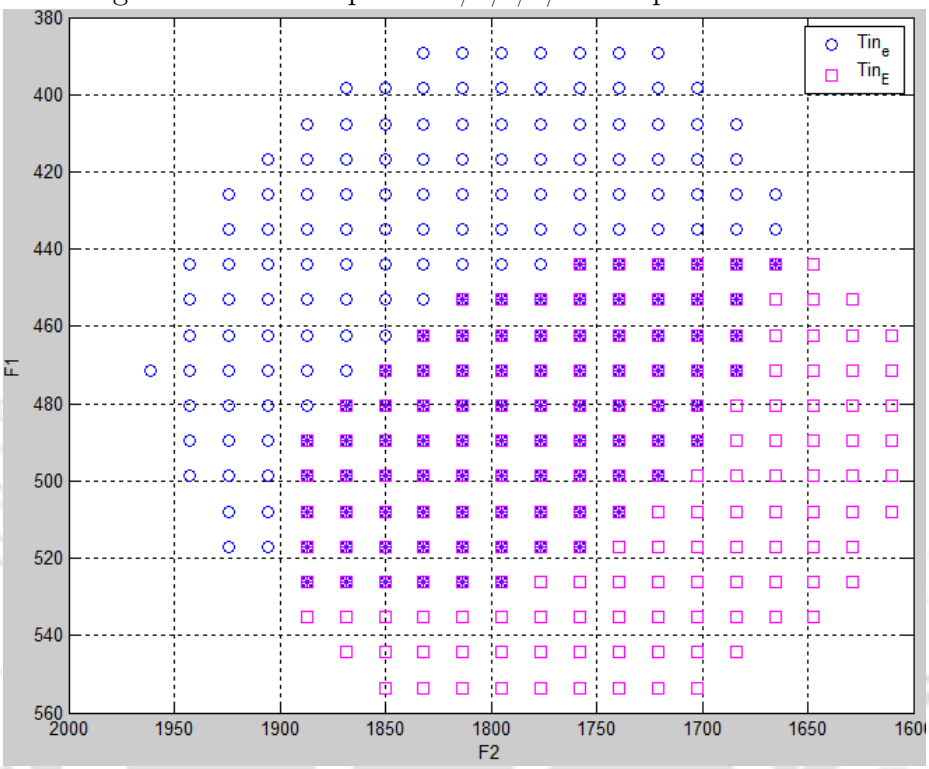


Figure 3.3: SOAM plot for /ɛ/-/e/ overlap for Jorhat.

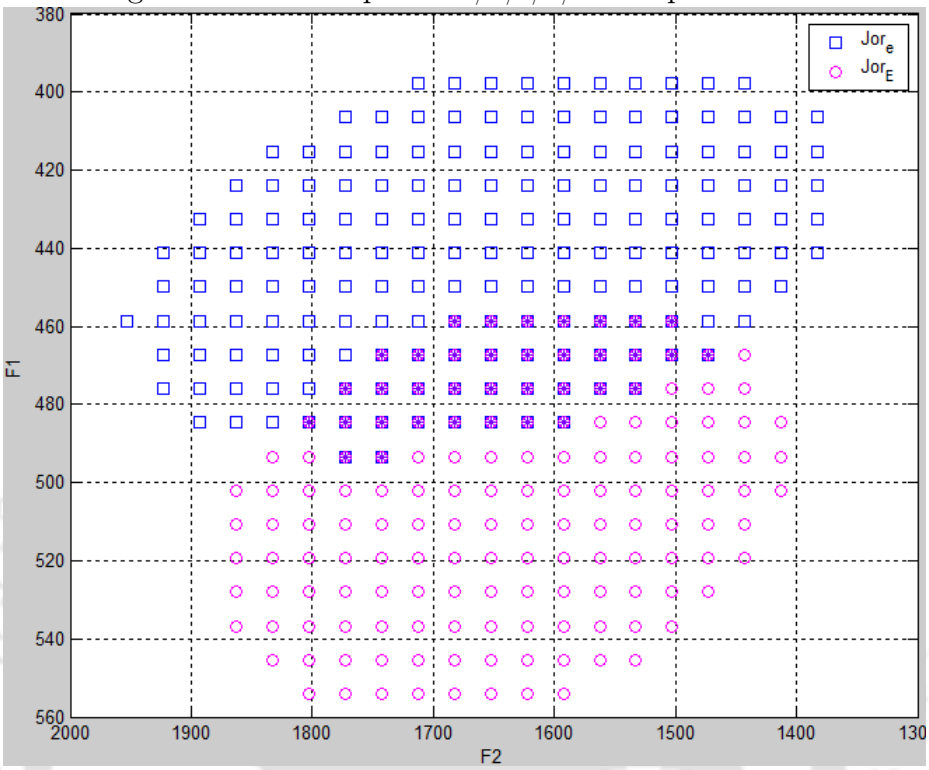


Figure 3.4: SOAM plot for /ɛ/-/e/ overlap for Nagaon.

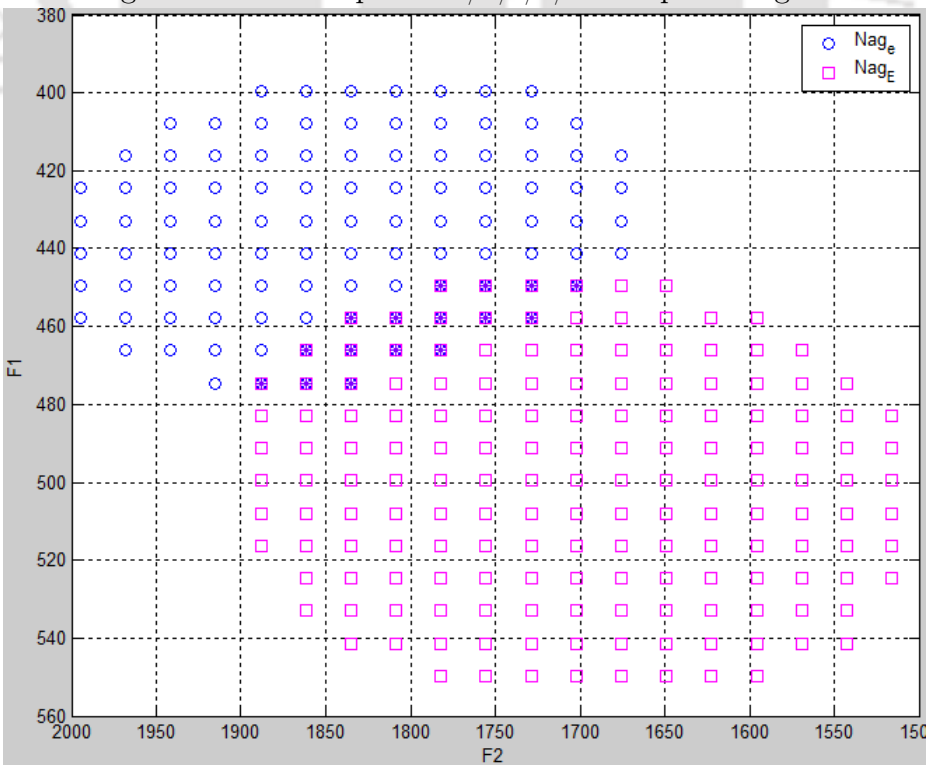


Figure 3.5: SOAM plot for /ɛ/-/e/ overlap for Nalbari.

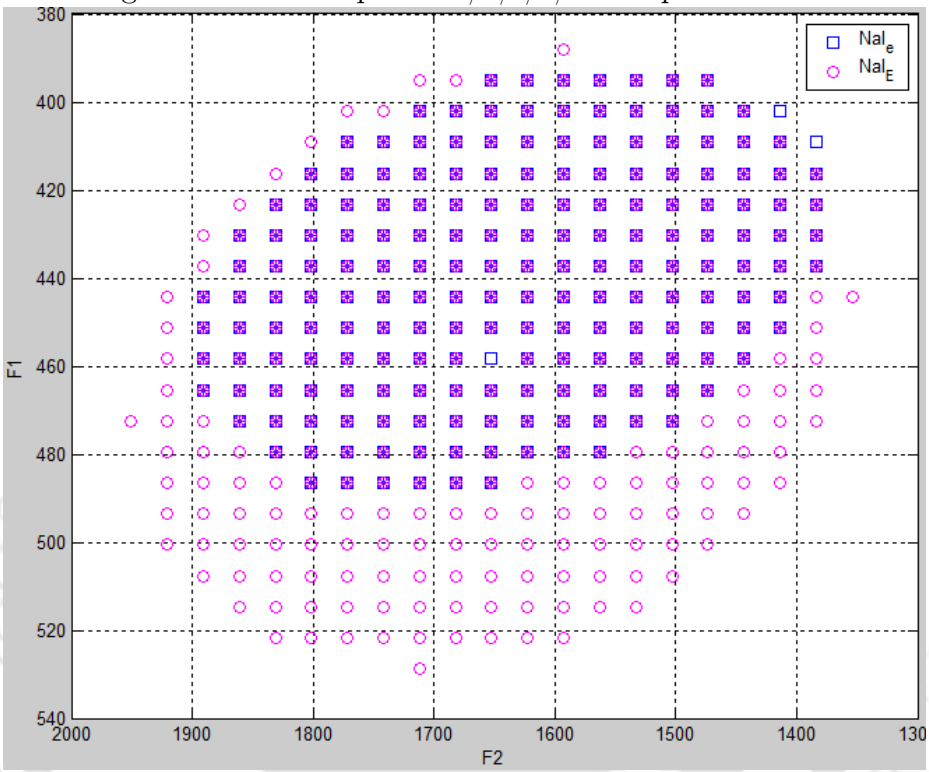
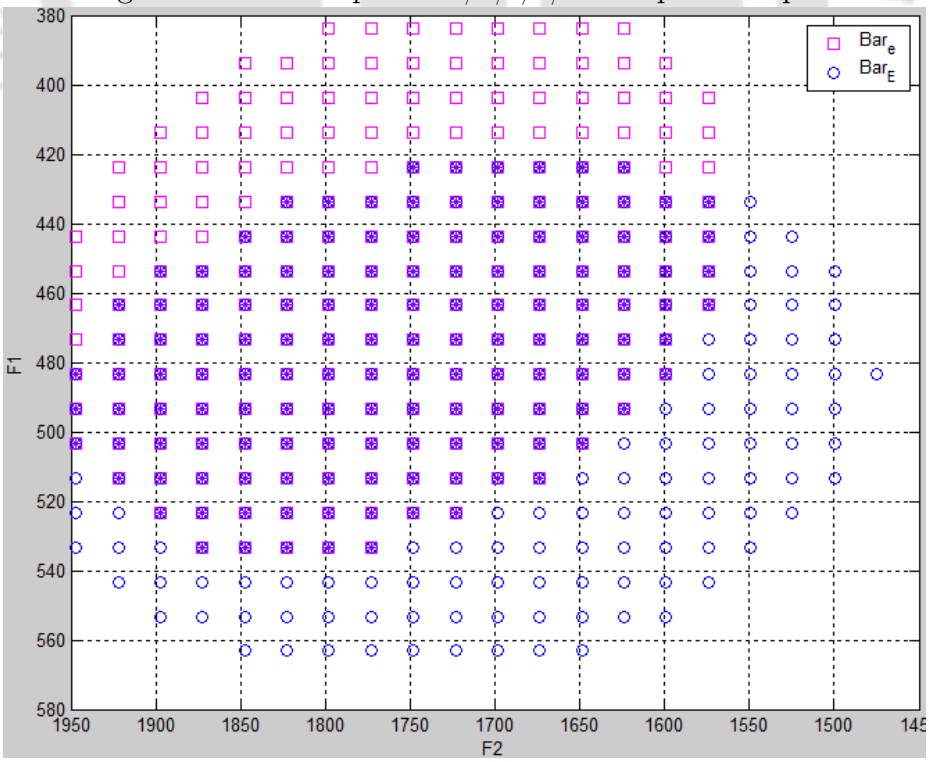


Figure 3.6: SOAM plot for /ɛ/-/e/ overlap for Barpeta.



As seen in the figures and as evidenced by Table 3.5, for the /ɛ/-/e/ pair Nalbari has the highest score with 99.4% indicating complete overlap. Barpeta with a score of 77.8% also indicates complete overlap. Nagaon shows complete distinctiveness with a score of 8.9%, whereas Tinsukia and Jorhat with 60.8% and 40.7% respectively, show partial overlap.

3.2.3.2 SOAM overlap of /u/-/ʊ/

Figure 3.7: SOAM plot for /u/-/ʊ/ overlap for Tinsukia.

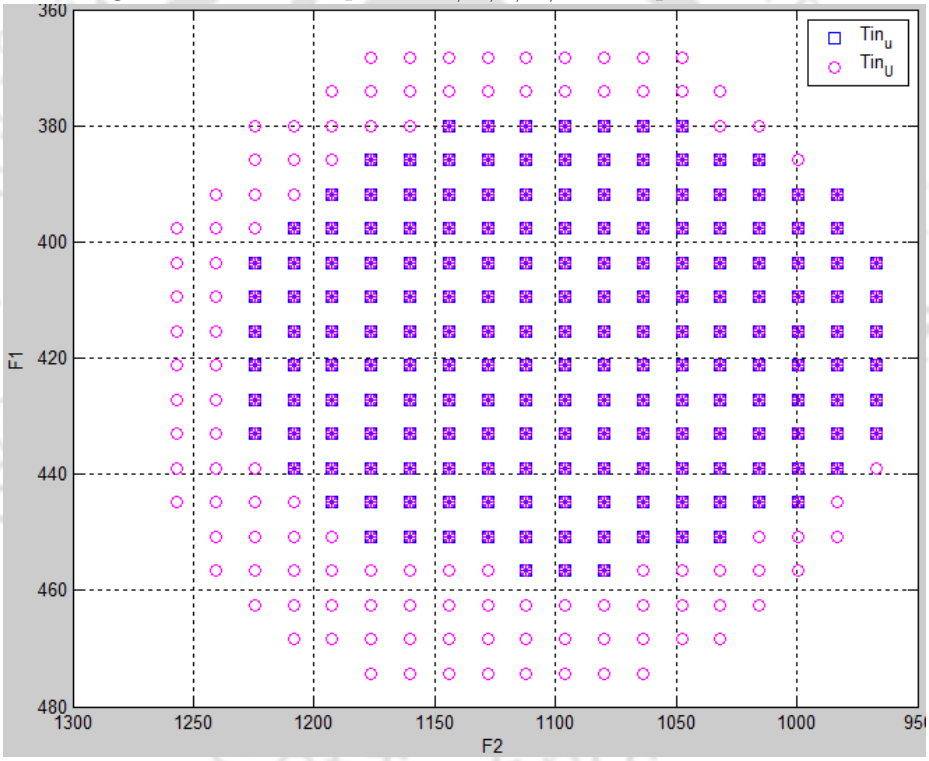


Figure 3.8: SOAM plot for /u/-/ʊ/ overlap for Jorhat.

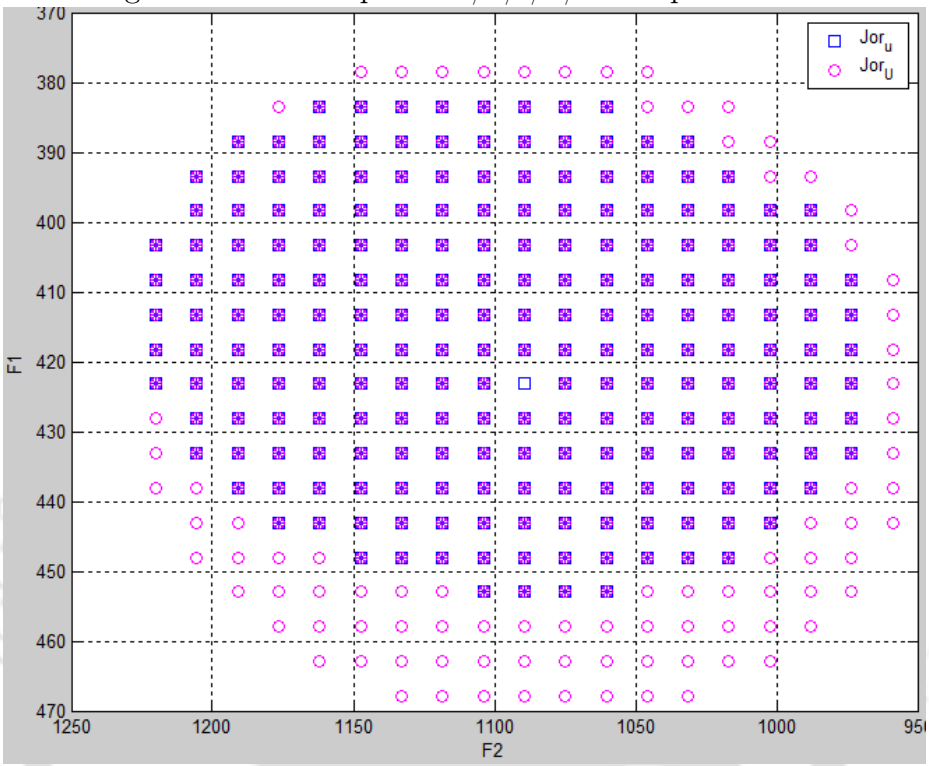


Figure 3.9: SOAM plot for /u/-/ʊ/ overlap for Nagaon.

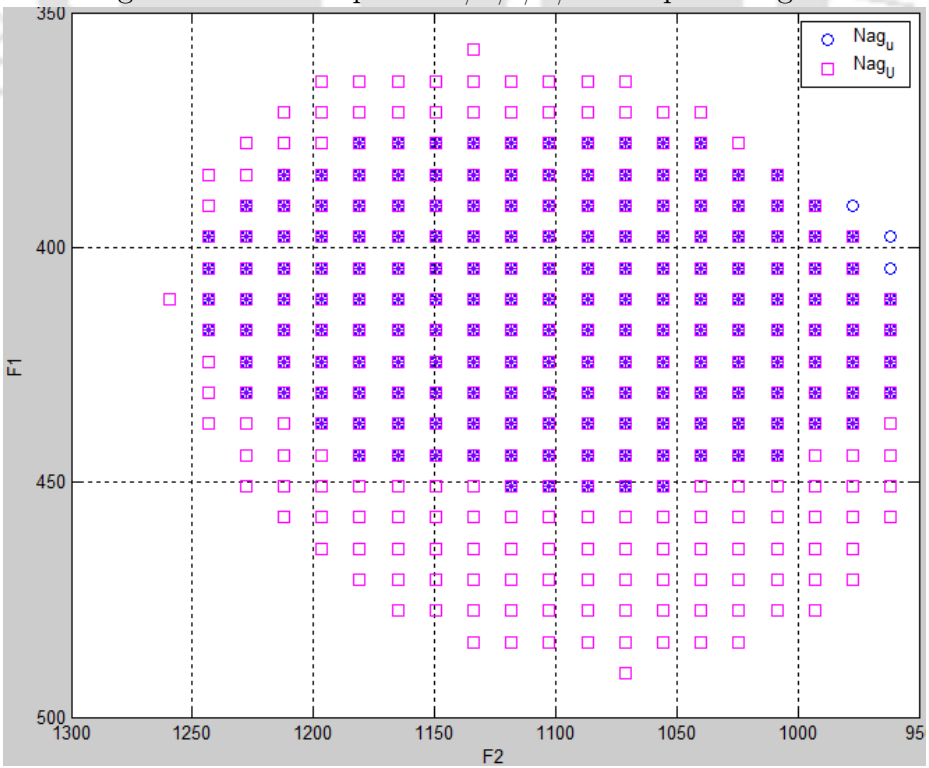


Figure 3.10: SOAM plot for /u/-/ʊ/ overlap for Nalbari.

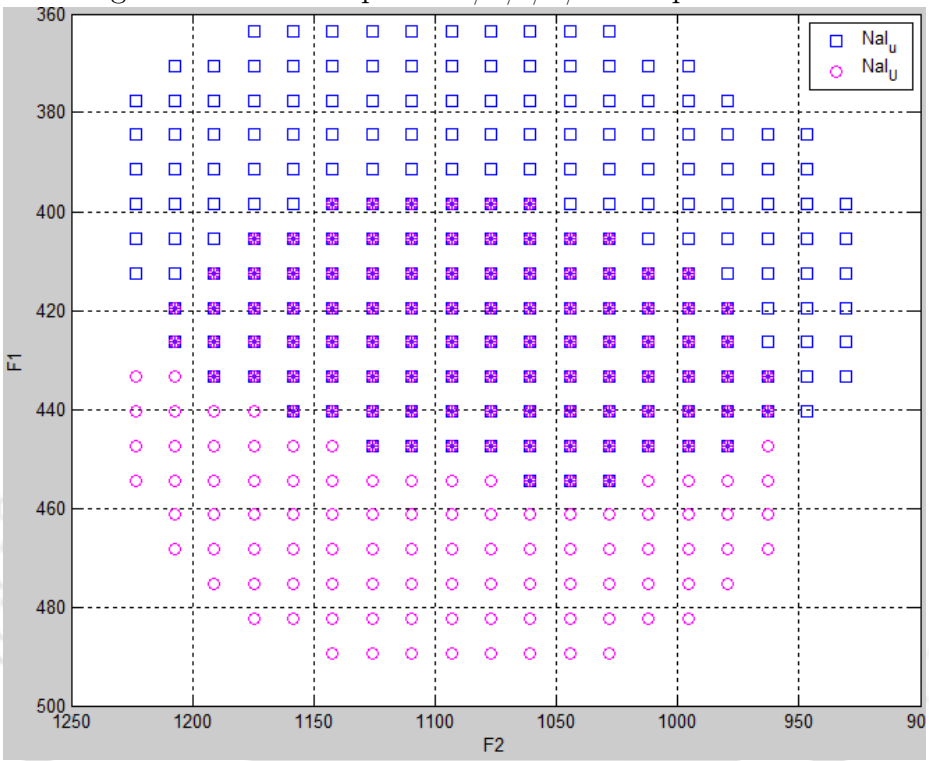
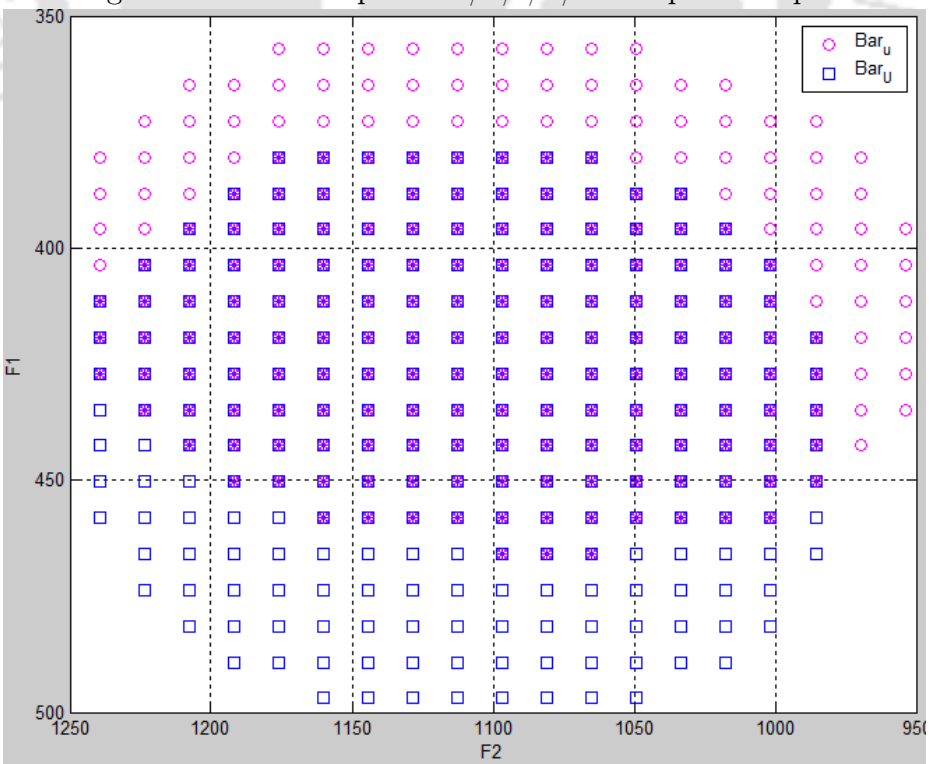


Figure 3.11: SOAM plot for /u/-/ʊ/ overlap for Barpeta.



The vowel pair /ʊ/-/u/ shows complete overlap of the vowels in Tinsukia, Jorhat and Nagaon, with a score of 96.4%, 99.2% and 92.1% respectively. On the other hand, Nalbari and Barpeta shows partial overlap with scores of 52.5% and 69.1% respectively.

3.2.3.3 SOAM overlap of /ʊ/-/o/

Figure 3.12: SOAM plot for /ʊ/-/o/ overlap for Tinsukia.

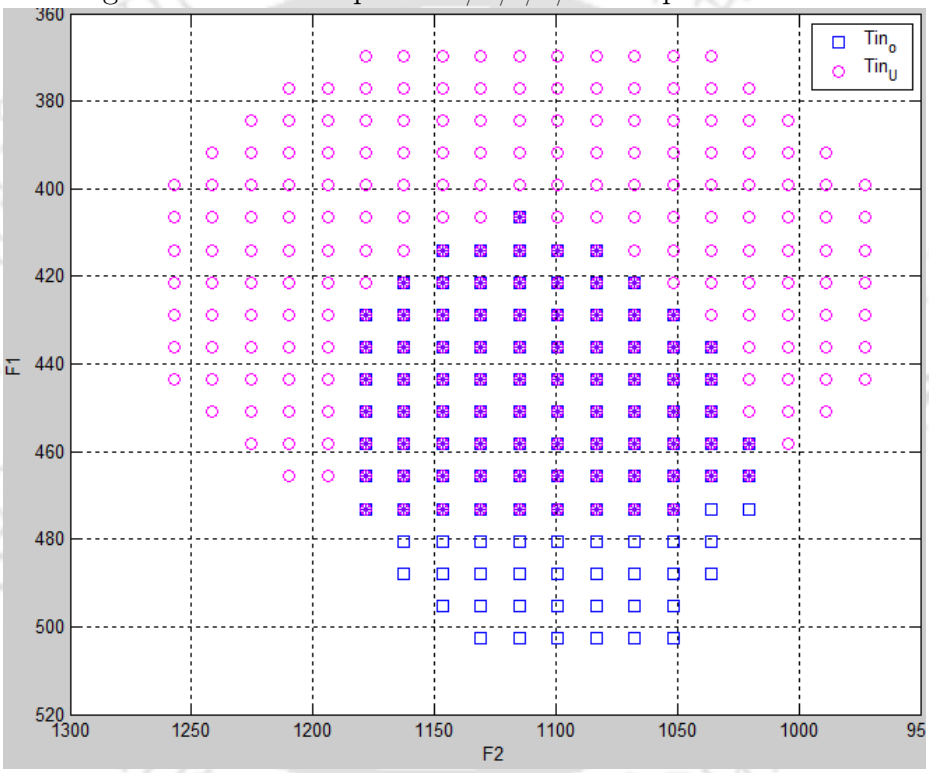


Figure 3.13: SOAM plot for /ʊ/-/o/ overlap for Jorhat.

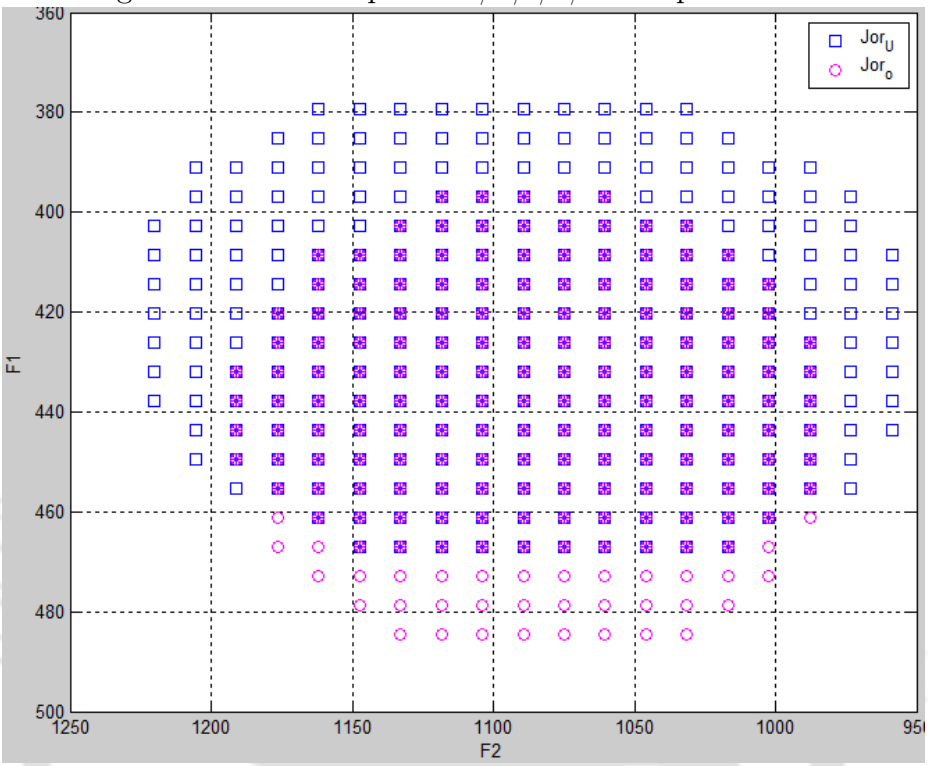


Figure 3.14: SOAM plot for /ʊ/-/o/ overlap for Nagaon.

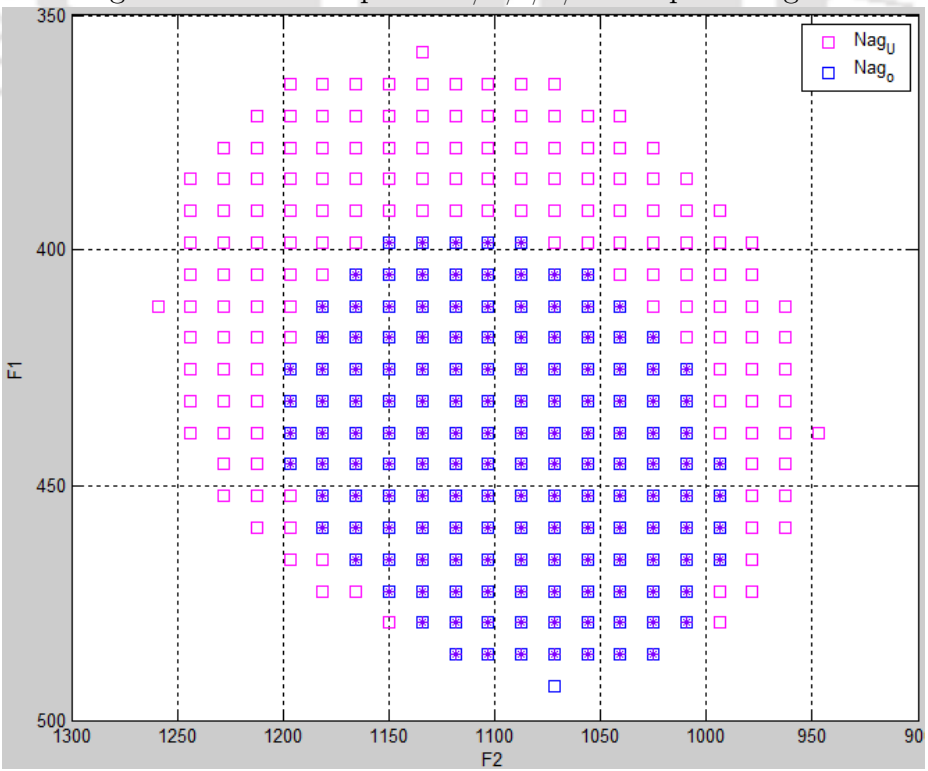


Figure 3.15: SOAM plot for /ʊ/-/o/ overlap for Nalbari.

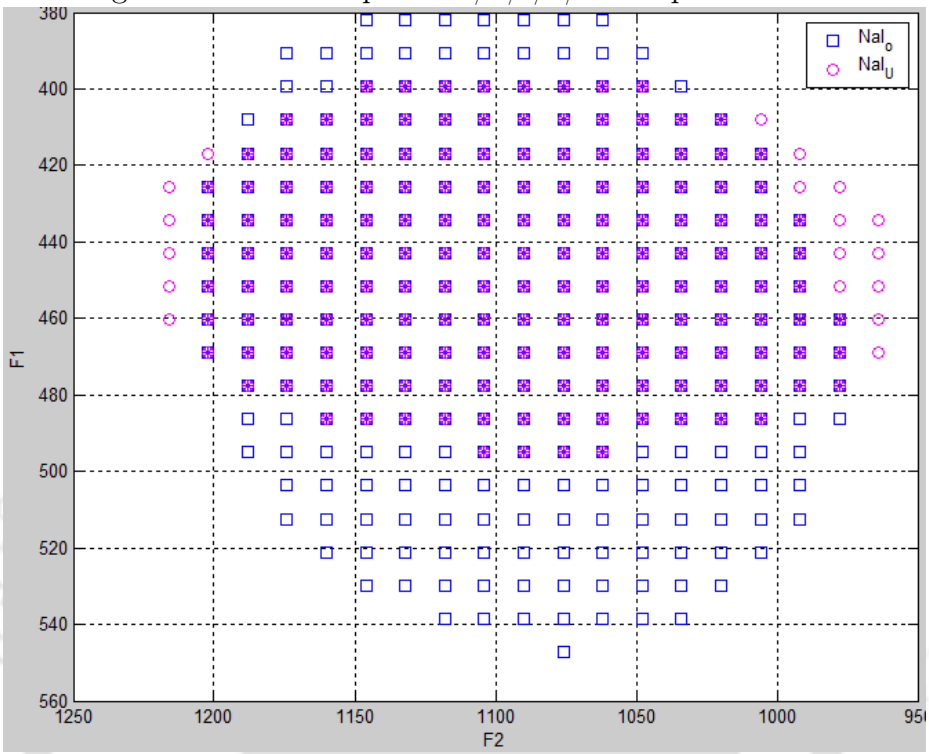
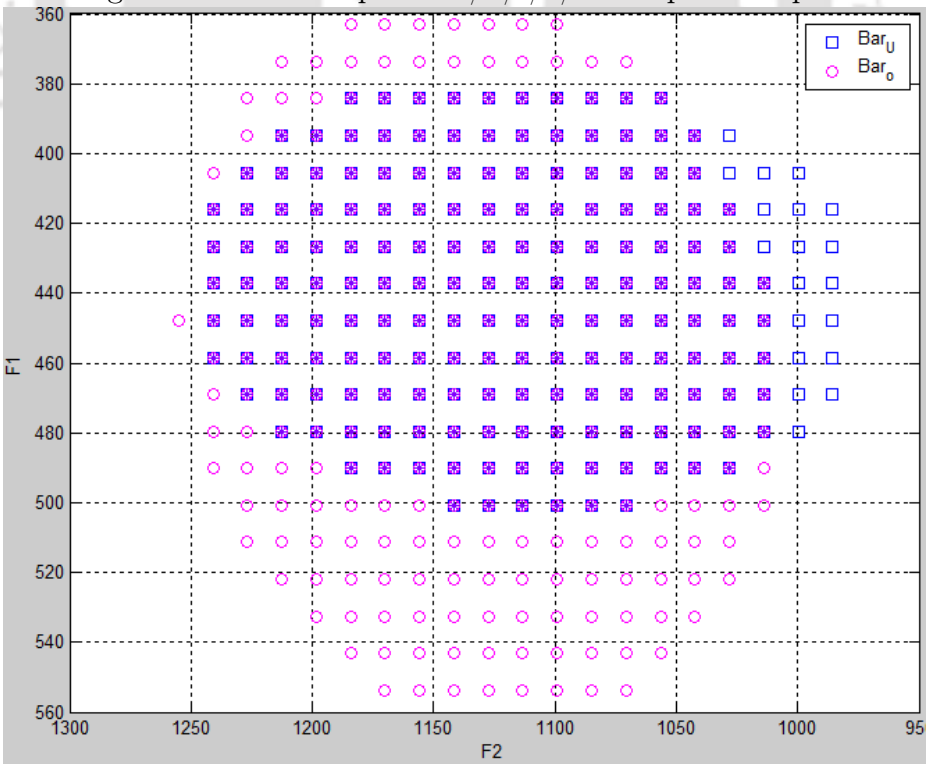


Figure 3.16: SOAM plot for /ʊ/-/o/ overlap for Barpeta.



The vowel pair /ʊ/-/o/ shows considerable overlap between the two vowels in all the varieties. Interestingly, Tinsukia and Nagaon with the highest percentages (92.3% and 92.7% respectively) show complete overlap of the vowels.

3.2.3.4 SOAM overlap of /o/-/ɔ/

Figure 3.17: SOAM plot for /o/-/ɔ/ overlap for Tinsukia.

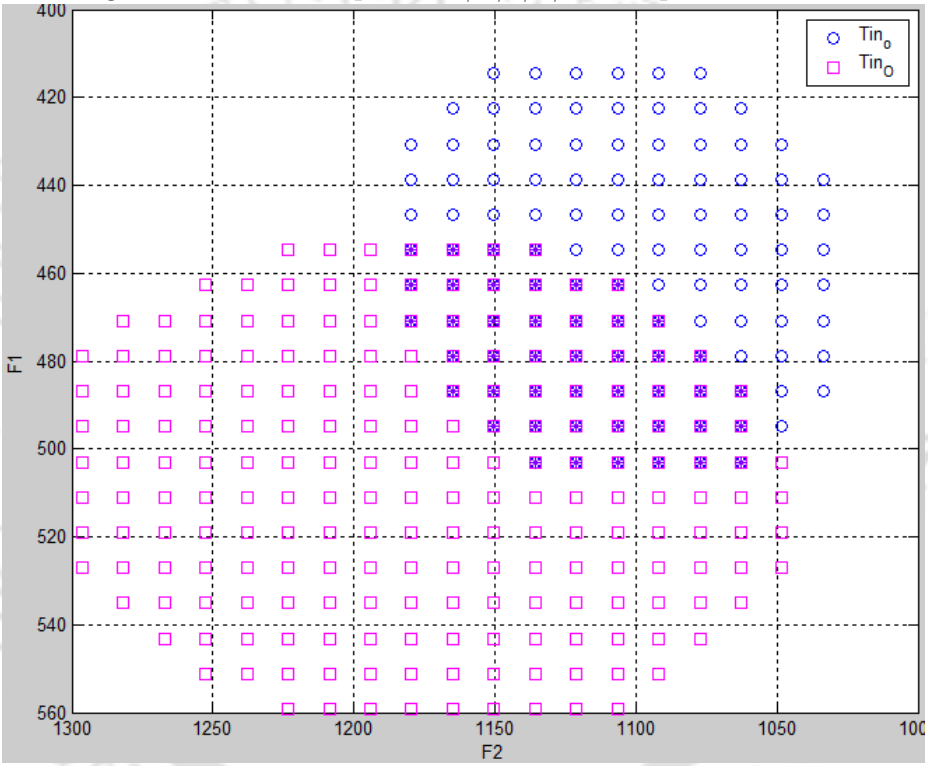


Figure 3.18: SOAM plot for /o/-/ɔ/ overlap for Jorhat.

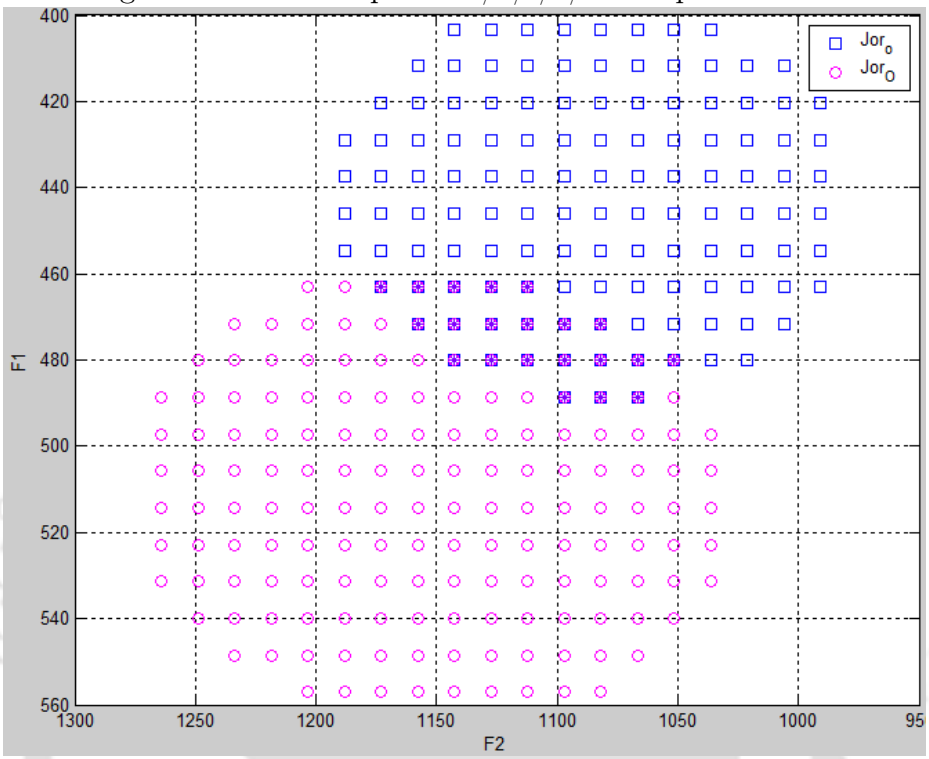


Figure 3.19: SOAM plot for /o/-/ɔ/ overlap for Nagaon.

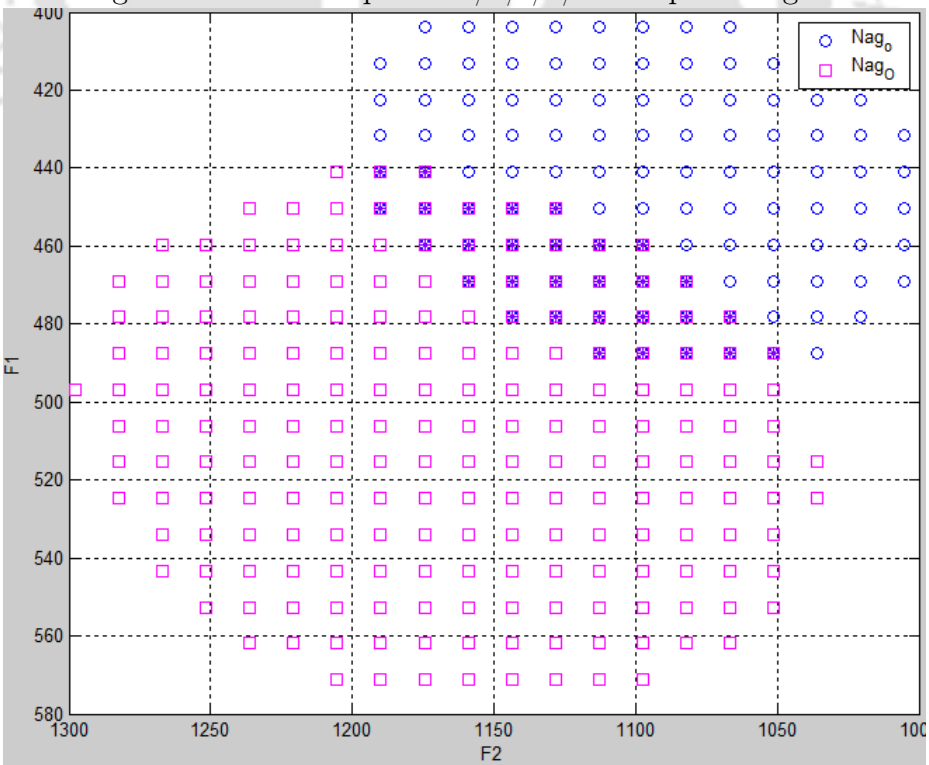


Figure 3.20: SOAM plot for /o/-/ɔ/ overlap for Nalbari.

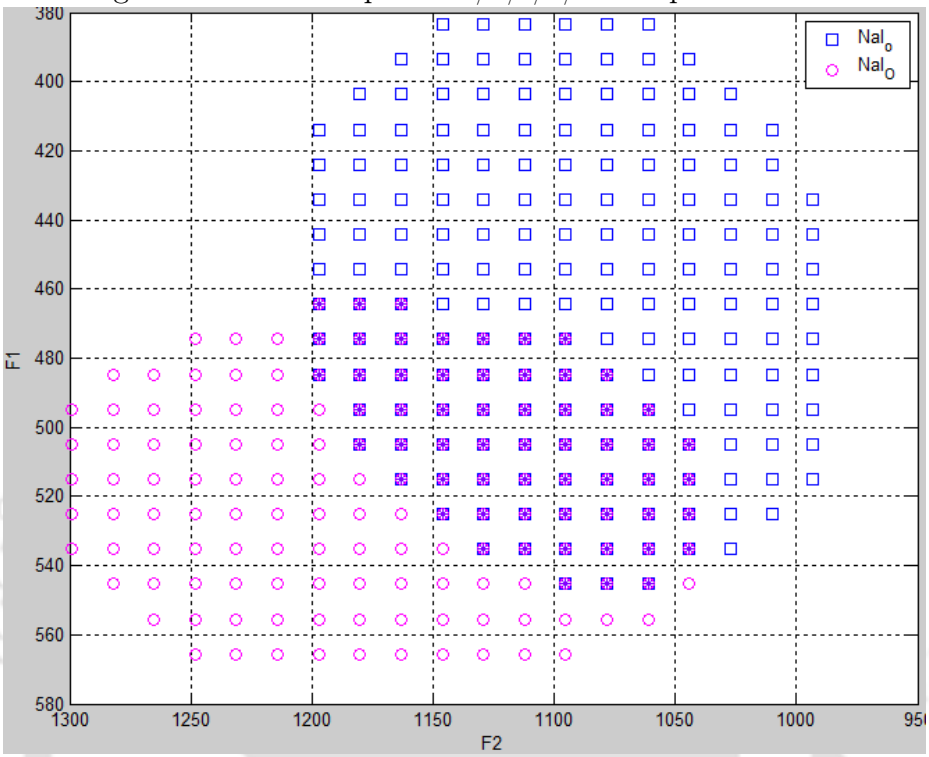
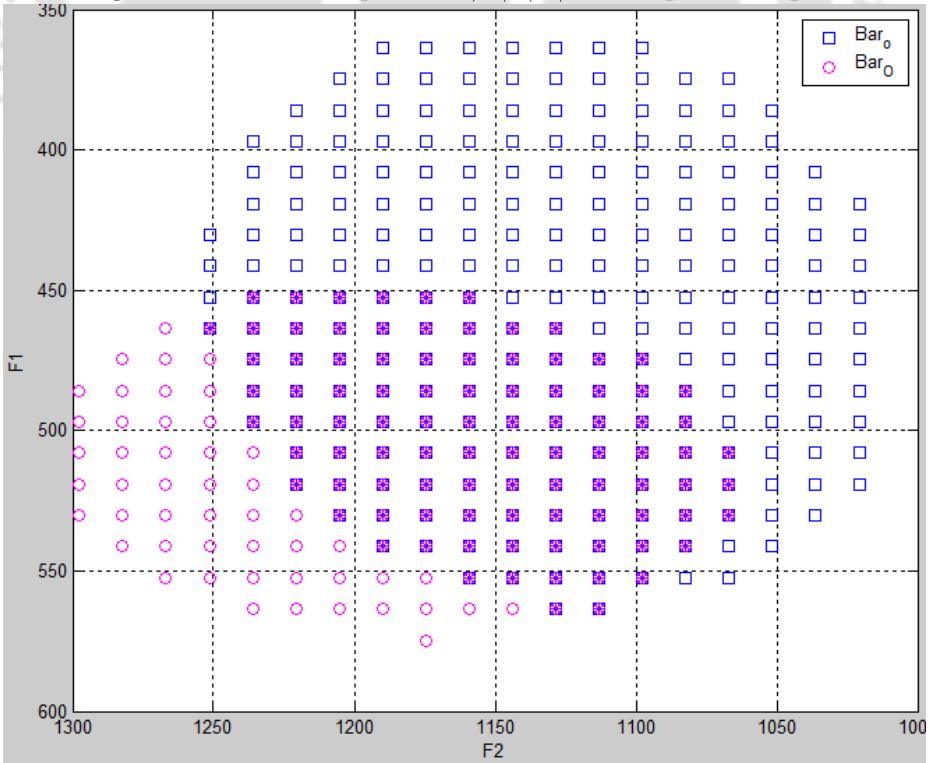


Figure 3.21: SOAM plot for /o/-/ɔ/ overlap for Barpeta.



The vowel pair /ɔ/-/o/ shows only marginal overlap in Jorhat, Nagaon, Nalbari and Barpeta varieties with 27.3%, 21.3%, 48.4% and 55.7% respectively and no overlap in case of Tinsukia with 6%.

Table 3.5: SOAM overlap measures for four vowel pairs by Assamese varieties.

Vowel pair	Tinsukia	Jorhat	Nagaon	Nalbari	Barpeta
/ɛ-e/	50.5%	23.8%	15.5%	98.4%	67.3%
/u-ʊ/	100%	99.5%	98.3%	51.8%	67.8%
/ʊ-o/	92.3%	87.7%	92.7%	84.8%	82.0%
/o-ɔ/	06.0%	27.3%	21.3%	48.4%	55.7%

3.2.4 Vowel Overlap Assessment with Convex Hulls (VOACH)

Following the procedure described in Haynes & Taylor (2014), for vowel pairs in the five Assamese varieties, the convex hulls for the vowels to be assessed are first created using the *chull* function in R. Spatial polygons are created using the *SpatialPolygons* function of the *sp* package. Thereafter, the interaction of the polygons is calculated using *ginteraction* function of the package *rgeos*. Finally, the convex hull area is calculated using the *convexHullArea* function of package *phonR* (Daniel McCloy, 2020). Table 3.6 presents the overlaps for the vowel pairs /ɛ-e/ and /ʊ-o/, /o-ɔ/ and /u-ʊ/ in each of the varieties. While doing the assessment, vowels that do not overlap at all do not return any score. However if we look in the areas among the overlapping vowels, the biggest scores are shown by the vowel pair /ɛ-ɔ/. These scores are provided in the table as a comparison to the scores of the most overlapping vowel pair. The convex hull plots for the vowel pairs are created using the *phonR* package. Figures 3.22 to 3.25 show the Convex Hull overlap of the four vowels pairs for 5 varieties of Assamese.

Table 3.6: VOACH for four vowel pairs by Assamese varieties.

Contrast	Tinsukia	Jorhat	Nagaon	Nalbari	Barpeta
/e-ɛ/	1.63	2.09	1.09	1.05	1.03
/u-ʊ/	1.06	1.03	1.38	1.14	1.28
/ʊ-o/	1.00	1.03	1.00	1.25	1.25
/o-ɔ/	1.26	2.11	1.27	1.04	1.11
/ɛ-ɔ/	10.9	21.38	3.75	3.00	6.54

As seen in table 3.6, Tinsukia and Jorhat show the lowest VOACH scores for the vowel pairs /ʊ-o/ and /u-ʊ/ indicating that the most overlap in the two varieties are happening in these two pairs. The convex hull plots in Figure 3.23 shows the tokens for the two vowels in both the varieties cluster together, as opposed to Nalbari in which the clusters are separate. Similarly, Nalbari and Barpeta varieties showed low scores for the vowel pair /e-ɛ/. The tokens for the two vowels as seen in Figure 3.22 are clustered together in both the varieties unlike Jorhat and Tinsukia where the clusters are clearly distinct. The Nagaon variety showed the lowest score for /ʊ-o/. It is also observed that the lowest score for Nalbari is for the vowel pair /o-ɔ/ indicating the most overlap in the variety. The /ʊ-o/ pair for Nalbari and Barpeta varieties which showed high overlap in the previous merger measures, shows comparatively high scores or low overlap in the two varieties. Even though /o-ɔ/ pair shows low VOACH scores for Nalbari and Barpeta, Figure 3.25 reveals that the vowel tokens are clustered separately.

Figure 3.22: Convex Hull overlap of /e-ɛ/ vowels for 5 varieties of Assamese.

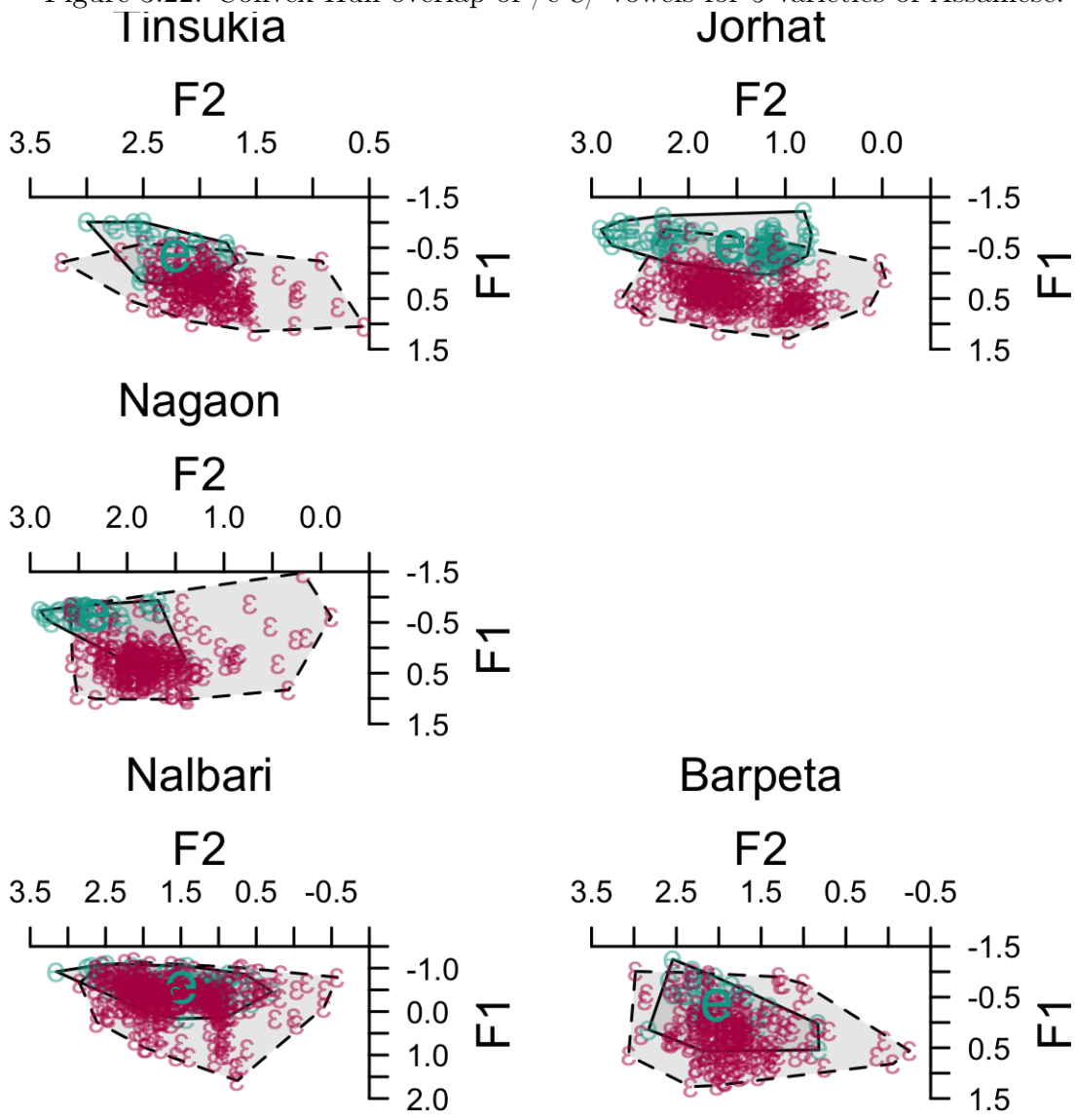


Figure 3.23: Convex Hull overlap of /u-ʊ/ vowels for 5 varieties of Assamese.

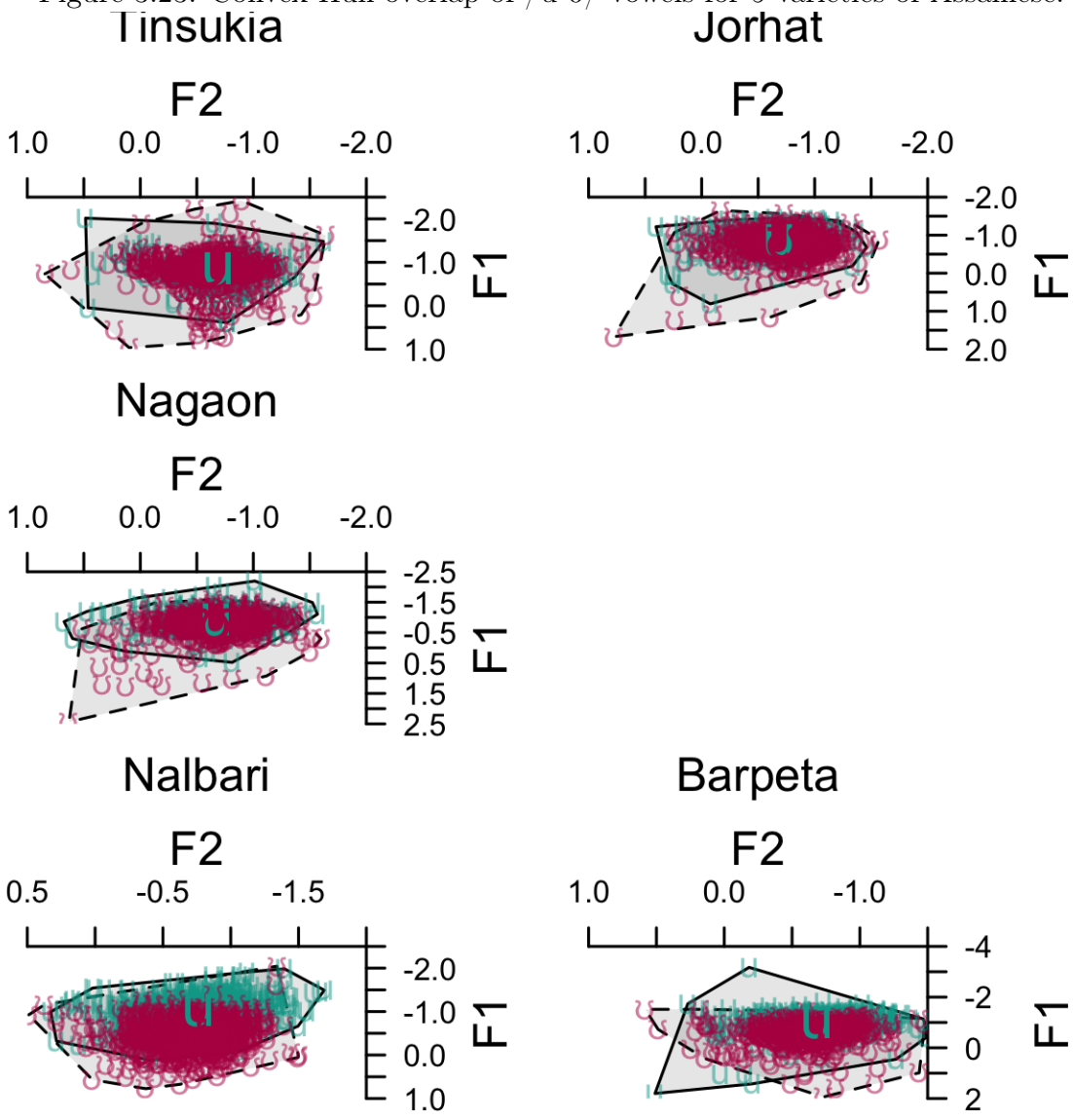


Figure 3.24: Convex Hull overlap of /ʊ-o/ vowels for 5 varieties of Assamese.

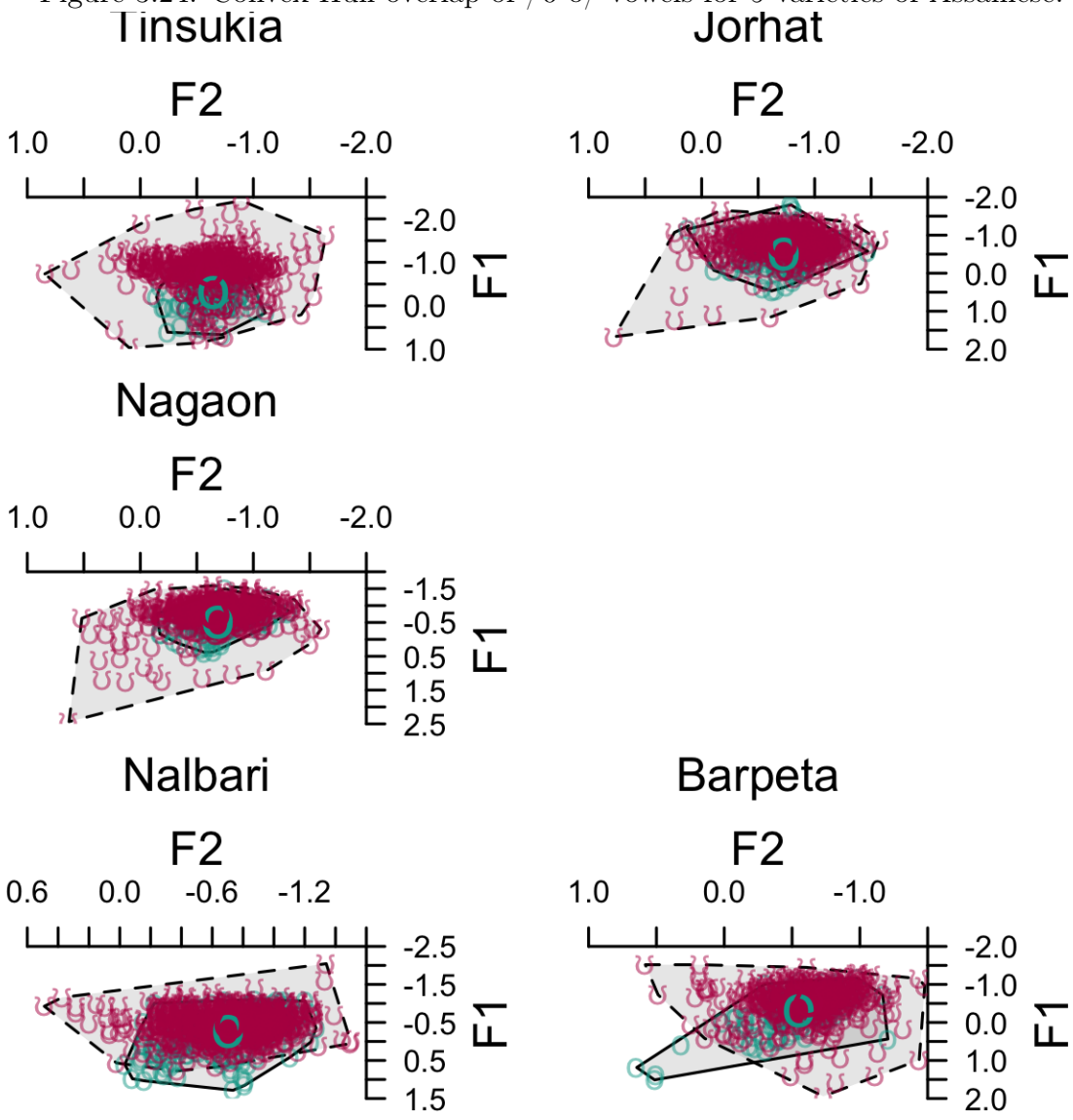
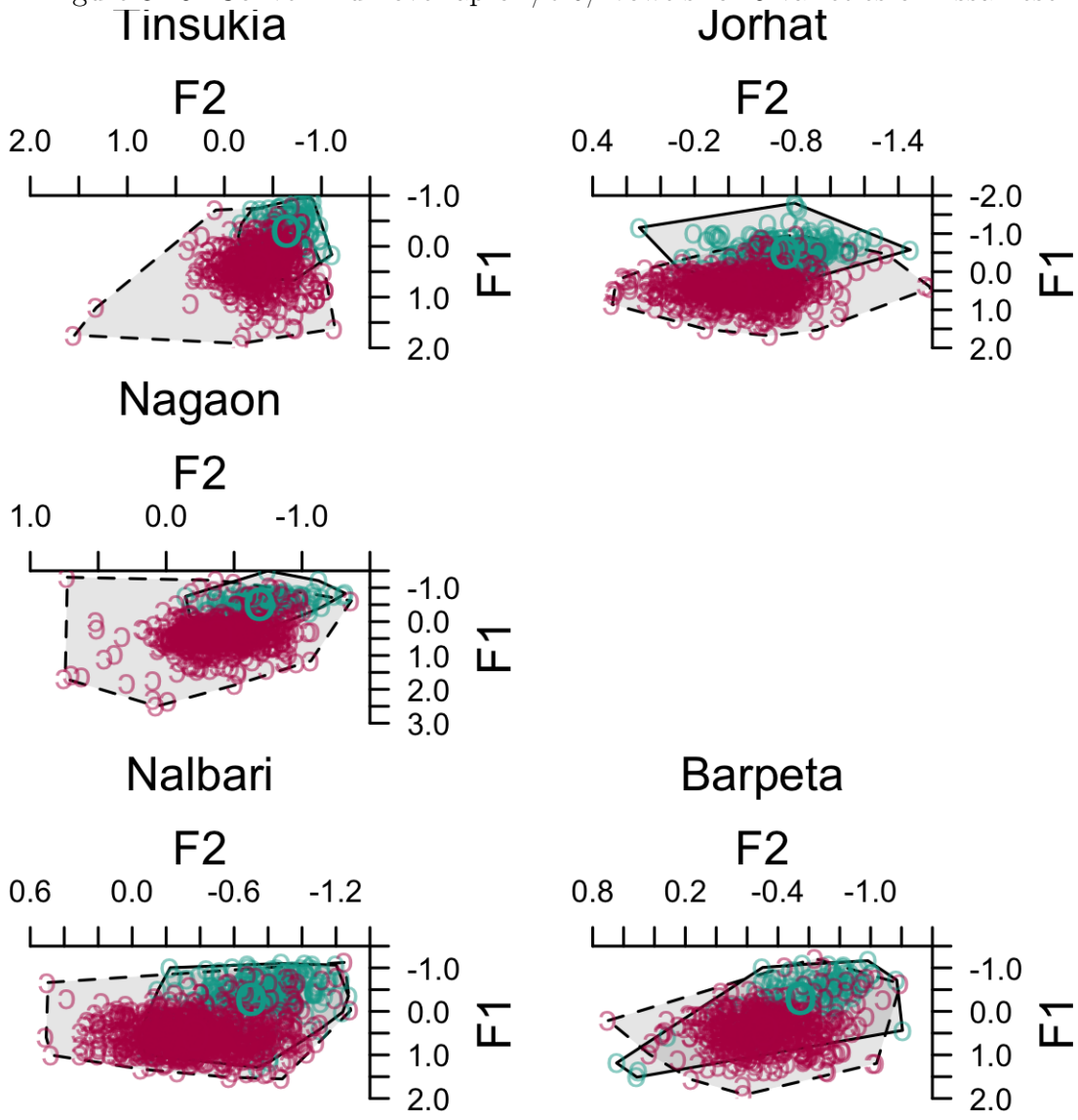


Figure 3.25: Convex Hull overlap of /o-ɔ/ vowels for 5 varieties of Assamese.



3.3 Conclusions

This chapter discussed the vowel overlaps in Assamese varieties in light of four measures for vowel mergers viz. Euclidean Distance, Pillai Bartlett Trace, Spectral Overlap Assessment Metric (SOAM) and Vowel Overlap Assessment with Convex Hulls (VOACH). In Section 3.1.1 we see that at least three types of segment merging mech-

anisms have been proposed in the literature. These are: merger by approximation, merger by transfer and merger by expansion. Section 3.1.2 discussed four measures of vowel mergers employed. These are: Euclidean Distance (ED), Pillai-Bartlett score, SOAM and VOACH. The results of these merger analysis on Assamese varieties confirm that all varieties show merger in at least one pair of vowels.

Figure 3.26: Normalized formant frequency plots for Tinsukia speakers with 1 standard deviation ellipses.

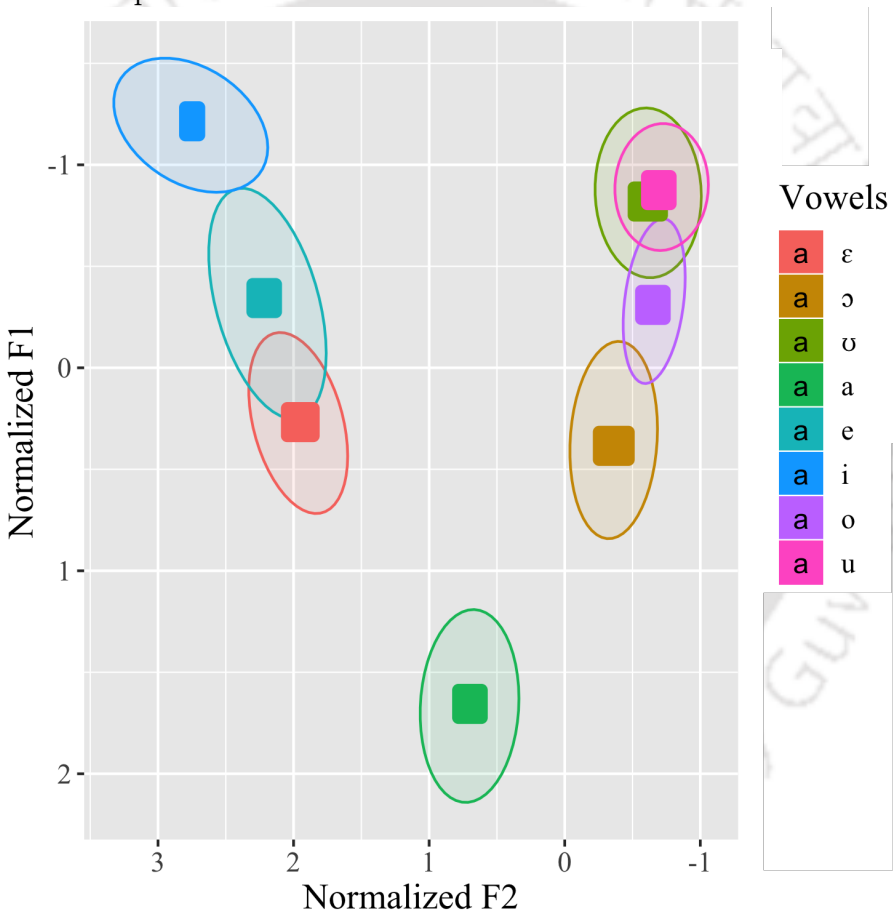


Table 3.7: Overlap measures for four vowel pairs in Tinsukia

Vowel pair	Euclidean Distance	Pillai Score	SOAM	VOACH
/ɛ-e/	89	0.29	50.5%	1.63
/u-ʊ/	**31**	**0.02**	**100%**	1.06
/ʊ-o/	50	0.24	92.3%	1.00
/o-ɔ/	124	0.41	6.0%	1.26

Figure 3.27: Normalized formant frequency plots for Jorhat speakers with 1 standard deviation ellipses.

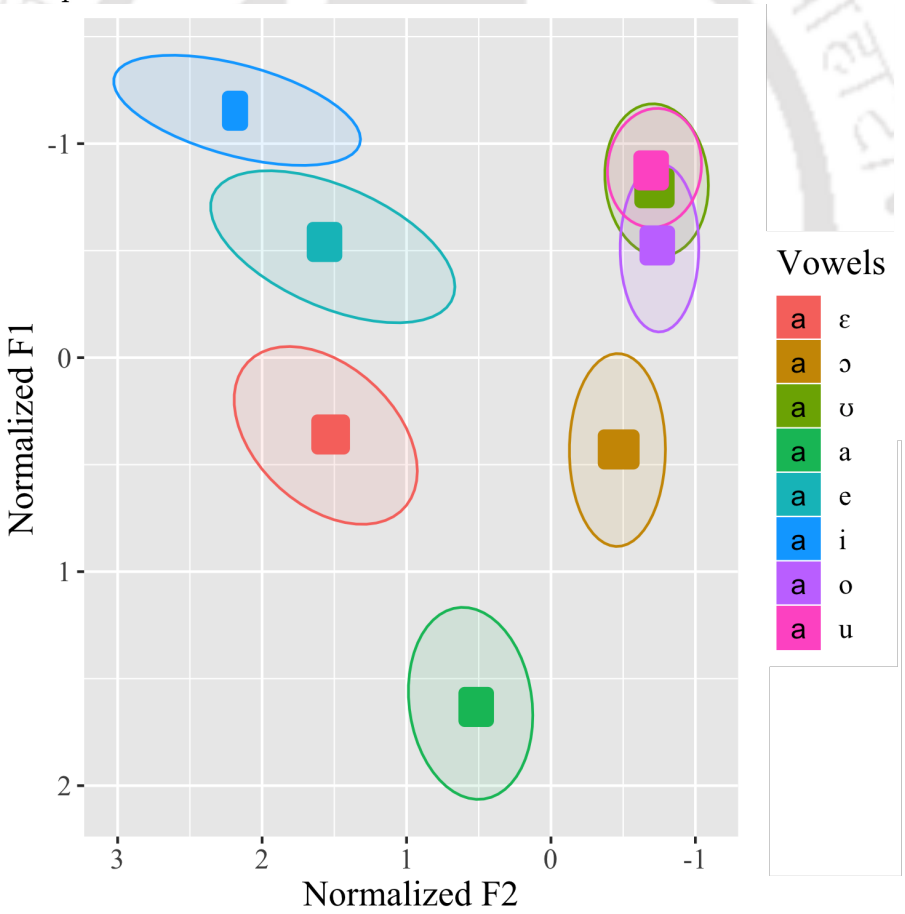


Table 3.8: Overlap measures for four vowel pairs in Jorhat

Vowel pair	Euclidean Distance	Pillai Score	SOAM	VOACH
/ɛ-e/	92	0.64	23.8%	2.09
/u-ʊ/	**12**	**0.02**	**99.5%**	1.03
/ʊ-o/	26	0.12	87.7%	1.03
/o-ɔ/	133	0.60	27.3%	2.11

Figure 3.28: Normalized formant frequency plots for Nagaon speakers with 1 standard deviation ellipses.

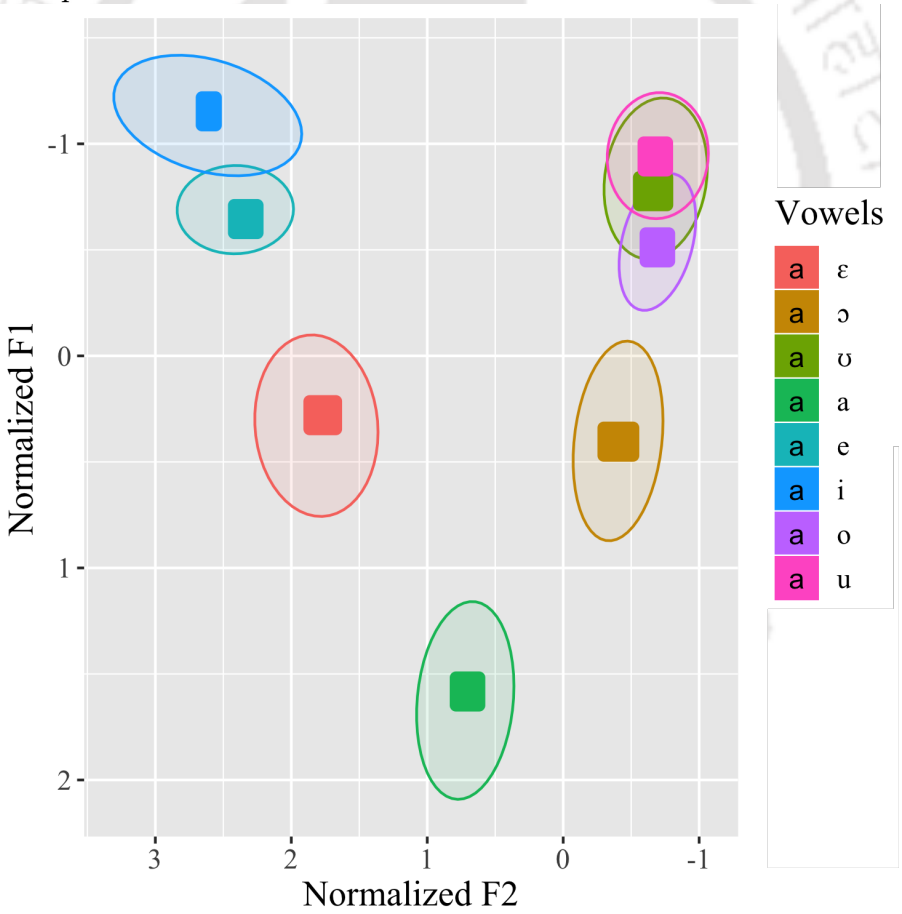


Table 3.9: Overlap measures for four vowel pairs in Nagaon

Vowel pair	Euclidean Distance	Pillai Score	SOAM	VOACH
/ɛ-e/	168	0.51	15.5%	1.09
/u-ʊ/	**17**	**0.05**	**98.3%**	1.28
/ʊ-o/	**28**	**0.09**	**92.7%**	1.25
/o-ɔ/	137	0.45	21.3%	1.11

Figure 3.29: Normalized formant frequency plots for Nalbari speakers with 1 standard deviation ellipses.

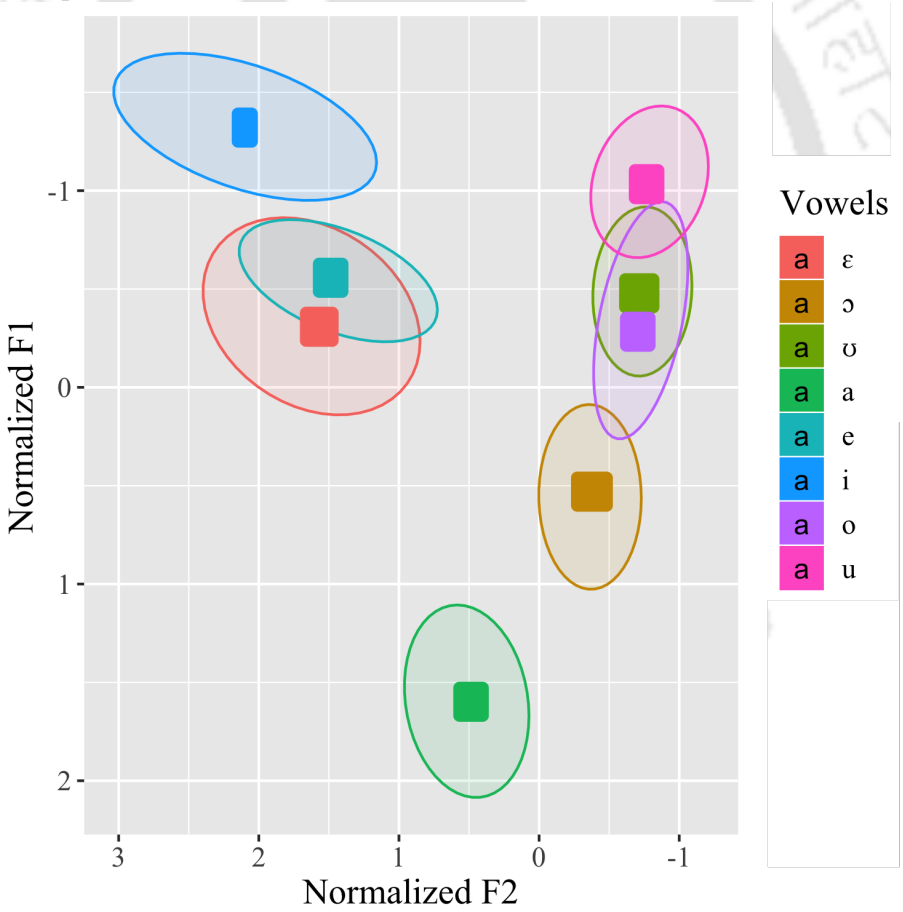


Table 3.10: Overlap measures for four vowel pairs in Nalbari

Vowel pair	Euclidean Distance	Pillai Score	SOAM	VOACH
/ɛ-e/	**36**	**0.08**	**98.4%**	1.05
/u-ʊ/	58	0.42	51.8%	1.14
/ʊ-o/	**19**	**0.05**	**84.8%**	1.25
/o-ɔ/	144	0.46	48.4%	1.04

Figure 3.30: Normalized formant frequency plots for Barpeta speakers with 1 standard deviation ellipses.

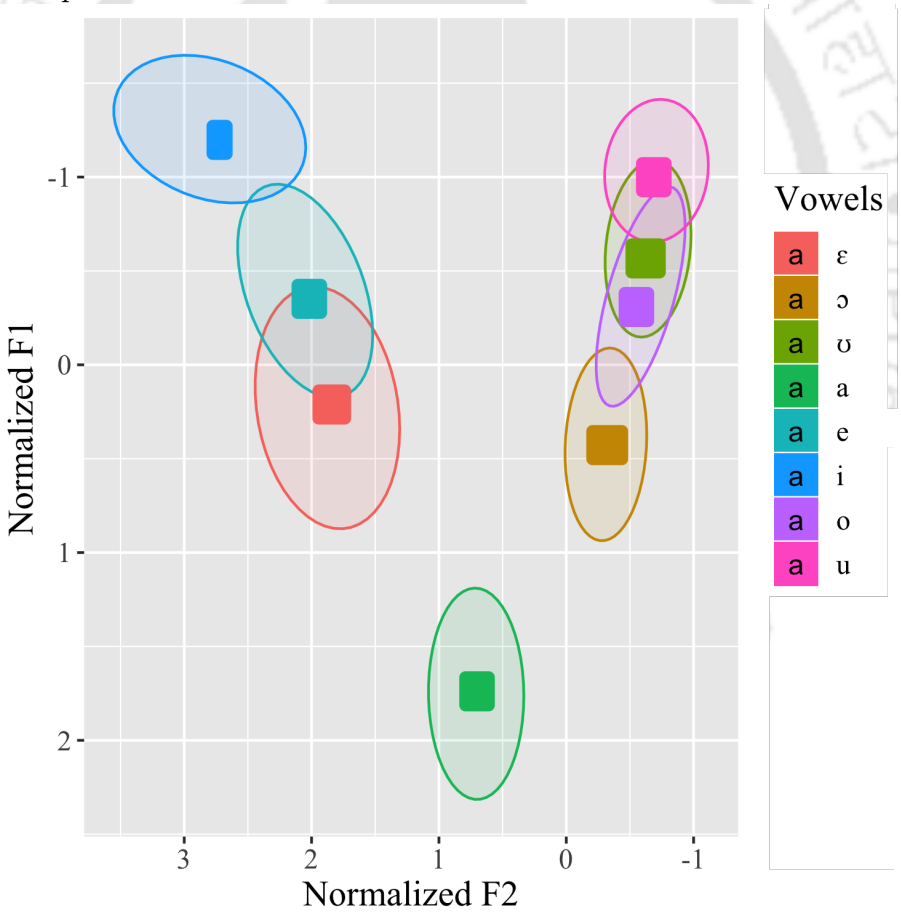


Table 3.11: Overlap measures for four vowel pairs in Barpeta

Vowel pair	Euclidean Distance	Pillai Score	SOAM	VOACH
/ɛ-e/	**70**	**0.15**	**67.3%**	1.03
/u-ʊ/	49	0.21	67.8%	1.28
/ʊ-o/	**37**	**0.06**	**82.0%**	1.25
/o-ɔ/	110	0.33	55.7%	1.11

Drawing from the conclusions of the overlap estimations using ED, Pillai score SOAM and VOACH, robust merger of vowels in the varieties of Assamese has been seen. In case of back vowels, the overlaps are asymmetrical, in the sense that the /ʊ/ vowel shows two opposite directions of vowel merger in the varieties. Statistical tests in Chapter 2 concluded that Tinsukia and Jorhat varieties do not distinguish between the vowels /u/ and /ʊ/. The merger of these two vowels in the two varieties is confirmed by the overlap measures calculated in this chapter. The /ʊ/ vowel moves to the /u/ vowel space and is completely absorbed by /u/ as observed in Figures 3.31 and 3.32. Recalling from Section 3.1.1, merger by approximation occurs when one vowel gradually moves to the phonetic space of the other vowel and the former category is completely absorbed by the latter. Phonetic space between the two phonemes become small enough to no longer maintain a contrast between them (Trudgill & Foxcroft, 1978). Hence, it can be concluded that /ʊ-u/ merger in the two Eastern Assamese varieties is a merger by approximation.

Figure 3.31: Scatter plot of /u/ and /ʊ/ vowels in Tinsukia variety.

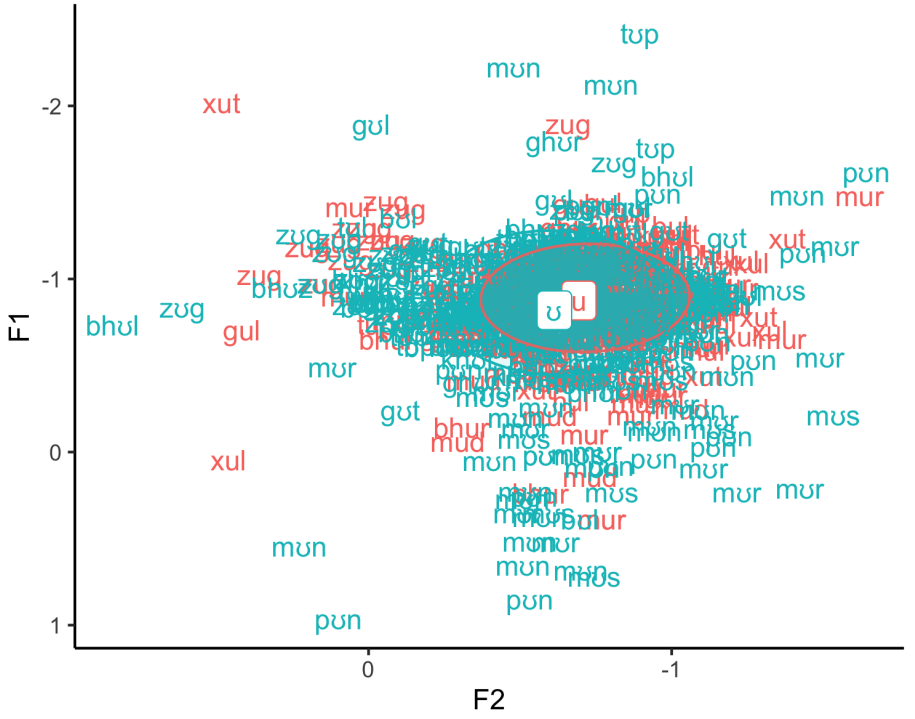
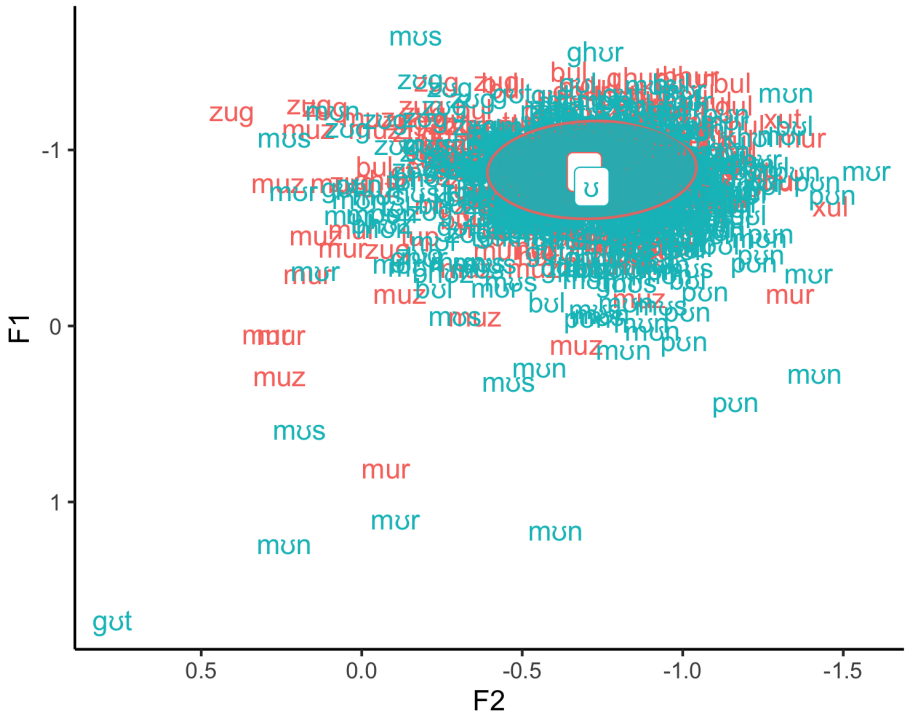


Figure 3.32: Scatter plot of /u/ and /ʊ/ vowels in Jorhat variety.



On the other hand, in case of the Western Assamese varieties Nalbari and Barpeta, the vowel /ʊ/ shows an opposite direction of movement and merges with vowel /o/. Statistical tests in Chapter 2 showed no significant distinction between the two vowels only for backness but not for height. Euclidean distance for Nalbari shows a value of 19 Mels which indicates that the two vowels are closer in proximity. In case of Barpeta however, ED value is 37 Mels. The Pillai scores of 0.05 and 0.06 for Nalbari and Barpeta respectively indicate complete overlap of the two vowels. So does the SOAM measures of 84.8% and 82% respectively. Scatter plots of the /ʊ/ and /o/ vowels in Figures 3.33 and 3.34 as well as a visual examination of the vowel plots in Figures 3.29 and 3.30 show that the ellipses of the two vowels for both the varieties are more spread and overlapped along F1. Hence, this looks more like a merger by expansion where the categorical boundaries between the two vowels have disappeared constituting a larger categorical space.

Figure 3.33: Scatter plot of /ʊ/ (denoted by "Uu") and /o/ vowels in Nalbari variety.

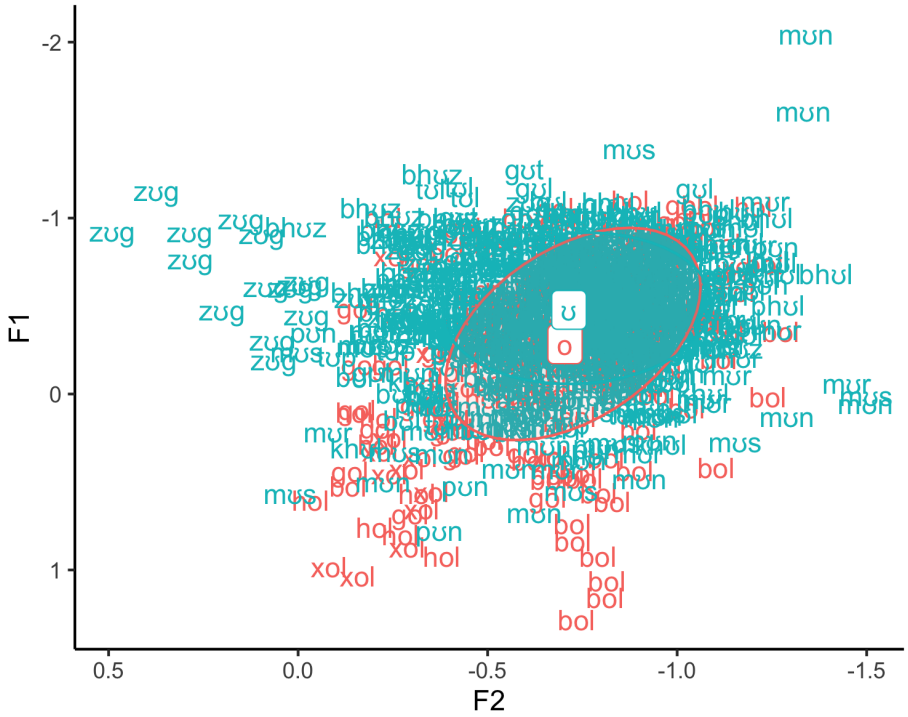
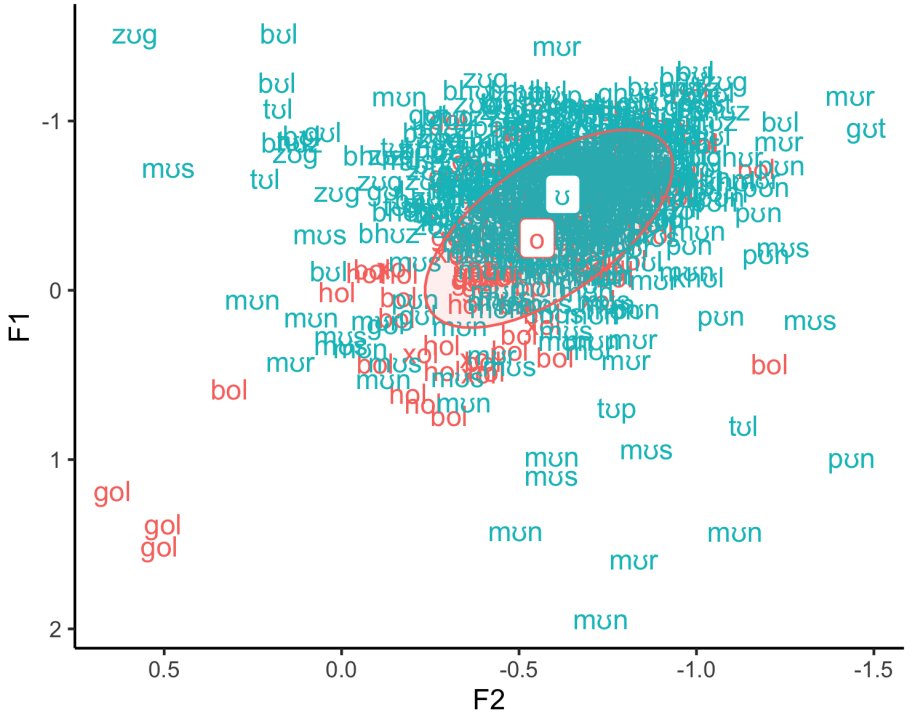
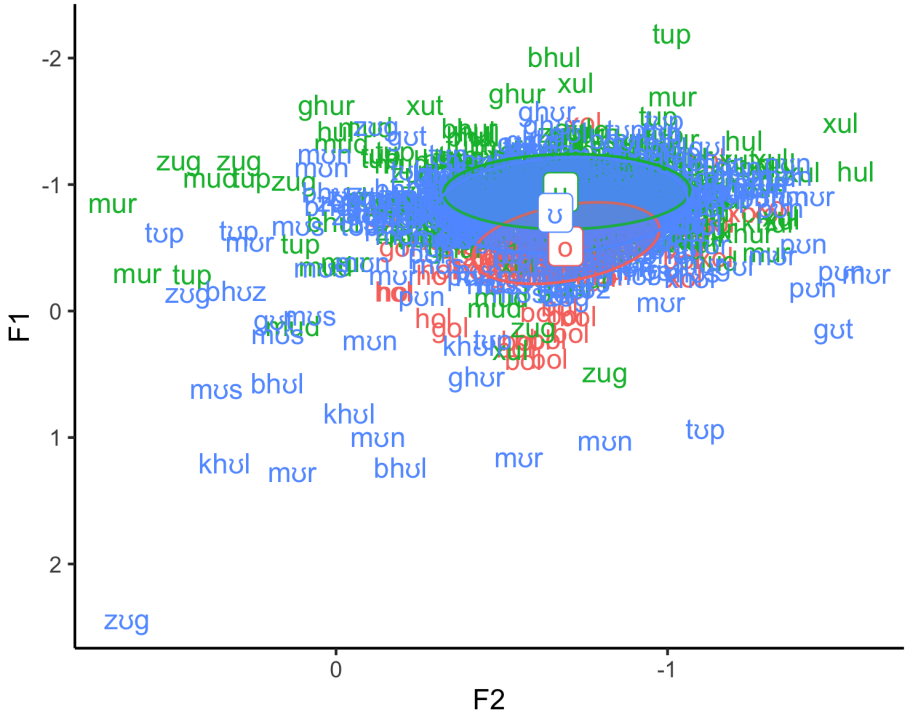


Figure 3.34: Scatter plot of /ʊ/ and /o/ vowels in Barpeta variety.



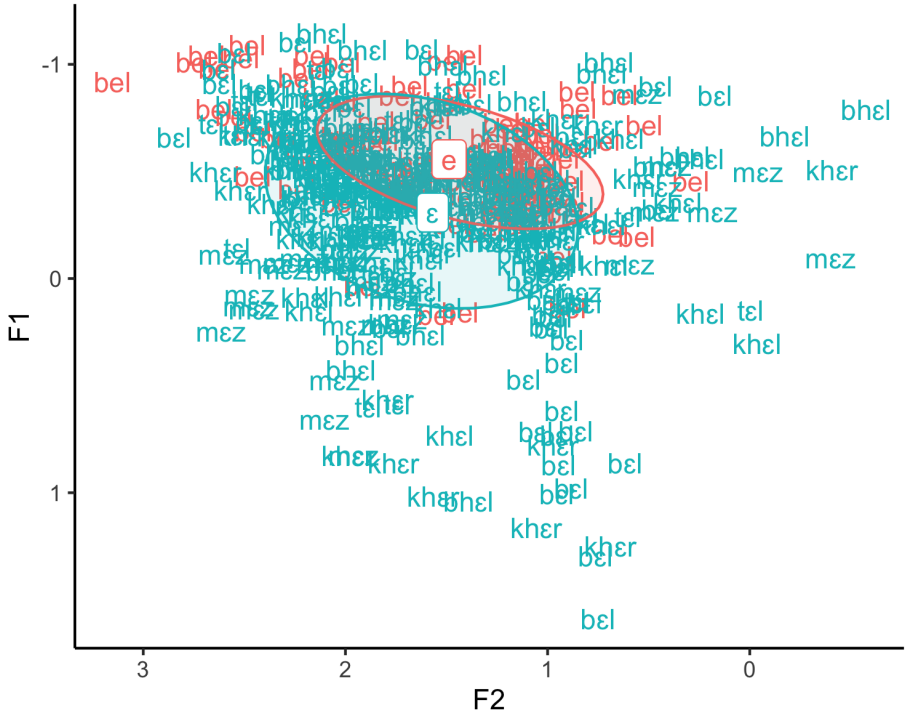
In case of Nagaon, visual examination of the vowel plot, formant frequencies as well as overlap measures indicate the merger of /ʊ-u/ vowels by approximation although statistical results show no distinction between the two vowels only in backness. The results also show considerable overlap of the vowel /ʊ/ with the vowels /o/. In chapter 2 it was pointed out that the overlap of three back vowels in Nagaon could be due to the variety's intermediate position between Eastern and Western Assamese, with influence from both the varieties. However, if we look at the scatter plot in Figure 3.35, we observe that although /ʊ/ seems to have a larger phonetic space sharing the same with both /u/ and /o/, more tokens of vowel /ʊ/ share the space with /u/ than with /o/. It can be postulated that although the vowel /ʊ/ has moved to the phonetic space of /u/, the vowel sometimes tend to share space with /o/.

Figure 3.35: Scatter plot of /u/, /ʊ/ and /o/ vowels in Nagaon variety.



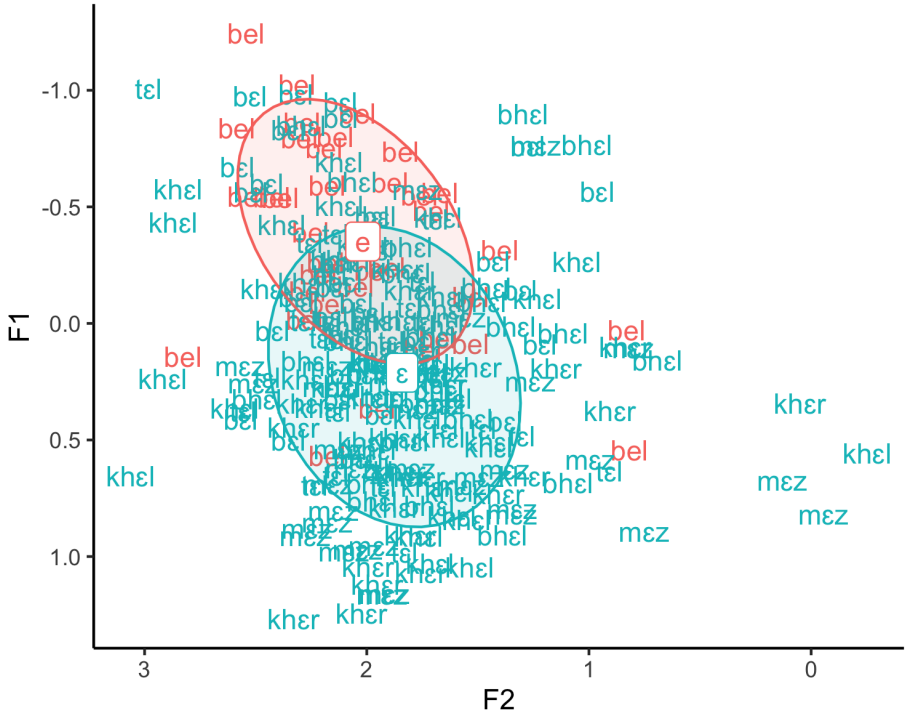
For the front vowels, both EA and CA do not indicate any merger. Although Tinsukia shows slight overlap of /e/ and /ɛ/ vowels as seen in Figure 3.26, none of the overlap measures indicates any significance of the overlap. However, WA varieties show definite overlap of the /e/ and /ɛ/ vowels. We have already seen in Chapter 2 that the vowel /ɛ/ in both Nalbari and Barpeta are higher in the vowel space (lower F1) than the EA and CA varieties. As observed in the scatter plot in Figure 3.36, the vowel /ɛ/ seems to have moved to the phonetic space of /e/ and is in the process of being completely absorbed by the latter. From both visual examination of the plots as well as overlap measures, it can be concluded that the merger in the Nalbari variety is a merger by approximation nearing completion.

Figure 3.36: Scatter plot of /e/ and /ɛ/ vowels in Nalbari variety.



In case of Barpeta, Euclidean distance shows a value of 70 Mels which does not indicate very close proximity. However, Pillai score of 0.15 and SOAM percentage of 67.3% indicates that the overlap is significant. Lower F1 of the vowel /ɛ/ and as well as its movement to the phonetic space of /e/ as seen in the Figure 3.37 indicates a merger by approximation similar to Nalbari, where the merger is yet to be complete.

Figure 3.37: Scatter plot of /e/ and /ɛ/ vowels in Barpeta variety.



Hence, we summarize a few points in conclusion. Firstly, vowels /ʊ/ and /u/ merge by approximation in the Tinsukia and Jorhat varieties whereby /ʊ/ moves to the phonetic space of /u/ and is completely absorbed by the latter. Secondly, vowels /ʊ/ and /u/ merge by approximation in the Nagaon variety, however /ʊ/ sometimes has a tendency to share space with /o/, most probably due to the region's intermediate position between EA and WA and the consequent influences from both varieties. Third, vowels /ʊ/ and /o/ merge by expansion in Nalbari and Barpeta varieties whereby both /ʊ/ and /o/ share a larger phonetic space. Next, vowels /e/ and /ɛ/ merge by approximation in Nalbari and Barpeta varieties whereby /ɛ/ shows incomplete movement towards the phonetic space of /e/. Lastly, with the complete mergers of back vowels, as well as the merger in progress of the front vowels, the canonical eight-vowel system of Assamese reduces to a seven-vowel one in Tinsukia, Jorhat and Nagaon varieties and to a six-vowel one in Nalbari and Barpeta varieties.

Scholars have pointed out that there is often a production-perception anomaly in

situations of mergers (Hay et al., 2010; W. Maguire et al., 2013). While the mergers are quite robust in terms of production, whether they are perceived the same can only be determined after a perception test is conducted to examine the speakers' ability to discriminate between the vowel pairs. Chapter 4 discusses the results of a set of perception experiments conducted to ascertain vowel mergers in the varieties of Assamese.

Chapter 4

Perceptual evidence

4.1 Introduction

The analysis of the production data in chapters 2 and 3 revealed a few insights about the production of Assamese vowels. Speakers of Tinsukia, Jorhat and Nagaon tend to merge the /u/ and /ʊ/ vowels and the Nalbari and Barpeta speakers tend to merge the /ʊ/-/o/ and /ɛ/-/e/ vowels. As has been pointed in studies of mergers that there is often a production-perception anomaly in situations of mergers (Hay et al., 2010; W. Maguire et al., 2013) this chapter attempts to test if the results of the production experiments from chapters 2 and 3 are also attested perceptually or is there a production-perception inconsistency in Assamese dialects. Two listening experiments, viz discrimination test and identification test are used to analyze the perception of vowels or mergers among speakers of the dialects. The results of the perception experiments are analyzed using D-prime, a statistic to measure sensitivity or discriminability.

4.2 Methodology

This section discusses the methodology adopted for the analysis of perception experiments conducted on participants from the five Assamese varieties. Section 4.2.1

presents the materials used as stimuli for the perception tests. Section 4.2.2 discusses the method of data collection and gives brief information of the participants in the study. Sections 4.2.3 and 4.2.4 describes the procedures of the two perception tests conducted. Finally, Section 4.2.5 discusses the D-prime analysis adopted in this study to account for the sensitivity of the participants to the stimuli.

4.2.1 Materials

As mentioned earlier, this chapter tries to test if the results of the production experiments from chapters 2 and 3 are also attested perceptually. For this purpose a list of words was prepared with the Assamese vowels in CVC contexts as presented in Table 4.1. The list consisted of both meaningful as well as nonsense words.

Table 4.1: List of words for perception tests

context	a	e	ɛ	i	ɔ	o	ʊ	u
b_l	bal	bel	bɛl	bil	bɔl	bol	bʊl	bul
g_l	gal	gel	gɛl	gil	gɔl	gol	gʊl	gul
m_h	mah	meh	mɛh	mih	mɔh	moh	mʊh	muh
s_k	sak	sek	sɛk	sik	sɔk	sok	sʊk	suk
z_r	zar	zer	zɛr	zir	zɔr	zor	zʊr	zur
z_t	zat	zet	zɛt	zit	zɔt	zot	zʊt	zut

The words were recorded from one male native speaker of Assamese who is also a phonetician and is aware of all the distinct eight vowels of standard Assamese. The recordings were done in a noiseless environment. For recording the data, a Tascam linear PCM recorder was used. A Shure unidirectional head-worn microphone was connected to the recorder via an xlr jack. The sampling frequency was kept at 44.1 kHz and 24 bit in .wav format. The words recorded from the list served as the stimuli for the multiple forced choice (MFC) discrimination and identification tests.

4.2.2 Data collection and participants

The present perception test was conducted in two phases. In the first phase, data was collected in the field and the tests were conducted using a Praat script on a laptop. Participants in the first phase included three native speakers of Assamese each from Tinsukia, Jorhat, Nagaon, Nalbari and Barpeta. They were made to sit in front of a computer screen and the sounds were played out to them using a Shure 1440 headphone. The second phase of data collection was carried out at a time when restrictions were imposed on physical contact with people due to the COVID-19 pandemic. Hence, the tests were re-designed to be conducted online using Google forms. Details about the procedures are given in the following sections. Probable candidates were first asked to attempt a mock test designed to acquaint them with the actual discrimination and identification experiments. Once they completed the mock tests and agreed to participate in the actual tests, the links to the same along with detailed instructions were sent to them via emails. Participants could use either a laptop or a mobile phone and were requested to use headphones/earphones and be at a quiet place while attempting the tests. They were urged to complete the tests in one sitting. For the online experiment, a total of 74 participants from the 5 regions attempted the mock test, out of which 41 participated in the actual perception test. 4 participants were rejected since they revealed to have lived in multiple locations until the age of 15 years. The responses of 1 more participant were rejected since she was a phonetician herself. Hence, online responses from a total of 36 participants were included in the final analysis. The total number of participants thus came to be 51. Details of the participants for the perception tests are provided in Tables A.3 and A.4 in Appendix A. They were in the age group of 20-50 years and reported no speech or hearing disorders. Apart from Assamese, the speakers could also understand Hindi and English in varying degrees.

4.2.3 Perception test 1: Discrimination test

In discrimination tests, participants are presented with two sounds one after the other and are asked to judge if the sounds they hear are similar or different. The present experiment aims to check if the participants distinguished between different vowels when presented in pairs. In our production analysis it was seen that all vowels except /a/ showed some amount of overlap with their adjacent vowels. Therefore for this experiment the seven vowels taken for analysis were /i/, /e/, /ɛ/, /u/, /ʊ/, /o/, and /ɔ/. The front vowels /i/, /e/ and /ɛ/ were paired with each other and with themselves. Similarly, the back vowels /u/, /ʊ/, /o/, and /ɔ/ were paired with each other and with themselves. Considering there are six consonantal contexts as seen in Table 4.1, the total number of word pairs taken as stimuli for each of the front and back vowels were 18 and and 24 respectively. It is to be noted that the word pairs always had the same consonantal contexts, the difference being only in the vowel. For example, the pairs /bel/-/bɛl/ or /gul/-/gʊl/ and never a /bel/-/gel/ pair. Therefore, in total 150 word pairs were taken as stimuli for the discrimination test.

In this test, two sounds were played one after the other. Along with the sounds, there appeared on the computer/mobile screen, the options "Same" and "Different" for the participants to click. After hearing the two sounds they were required to click on either of the options based on their judgement of the words. Before each stimuli there was an initial silence of 0.5 seconds and between each words of the pair there was a medial silence of 0.8 seconds. Before the experiment the participants were instructed on how to proceed with the test and were also given a trial session to acquaint them with the experiment. The results of the test were then collected on an excel sheet.

4.2.4 Perception test 2: Identification test

In identification tests participants were presented with a sound and were asked to choose the correct meaning of the word they heard, from a list of options. The

experiment aimed to check if the participants were able to correctly identify the words presented to them. Stimuli consisted of 22 words presented to the participants using a Shure 1440 headphone. The stimuli were randomly repeated 3 times, making the total number of stimuli 66. With each sound, there appeared a list of options on the screen from which the participants were to select the option they thought was the correct meaning of the word. If, for example, the word played to the participant is /bɛl/ meaning "woodapple", since the target vowel being tested here is /ɛ/, a front vowel, the options included the meanings of the words containing all the front vowels that the participant could confuse.

Similarly, while testing each of the back vowels, the options included the meanings of the words containing all the back vowels that a participant could confuse. Before each stimuli there was an initial silence of 0.5 seconds. The participants had the option to go back to the previous stimuli if they thought they clicked on the wrong option by mistake. They also could choose more than one option for each sound if they thought the word fitted in two or more sentences from the given options. Before the experiment the participants were explained how to proceed with the test and were also given a trial session to familiarize them with the experiment. The results of the test were then collected on an excel sheet.

4.2.5 Calculation of D-prime

D-prime (d′) or Discriminability index is a statistic used in signal detection theory to measure sensitivity or discriminability that is unaffected by response biases. Simply, it is a measure of an individual's ability to detect signals. It provides the separation (in standard deviation units) between the means of the noise and signal plus noise distributions. In a presentation following Macmillan & Creelman (2004), Keating (2005) uses detection theory to frame sensitivity as detecting a signal (for example, against background noise or as compared to another signal). She models how a per-

ceiver decides whether a signal is present or not. An AX (same-different) experiment is designed which presents signals and non-signals to the subjects. The subjects are then asked to detect all and only the signals. The following Table 4.2 presents the way an AX experiment is traditionally viewed. The possible outcomes of the experiment are as follows: "Yes" represents the presence of a signal or the difference to be detected, "No" represents the absence of a difference.

Table 4.2: An AX experiment as presented by Keating (2005).

	Response: Different (yes)	Response: Same (no)
Stimuli: YES (different)	HIT	MISS
Stimuli: NO (same)	FALSE ALARM	CORRECT REJECTION

If calculated as proportions of the row totals, the alarms can be viewed as estimates of probabilities of responses:

Hit rate H: proportion of YES trials to which subject responded YES = P("yes" | YES)

False alarm rate F: proportion of NO trials to which subject responded YES = P("yes" | NO)

If the table is rewritten with these and the other two rates, each row will total to 1.0. However, what is of interest is the results of the pair (H,F). We can then consider the performance of three subjects: the perfect subject will have a performance of (1,0), a random subject will have H=F and a subject who always answers YES will have a performance of (1,1). That is to say, that the best subject maximizes H (thereby minimizing the Miss rate) and minimizes F (thereby maximizing the Correct Rejection rate). Therefore, the larger the difference between H and F, the better is the subject's sensitivity. This difference is measured using the statistic d′ ("d-prime"). It is the distance between the Signal and the Signal + Noise. However, d′ is the difference between the z-transforms of these 2 rates, and not simply H-F, as given in equation

4.1:

$$d' = z(H) - z(F) \tag{4.1}$$

Neither H nor F in the equation can be 0 or 1 (values are adjusted slightly up or down, if so).

D-prime for the AX (discrimination) experiment in this study was calculated following Macmillan & Creelman (2004)'s method. For ease of reference in the discussion, the same vowel stimuli (like /i-i/) are referred to as A-A stimuli and the different vowel ones are referred to as A-B stimuli. Responses of the participants were categorized as follows:

A-B stimuli: "Hit" if response was "different" and "Miss" if response was "same".

A-A stimuli: "False alarm" if response was "different" and "Correct rejection" if response was "same".

The following example of discrimination test results for the /ʊ-o/ pair by participants from Barpeta shows the d′ calculation procedure used in this work.

Table 4.3: Example of d′ calculation.

Stimuli	Hit	Miss	False alarm	Correct Rejection	Total
/ʊ-ʊ/	0	0	2	64	66
/ʊ-o/	92	40	0	0	132
/o-o/	0	0	4	62	66

Hit rate, H = No. of "Hit" responses for /ʊ-o/ stimuli / Total no. of stimuli
= 92/132 = 0.69

False alarm rate, F = (No. of "False alarm" responses for /ʊ-ʊ/ stimuli + No. of "False alarm" responses for /o-o/ stimuli) / (Total no. of /ʊ-ʊ/ stimuli + Total no.

of /o-o/ stimuli)

= (2+4)/(66+66) = 6/132 = 0.04

z-transforms of H and F were calculated using the Excel formula = $NORMINV(H, 0, 1)$ where, *NORMINV* returns the inverse of the normal cumulative distribution for the specified mean and standard deviation (SD), 0 being the specified arithmetic mean and 1 being the specified SD. It is to be noted that neither H nor F can be 0 or 1. In such cases, the values of H and F were slightly adjusted to 0.99 and 0.01 respectively.

Thus, z(H) = 0.5 and z(F) = -1.7

and d′ = z(H) - z(F) = 0.5 - (-1.7) = 2.2

4.3 Results

This section presents the results of the perception tests conducted on participants from Tinsukia, Jorhat, Nagaon, Nalbari and Barpeta. Discrimination test results are calculated in terms of four responses: Correct rejection, False alarm, Hit and Miss and d′ scores for vowel contrasts were calculated. Table 4.4 provides the d′ scores and percentage of Hit responses for five vowel contrasts for each of the regions. Figure 4.1 plots the corresponding d′ scores. Higher d′ scores means higher sensitivity in discriminating the contrasts. Figure 4.2 shows the results of the correct identification of vowels in each of the regions. Detailed responses for the perception tests are provided in Appendix C. The following subsections discuss the results of the perception experiments for each variety. Bar plots of the responses (percentage values are rounded to the nearest whole number) for the two tests are provided in subsequent figures from 4.3 to 4.22.

Table 4.4: Discrimination results for vowel contrasts in five regions.

Region	/i-e/		/e-ɛ/		/u-ʊ/		/ʊ-o/		/o-ɔ/	
	d′	% Hit	d′	% Hit	d′	% Hit	d′	% Hit	d′	% Hit
Tinsukia	4.0	94.2	3.1	75.8	2.2	59.2	2.7	83.3	3.2	90.8
Jorhat	3.6	91.7	2.7	68.1	1.4	38.9	2.7	84.0	3.5	95.8
Nagaon	4.7	100.0	3.6	89.8	2.2	54.6	2.6	80.6	4.4	99.1
Nalbari	4.4	100.0	2.3	66.7	2.2	61.1	2.1	63.0	3.6	93.5
Barpeta	4.0	98.5	1.6	48.5	2.4	57.6	2.2	69.7	3.0	92.4

Figure 4.1: d′ scores for vowel contrasts in the five regions.

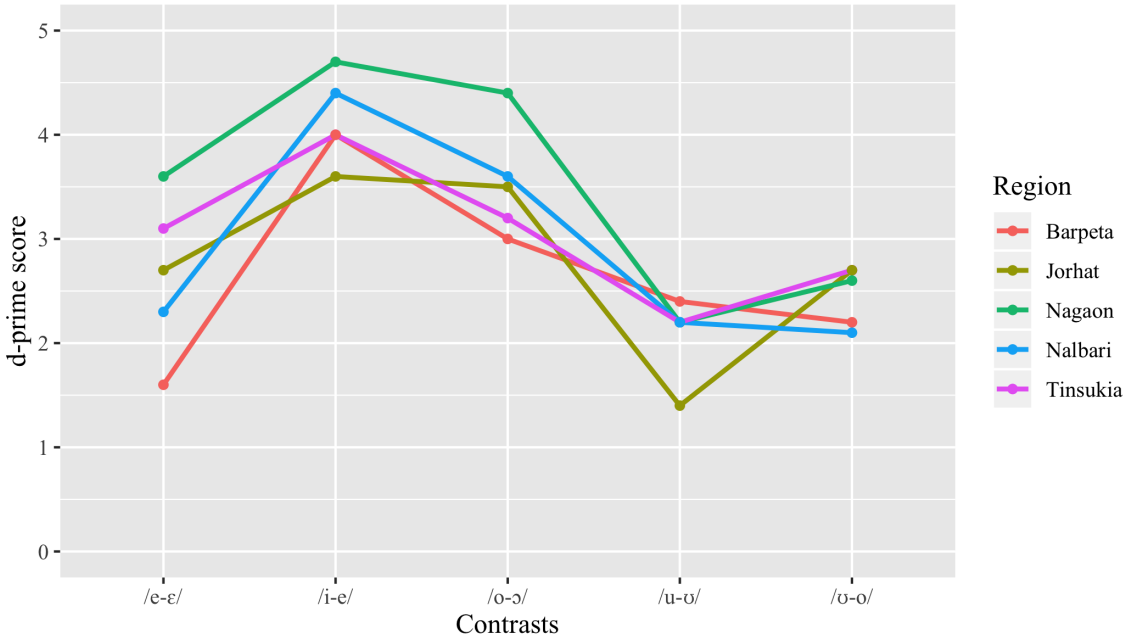
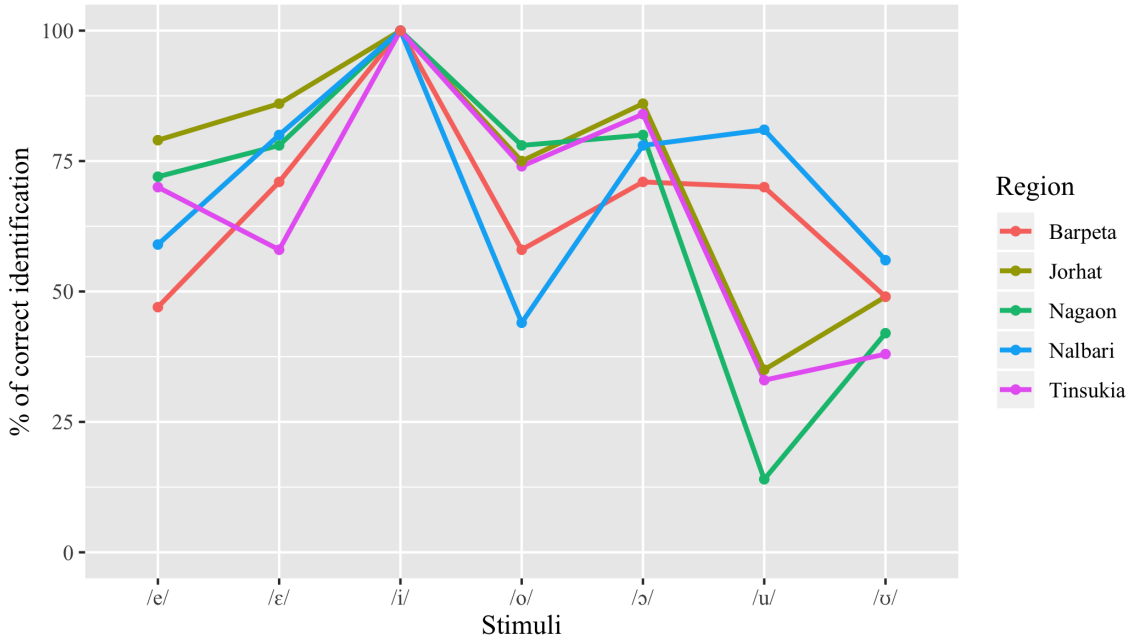


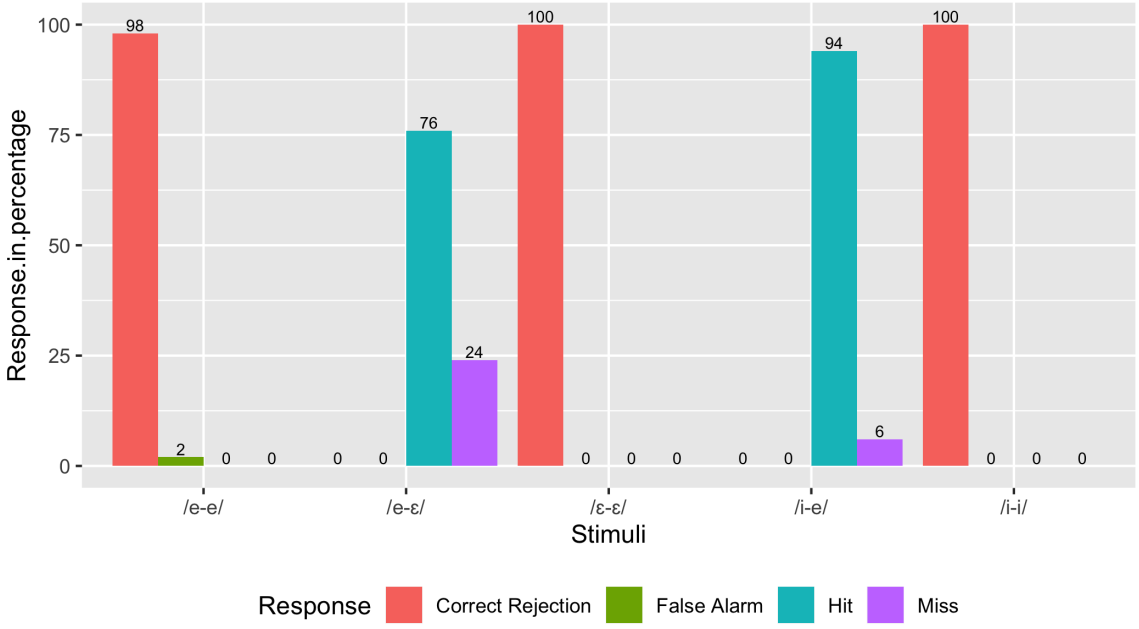
Figure 4.2: % of correct identification of vowels in the five regions.



4.3.1 Perception results for Tinsukia

Figure 4.3 shows the discrimination of front vowels by participants from Tinsukia. As seen in the figure, Correct rejection for the A-A stimuli is very high (100% for /i-i/, and /ɛ-ɛ/, 98% for /e-e/). That is, when the stimuli consisted of the same word consecutively like /i-i/, /e-e/ and /ɛ-ɛ/, the participants are able to correctly reject any difference in the stimuli. For the pair /i-e/, the percentage of Hit response is 94% and that of Miss response is 6%. That is, the participants were able to correctly discriminate between the two words 94% of the time. Similarly, the percentage of Hit response for /e-ɛ/ pair is 76% and Miss response for the same pair is 24%. That means, 76% of the times the participants were able to discriminate between the two words in the stimuli and hit the "Different" button and 24% of the times they hit the "Same" button.

Figure 4.3: Discrimination of front vowels by participants from Tinsukia



In case of the discrimination of back vowels we see a high percentage of Correct rejection for the A-A stimuli. As seen in figure 4.4 the percentage for /u-u/ is 100%, /ʊ-ʊ/ is 95%, and both /o-o/ and /ɔ-ɔ/ pairs have a percentage of 97. This indicates that the participants were able to reject distinction between the two words in the stimuli. In case of the A-B stimuli, the /o-ɔ/ pair has a Hit response of 91% and a Miss response of 9%. The /ʊ-o/ pair has a Hit response of 83% and a Miss response of 17%. This indicates that participants were able to distinguish between the two sounds in the pairs. Finally, for the /u-ʊ/ pair we see that the Hit response is 59% whereas the Miss response is 41%. This again indicates that participants were not able to distinguish between the two sounds.

Figure 4.4: Discrimination of back vowels by participants from Tinsukia

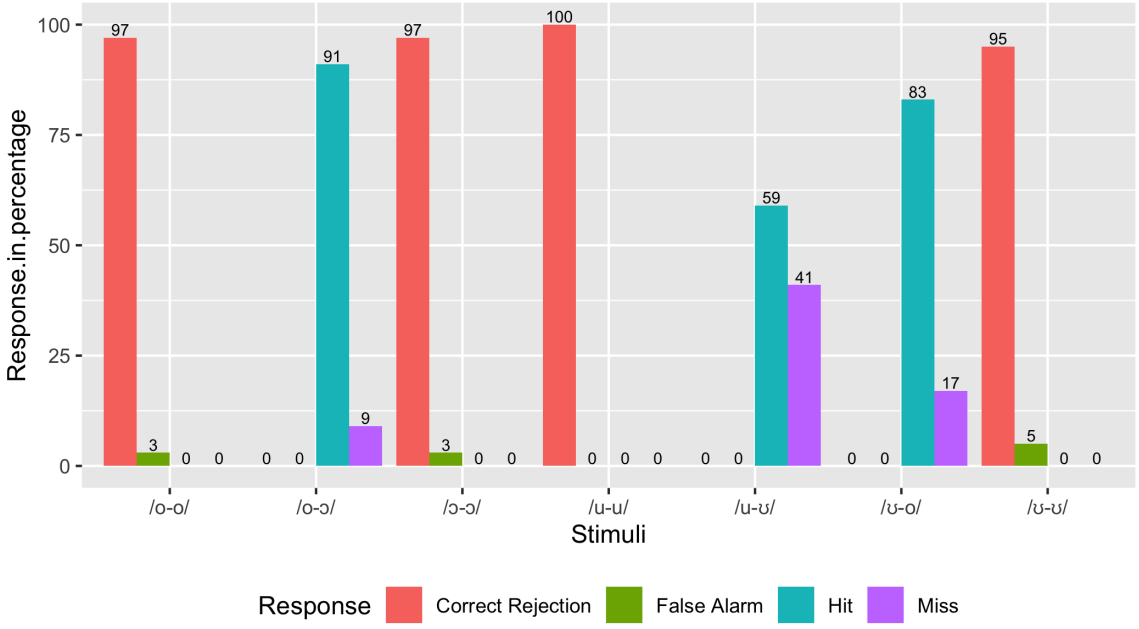


Figure 4.5 shows the identification of front vowels by participants from Tinsukia. As seen in the figure, when the /i/ stimuli was presented to the participants, they identified the word as /i/ 100% of the time. In case of the vowel /e/, 70% of the responses were for /e/, and 10% for /e, ɛ/ and 18% for /ɛ/ and 18% for /i/. In case of the vowel /ɛ/, participants identified the vowel as /e/ 58% of the time, 28% as /e/ and 12% as both /e/ and /ɛ/ .

Figure 4.5: Identification of front vowels by participants from Tinsukia

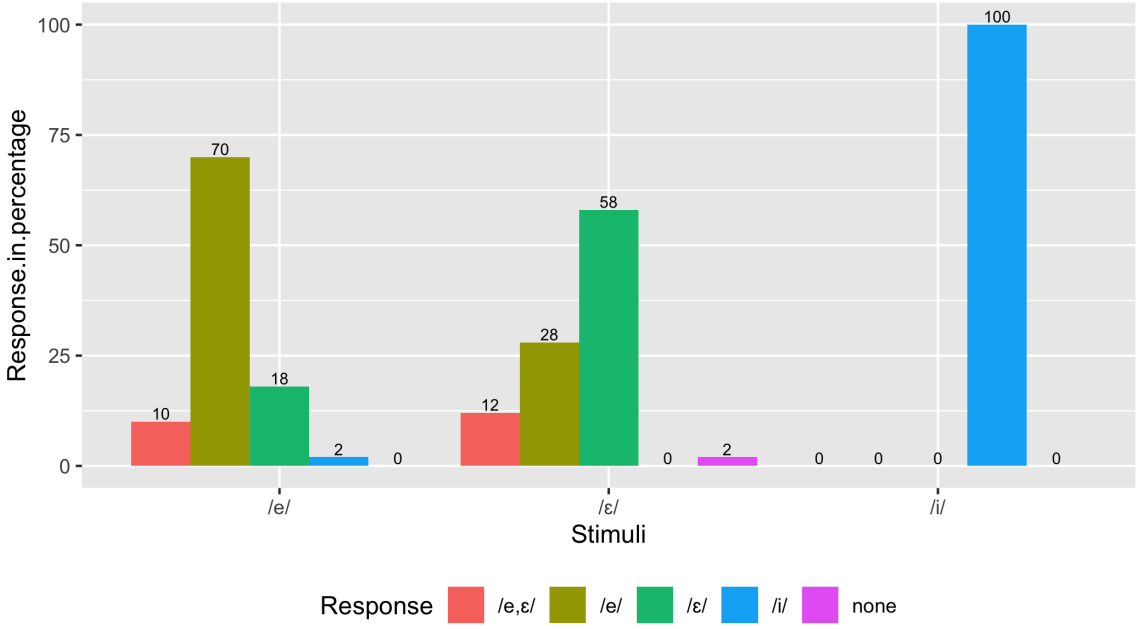
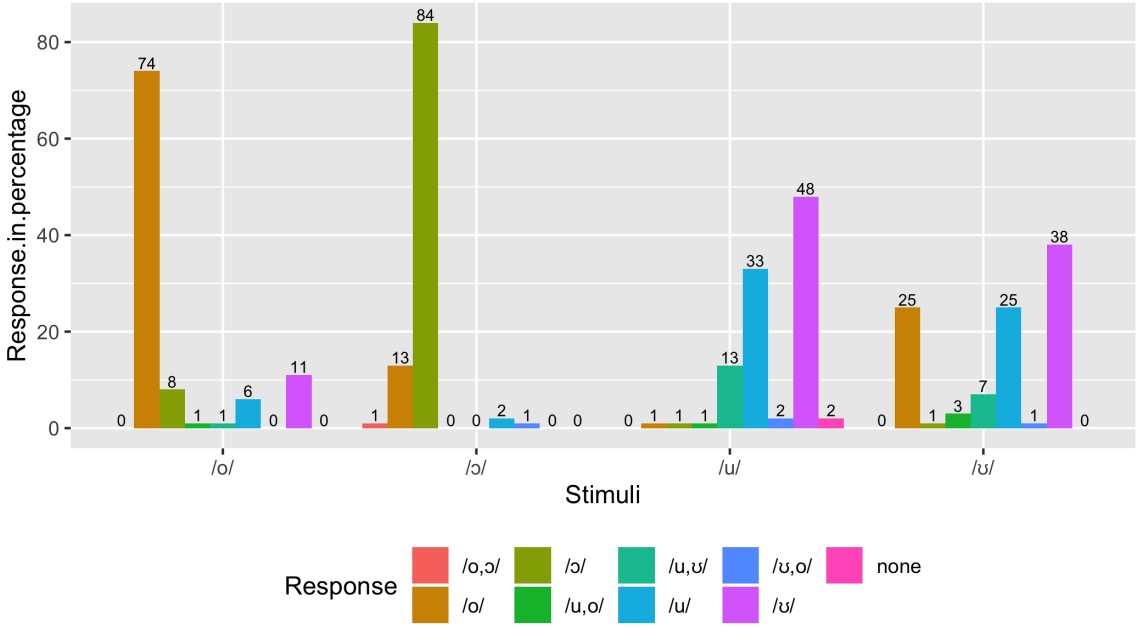


Figure 4.6 shows the identification of back vowels by participants from Tinsukia. We see a high percentage of correct identification for the vowels /o/ and /ɔ/ with 74% and 84% respectively. This indicates that the participants from Tinsukia did not confuse the two stimuli with other vowels and correctly identified them for most of the time. Vowels /u/ and /ʊ/ however, had mixed responses. For the stimuli for /u/, identification responses were 33% for /u/, 48% for /ʊ/ and 13% for /u,ʊ/. Similarly for the stimuli for /ʊ/ identification responses were 25% for /u/, 38% for /ʊ/, 25% for /o/ and 7% for /u,ʊ/. The varied responses for these two vowels indicate that the participants were confused while correctly assigning the responses to the stimuli.

Figure 4.6: Identification of back vowels by participants from Tinsukia



Chapter 3 discussed the measures determining the mechanisms of vowel mergers in Assamese varieties. It was observed, as seen in Figure 2.6 that Tinsukia speakers show an overlap of the vowels /u/ and /ʊ/ in production. Table 3.7 showed that the results of Euclidean distance, Pillai score, SOAM and VOACH confirm these observations and evidenced merger by approximation of the vowels /u/ and /ʊ/ whereby the vowel /ʊ/ moves to the phonetic space of vowel /u/ and the former is completely absorbed by the latter. Results of the discrimination test for the /u-ʊ/ pair show that the participants were less sensitive to the distinction between the two sounds. Identification test results show that in case of the vowel /u/, there were more hits for /ʊ/ than for /u/. In case of identification of /ʊ/ vowel 25% of hits were for /u/. This indicates that participants categorized the two vowels as one. Again from Table 3.7, the low Euclidean distance of 31 Mel between /u/ and /ʊ/ indicates a merger. It is also noted that the vowel /ʊ/ has been identified as /o/ 25% of the time. Discrimination of /ʊ-o/ shows 17% Miss response for the pair, indicating that

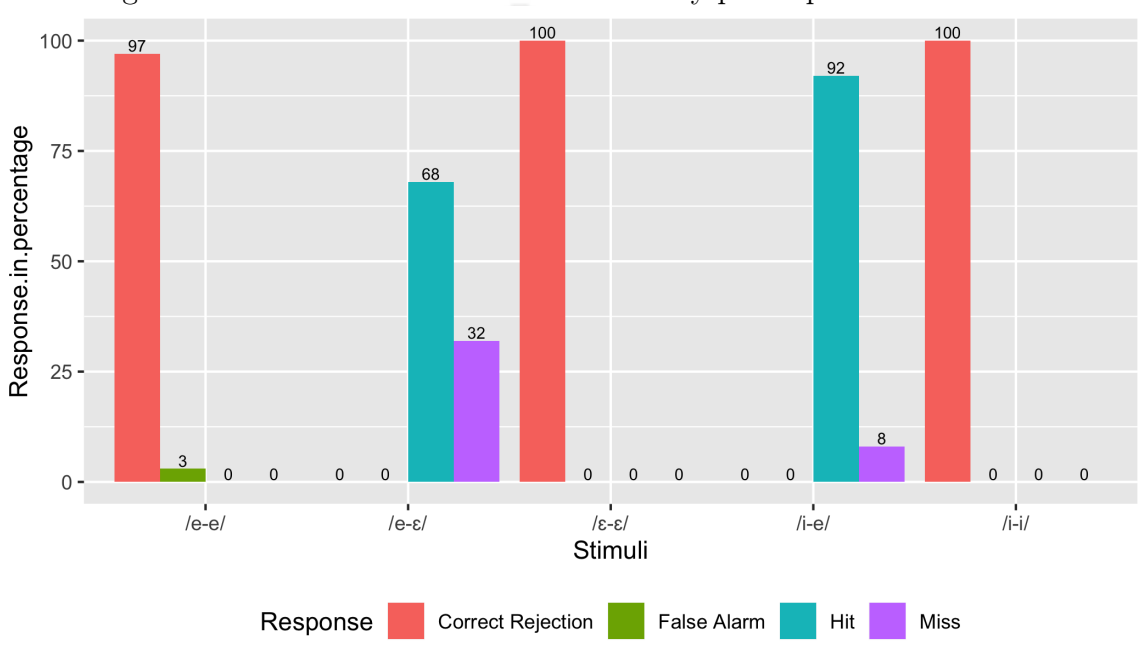
the participants categorized /ʊ/ as /o/ in a few cases. Fig 2.6 also shows some amount of overlap between the two vowels and overlap measures show a Pillai score of 0.24 and a SOAM score of 92.3%. which can be considered to be indicative of complete overlap. However, the results of all tests considered, the complete overlap of the vowels /u/ and /ʊ/ is more robust than the overlap of /ʊ/ and /o/. An explanation of the /ʊ/-/o/ overlap can be a probable movement of /o/ to a lower F1 to compensate for the space left behind by the merger of /ʊ/ with /u/. In perception, the confusion could arise from the higher F1 of a canonical /ʊ/ which the participants have categorized as an /o/ rather than as an /u/. Similarly, we see that 28% of the stimuli /ɛ/ has been identified as /e/. Discrimination tests also show some confusion among the participants in categorizing the two vowels. Our production results and merger measures on the other hand, do not indicate robust overlap of the /e/ and /ɛ/ vowels. The contrasting perception test results indicate slight asymmetry between production and perception. While speakers from Tinsukia did distinguish between the two vowels in production, they were unable to make the same distinction fully when it came to perceiving the two vowels. d′ scores calculated for the A-B vowel pairs reveal that Tinsukia participants have highest sensitivity in discriminating /i-e/ vowel pair (d′=4.0) and lowest in discriminating /u-ʊ/ vowel pair (d′=2.2). The results confirm that /u-ʊ/ vowel pair in Tinsukia merge both in production and in perception.

4.3.2 Perception results for Jorhat

Figure 4.7 shows the discrimination of front vowels by participants from Jorhat. As seen in the figure, Correct rejection for the A-A stimuli is 100% for /i-i/, /ɛ-ɛ/ and 97% for /e-e/. That is, participants are able to correctly reject any difference between the words in the stimuli for /i-i/, /e-e/ and /ɛ-ɛ/ pairs. For the pair /i-e/, the percentage of Hit response is 92% and that of Miss response is 8%. That is, the participants were

able to correctly discriminate between the two sounds. Similarly, the percentage of Hit response for /e-ɛ/ pair is 68% and Miss response for the same pair is 32%. That means, the participants did have some confusion when discriminating between the two words in the stimuli.

Figure 4.7: Discrimination of front vowels by participants from Jorhat



In case of the discrimination of back vowels we see high percentage of Correct rejection for the A-A stimuli. As seen in figure 4.8 the percentages for /u-u/ and /ʊ-ʊ/ is 96%, /o-o/ is 94% and /ɔ-ɔ/ is 97%. This indicates that the participants were able to reject distinction between the two sounds in the A-A stimuli. In case of the A-B stimuli, /o-ɔ/ pair has a Hit response of 96% and a Miss response of 4%, the /ʊ-o/ pair has a Hit response of 84% and a Miss response of 16% indicating that participants were able to distinguish between the two sounds in the pairs. Finally, for the /u-ʊ/ pair we see that the Hit response is 39% whereas the Miss response is 61%. This indicates that participants were not able to distinguish between the two sounds and hit the "Same" button 61% of the times.

Figure 4.8: Discrimination of back vowels by participants from Jorhat

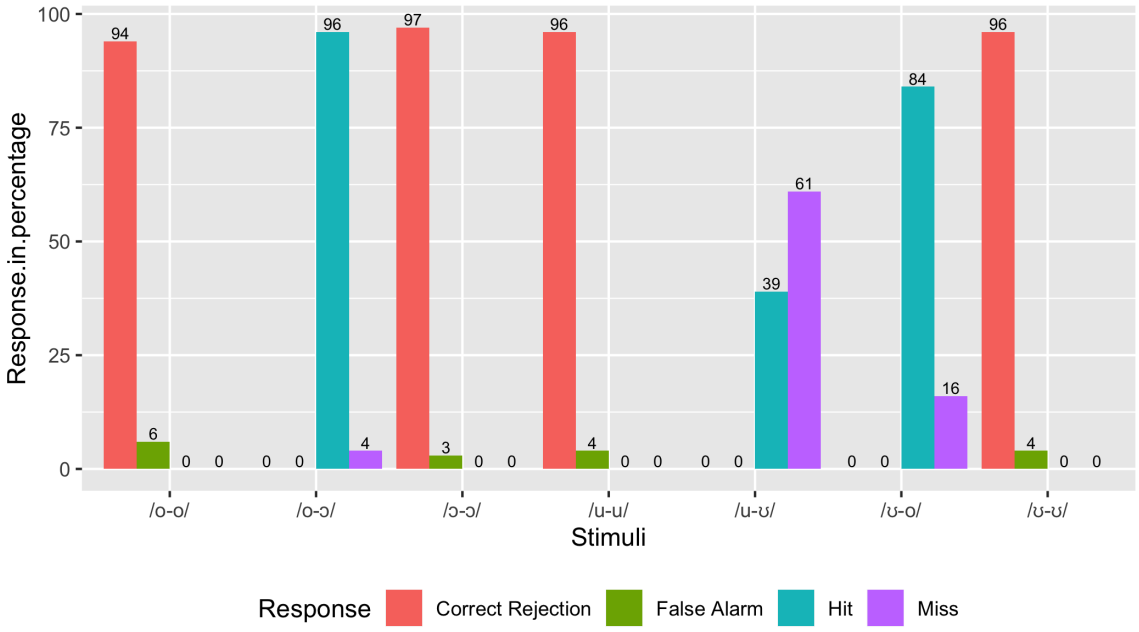


Figure 4.9 shows the identification of front vowels by participants from Jorhat. As seen in the figure, when the /i/ stimuli was presented to the participants, they identified the word as /i/ 100% of the time. In case of the vowel /e/ and /ɛ/ percentages of correct identification are 79% and 86% respectively.

Figure 4.9: Identification of front vowels by participants from Jorhat

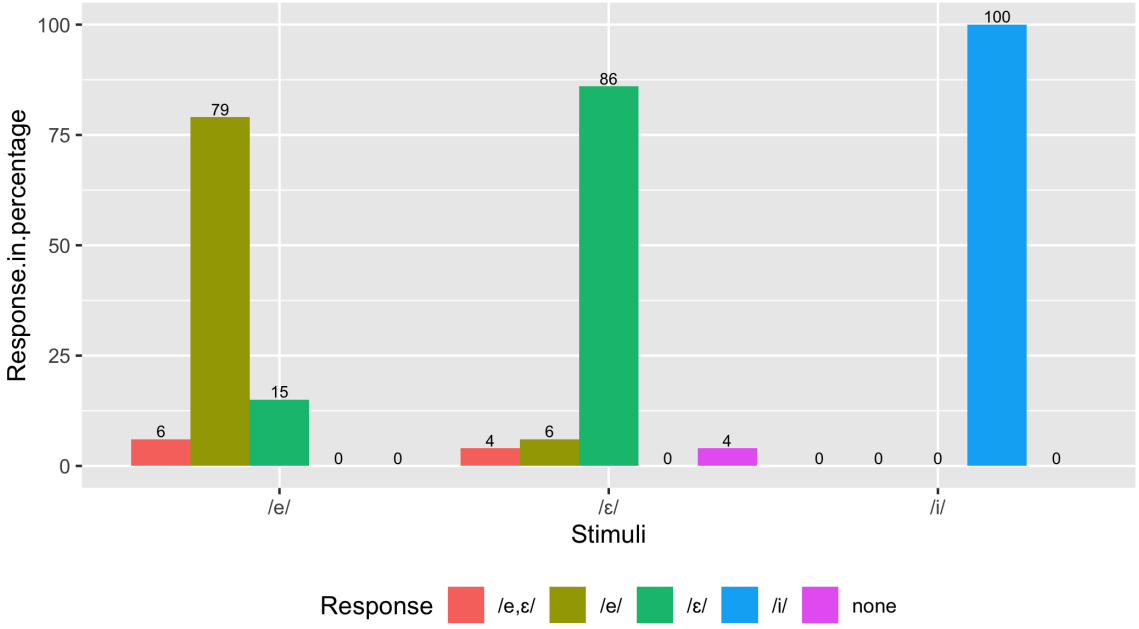
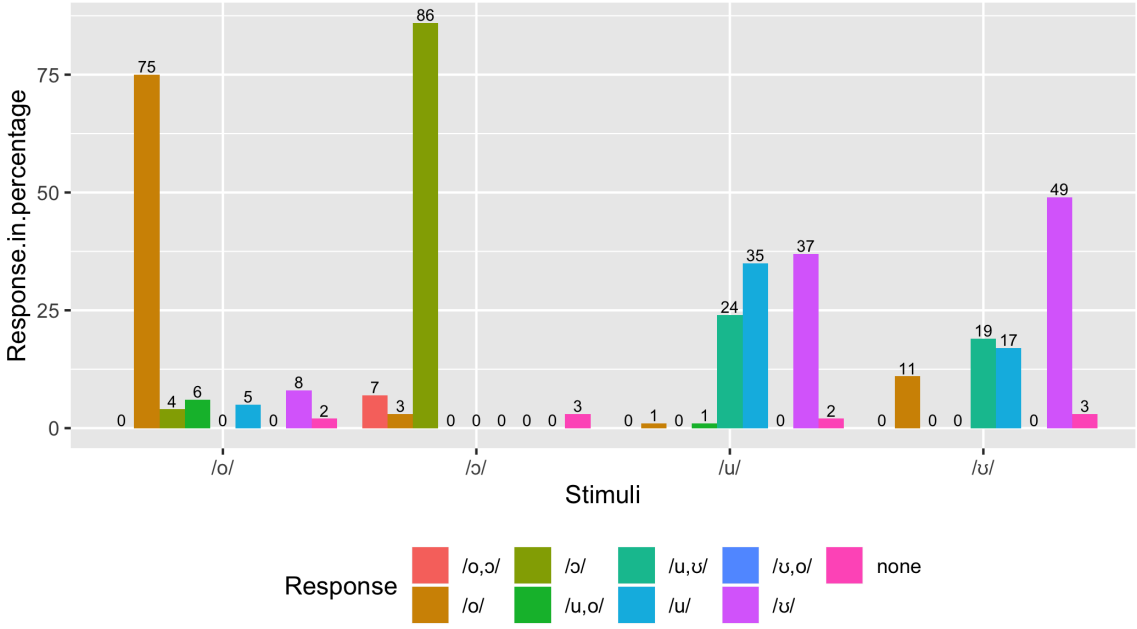


Figure 4.10 shows the identification of back vowels by participants from Jorhat. We see two vowels with a high percentage of correct identification. These are the vowels /o/ and /ɔ/ a response of 75% and 86% respectively. This indicates that the participants from Jorhat did not confuse the two stimuli with other vowels and correctly identified them for most of the time. Vowels /u/ and /ʊ/ however, had mixed responses. For the stimuli for /u/, identification responses were 35% for /u/, 24% for /u,ʊ/ and 37% for /ʊ/. Similarly for the stimuli for /ʊ/ identification responses were 17% for /u/, 19% for /u,ʊ/, 11% for /o/ and 49% for /ʊ/. The varied responses for these two vowels indicate that the participants were confused while correctly assigning the responses to the stimuli.

Figure 4.10: Identification of back vowels by participants from Jorhat



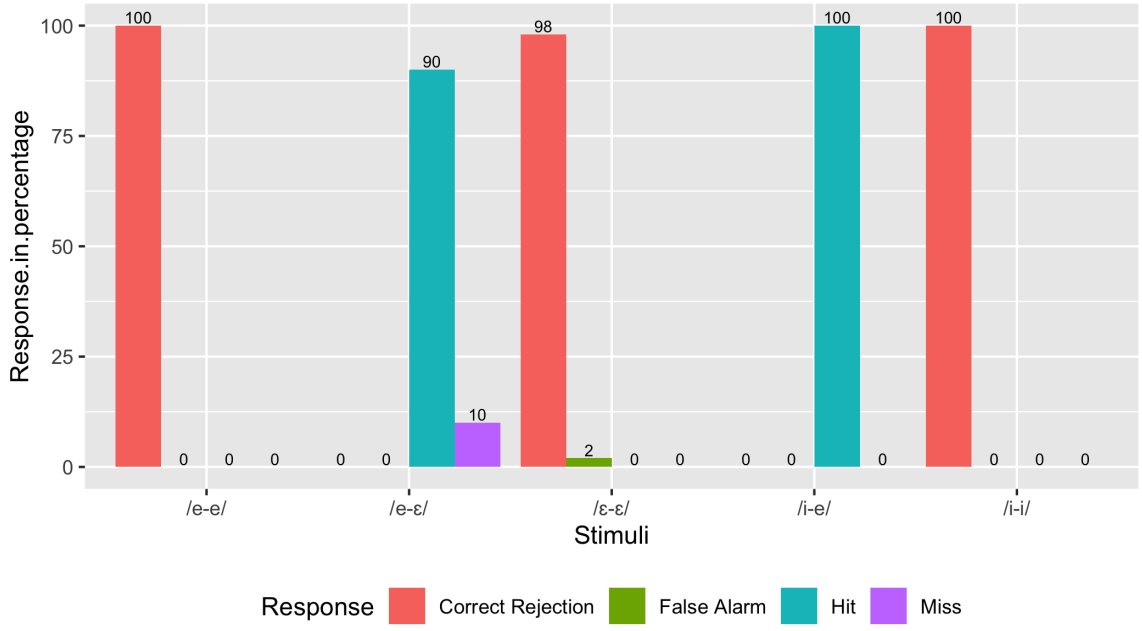
As seen in Figure 2.11 in Chapter 3, Jorhat speakers show an overlap of the vowels /u/ and /ʊ/ in production. Table 3.8 showed that the results of Euclidean distance, Pillai score and SOAM confirm these observations and evidenced merger by approximation whereby the vowel /ʊ/ moves to the phonetic space of the vowel /u/ and is completely absorbed by the latter. Results of the discrimination tests for the /u-ʊ/ pair shows that participants from Jorhat were less sensitive to the between the two vowels. Again, identification test results show that in case of boh vowels /u/ and /ʊ/, Tinsukia participants could not categorize the vowels into a single category. Their responses were distributed across only /u/, only /ʊ/, and both /u/ and /ʊ/. This indicates that participants categorized the two vowels as one and had difficulty distinguishing them as two separate vowels. Again from Table 3.8, the low Euclidean distance of 12 Mel between /u/ and /ʊ/ indicates a merger by approximation, as also seen in case of Tinsukia. We also see some confusion among the participants in the discrimination of /e-ɛ/ pair. However, our production results and merger measures,

do not indicate robust overlap of the two vowels. Additionally, participants did not show any confusion when it came to identifying the front vowels. Hence, this could be an isolated case of missed judgement on the part of the Jorhat participants. d′ scores calculated for the A-B vowel pairs reveal that Jorhat participants have the highest sensitivity in discriminating /i-e/ vowel pair (d′=3.6) and lowest in discriminating /u-ʊ/ vowel pair (d′=1.4). The results confirm that /u-ʊ/ vowel pair in Jorhat merge both in production and in perception.

4.3.3 Perception results for Nagaon

Figure 4.11 shows the discrimination of front vowels by participants from Nagaon. as seen in the figure, in case of A-A stimuli correct rejection is 100% for /i-i/ and /e-e/ pairs and 98% for /ɛ-ɛ/ pair. For the A-B pairs, the Hit percentage for /i-e/ is 100% and for the /e-ɛ/ stimuli it is 90%. This indicates that the participants from Nagaon were able to correctly perceive the differences between the front vowels.

Figure 4.11: Discrimination of front vowels by participants from Nagaon



In case of the discrimination of back vowels we again see high percentage of correct rejection for the A-A stimuli, as seen in figure 4.12. The percentage for /u-u/, and /ɔ-ɔ/ pairs is 100%. Both /ʊ-ʊ/ and /o-o/ pairs have a Correct rejection response of 96%. This indicates that the participants were able to reject distinction between the two sounds in the stimuli. In case of the A-B stimuli we see that participants distinguished between the two sounds in the /o-ɔ/ stimuli with a Hit response of 99%. The Hit response for the /ʊ-o/ pair is however 81% which means that participants regarded the stimuli as "same" a few times (19%). /u-ʊ/ has a Hit response of 55% and a Miss response of 45% which indicates that about half of the time participants did not distinguish between the two words presented to them.

Figure 4.12: Discrimination of back vowels by participants from Nagaon

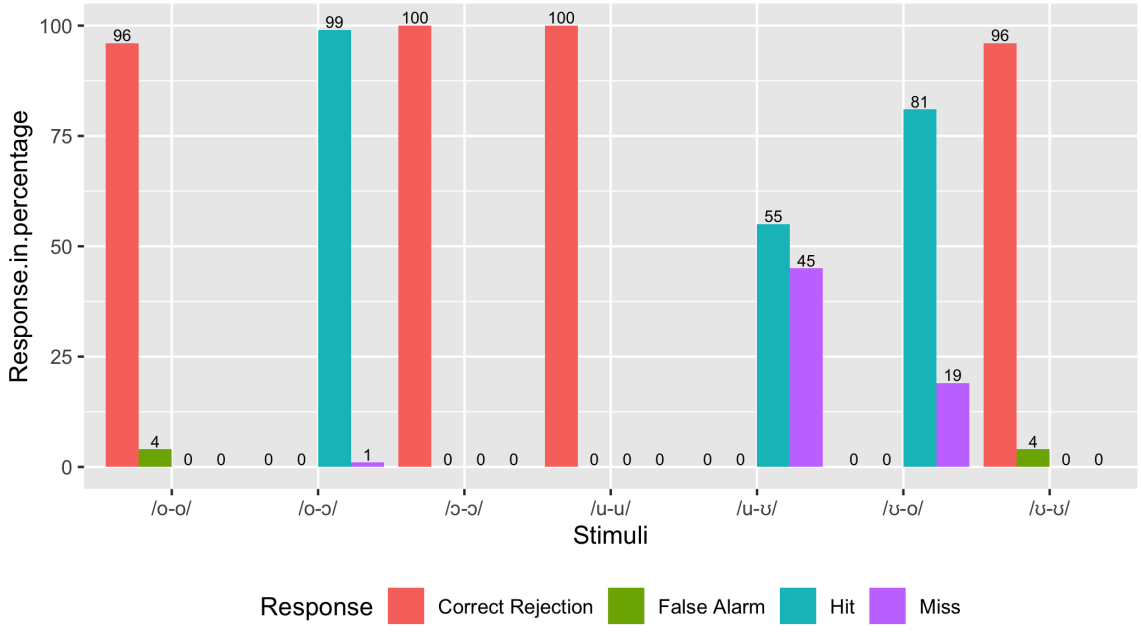


Figure 4.13 shows the identification of front vowels by participants from Nagaon. Participants' response shows 100% correct identification for the /i-i/ vowel pair. In case of the vowel /e/ and /ɛ/ percentages of correct identification are 72% and 78% respectively.

Figure 4.13: Identification of front vowels by participants from Nagaon

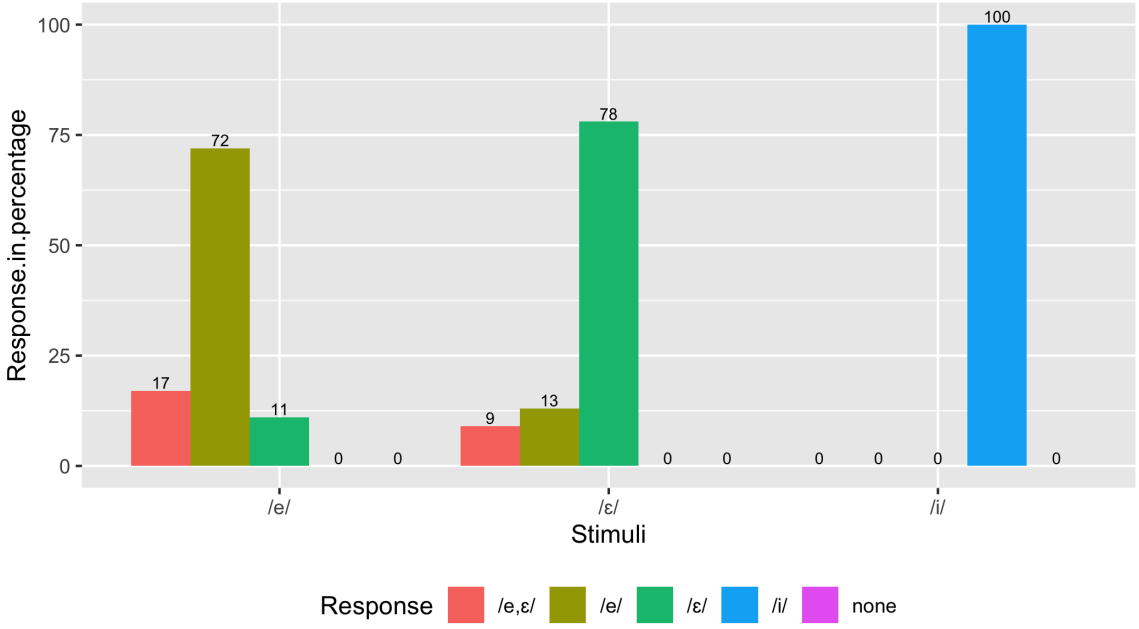
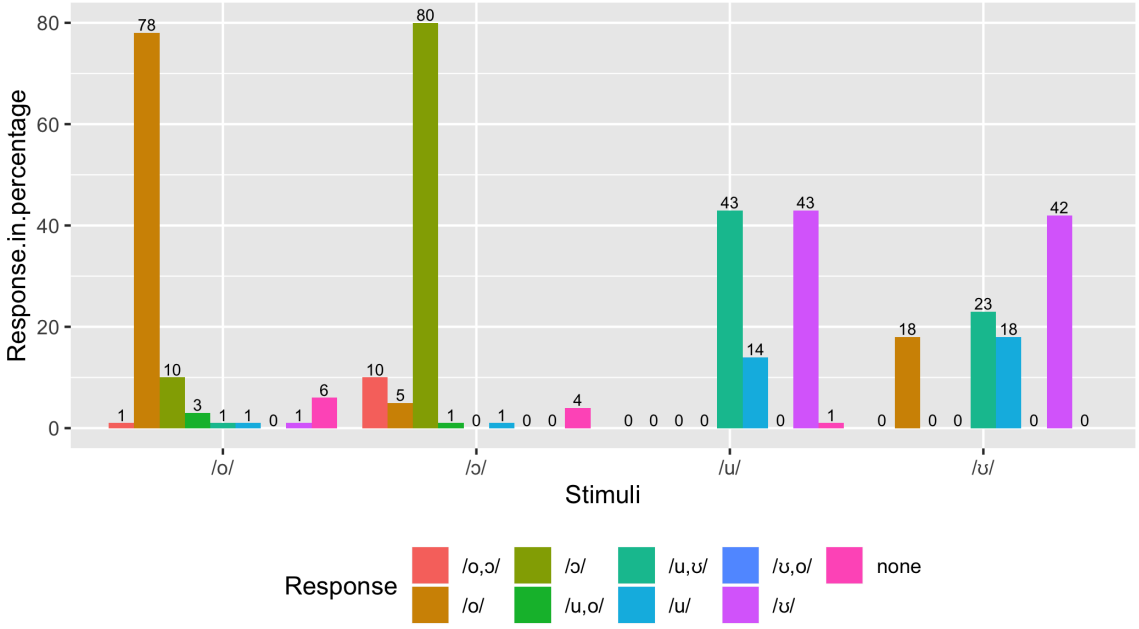


Figure 4.14 shows the identification of back vowels by participants from Nagaon. We see two vowels with a high percentage of correct identification. These are the vowels /o/ and /ɔ/ with a response of 78% and 80% respectively. The high percentage indicates that the participants from Nagaon did not confuse the stimuli with other vowels and correctly identified them for most of the time. Stimuli for vowel /u/ showed identification response of only 14% for /u/. Both /u,ʊ/ and /ʊ/ received 43% of responses each. Similarly for the stimuli for /ʊ/, identification responses were 18% for /u/ and /o/, 23% for /u,ʊ/ and 42% for /ʊ/. The varied responses for these two vowels indicate that the participants were confused while correctly assigning the responses to the stimuli.

Figure 4.14: Identification of back vowels by participants from Nagaon



Perception tests for Nagaon shows that like Tinsukia and Jorhat varieties, participants could not distinguish between the vowels /u/ and /ʊ/. It was seen Figure 2.16 that there is considerable overlap of ellipses between the vowels /u/ and /ʊ/. In Table 3.9 the results of Euclidean distance, Pillai score and SOAM evidenced complete merger of the two vowels. Results of the discrimination test for the /u-ʊ/ pair show that participants from Nagaon were less sensitive to the distinction between the two vowels. Again, identification test results show that in case of the vowel /u/, there were higher responses /ʊ/ as well as when participants opted for both /u/ and /ʊ/ as the correct answer (43%). Similar confusion was observed in the identification of /ʊ/ stimuli where 23% of the time the participants responded that the sound they heard could be both /u/ and /ʊ. In case of production, vowel /ʊ/ shows similar merger with vowel /o/, with considerable overlap of ellipses with supporting results of merger measures as seen in Chapter 3. Perception results, however did not fully corroborate the production results. No major confusion was observed in the discrim-

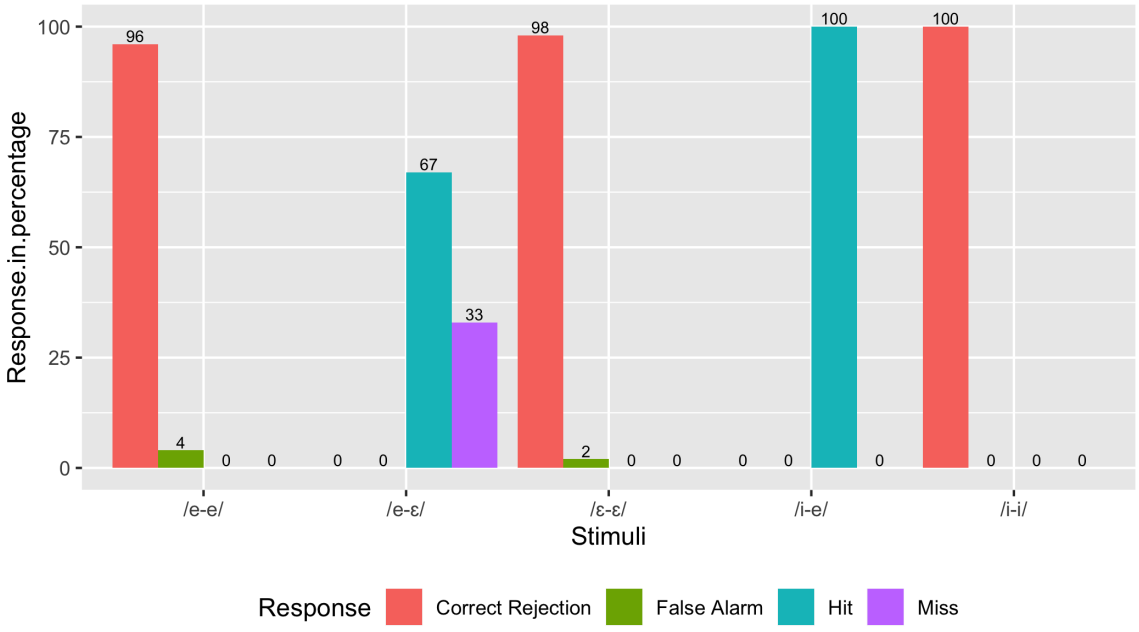
ination of the /ʊ-o/ pair. While 18% of identification responses were for /o/, this can not be considered to be indicative of a merger. Impressionistic observation of production results indicated overlap of three vowels in case of Nagaon: /u/, /ʊ/ and /o/. However, we ruled out a merger of three vowels in Chapter 3. As in the case of Tinsukia the /ʊ-o/ overlap in production in Nagaon may be due to a probable movement of /o/ to a lower F1 to compensate for the space left behind by the merger of /ʊ/ with /u/. The same can be said in case of perception. The /ʊ-o/ overlap can be accounted to the perceptual space left behind by the merger of the /ʊ-u/ vowels in perception. The confusion could arise from the higher F1 of a canonical /ʊ/ which the participants have categorized as an /o/ rather than as an /u/. It is important to note here that Nagaon belongs to Middle Assam which is geographically between Eastern and Western Assam and has influences from both varieties. This could be another reason for the overlap of three vowels. In our production analysis we have seen a two way distribution of the vowel /ʊ/. While in Eastern Assam the vowel moves up and merges with /u/, the same vowel in Western Assam moves down and merges with /o/. Now Nagaon being in between these two regions, we might be observing an intermediate outcome of these two opposite mergers and hence /ʊ/ shows overlap with both /u/ and /o/ in production. d′ scores calculated for the A-B vowel pairs reveal that Nagaon participants have the highest sensitivity in discriminating /i-e/ vowel pair (d′=4.7) and lowest in discriminating /u-ʊ/ vowel pair (d′=2.2). The results confirm that /u-ʊ/ vowel pair in Nagaon merge both in production and in perception.

4.3.4 Perception results for Nalbari

Figure 4.15 shows the discrimination of front vowels by participants from Nalbari. All the A-A stimuli show a high percentage of Correct rejection response. While for /i-i/ the participants are able to correctly reject any difference in the stimuli 100% of

the time, in case of /e-e/ and /ɛ-ɛ/ the response was 98% and 96% respectively. In case of A-B stimuli, for the /i-e/ pair, the percentage of Hit response is 100%. That is, the participants were able to correctly discriminate between the two words all of the time. However, /e-ɛ/ has a Hit response of only 67%.

Figure 4.15: Discrimination of front vowels by participants from Nalbari



In case of the discrimination of back vowels as seen in figure 4.16, we see 100% Correct rejection response for /u-u/ pair. For other A-A stimuli the Correct rejection responses were as follows: /ʊ-ʊ/ 94%, /o-o/ 98% and /ɔ-ɔ/ 98%. In case of the A-B stimuli the /o-ɔ/ pair has a Hit response of 94%. The /ʊ-o/ pair has a Hit response of 63% and a Miss response of 37%. The low percentage of Hit response for /ʊ-o/ stimuli indicates the participants had confusion in distinguishing between the two sounds. Again we see a similar response of Hit and Miss (61% and 39%) for /u-ʊ stimuli indicating confusion among the participants regarding the two vowels.

Figure 4.16: Discrimination of back vowels by participants from Nalbari

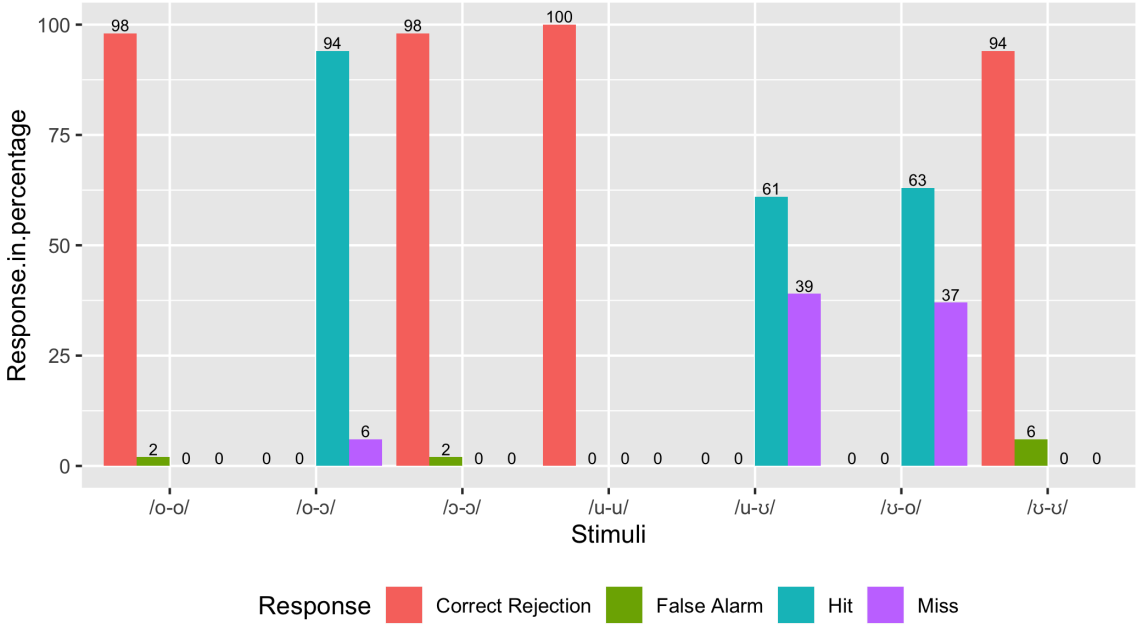
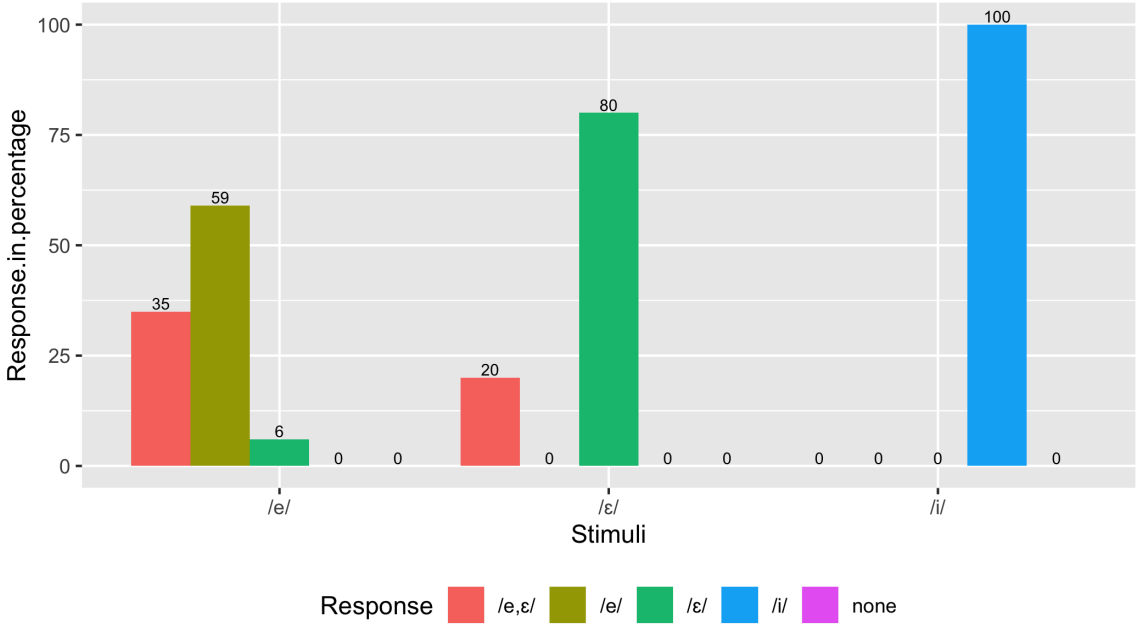


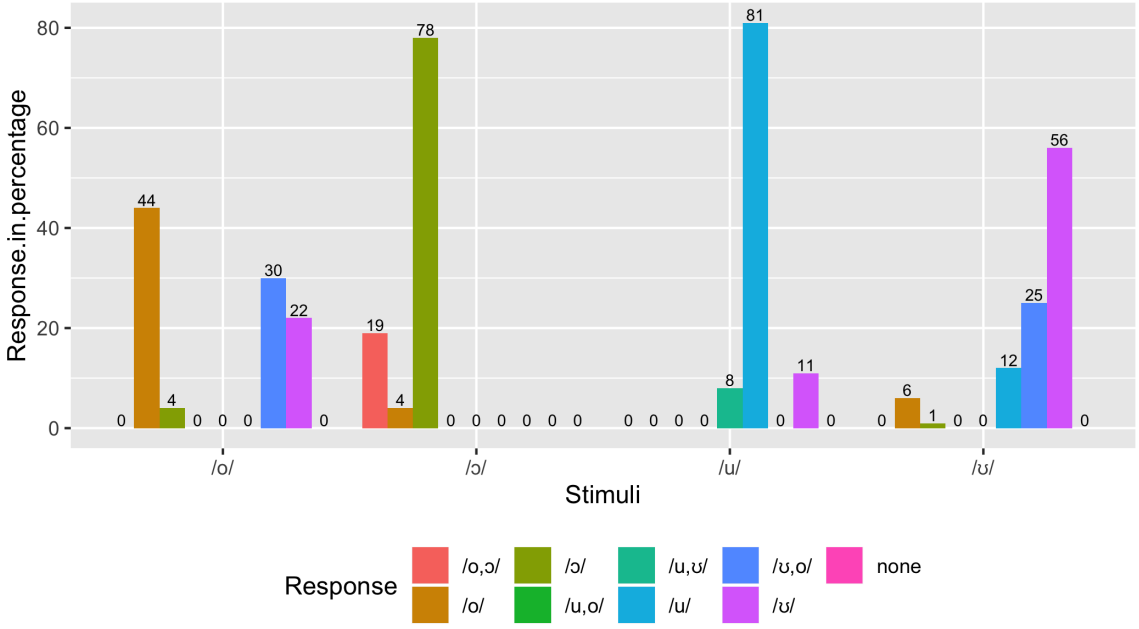
Figure 4.17 shows the identification of front vowels by participants from Nalbari. As seen in the figure, when the /i/ stimuli was presented to the participants, they identified the word as /i/ 100% of the time. In case of the vowel /e/, 59% of the responses were for /e/, 35% for /e, ɛ/ and 6% for /ɛ/. In case of the vowel /ɛ/ again, 20% of the times the vowel was identified as /e, ɛ/ and 80% as /ɛ/.

Figure 4.17: Identification of front vowels by participants from Nalbari



For back vowels, as seen in Figure 4.18, the highest identification for the stimuli /u/ was 81% for the response /u/. A lower response was recorded for /u, ʊ/ at 8%, and /ʊ/ at 11%. The stimuli /ʊ/ had an identification response for /ʊ/ at 56%, /ʊ,o/ at 25% and 12% for the option /u/. The stimuli for vowel /o/ had a response of 44% for /o/, 30% for /ʊ-o/ and only 22% for /ʊ. Again the stimuli for the vowel /ɔ/ had a response of 78% for /ɔ/ and 19% for /o-ɔ/.

Figure 4.18: Identification of back vowels by participants from Nalbari



As seen in Figure 2.21 in Chapter 3 Nalbari speakers show an overlap of the vowels /ʊ/ and /o/ in production. Table 3.10 showed that the results of Euclidean distance, Pillai score and SOAM confirm these observations and evidenced merger by expansion where the categorical boundaries between the two vowels have disappeared, constituting a larger categorical space. Results of the discrimination test for the /ʊ-o/ pair although did not show robust discrimination of the two vowels, the percentage Hit responses was above 60% indicating that the participants were able to distinguish between the two sounds most of the time. Identification test results show that in case of the vowel /ʊ/, 25% responses were for /ʊ,o/ and 12% for /u/. Similar results were seen in case of identification of /o/ vowel where 30% of the responses were for /ʊ,o/. This indicates that participants had some confusion while assigning the correct category to the two vowels. Again from the Figure 2.21, we observe the overlapping ellipses occupying a larger categorical space even though a low Euclidean distance of 19 Mel is observed between /ʊ/ and /o/, indicating a merger by expan-

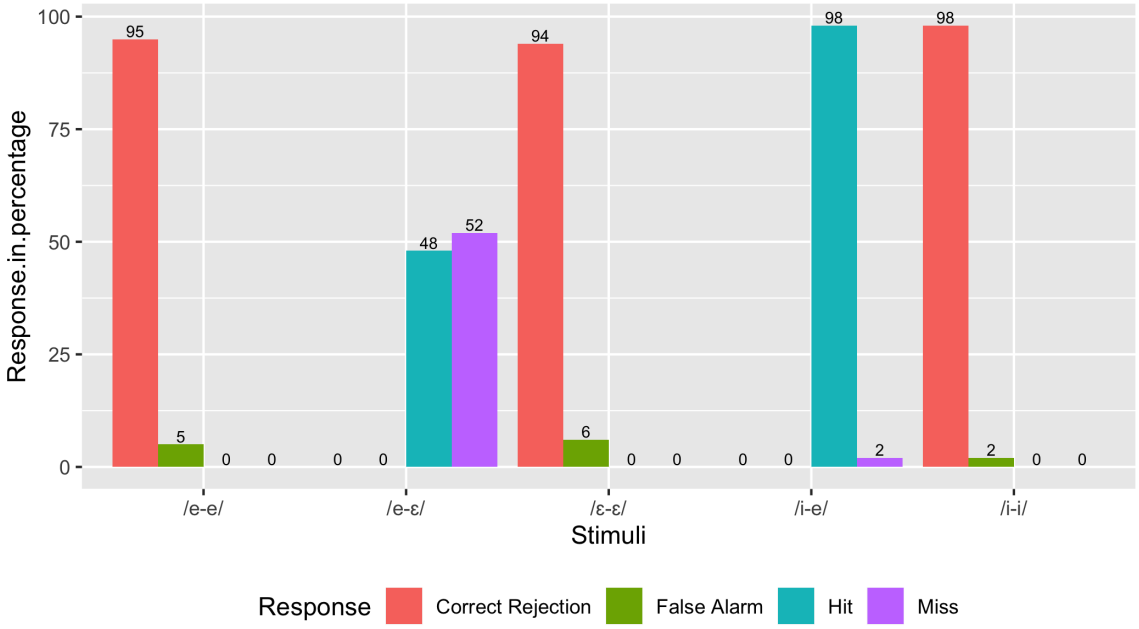
sion of the two vowels. An important observation to note is that participants have a high response of identification of the /u/ vowel suggesting that they considered the vowel as completely distinct. Secondly, we also see slight confusion among the participants in the discrimination of the /e-ɛ/ pair. Our production results and merger measures, indicate robust overlap of the /e/ and /ɛ/ vowels. Perception results show 67% Hit responses while discriminating the two vowels and high identification scores particularly for the vowel /ɛ/ (80%). The movement of the vowel /ɛ/ towards /e/ and a complete overlap of the two vowels as seen in Figure 2.21, on the other hand, evidences a merger by approximation in case of /e/ and /ɛ/ Thus we see a slight mismatch between production and perception, which according to Hay et al. (2010); W. Maguire et al. (2013) is expected in cases of near-mergers. d′ scores calculated for the A-B vowel pairs reveal that Nalbari participants have the highest sensitivity in discriminating /i-e/ vowel pair (d′=4.4) and lowest in discriminating /ʊ-o/ vowel pair (d′=2.1). However, low d′ is also seen in case if /u-ʊ/ (d′=2.2) and /e-ɛ/ (d′=2.3). Hence, we see two kinds of mergers in case of Nalbari: merger by expansion for /ʊ/ and /o/ and merger by approximation for /e/ and /ɛ/. The low d′ score for /u-ʊ/ pair also indicates that participants' low sensitivity in discriminating between the two sounds hinting at a mis-match between production and perception. Production results showed no evidence of merger between the two vowels. However, we have seen that /ʊ/ merges with /o/. As a result, native speakers of Nalbari variety have an /ʊ/ phoneme which is higher in F1 as opposed to the canonical /ʊ/ phoneme with a lower F1, thus preventing them from correctly discriminating between the two lower phonemes. Hence the production-perception anomaly for the /u-ʊ/ contrast.

4.3.5 Perception results for Barpeta

Figure 4.19 shows the discrimination of front vowels by participants from Barpeta. As seen in the figure, Correct rejection response is high for the A-A stimuli. For /i-i/

the response is 98%, for both /e-e/ it is 95% and for /ɛ-ɛ/ the response is 94%. That is, the participants are able to correctly reject any difference in the stimuli. For the pair /i-e/, the percentage of Hit response is 98%. However for the /e-ɛ/ pair the Hit response is 48% and Miss response is 52% indicating that participants had confusion while distinguishing the two sounds.

Figure 4.19: Discrimination of front vowels by participants from Barpeta



In case of the discrimination of back vowels we see high percentage of Correct rejection for the A-A stimuli. As seen in figure 4.20, /u-u/ pair has a Correct rejection response of 100%. The /ʊ-ʊ/ has 97% and both /o-o/ and /ɔ-ɔ/ pairs have a Correct rejection response of 94%. This indicates that the participants were able to reject distinction between the two words in the stimuli. In case of the A-B stimuli however, the /o-ɔ/ pair has a Hit response of 92% and a Miss response of 8% indicating that participants were able to distinguish between the two sounds. The /ʊ-o/ pair has a Hit response of 70% and a Miss response of 30%. This indicates that participants had slight difficulty distinguishing the two words. Finally, for the /u-ʊ/ pair we see

that the Hit response is 58% whereas the Miss response is 42%. This indicates that participants were not able to distinguish between the two words most of the times.

Figure 4.20: Discrimination of back vowels by participants from Barpeta

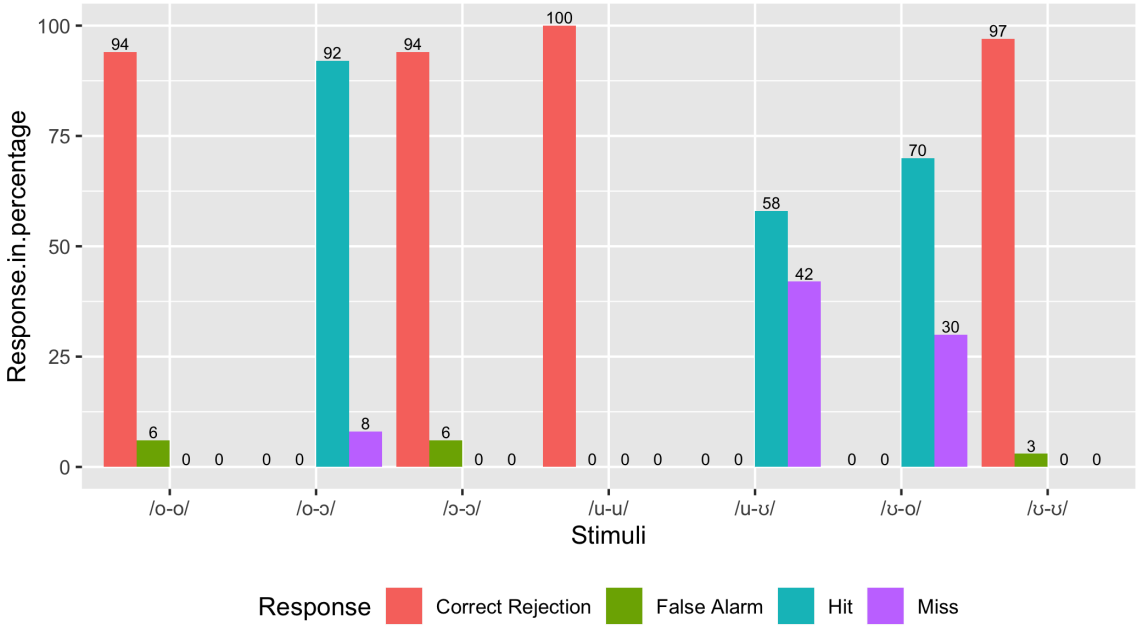
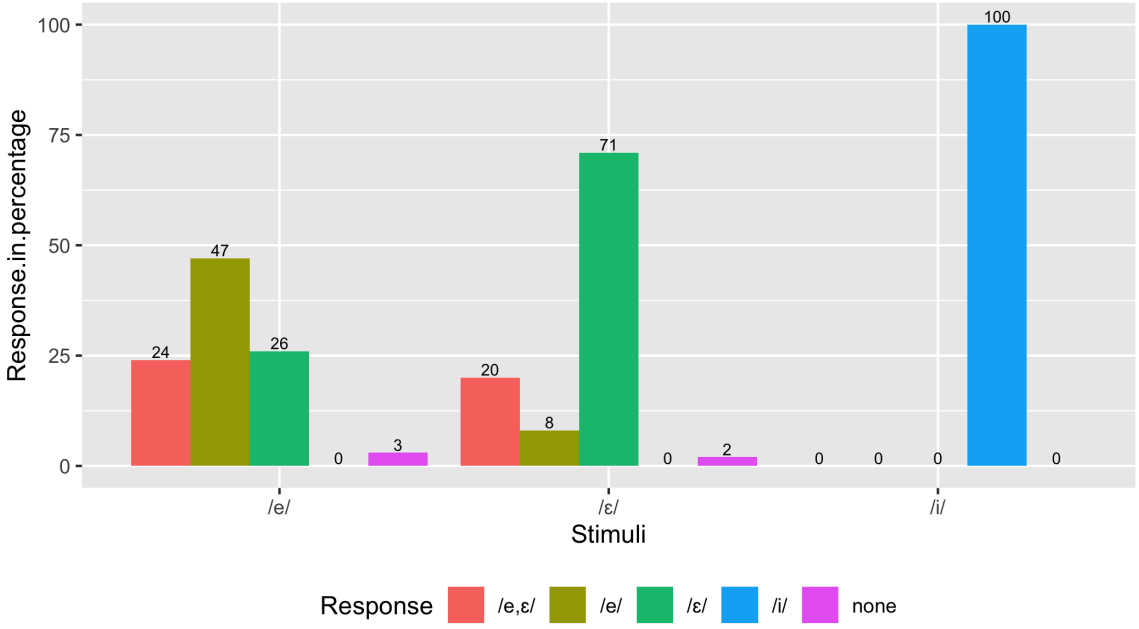


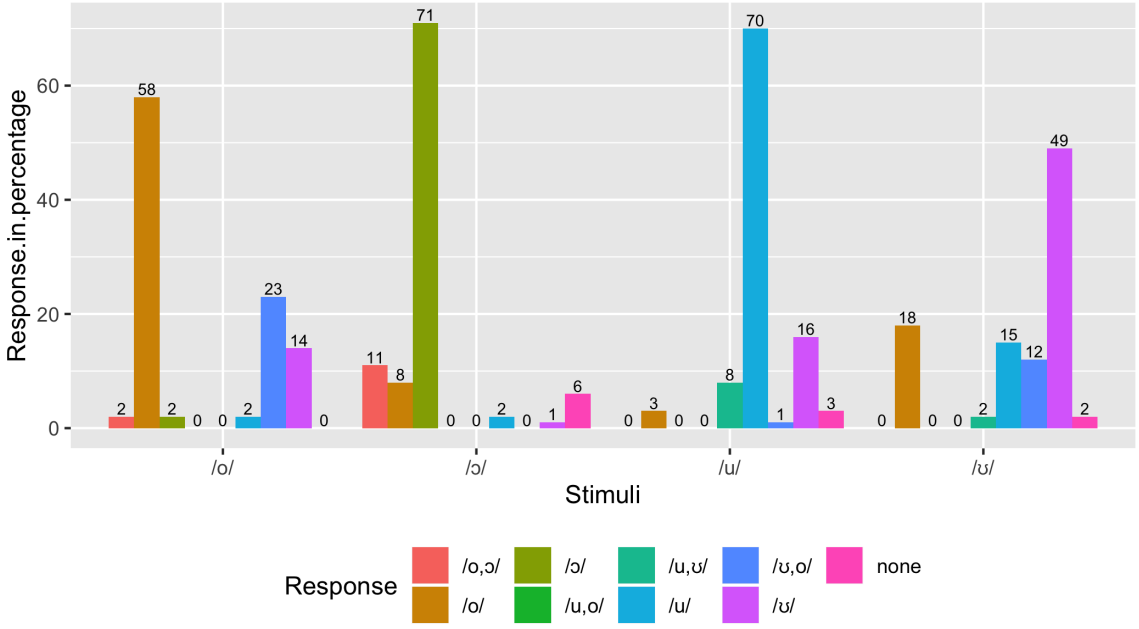
Figure 4.21 shows the identification of front vowels by participants from Barpeta. As seen in the figure, when the /i/ stimuli was presented to the participants, they identified the vowel as /i/ 100% of the time. In case of the vowel /e/, 47% of the responses were for /e/, 24% for /e,ɛ/ and 26% for /ɛ/. In case of the vowel /ɛ/, participants identified the vowel as /ɛ/ 71% of the time. 20% was for /e, ɛ/ and 8% was for /e/.

Figure 4.21: Identification of front vowels by participants from Barpeta



For back vowels, as seen in Fig 4.22, the highest identification for the stimuli /u/ was 70% for the response /u/. A lower response was recorded for /u, ʊ/ at 8% and/ʊ/ at 16%. The stimuli /ɔ/ had the highest identification for the response /ɔ/ at 71%, /o,ɔ/ at 11% and a minimal 8% for the response /o/. This indicates that the participants from Barpeta did not confuse the /u/ and /ɔ/ stimuli with other vowels and much correctly identified them for most of the time. The vowels /ʊ/ and /o/ however had a more varied response. While 49% of the times the stimuli /ʊ/ was correctly identified as /ʊ/, 18% was for /o/, 15% for /u/ and 12% for /u,ʊ/. Similarly, stimuli /o/ had an identification response of 58% for /o/, 23% for /ʊ,o/, 14% and for /ʊ/ along with minimal percentages of other responses. The varied response of the /ʊ/ and /o/ vowels indicate that the participants from Barpeta had some confusion identifying those vowels.

Figure 4.22: Identification of back vowels by participants from Barpeta



In Chapter 3 we have seen that Barpeta speakers show a merger of two pairs of vowels: the front vowels /e/ and /ɛ/ and the back vowels /ʊ/ and /o/. Table 3.11 showed that the results of Euclidean distance, Pillai score and SOAM confirm these observations and evidenced merger by expansion in case of /ʊ/-/o/ where the categorical boundaries between the two vowels have disappeared constituting a larger categorical space. Results of the discrimination test for the /ʊ-o/ pair with a Hit response of 70% and a Miss response of 30%, show that participants had some difficulty in distinguishing the two vowels. Again, in Figure 2.26, we observe the overlapping ellipses occupying a larger categorical space, with an Euclidean distance of 37 Mel between /ʊ/ and /o/. This indicates a merger by expansion of the two vowels. As in the case of Nalbari, participants have a high response of identification of the /u/ vowel suggesting that the vowel is considered to be completely distinct in the variety. Participants however showed more confusion while distinguishing the /u-ʊ/ vowel pair, with a Hit response of 58%. We also see similar confusion among the participants

in the discrimination of the /e-ɛ/ pair with 52% Miss response. Identification test results also indicate that participants had difficulty identifying the two vowels. Our production results and merger measures, indicate robust overlap of the /e/ and /ɛ/ vowels. However, unlike Nalbari, which has a smaller Euclidean distance of 36 mels between the two vowels indicating merger by approximation, we see a comparatively larger distance of 70 mels, although we see a Pillai score of 15 and SOAM score of 67.3% indicating that the merger of these two vowels could be in process currently. d′ scores calculated for the A-B vowel pairs reveal that Barpeta participants have the highest sensitivity in discriminating /i-e/ vowel pair (d′=4.0) and lowest in discriminating /e-ɛ/ vowel pair (d′=1.6). However, low d′ is also seen in case if /ʊ-o/ (d′=2.2) and /u-ʊ/ (d′=2.4). The results confirm that vowels in the /ʊ-o/ and /e-ɛ/ pairs in Barpeta merge both in production and in perception. Again, the low d′ score for /u-ʊ/ pair also indicates at a mis-match between production and perception. Like in the case of Nalbari, this anomaly is due to the higher F1 of the /ʊ/ vowel which is result of its merging of with the vowel /o/.

4.4 Conclusions

This chapter looked into the perception of vowels and vowel mergers in Assamese by native speakers of the language. Production experiments and subsequent overlap measures conducted on the Assamese vowels spoken in the five regions confirmed two way mergers in Assamese varieties. Therefore, a perception experiment was done to look into the native speakers' perception of these mergers. Two types of perception tests were conducted: (a) the Discrimination test, where participants were asked to determine whether they hear a difference between two sounds, and (b) the Identification test, where participants were asked to identify the sound they hear.

All regions showed high sensitivity in the discrimination of /i-e/ vowels and high

rates of correct identification of the /i/ vowel indicating that the two vowels are distinct in perception. Lowest sensitivity in discriminating /e-ɛ/ vowels was shown by Nalbari and Barpeta, although Nalbari had a slightly higher sensitivity. Identification results show that Nalbari and Barpeta had the lowest correct identification rates for /e/ among all the regions. However, identification rates for /ɛ/ were higher in these two varieties. We observed in Chapter 3 that although vowels /e/ and /ɛ/ merge in Nalbari and Barpeta varieties, vowel /ɛ/ showed incomplete movement towards the phonetic space of /e/ evidencing a near-merger. This seems to reflect in the perception results as well. Interestingly, Tinsukia with a percentage of 58%, showed the lowest identification for /ɛ/ among all the five regions. Although production results showed slight overlap of the vowel with /e/, merger measures confirmed no movement of the /ɛ/ vowel in Tinsukia. It would be interesting to examine if this low percentage translates to a probable perceptual merging in a few years.

For the /u-ʊ/ vowel contrast, lowest sensitivity (d′ scores) was shown by Jorhat participants indicating that the two vowels were perceptually merged in the variety. Tinsukia, Nagaon and Nalbari had the same sensitivity in discriminating the vowels and Barpeta had slightly higher sensitivity. Tinsukia, Jorhat and Nagaon showed low scores in correctly identifying both /u/ and /ʊ/. Results from the production experiments showed the merger of /u-ʊ/ vowels in the three varieties. Perception test results confirm that the two vowels are merged both in production and perception.

For the /ʊ-o/ vowel contrast, lowest sensitivity (d′ scores) was shown by Nalbari and Barpeta participants. Tinsukia, Jorhat and Nagaon were only slightly higher. Identification test results show considerable confusion between the two vowels among the Nalbari and Barpeta participants. Production experiments showed the merger of /ʊ-o/ vowels in the two regions and this is reflected in the perception tests as well.

All regions also showed comparatively higher sensitivity in the discrimination of /o-ɔ/ vowels. Correct identification rates of the /ɔ/ vowel were highest for Tinsukia

and Jorhat indicating distinctiveness in the two varieties. The lowest rate for the same was seen for Barpeta with participants showing confusion in a few cases and opting for both /o/ and /ɔ/.

The results of the perception experiments thus reveal two kinds of perception interpretation with respect to mergers in Assamese varieties. Firstly, participants who did not categorize between two vowels phonemically, also did not categorize the vowels as different in perception. Secondly, participants who were aware of the phonetic distinction between two vowels that merged phonemically in production categorized one of the vowels with the nearest phonetically similar sounding vowel. The results of the perception experiments confirm the merger of /u-ʊ/ vowels in Tinsukia, Jorhat and Nagaon varieties both perceptually and production wise. The same is confirmed for /ʊ-o/ vowels in Nalbari and Barpeta. As has been reported in Hay et al. (2010) and W. Maguire et al. (2013), there is often a production-perception anomaly in situations of mergers, particularly in cases of incomplete or near-mergers. Our results also provide evidence of a production-perception dichotomy existing in Assamese varieties. This anomaly is evident in both Nalbari and Barpeta where the vowels /e-ɛ/ have shown evidences of near-mergers by approximation. Additionally, a production-perception anomaly has also been seen in case of the vowel /ʊ/. This is due to the higher F1 value of the vowel as a result of its merger with the /o/ vowel in the two varieties.

Production-perception anomaly in situations of mergers have been attested in several studies. Labov et al. (1991) identify a number of near-mergers in which, despite there being a consistent difference in production, speakers report no difference in perception. Contrasting these findings is Hay et al. (2006)'s study on merger-in-progress in New Zealand English, which report a high accuracy of individuals in identifying sounds undergoing merger, even though these same individuals merge the sounds in production. Hay et al. (2010) outline a range of tests in which individuals who merge

two sounds to some degree were exposed to dialects in which the distinction between the sounds is maintained. Results showed that a wide range of factors influence accuracy in the perception task, including characteristics specific to participants, word, context and the perceived speaker. The limitations of the discrimination and identification tests may not be able to uncover deeper linguistic decisions that speakers make in responding to the stimuli. For example, the perception tests conducted do not tell us if the participants of the perception test considered the stimuli provided to them to be from their own variety or from another variety. For varieties that are not similar to standard Assamese, the responses may be different depending on the participants' perception of the variety of the speaker who produced the stimuli for the perception test.

Chapter 5

Rhythm and speech rate

5.1 Introduction

In the previous chapters on production of Assamese vowels, evidence of two kinds of mergers in Assamese varieties have been found. While Tinsukia, Jorhat and Nagaon varieties reduce the eight-vowel system of Assamese to a seven-vowel one by the merger of /ʊ/ and /u/ vowels, Nalbari and Barpeta reduce it to a six-vowel one by merging two vowel pairs: /ʊ/ with /o/ and /ɛ/ with /e/. The motivation for looking into the rhythm and speech rate in Assamese dialects stems from a number of factors. During the course of collecting the data for this study, most of the participants revealed that they are able to distinguish between an "Upper Assam" (Eastern Assam) speaker and a "Lower Assam" (Western Assam) speaker. Among several criteria that were mentioned, and that hinted at mostly lexical and morphological features, most participants also reported that they perceived the Western Assamese varieties to be faster than the Eastern Assamese varieties.

While it has been seen that cross linguistically certain languages are rightly perceived as faster than others (Pellegrino et al., 2011), on the other hand, in a given language one variety may be faster than the other (Jacewicz, Fox, O'Neill, & Salmons, 2009) in terms of syllables per second and similar measures. Apart from the rate of speaking, the perceived temporal differences in a language can also be prompted

by rhythmic differences (Nespor, 1990) popularly cited as the difference between a 'machine-gun firing type' and a 'morse-code type' language. A similar instance of rhythm variation was observed in the case of Arabic dialects where participants mentioned that North African Arabic sounded 'faster and jerky or more halting' than Eastern Arabic (Ghazali et al., 2002). Based on how speech units are organized in time, the world's languages are said to fall into three rhythm categories, namely, syllable-timed, stress-timed (Pike, 1945; Abercrombie, 1967) and mora-timed (Han, 1962; Port et al., 1980, 1987). Assamese has been reported to be a syllable timed language, based purely on auditory impression (Ray, 1966). However, another quantitative study with a single Assamese speaker claims that Assamese rhythm is mora-timed (Savithri, 2009).

A holistic typological study of language structures (Donegan & Stampe, 1983), suggests that there could be a possible link between rhythm and vowel quality. They note that stress-timed languages typically have larger vowel inventories while non-stress languages have simpler systems. Other studies, for example G. Schwartz (2010) have reported that phonological features associated with vowels contribute to the perception of a language's rhythm. Although an exhaustive analysis of rhythm and speech rate in Assamese varieties based on the phonological features associated with vowels would be beyond the purview of this thesis, an attempt has been made to explore the varieties of the language on individual temporal and rhythm properties. This chapter provides rhythm metrics for Assamese as it is spoken in the five different geographical regions of Assam. The correlation among rhythm metrics, rate of articulation and the number of syllables in a measured utterance are also examined. Finally, a quadratic discriminant analysis with the rhythm metrics and articulation rate is conducted to determine the accuracy in classifying Assamese spoken in five different geographic areas, comprised of three dialect regions. Considering the relationship proposed between rhythm types and vowel quality in languages (Donegan

& Stampe, 1983; Dellwo, 2006; G. Schwartz, 2010), in this chapter we would like to investigate if the classification of the five varieties of Assamese, based on rhythm properties also reflect the accuracy in classification of the varieties using vowel quality. If the classification accuracy using rhythm is comparable to the classification accuracy using formant features, that may point towards a direct relationship between vowel quality and rhythm characteristics.[1]

5.1.1 Rhythm metrics

As mentioned earlier, the world's languages fall into three rhythm categories, namely, syllable-timed, stress-timed (Pike, 1945; Abercrombie, 1967) and mora-timed (Han, 1962; Port et al., 1980, 1987). However, studies based on acoustic evidence have claimed that rhythm classes are not categorical but a continuum (Roach, 1982; Dauer, 1983; Hamdi et al., 2004). It was also observed that languages perceived as stress-timed have a greater variety of syllable types than syllable-timed languages and exhibit vowel reduction in unstressed syllables (Dauer, 1983). Hence, temporal measures like ΔC (standard deviation of consonantal intervals), ΔV (standard deviation of vocalic intervals) and %V (percentage of vocalic intervals in an utterance) are taken as metrics to account for rhythm in languages (Ramus et al., 1999).

Stress-timed languages can allow a greater variety of consonant clusters, therefore they were found to have a higher ΔC, whereas %V is lower due to vowel reduction. However, ΔV can be affected by a variety of language specific or contextual factors and therefore can not be interpreted clearly. Among these three absolute temporal measures, the combination of %V and ΔC was found to be the best way of distinguishing rhythm classes (Ramus et al., 1999). When %V and ΔC are plotted on an x-y plane, the stressed-timed and syllable-timed languages cluster into different groups.

[1]This chapter is a slightly modified version of the paper "Speech rate measures in five geographical varieties of Assamese" published in the Proceedings of Speech Prosody, 2020.

It was also seen that the speaking rate can affect measures like %V, ΔV and ΔC in varying degrees and thus, are not always effective in differentiating rhythm classes (Grabe & Low, 2002). Hence, rhythm measures based on the Pairwise Variability Index (PVI) was proposed to minimize the effect of the speaking rate (Grabe & Low, 2002). This approach categorizes languages based on durational variability of successive units of speech and can represent both rate normalized or raw values. Similarly, another measure, VarcoΔC, was proposed to minimize the effect of speech rate on ΔC, as it measures relative variation of consonantal intervals instead of absolute variation. The effect of speaking rate on the vocalic measure %V and intervocalic measure ΔC was analyzed by (Dellwo & Wagner, 2003). Though it was found that ΔC is highly affected by speech rate and %V remains rather stable, the results still corresponded to the claim that languages of the two different rhythm classes are distinguishable by %V and ΔC (Ramus et al., 1999).

Studies have shown that the successful identification and classification of languages and their varieties, based on rate of speaking and rhythm, is possible. For example, significant speaking rate differences have been observed between varieties of American English (Jacewicz, Fox, O'Neill, & Salmons, 2009; Jacewicz et al., 2010); New Zealand English and American English (Robb & Gillon, 2007) and Dutch spoken in the Netherlands and Belgium (Verhoeven et al., 2004). Also, as mentioned earlier, rhythm has been found to be different for different varieties of spoken Arabic (Hamdi et al., 2004; Biadsy & Hirschberg, 2009; Ghazali et al., 2002). Using these phonetic correlates, dialect identification was also attempted (Biadsy & Hirschberg, 2009). Even though rhythm has been found to be distinctive for many languages, there are claims that it is highly affected by the rate of speaking, depending on the way it is measured (Grabe & Low, 2002). On the other hand there are also claims that the rhythm metrics proposed so far in the literature cannot actually classify languages into separate rhythmic classes (Arvaniti, 2012). Nevertheless, it is important that the

measurements are also conducted on languages hitherto unanalyzed for rhythm.

Section 5.2 presents the methodology for the rhythm and speech rate analysis. In Section 5.3, the results obtained for rate of speaking and rhythm for Assamese spoken in five regions are discussed. Apart from that, a co-variation matrix obtained by comparing the rate of speaking and rhythm measures using Pearson's correlation method is presented. The section also reports the results of a quadratic discriminant analysis that was conducted to see the effectiveness of articulation rate and rhythm in discriminating Assamese spoken in five different geographic regions. Finally a series of statistical tests are conducted to estimate the effect of different geographical regions on the articulation rate and rhythm measures obtained.

5.2 Methodology

5.2.1 Materials and data collection

Native speakers of Assamese from the five geographical locations Tinsukia, Jorhat, Nagaon, Nalbari and Barpeta were recorded reading an Assamese translation of the story 'The North Wind and the Sun'. The passage consisted of 103 words, 252 syllables and 447 phonemes. As with previous data collection, the recordings were done in a noiseless environment. A Shure unidirectional head-worn microphone and a Tascam linear PCM recorder were used for the recordings. The two were connected via an xlr jack. The sampling frequency was fixed at 44.1 kHz and 24 bit in .wav format. The data recorded was segmented phoneme-wise in Praat 5.4. In order to ensure that pauses and fillers do not affect the measurements, only breath groups with at least 7 continuous syllables were considered for analysis.

5.2.2 Measurements

For speaking rate, syllables per second and segments per second measures were calculated following (Jacewicz et al., 2010). To calculate syllable per second, the number of spoken syllables is divided by the duration of each phrase. Likewise, segments per second are calculated by dividing the number of spoken segments (phoneme) by the duration of each phrase. To analyze the rhythm variation, interval measures that are evaluated are: %V, percent-age of vocalic intervals in the complete utterance; ΔC, standard deviation of duration of consonantal intervals; ΔV, standard deviation of the duration of vocalic intervals in a breath group (Ramus et al., 1999).

% V is calculated using Equation 5.1 d_k is the duration of the k^{th} vocalic interval, m is the total number of intervals and d_u is the total duration of the utterance.

$$\%V = 100X\left[\sum_{k=1}^{m} d_k\right]/d_u \tag{5.1}$$

It is seen that speaking rate can affect these measures and hence, rhythm measures based on Pairwise Variability Indices (PVI) was proposed to minimize the effect of speaking rate (Grabe & Low, 2002). The measures evaluated are: nPVI-V (normalized value of the durational variation of two consecutive vocalic intervals) and VarcoΔC (relative variation of consonantal intervals). VarcoΔC is calculated using Equation 5.2 and is defined as the percentage of standard deviation of consonantal interval duration (ΔC) of the average duration of consonantal intervals (mean C) (Dellwo, 2006).

$$Varco\Delta C = \frac{\Delta C * 100}{meanC} \tag{5.2}$$

Similarly, VarcoΔV is calculated from the vocalic intervals.

$$Varco\Delta V = \frac{\Delta V * 100}{meanV} \tag{5.3}$$

Two Pairwise Variability Index measures, nPVI-V and rPVI are evaluated following Grabe & Low (2002). nPVI-V is the rate normalized measures of the durational variation of two consecutive vocalic intervals and rPVI is the raw measure of the durational variation of two consecutive consonantal intervals. In Equation 5.4 for nPVI, d_k is the duration of the k^{th} interval, m is the total number of intervals.

$$nPVI = 100X\left[\sum_{k=1}^{m-1}\left|\frac{d_k - d_{k+1}}{\frac{d_k+d_{k+1}}{2}}\right|/(m-1)\right] \tag{5.4}$$

In Equation 5.5 m is number of intervals, vocalic or intervocalic, in the text and d_k is the duration of the k^{th}th interval.

$$rPVI = \left[\sum_{k=1}^{m-1}\left|d_k - d_{k+1}\right|/(m-1)\right] \tag{5.5}$$

5.3 Results

5.3.1 Correlation matrix

Rhythm measures have been found to be influenced by the number of syllables in the measured utterances. Similarly, rate of speaking affects some of the rhythm measures. In order to see how utterance length, speaking rate and rhythm interact with each other, a Pearson's correlation test was conducted. The results obtained are presented in form of a matrix in Figure5.1. As seen in the figure, ΔC, rPVI, ΔV and Varco-ΔV are strongly correlated with the rate of articulation (as indicated by Sylps and Segps). As the rate of speaking increases, these measures are lowered. Length of breath group

(as indicated by nv, number of syllables in a breath group) is strongly correlated to Varco-ΔC, ΔC, Varco-ΔV and ΔV. As the length of an utterance increases, Varco-ΔC, ΔC, Varco-ΔV and ΔV decrease. Considering these results it can be concluded that, overall, nPVI-V and %V are least affected by both rate of articulation and length of utterance. Interestingly, while Varco-ΔC is least affected by rate of articulation, it is affected by the length of utterance.

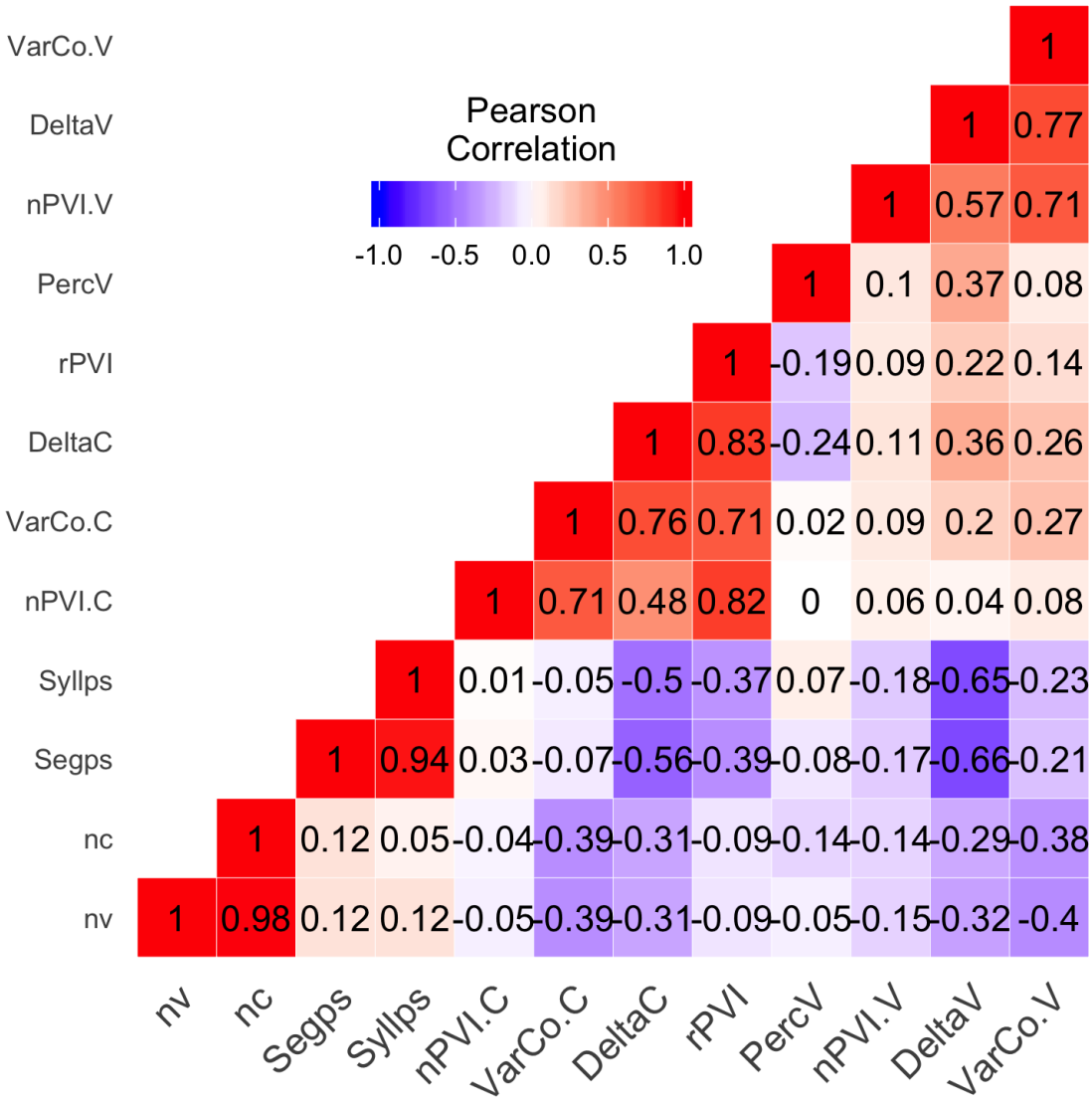


Figure 5.1: Pearson's correlation matrix

5.3.2 Rate of Articulation

The rate of articulation of speakers of Assamese from five regions of Assam are presented in Figure 5.2. As seen in the figure, there is a noticeable difference in the rate of articulation by regions.

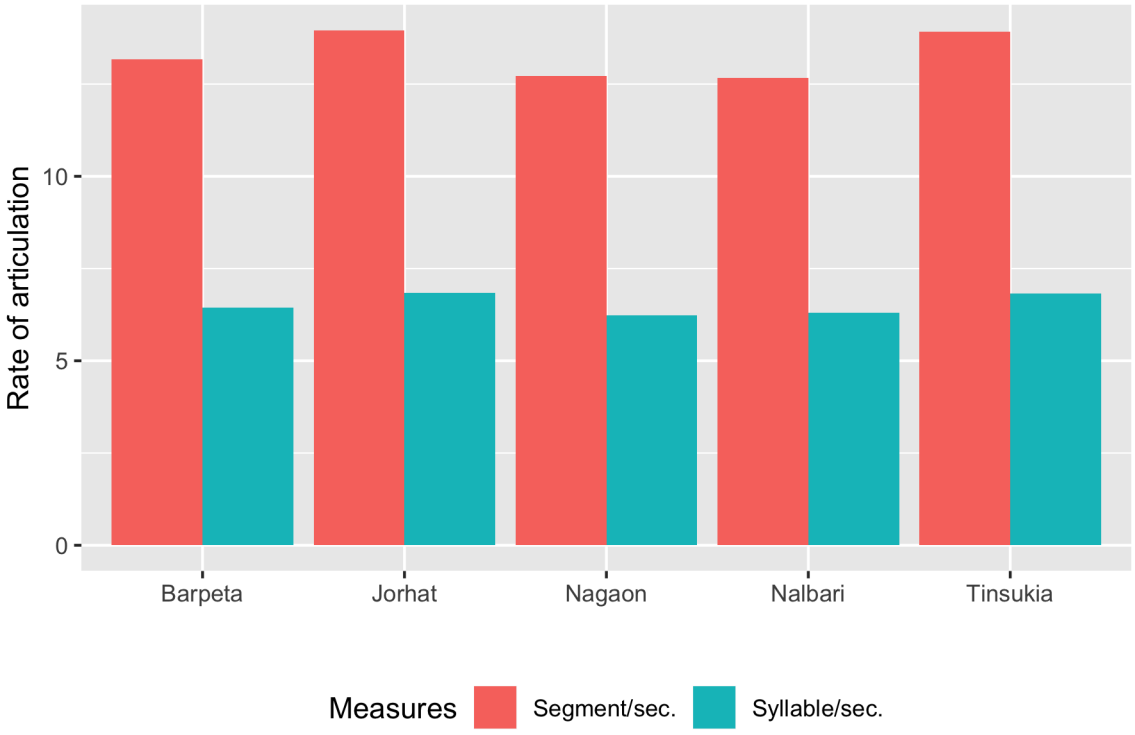


Figure 5.2: Rate of articulation in five Assamese regions

Assuming articulation rate difference by region, a couple of Linear Mixed Effect (LME) models were fitted using the *lme4* package on R (Bates et al., 2015). In both models, syllable/second and segment/second were the dependent variables and region was the fixed factor. Speaker and gender were considered random variables. After the models were built, Wald chisquare tests were conducted to see the effect of region on rate of articulation. The Wald chisquare tests conducted on both measures of articulation rate yielded non-significant p-values, (χ^2=6.4, p = 0.17, for segment/ second; χ^2=5.7, p = 0.22, for syllable/ second) indicating a weak correlation between

region and rate of articulation.

Subsequently, the five regions were subsumed under the three dialect areas they are said to belong to, and again a couple of LME models were built to see if dialect areas differ significantly in terms of their rate of articulation. Dialect area was considered fixed factor and speaker and gender were considered random factors. As expected, dialect areas showed no significant interaction with rate of articulation. For dialect areas, Wald chisquare tests yielded non-significant p-values for segment/second (χ^2=5.5, p = 0.06) and syllable/second (χ^2=5.4, p = 0.07).

5.3.3 Rhythm metrics

The rhythm measurements obtained for Assamese spoken in five regions of Assam are provided in Table 5.1. The table also presents rhythm measures of Japanese, English and Spanish for comparison, obtained from previous observations (Grabe & Low, 2002; Grenon & White, 2008; White & Mattys, 2007).

Table 5.1: Rhythm measure comparisons

Region	%V	nPVI-V	r-PVI	ΔV	ΔC	Varco-V	Varco-C
Barpeta	47.4	51.8	56.5	43.7	53.3	55.2	59.1
Jorhat	48.1	48.9	43.4	41.4	45.9	53.8	54.6
Nagaon	48.6	51.3	49.0	46.2	50.6	54.3	55.2
Nalbari	50.3	55.5	48.2	52.5	51.3	58.4	57.4
Tinsukia	50.0	52.4	46.4	42.6	47.0	54.5	58.3
British English	41.1	57.2	64.1	46.6	56.7	64.0	47.0
Spanish	50.8	29.7	57.7	20.7	47.5	41.0	46.0
Japanese	45.5	40.9	62.5	53.0	55.5	56.0	-

In Section 5.3.1 it has been observed that in case of Assamese, all measures, except %V and nPVI-V, are significantly affected by rate of articulation and length of utterance. Hence, we initially plotted the %V and nPVI-V values obtained from Assamese speakers, separated by regions. In order to compare them with prototypical syllable-timed, stress-timed and mora-timed languages, the measures from Assamese speakers were plotted along with that of British English, Spanish and Japanese for

comparison, obtained from previous observations (Grabe & Low, 2002; Grenon & White, 2008; White & Mattys, 2007). Figure 5.3 shows the distribution of languages on a %V and nPVI-V plane. As observed from the figure, the Assamese regions get clustered closer to Japanese.

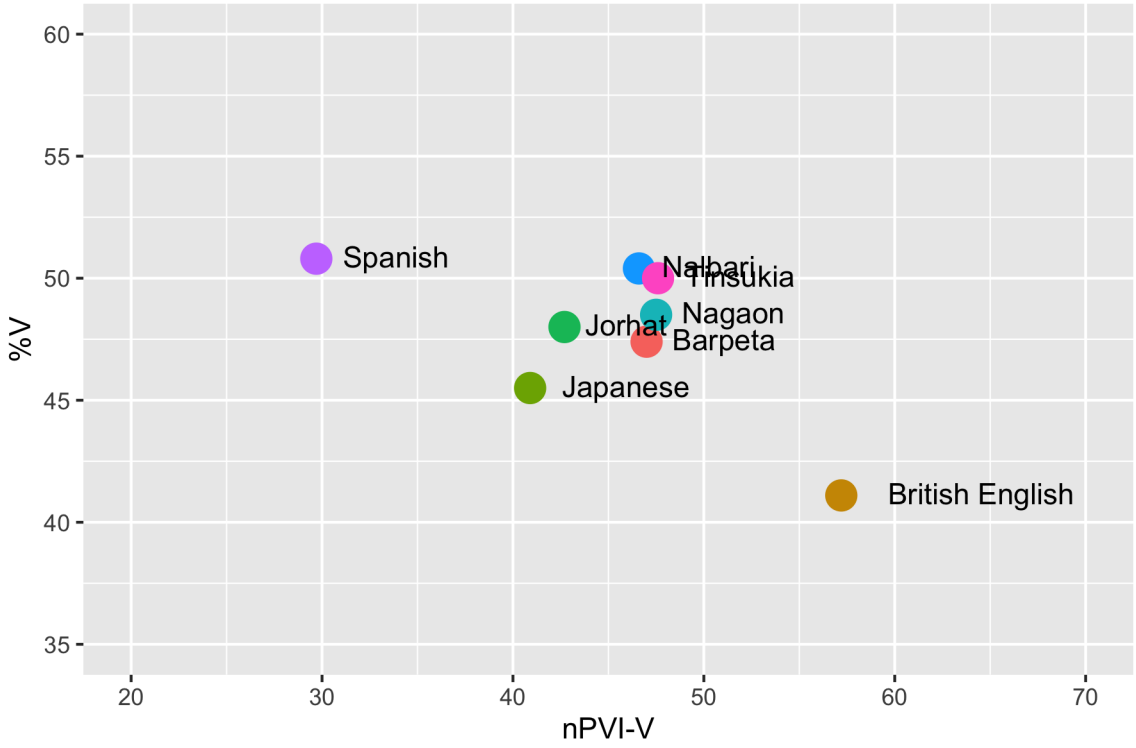


Figure 5.3: nPVI-V and %V in Assamese speakers

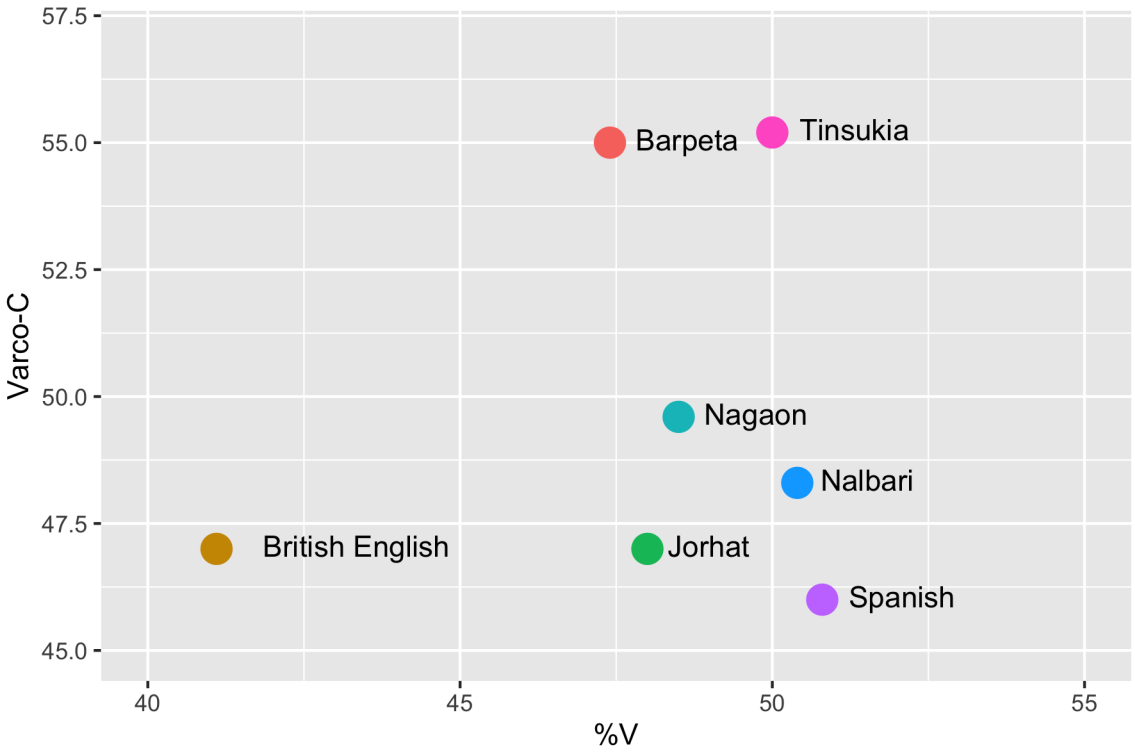


Figure 5.4: %V and Varco-C in Assamese speakers

It has been suggested that Varco-ΔC is not affected by speech rates and is considered better in discriminating stress-timed languages from syllable-timed languages (Dellwo, 2006). In Section 5.3.1 it has been noticed that in case of Assamese, even though speech rate does not affect Varco-ΔC, utterance length does. Nevertheless, it was decided to plot the two measures on an x-y plot as seen in Figure 5.4. While data for Japanese is not plotted in this plot, no coherent pattern of Varco-ΔC distribution of Assamese varieties has been seen. Hence, separation between rhythm type is best, with the given data, %V and nPVI-V are plotted on an x-y plot.

In order to see if the effect of region and dialect areas on rhythm measures, LME models were built for each measure where region or dialect area was considered fixed effect and gender and speaker were considered random effects. The models were subjected to Wald χ^2 test and the results obtained are presented in Table 5.2. Table

5.2 shows that all the rhythm measures, except ΔC and nPVI-V, show a significant effect of region. On the other hand, for dialect areas, %V, Varco-ΔC and nPVI-V do not show any effect of dialect areas.

Table 5.2: Wald χ^2 tests on LME models for rhythm measures

Measure	Region		Dialect area	
	χ^2	**p-value**	χ^2	**p-value**
%V	10.08	0.039	0.81	0.66
ΔV	17.32	0.002	7.48	0.02
ΔC	07.23	0.124	6.10	0.04
Varco-ΔV	14.28	0.006	8.48	0.01
Varco-ΔC	12.81	0.012	4.83	0.09
nPVI-V	08.63	0.070	3.67	0.16
rPVI	15.77	0.003	6.84	0.03

5.3.4 QDA results

In order to see the discriminability of Assamese spoken in different regions or different dialectal areas by means of rhythm and rate of articulation measures, a Quadratic Discriminant Analysis (QDA) was conducted by combining the rhythm measures and rate of articulation measures as features. Table 5.3 presents the results of QDA. Overall, Assamese spoken in five regions is discriminated with an accuracy of about 42.1% with all the rhythm and speaking rate measures. However, the best result of 42.8% is achieved when ΔC is removed from the analysis. In case of dialect identification, best accuracy of 47.4% is achieved when rPVI is removed from the analysis.

Table 5.3: Results of Quadratic Discriminant Analysis

Sl.	Features	Region ID	Dialect ID
1.	Seg/sec.	24.2%	39.7%
2.	Syl/sec.	24.0%	39.3%
3.	%V	22.2%	33.3%
4.	ΔV	25.8%	40.0%
5.	Seg/sec.+/sec.	25.9%	41.5%
6.	%V+ΔV+VarCo-V+ΔC+VarCo-C+nPVI-V+rPVI	37.8%	42.6%
7.	Seg/sec.+Syl/sec.+%V+ΔV+VarCo-V+ΔC+VarCo-C+nPVI-V+rPVI	42.1%	47.2%
8.	Seg/sec.+Syl/sec.+%V+ΔV+VarCo-V+VarCo-C+nPVI-V+rPVI	**42.8**%	46.9%
9.	Seg/sec.+Syl/sec.+%V+ΔV+VarCo-V+ΔC+VarCo-C+nPVI-V	41.5%	**47.4**%

Table 5.4: Region-wise accuracy in QDA for region ID

Tinsukia	Jorhat	Nagaon	Nalbari	Barpeta
59.5%	38.5%	28.1%	47.2%	40.7%

Table 5.5: Dialect-wise accuracy in QDA for Dialect ID

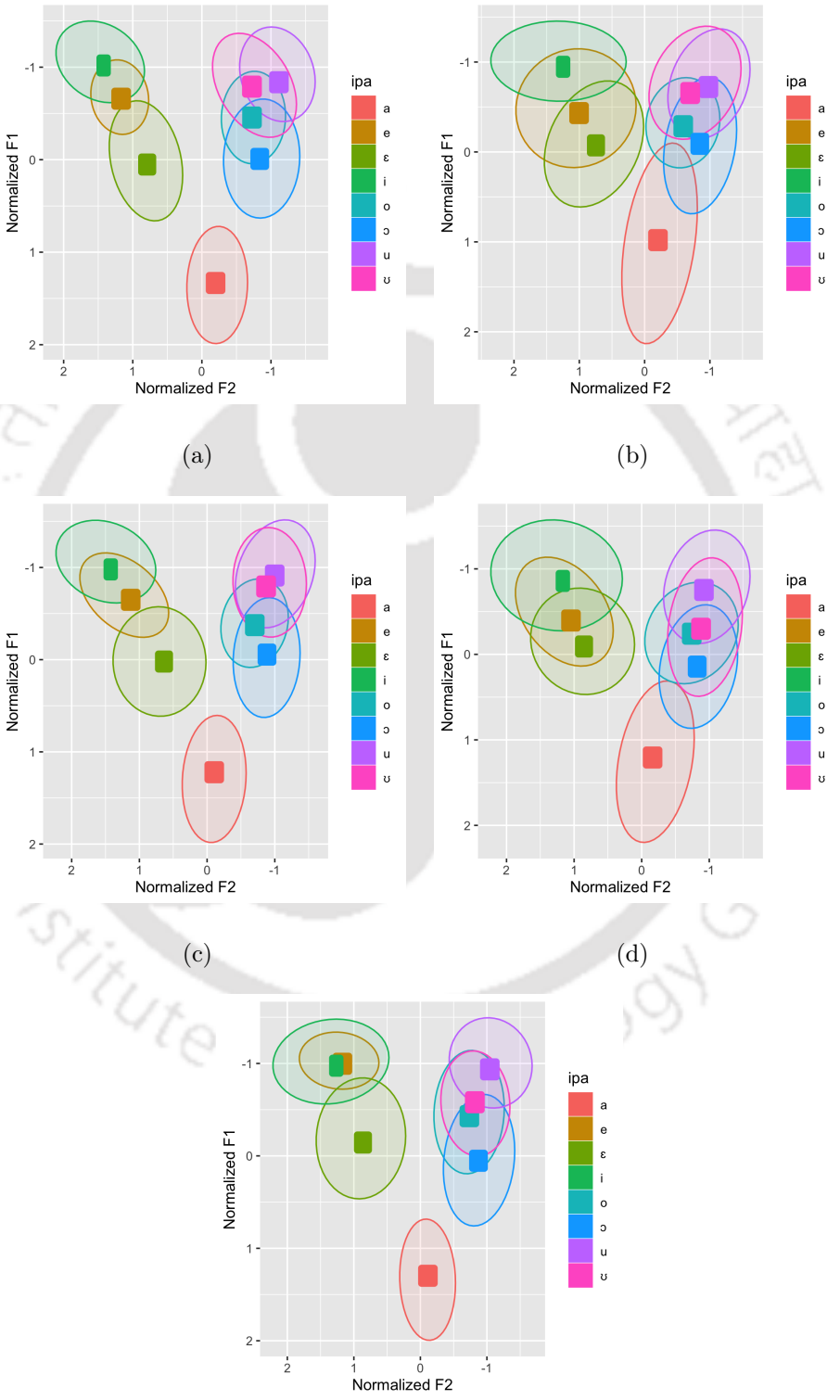
EA	CA	WA
71.3%	15.4%	55.6%

Table 5.4 and Table 5.5 provide the accuracy by region and dialect areas for the best discrimination accuracy. As seen in Table 5.4, in the best case, all areas, except Nagaon, are correctly categorized with an average accuracy of 46%. Similarly, Eastern and Western Assamese dialect areas are correctly categorized with an average accuracy of 63%. However, in case of the Nagaon region, which also is the sole member in the CA category, in this study, the accuracy in the lowest. It is important to note here that CA dialect division which is geographically nestled between EA and WA speaking regions and has been considered as an intermediate dialect (G. C. Goswami & Tamuli, 2003). This can be assumed to be the reason for the lower accuracy in categorizing the Nagaon region.

5.4 Conclusion

As mentioned in Section 5.1, a possible link between rhythm and vowel quality has been reported by Donegan & Stampe (1983) stating that stress-timed languages typically have larger vowel inventories while non-stress languages have simpler systems. Further, it is known that phonological features associated with vowels contribute to the perception of a language's rhythm (G. Schwartz, 2010). The evidence of vowel variation in Assamese varieties and previous claims by Ray (1966) and Savithri (2009) about rhythm in Assamese prompted a look into the phonetics of prosody in Assamese in this chapter. Accordingly, this chapter provided an analysis of speech rate and rhythm measures in five Assamese speaking regions belonging to three broad dialect areas. From the results of the analysis, rate of articulation measures indicated no strong correlation between region or dialect with speaking rate. In case of rhythm measures nPVI-V and %V are found to be least affected by speaking rates and length of breath groups. However, as suggested in the literature, VarcoΔC was not affected by rate of speaking however, it highly correlated with the length of utterance. When plotted on an nPVI-V-%V plane, Assamese varieties clustered close to Japanese, which is a mora-timed language.

Figure 5.5: F1-F2 plots of Assamese vowels produced in read passage of (a) Tinsukia, (b) Jorhat, (c) Nagaon, (d) Nalbari and (e) Barpeta areas.



(a) (b) (c) (d) (e)

Figure 5.5 presents the F1-F2 plots of Assamese vowels produced in read passage as listed as: (a) Tinsukia, (b) Jorhat, (c) Nagaon, (d) Nalbari and (e) Barpeta areas. As seen in the figures, considerable overlap of vowel ellipses with comparatively less separation among vowels is observed. However, vowel means indicate similar trends of merger observed for vowels produced in isolation and sentence frames as discussed in Chapter 2 and Chapter 3.

Previous works differ in concluding the rhythm class for Assamese (Ray, 1966; Savithri, 2009). Considering the analysis of rhythm class in Assamese, evidence of Assamese rhythm being more akin to mora-timed rhythm class has been found. Mahanta (2001) claims that in Assamese a foot is always bimoraic and contains two moras. Hence, rhythm measures clustering Assamese with mora-timed languages, is not surprising after all. Finally, native speakers who participated in the data collection for this work had indicated that one of the distinguishing features that differentiated between Eastern Assamese and Western Assamese is the speed or tempo of speech. This work reports that rhythm and rate of articulation measures can discriminate Assamese spoken in different regions, at least above the chance level. Using rhythm metrics and rate of speech measures a QDA classification was conducted. The rhythm features based region classification yielded moderate accuracy. In Chapter 6 we will compare these classification results with that of classification using vowel features to explore any correlation between vowel quality and rhythm measures.

Chapter 6

Conclusion

The primary objective of this work was to look into the nature of vowel variation in different varieties of Assamese. Accordingly, the acoustic properties of vowels in Assamese spoken in five different geographical locations of Assam were studied. These five regions were Tinsukia, Jorhat, Nagaon, Nalbari and Barpeta. Using acoustic and statistical measurements this work attempted to find what significant changes the eight-vowel system of standard Assamese has undergone in these varieties. This work also looked into the nature of vowel mergers in Assamese using different methods for calculating the same. Additionally, a study of the Assamese native speakers' perception of the vowels in merger was also conducted. In the following sections the major findings of this work have been briefed.

6.1 Major findings

6.1.1 Vowel variation in Assamese varieties

As described in Chapter 1, Assamese is broadly divided into three dialect groups based on geographical locations and each of these three varieties have several sub varieties which are distinct from each other. These distinctions are based on several factors, vowel variation being one of them (Taid, 1988; G. C. Goswami & Tamuli, 2003; Deka, 2007). As such, an objective of this work was to study the acoustic

properties of Assamese vowels as spoken in Tinsukia, Jorhat, Nagaon, Nalbari and Barpeta and look into the geographical progression of vowel change in Assamese dialects and present an understanding of the vowel space in each of these varieties.

The results of the production experiment have concluded that there are significant vowel changes that affect variation among these varieties. Among the five geographical locations taken for the study, Tinsukia and Jorhat are representatives of the Eastern Assamese variety, Nagaon represents the Central Assamese variety and Nalbari and Barpeta represent the Western Assamese variety. The study concludes that Tinsukia, Jorhat and Nagaon varieties do not distinguish between the vowels /u-ʊ/ and maintain significant acoustic separation only among seven vowels. In case of Nalbari and Barpeta varieties, on the other hand, no distinction is made between the vowels /o-ʊ/. Additionally, no distinction is made between the front vowels /ɛ/ and /e/, reducing the eight-vowel system to a six-vowel one. In the case of Nagaon however, we also see considerable overlap of the back vowels /u,ʊ,o/, as also confirmed by the Pillai scores and SOAM measurements in Chapter 3. Thus we see a gradual change in the movement of a vowel which has been marked as /ʊ/ in Standard Assamese and we can almost trace its path as we go from Tinsukia to Jorhat onto Nagaon and finally to Nalbari and Barpeta.

6.1.2 Vowel merger in Assamese varieties

Assamese varieties displayed considerable overlap between vowel ellipses indicating vowel mergers in the varieties. To draw conclusive evidence of the mergers, the vowels underwent a series of tests to determine the nature of the mergers, viz. Euclidean Distance, Pillai Bartlett Trace, Spectral Overlap Assessment Metric(SOAM) and Vowel Overlap Assessment with Convex Hulls (VOACH). Results of the merger tests confirm the findings of the production experiments. While Tinsukia, Jorhat and Nagaon varieties showed merger of /u-ʊ/ vowels, Nalbari and Barpeta showed merger

of /o-ʊ/ along with merger of front vowels /e-ɛ/.

As seen in this work, the /ʊ/ vowel shows two opposite directions of merger in the varieties. In the Tinsukia, Jorhat and Nagaon varieties, the vowel raises to the vowel space of the /u/, completely absorbed by the latter, thus showing evidence of merger by approximation. On the other hand, in Nalbari and Barpeta varieties the categorical boundaries between the /ʊ/ and /o/ vowels have disappeared constituting a larger categorical space and indicating merger by expansion. Again, in the Nalbari and Barpeta varieties, the /e/ and /ɛ/ overlap seems like a near-merger by approximation. Thus, this work concludes that vowel merger is a distinct feature of Assamese vowel system and the asymmetrical nature of merger is an important feature in distinguishing Assamese varieties.

Table 6.1: Summary of mergers in Assamese varieties

Region	Vowels	Merger type
Tinsukia	/u-ʊ/	Approximation
Jorhat	/u-ʊ/	Approximation
Nagaon	/u-ʊ/	Approximation
Nalbari	/ʊ-o/	Expansion
	/e-ɛ/	Near-merger by approximation
Barpeta	/ʊ-o/	Expansion
	/e-ɛ/	Near-merger by approximation

6.1.3 Perception of mergers

One of the main objectives of this work was to see if speakers across varieties are able to categorically perceive the Assamese vowels. The results of the production experiments and subsequent calculation of overlap measures conducted on the Assamese vowels spoken in the five regions confirmed two way mergers in Assamese varieties. A perception experiment was done to look into the native speakers' categorical perception of these mergers and two types of perception tests were conducted:

Discrimination test, where participants were asked to determine whether they hear a difference between two sounds and Identification test where participants were asked to identify the sound they hear.

The results of the experiments confirm perceptual merger of /u-ʊ/ vowels in the Tinsukia, Jorhat and Nagaon varieties. The same is confirmed for /ʊ-o/ vowels in Nalbari and Barpeta. The results also provide evidence of a production-perception mismatch existing in Assamese varieties in situations of near-mergers of /e-ɛ/ vowels in both Nalbari and Barpeta. Additionally, production-perception anomaly has also been seen in case of the vowel /ʊ/. This is due to the higher F1 value of the vowel as a result of its merger with the /o/ vowel in the two varieties.

6.1.4 Rhythm and speech rate

Apart from lexical and segmental differences among the Assamese varieties, native speakers of Assamese also report perceiving the Western varieties as being faster than the Eastern and Central varieties. Hence, in this study the five geographical varieties are analysed for speaking rate and rhythm in read speech. For speaking rate, the measures calculated are syllable per second and segments per second. Table 6.2 below summarizes the result of the speech rate measures in the five varieties.

Table 6.2: Speech rate measures in five geographical varieties of Assamese

Region	Segment per sec	Syllable per sec
Tinsukia	13.9	6.7
Jorhat	8.7	4.3
Nagaon	13.2	6.3
Nalbari	11.6	5.8
Barpeta	13.5	6.5

Results show that Tinsukia has the highest segment per second and syllable per

second measure among all the five varieties indicating faster speech. Jorhat has the lowest measures for both indicating slower speech. To analyze the rhythm variation, interval measures that are evaluated are: %V, percent-age of vocalic intervals in the complete utterance; ΔC, standard deviation of duration of consonantal intervals; ΔV, standard deviation of the duration of vocalic intervals in a breath group and Varco V (Ramus et al., 1999).

The results obtained show that the rhythm measures of Assamese are comparable to that of the mora-timed languages, such as Japanese. The results also showed that the five regions differed significantly from each other in terms of rhythm measures and rate of speaking. A quadratic discriminant analysis showed that the Assamese spoken in the five regions can be discriminated with an accuracy of about 42% with rhythm measures and speaking rate. A tertiary finding also showed that %V and nPVI-V are the least affected measure by rate of speaking.

6.2 QDA classification of vowels

A quadratic discriminant analysis was conducted to see the accuracy in classifying vowels using the first and second formant features (F1 and F2). The results of the QDA classifications are presented in Table 6.3 to Table 6.7.

Table 6.3: Accuracy in % in QDA classification of vowels of Tinsukia variety.

	ɛ	ɔ	ʊ	a	e	i	o	u
ɛ	**93.65**	0	0	2.12	4.23	0	0	0
ɔ	0	**92.24**	1.94	0.88	0	0	4.76	0.18
ʊ	0	6.6	**31**	0	0	0	1.4	61
a	0.24	1.95	0.73	**97.08**	0	0	0	0
e	55.26	0	0	0	**26.32**	18.42	0	0
i	0	0	0	0	2.99	**97.01**	0	0
o	0	23.65	26.35	0	0	0	**30.41**	19.59
u	0	1.54	21.27	0	0	0	1.97	**75.22**

Table 6.3 presents the percentage of accuracy in QDA classification of vowels of Tinsukia variety. As seen in the table, among the back vowels, vowel /ɔ/ shows the highest accuracy in classification. Vowel /ʊ/ shows 61% classification as /u/. Vowel /u/ on the other hand shows 75.22% classification as itself and 21.27% as vowel /ʊ/. Classification of vowel /o/ on the other hand, is distributed among the other back vowels. Production results have shown the merger of the vowels /ʊ-u/ in Tinsukia with /ʊ/ raising to the position of /u/ and being completely absorbed by the latter. This is evident also in the QDA classification results /ʊ/ is more classified as /u/ than /u/ as /ʊ/. Among the front vowels, while vowels /i/ and /ɛ/ show accurate classification, vowel /e/ is classified more as /ɛ/. Production results confirmed no merger of the two front vowels, however, it is to be noted that tokens of the /e/ vowel were lesser in comparison to the other vowels. Since QDA classification is sensitive to the number of data points, the low number of tokens could be a reason of this discrepancy.

Table 6.4: Accuracy in % in QDA classification of vowels of Jorhat variety.

	ɛ	ɔ	ʊ	a	e	i	o	u
ɛ	**93.52**	1.02	0	1.71	3.41	0.34	0	0
ɔ	0	**94.88**	2.11	0.6	0	0	1.51	0.9
ʊ	0	3.21	**29.39**	0.34	0.17	0	3.21	63.68
a	0.17	1.9	0	**97.75**	0	0	0.17	0
e	8	0	0	0	**77**	15	0	0
i	0	0	0	0	4.52	**95.48**	0	0
o	0	10.57	45.81	0	0	0	**18.5**	25.11
u	0	1.25	19.5	0	0.18	0.18	0.89	**78**

Table 6.4 presents the percentage of accuracy in QDA classification of vowels of Jorhat variety. As in the case of Tinsukia, vowel /ɔ/ shows the highest accuracy among the back vowels. Vowel /ʊ/ shows 63.68% classification as /u/. Again, Vowel /u/ shows 78% classification as itself and 19.5% as /ʊ/. Vowel /o/ shows the lowest accuracy with the classification distributed among the other back vowels. Production results of Jorhat vowels have shown the raising of /ʊ/ to the position of /u/ and merging with the latter. This is evident also in the QDA classification results /ʊ/ is more classified as /u/ than /u/ as /ʊ/. Among the front vowels, vowels /i/ and /ɛ/ show accurate classification. Vowel /e/ also shows high accuracy of 77%.

Table 6.5: Accuracy in % in QDA classification of vowels of Nagaon variety.

	ɛ	ɔ	ʊ	a	e	i	o	u
ɛ	**93.07**	0	1.3	0.87	3.46	0.87	0	0.43
ɔ	0.33	**93.09**	2.8	1.64	0	0.16	0	1.97
ʊ	0	4.68	**28.46**	0.75	0	0	0	66.1
a	1.06	3.18	0.42	**95.34**	0	0	0	0
e	7.89	0	0	0	**81.58**	10.53	0	0
i	0	0	0	0	5.19	**94.81**	0	0
o	0	10.74	69.13	0	0	0	**0**	20.13
u	0.41	1.45	17.43	0	0	0	0	**80.71**

The percentage of accuracy in QDA classification of vowels of Nagaon variety are presented in Table 6.5. As seen in the table, vowels /ɔ/ and /u/ show the highest accuracy in classification. Vowel /ʊ/ also shows high accuracy with a small percentage classification as /u/. Vowel /o/ on the other hand, shows no classification as itself and 69.13% as /ʊ/. Although production results showed merger of the /u-ʊ/ vowels, considerable overlap of the /ʊ-o/ vowels was also observed. Hence the predictability of its classification as /ʊ/ can not be ruled out. Among the front vowels, all three vowels show highest accuracy in classification.

Table 6.6: Accuracy in % in QDA classification of vowels of Nalbari variety.

	ɛ	ɔ	ʊ	a	e	i	o	u
ɛ	**89.69**	0.84	1.39	4.18	0	3.62	0	0.28
ɔ	0.24	**92.6**	5.97	0.84	0	0	0	0.36
ʊ	0.27	8.14	**73.7**	0	0	0.27	0	17.62
a	0.58	3.47	0	**95.95**	0	0	0	0
e	91.3	0	0	0	**0**	8.7	0	0
i	7.8	0	0	0	0	**92.2**	0	0
o	0	23.72	61.68	0	0	0	**0**	14.6
u	0.16	0.48	19.21	0	0	0.16	0	**80**

Table 6.6 presents the percentage of accuracy in QDA classification of vowels of Nalbari variety. Among the back vowels, vowel /ɔ/ and /u/ show the highest accuracy in classification. Vowel /ʊ/ also shows a high accuracy with 17.62% classification as /u/. As in the case of Nagaon, vowel /o/ shows no classification as itself and 61.68% as /ʊ/. Since production results have shown the merger of the vowels /ʊ-o/ in Nalbari, this classification is predictable. Among the front vowels, vowels /i/ and /ɛ/ show accurate classification. However vowel /e/ shows 91.3% classification as /ɛ/. Again, since Nalbari showed merger of /e/ and /ɛ/ vowels in production, this result is highly predictable.

Table 6.7: Accuracy in % in QDA classification of vowels of Barpeta variety.

	ɛ	ɔ	ʊ	a	e	i	o	u
ɛ	**93.17**	1.95	0	0.49	0	4.39	0	0
ɔ	0.18	**92.74**	4.9	1.09	0	0	0	1.09
ʊ	0	9.52	**60.66**	0	0	0	0	29.81
a	1.51	2.11	0	**96.37**	0	0	0	0
e	77.78	0	0	0	**0**	22.22	0	0
i	6.06	0	0	1.52	0	**92.42**	0	0
o	0	28.26	56.52	2.17	0	0	**0**	13.04
u	0	1.79	18.34	0.22	0	0	0	**79.64**

The percentage of accuracy in QDA classification of vowels of Barpeta variety is shown Table 6.6. Vowel /ɔ/ shows the highest accuracy in classification. Vowel /u/ shows 79.64% accuracy in classification with 18.34% as /ʊ/. As in the case of Nalbari, vowel /o/ shows no classification as itself and 56.52% as /ʊ/. Since the vowels /ʊ-o/ merged in production, this result is not unpredictable. Among the front vowels, while vowels /i/ and /ɛ/ show accurate classification, vowel /e/ shows 77.78% classification as /ɛ/. Again since Barpeta showed merger of /e/ and /ɛ/ vowels in production, this result is highly predictable.

As observed in the table, vowels that showed merger in production in the Assamese varieties displayed the lowest accuracy in classification. Vowels that showed distinctiveness on the other hand, showed high accuracy. It can thus be concluded from the results that acoustic similarity between overlapping vowel pairs are robust. These acoustic similarities also result in poor modelling of overlapping vowel categories making accurate classification difficult.

6.3 Dialect classification

Considering the evidence of vowel variation in Assamese presented in this work, we are motivated to explore if vowel acoustics can actually aid automatic dialect identification. If automatic dialect identification using vowel features yields high accuracy, it also substantiates our claim in the previous sections of this work that vowel acoustics of the five varieties of Assamese are distinct. Hence, a dialect classification using Random Forest (RF) classifiers was attempted to see if those characteristics are going to aid in automatic dialect identification in Assamese using vowel features. The importance of incorporating vowel specific features in automatic dialect identification was discussed in several early works (Higgins et al., 1999). In Ferragne & Pellegrino (2007), automatic dialect identification was attempted on 15 British English varieties using only vowels with an overall accuracy of 90%. Biadsy et al. (2009) also underlined the importance of vowels, specifically the emphatic vowels in Arabic in achieving higher accuracy in Arabic dialect classification. Chittaragi & Koolagudi (2019) used acoustic-phonetic features, namely, formant frequencies (F1–F3), and prosodic features such as, energy, pitch (F0), and duration etc. In classifying five Kannada dialects, an accuracy of 76% is reported by the authors. Recently, Devi & Thaoroijam (2020) also used F1, F2, F3, duration, energy and F0 features from vowels to classify three dialects of Meeteilon (Manipuri) language. They report the best accuracy of 61.57% in Meeteilon dialect identification. In case of Assamese, a previous attempt at automatically classifying Assamese dialects using vowels as input segments achieved a dialect recognition accuracy of 89.32% (Sarma & Sarma, 2016).

The classifier used in this study uses the acoustic features used in the previous chapters. G. C. Goswami (1982) has mentioned that certain monophthongs may be pronounced as diphthongs in the western varieties of Assamese. Considering that, in order to capture vowel-inherent dynamic changes, the formant frequency trajectory within a vowel was transformed using discrete cosine transform (DCT), and the first

three coefficients of the DCT, namely C0, C1 and C2 were extracted for each formant. Hence, the final set of features consisted of 14 features, namely, F1, F2, F3, and the first three DCT coefficients of the 3 formant trajectories (3 x 3 =9), F0 and duration. Apart from the feature vectors, the data grid included information about the speakers, vowels, gender, contexts (sentence or isolation), onset and coda consonants. The variable `region`, representing five regions of Assam, namely Tinsukia, Jorhat, Nagaon, Barpeta and Nalbari were the predictors in the classifier.

6.3.1 Random Forest classifier

Random Forest is an ensemble method used for classification or regression. When it is used for classification, the output of a trained RF is the class label of input data. On the other hand, when a RF is used for a regression task, the output is a value predicted corresponding to the given input. An RF employs many decision trees, each of which carries out the classification or regression task independently. In case of classification, the class label of a new instance of input is decided as that label set by a majority of the decision trees. In case of prediction, the output of the RF is the average of the outputs of all decision trees. Randomness in a RF is introduced through several schemes. In the following sections, the use of an RF for classifying an Assamese vowel as belonging to one of the 5 dialectal regions has been described.

In general, increasing the number of decision trees in a RF improves accuracy of classification; yet, the increase does not result in over-training of the data (Ho, 1995). However, the method would work the best if the trees are not correlated, because correlated trees tend to make similar errors in classification. Hence, the ability to predict classes accurately increases as the number of uncorrelated decision trees increase. Individual decision trees are sensitive to small changes in the training data. Hence, each tree is trained with a data set that is obtained by randomly sampling the entire training data with replacement. This method of sampling with

a replacement is known as bootstrap procedure, wherein generated data sets can contain several replicates of the same observations.

While the bootstrap procedure, as described above, introduces randomness in RF based classification, additional randomness is introduced by random feature selection during the construction of a decision tree. At every node of the tree, a subset of the entire feature set is randomly selected. The optimal feature to split the tree at this node is chosen from this randomly selected subset of features. Bootstrap procedure and random selection of features reduces correlation among decision trees, thereby increasing the probability of higher accuracy of classification. The size of this random subset of features is called a set by the user, and is known as the `mtry` parameter in the RF classification program of the R software toolset. The other parameters to be set by the user are `nodesize`, `maxnode` and `ntree`. The size of the tree is controlled by `nodesize` and `maxnode` parameters. The splitting of a node of tree is repeated until each cell of the tree contains less than `nodesize` data points. The parameter `maxnode` sets the maximum number of terminal nodes/cells in a tree.

Finally, the effectiveness of a random forest classification depends on the number of uncorrelated trees that are used in computing the aggregate results. Usually, this parameter is defined as `ntree` in the literature. To summarize, the architecture of an RF is characterized by four parameters, namely `mtry, nodesize, maxnode` and `ntree` (Scornet, 2017). The accuracy of the classifier and the computational cost depends on the values assigned to these parameters. While the trade-off between the two is not clear, it is generally true that larger values set for `mtry` result in better accuracy. Similarly, it is well known that high accuracies can be obtained with a greater number of trees, that is with higher `ntree` values (Oshiro et al., 2012). Hence, the values assigned to these parameters depend on the accuracy obtained and computational cost incurred. In the following section, we discuss how the optimal values for the four parameters mentioned in this section were determined for dialect

classification of Assamese using vowels.

6.3.2 Experimental Setup

The random forest classification experiment was set up using the R software (R Core Team, 2019) with the `randomForest`, `caret` and `e1071` packages (Liaw & Wiener, 2002; Kuhn, 2015; Dimitriadou et al., 2008). The classifier was trained on 80% of the data used to derive vowel acoustics in the previous chapters. The parameters, `mtry, nodesize` and `maxnode`, were changed stepwise to arrive at the optimum value for training the model. Finally the model was tested with the remaining 20% of the dataset.

6.3.3 Model parameters

For training, 80% of the data, i.e. 11431 vowels iterations were randomly selected. The remaining 20% were kept for testing the model. Initially, the training set was subjected to training using 5 fold validation. In the initial training, the `mtry` parameter was varied from 1 to 70, keeping `nodesize` at 14 and `maxnode` at 24. The optimal results were obtained when `mtry = 56`. Following that we varied the `maxnode` parameters from 20 to 2000 and found the highest accuracy when `maxnode = 452`. Finally, the `ntree` parameter was varied from 250 to 1100 at steps of 50, and at `ntree = 1000`, the best accuracy was obtained. Hence, for training the model, the following parameters were set:

- `nodesize` = 14
- `mtry` = 56
- `maxnode` = 452
- `ntree` = 1000

6.3.4 Results

6.3.4.1 Region-wise classification

The model trained using 80% of the data, i.e. 11431 vowel tokens was used to validate 20% of the randomly distributed data, i.e. 2858 vowel tokens. The overall accuracy that the Random Forest classifier reported is 93.4%. However, when overall accuracy is calculated as an average of the accuracy for the five regions, the number is reduced to 92.8%. Table 6.8 provides the dialect identification accuracy obtained for the five regions of Assam.

The system derived accuracy is calculated as the percentage of total number of tokens correctly identified out of the 2857 vowel tokens. However, we also calculated overall accuracy as an average of the individual accuracy for each of the five regions in the test set (we refer to as 'manually calculated' hereafter). When the model is trained with all the 14 features, both the system-generated and manually-calculated accuracy of dialect classification was 99.7%. As the training and test datasets were not mutually exclusive in terms of the speakers, we assumed that this high classification score is possibly due to speaker resemblance rather than dialect resemblance.

Considering the possibility of speaker effects, we conducted another classification with 80% of speakers in each dialect included in the training set and 20% of speakers in each dialect included in the test set. This resulted in 12644 vowel tokens in the training set and 1645 vowel tokens is the test set. The average accuracy of classification in the five regions reduced to 63.9% (system generated, 60.4%). As seen in Table 6.8, the classification results indicate that the Barpeta and Tinsukia regions have poor accuracy and they are confused with regions within their own dialectal regions, i.e., Western Assamese (WA) and Eastern Assamese (EA), respectively. The classification of Tinsukia is biased towards the Jorhat region and Barpeta is biased towards the Nalbari region, which is possibly due to a class imbalance arising due to the lesser

amount of data available for the Tinsukia and Barpeta regions.

Table 6.8: Confusion matrix for region identification (values in %)

	Barpeta	Jorhat	Nagaon	Nalbari	Tinsukia
Barpeta	**22.2**	0	0	77.8	0
Jorhat	0	**100**	0	0	0
Nagaon	0	0	**100**	0	0
Nalbari	2.6	0	0	**97.4**	0
Tinsukia	0	100	0	0	**0**

6.3.4.2 Dialect-wise classification

A classification attempt by dialect areas as identified for Assamese dialects in G. C. Goswami & Tamuli (2003) was also conducted. In this classification experiment, the Jorhat and Tinsukia regions were subsumed under the EA group, Nagaon under the CA group and Barpeta and Nalbari regions under the WA group. In order to run a 5-fold experiment, the speaker set in each of the 3 dialect regions was divided into 5 mutually exclusive subsets of speakers while maintaining gender balance. In fold1 experiment, data from the first subset of speakers of all regions formed the test set. The remaining data formed the training set. The fold1 experiment was run 10 times to explore the variation of classification accuracy due to randomness inherent in the RF classifier. The results of the 10 trials, (T1 – T10), of the first fold are shown in Table 6.9. The highest accuracy is 81.8%, while the average of 10 accuracy is 70.6%.

A series of experiments was conducted to determine the optimal values of the parameters. The parameters `nodesize`, `mtry`, `maxnode` and `ntree` were set at 14, 4, 34 and 248, respectively. Five-fold classification was conducted with alternating speaker exclusive train and test sets. The classification accuracies (in %) of the 5-fold experiments were 81.8, 68.0, 76.7, 72.4, 71.8; the average accuracy was 74.1%. Table

6.10 shows the confusion matrix of the classification in the first fold. While EA and WA dialects are classified with an average accuracy of 91.8%, the CA dialect has lower accuracy and tends to be misclassified as WA. The low accuracy in case of CA can be attributed to data imbalance arising due to lower sampling points.

Table 6.9: Classification accuracy (in %) in 10 trials of the first fold of training-testing

T1	T2	T3	T4	T5	T6	T7	T8	T9	T10
68.5	65.0	81.8	70.3	68.6	68.7	64.8	67.1	71.7	79.8

Table 6.10: Confusion matrix for Assamese dialect classification (values in %)

	Eastern	Central	Western
Eastern	**94.9**	00.0	05.1
Central	07.9	**61.7**	30.4
Western	11.4	00.0	**88.6**

6.3.4.3 Region-wise classification with balanced dataset

In Sections 6.3.4.2 and 6.3.4.1, it was assumed that imbalanced datasets have resulted in errors in classification. In order to confirm that, a final experiment was conducted, where, for each region, the number of speakers in the test and the train sets were kept constant, i.e. for each region, 2 speakers were in the test set and 8 speakers were in the training set. This resulted in 7684 and 1454 vowel tokens for training and testing, respectively. The training parameters were same as described in Section 6.3.4.1. When the test set was used to evaluate the model, it yielded an accuracy of 94.0% in correctly recognizing the region of the speakers. The confusion martrix in Table 6.11 shows that with the balanced dataset, the accuracies for region identification are higher. Compared to the accuracy in Table 6.8, the accuracy in classifying the Barpeta and Tinsukia regions is significantly high.

Table 6.11: Confusion matrix for region identification using balanced dataset (values in %)

	Barpeta	Jorhat	Nagaon	Nalbari	Tinsukia
Barpeta	**96.4**	0	0	3.6	0
Jorhat	0	**83.7**	0	0	16.3
Nagaon	0	0	**100**	0	0
Nalbari	9.4	0	0	**90.6**	0
Tinsukia	0	0.5	0	0	**99.5**

This study showed that vowels do have enough information that can be exploited for dialect or region classification. The results also showed that geographically closer areas have a greater tendency to be confused. Moreover, it was also noticed that Random Forest classification is prone to errors arising due to imbalanced classes, which need to be addressed for more reliable classification. This problem has been discussed extensively in the literatur, as in More & Rana (2017). A future direction of this work will address the issues arising due to data imbalance to achieve more robust classification.

6.4 Implications and future prospects

The findings of this study provide an understanding of the spectral and temporal features of vowels in Assamese varieties and the distinctive phonetic features of dialectal variation in Assamese. The study also provides an insight into the synchronic vowel variation in the language and attests to an asymmetrical pattern of vowel merging in Assamese varieties: in the Tinsukia, Jorhat and Nagaon varieties, the /ʊ/ vowel is raised and merged with the /u/ vowel by merger by approximation. In the Nalbari and Barpeta varieties, the same vowel is tends to expand its phonetic space and merge with the /o/ vowel. Again, in the same varieties the vowel /ɛ/ shows evidences of near-merger with the vowel /e/.

The results of the perception tests, while confirming the perceptual mergers of the back vowel contrasts, also hint at a production-perception dichotomy existing in Assamese varieties. A detailed work in the future on the perception of these mergers is expected to throw significant light on the understanding of vowel perception in the language and also validate the findings of this acoustic study.

The results from the dialect classification experiment in Chapter 5 showed weak evidence of rhythm variation across the five varieties explored in this work. On the other hand, the dialect identification experiment using vowel features in this chapter yields much better accuracy. Hence, we do not see a direct correspondence between dialect-specific rhythm features and dialect-specific vowel features in the five varieties investigated in this work. Additionally, as a secondary finding, this work has been able to show that Assamese is more akin to mora-timed languages and also provides the rhythm metrics for the language. However, an exhaustive analysis of rhythm and speech rate in Assamese varieties based on the phonological features associated with vowels would provide deeper insight into these temporal features.

Finally, as there is evidence that vowel characteristics are different for the five regions, we tested if the differences are also captured in automatic dialect identifi-

cation in Assamese using vowel features. A random forest classifier was trained to classify the five regions based on vowel characteristics. The accuracy of classification in unseen test data was 94.0%. A confusion matrix showed that geographically closer regions have a greater tendency to be confused. The results confirmed that the acoustic and perceptual evidence for vowel variation in Assamese spoken in the five regions is robust.

Overall, this study attempted to understand the nature of vowel variation in different varieties of Assamese from an acoustic phonetic perspective. This dissertation provided a detailed exploration of the acoustic features of vowels spoken in five geographically distinct regions of Assam. It is expected that the findings of this study will help motivate more detailed and expansive exploration of acoustic features of Assamese varieties.

Appendix A

Background information of participants

A.1 Participants for production experiment

Table A.1: Details of participants for the preliminary data

Region	Speaker	Age	Gender	Medium of instr.	Other languages known	Type of data collected
Jorhat	JF01	33	F	Assamese	Hindi, English	Wordlist and passage
	JF02	37	F	Assamese	Hindi	Wordlist and passage
	JF03	23	F	Assamese	English, Hindi	Wordlist and passage
	JF04	34	F	Assamese	Hindi, English	Only Wordlist
	JM01	44	M	Assamese	Hindi, English	Wordlist and passage
	JM02	35	M	Assamese	English, Hindi, Bengali	Wordlist and passage
	JM03	35	M	Assamese	Hindi, English	Only Wordlist
	JM04	33	M	Assamese	Hindi, English	Only Wordlist
	JM05	29	M	Assamese	English, Hindi	Only Wordlist
	JM06	46	M	Assamese	English, Hindi	Only Wordlist
	JM07	36	M	Assamese	English, Hindi	Only Wordlist
Nalbari	NF01	43	F	Assamese	English, Hindi	Wordlist and passage
	NF02	26	F	Assamese	Hindi	Wordlist and passage
	NF03	26	F	Assamese	Hindi	Wordlist and passage
	NM01	35	M	Assamese	English, Hindi	Wordlist and passage
	NM02	41	M	Assamese	English, Hindi	Wordlist and passage
	NM03	42	M	Assamese	English, Hindi	Wordlist and passage
	NM04	41	M	Assamese	English, Hindi, Bengali	Only Wordlist
	NM05	52	M	Assamese	English, Hindi	Only Wordlist
	NM06	26	M	English	Hindi, English, Bengali	Only Wordlist
	NM07	38	M	Assamese	Hindi	Only Wordlist

A.2 Participants for perception experiment

Table A.2: Details of participants from the field

Region	Speaker	Age	Gender	Education	Medium of instr.	Occupation	Other languages known	Type of data collected
Barpeta	BF01	35	F	HSLC	Assamese		Hindi	Wordlist and passage
	BF02	28	F	MA	Assamese		Hindi	Wordlist and passage
	BF03	32	F	BA	Assamese	Service	Hindi, English	Wordlist and passage
	BF04	26	F	MA	Assamese		Hindi, English	Wordlist and passage
	BF05	26	F	MA	Assamese	Teacher	Hindi, English	Wordlist and passage
	BF06	39	F	BA	Assamese		None	Wordlist and passage
	BM01	20	M	HS	Assamese	Teacher	Hindi, English	Wordlist and passage
	BM02	20	M	HSLC	Assamese		Hindi, English	Wordlist and passage
	BM03	27	M	HSLC	Assamese	Business	Hindi	Wordlist and passage
	BM04	24	M	BA	Assamese		Hindi	Wordlist and passage
Jorhat	JF05	36	F	BA	Assamese		Hindi	Only wordlist
	JF06	27	F	MA	Assamese	Student	Hindi, English	Only wordlist
	JF07	38	F	BA	Assamese	Housewife	None	Only wordlist
	JF08	40	F	HS	Assamese	Service	Hindi, English, Bengali	Only wordlist
	JF09	34	F	HS	Assamese	Housewife	None	Only wordlist
	JM08	45	M	BA	Assamese	Service	Hindi	Only wordlist
	JM09	30	M	HS	Assamese	Service	Hindi, Bengali	Only wordlist
	JM10	57	M	HS	Assamese	Village headman	Bengali	Only wordlist
	JM11	30	M	BA	Assamese	Service	Hindi, Bengali	Wordlist and passage
	JM12	32	M	BA	Assamese	Business	Hindi	Wordlist and passage
	JF10	53	F	BA	Assamese	Housewife		Only passage
	JF11	38	F	BA	Assamese	Service	Hindi, Bengali	Only passage
	JM13	46	M	BA	Assamese		Hindi	Only passage
Nagaon	GF01	30	F	BA	Assamese		Hindi	Wordlist and passage
	GF02	35	F	HSLC	Assamese		None	Wordlist and passage
	GF03	30	F	HS	Assamese	Teacher	None	Wordlist and passage
	GF04	25	F	MA	Assamese		Hindi, English	Wordlist and passage
	GF05	48	F	HS	Assamese		None	Wordlist and passage
	GM01	70	M	HS	Assamese	Teacher	None	Only wordlist
	GM02	29	M	MA	Assamese	Service	Hindi, English	Wordlist and passage
	GM03	40	M	HS	Assamese	Driver	None	Wordlist and passage
	GM04	25	M	MA	Assamese		Hindi, English	Wordlist and passage
	GM05	28	M	BA	Assamese		Hindi	Wordlist and passage
	GM06	39	M	HS	Assamese	Farmer		Only passage
Nalbari	NF04	36	F	HS	Assamese		Hindi	Only Wordlist
	NF05	28	F	MA	Assamese	Teacher	Hindi,	Only Wordlist
	NF06	25	F	BA	English	Student	Hindi, English	Only Wordlist
	NF07	40	F	VIII	Assamese	Housewife	None	Only Wordlist
	NF08	25	F	MSc	Assamese		Hindi,English	Wordlist and passage
	NF10	26	F	MA	Assamese	Farmer	None	Only Wordlist
	NM08	31	M	HS	Assamese	Business	Hindi	Only Wordlist
	NM09	45	M	BSc	Assamese	Business	Hindi, English	Wordlist and passage
	NM10	45	M	MA	Assamese	Teacher	None	Wordlist and passage
	NM11	42	M	HS	Assamese	Service	Hindi	Only passage
Tinsukia	TF01	45	F	BA	Assamese		Hindi, Bengali	Wordlist and passage
	TF02	43	F	HSLC	Assamese	Housewife	None	Wordlist and passage
	TF03	27	F	BA	Assamese	Housewife	Hindi, English, Bengali	Wordlist and passage
	TF04	48	F	HS	Assamese	Housewife	None	Wordlist and passage
	TF05	26	F	BA	Assamese	Teacher	Hindi, English	Wordlist and passage
	TM01	25	M	BCom	Assamese	Service	Hindi	Wordlist and passage
	TM02	49	M	HSLC	Assamese	Service	Hindi, Bengali	Wordlist and passage
	TM03	21	M	BCom	Assamese		Hindi, English, Bengali	Wordlist and passage
	TM04	72	M	HSLC	Assamese	Service	English, Hindi	Wordlist and passage
	TM05	29	M	BA	Assamese	Service	Hindi, Bengali	Wordlist and passage

Table A.3: Details of participants for perception tests (Phase I)

Region	Participant	Age	Gender	Medium of inst.	Education	Occupation
Barpeta	B01	25	M	Assamese	MA	Student
	B02	27	M	Assamese	MA	Student
	B03	30	F	Assamese	MA	Student
Jorhat	J01	27	F	English	MA	Student
	J02	28	F	Assamese	MA	Student
	J03	29	F	Assamese	MA	Student
Nagaon	G01	24	M	Assamese	B.Tech	Student
	G02	27	M	Assamese	BA	Business
	G03	32	M	Assamese	PhD	Service
Nalbari	N01	28	F	Assamese	MA	Service
	N02	33	M	Assamese	PhD	Student
	N03	47	M	Assamese	MA	Service
Tinsukia	T01	25	M	English	B.Tech	Student
	T02	27	F	Assamese	B.Tech	Service
	T03	40	M	Assamese	MA	Service

Table A.4: Details of participants for perception tests (Phase II)

Region	Name	Age group	Gender	Educational	Medium of instr.
Barpeta	B04	15-20	Male	Bachelor's level	English
	B06	31-35	Female	PhD level	Assamese
	B06	31-35	Male	Master's level	Assamese
	B07	21-25	Male	Master's level	Assamese
	B08	26-30	Female	PhD level	Assamese
	B09	36-40	Male	Master's level	Assamese
	B10	26-30	Male	Master's level	English
	B11	26-30	Male	Master's level	Assamese
Jorhat	J03	15-20	Female	Bachelor's level	English
	J05	26-30	Male	Master's level	Assamese
	J06	30-35	Female	Bachelor's level	Assamese
	J07	31-35	Female	PhD level	Assamese
	J08	21-25	Female	Master's level	Assamese
	J09	31-35	Male	PhD level	Assamese
	J10	26-30	Female	Master's level	English
	J11	Above 50	Female	Bachelor's level	Assamese
	J12	26-30	Female	Master's level	English
Nagaon	G04	31-35	Female	PhD level	Assamese
	G05	26-30	Female	Master's level	Assamese
	G06	21-25	Female	Master's level	Assamese
	G07	21-25	Female	Master's level	Assamese
	G08	31-35	Male	Master's level	Assamese
	G09	21-25	Female	Master's level	Assamese
Nalbari	N04	25-30	Female	Master's level	Assamese
	N05	21-25	Male	Bachelor's level	Assamese
	N06	30-35	Female	PhD level	Assamese
	N07	26-30	Female	PhD level	Assamese
	N08	25-30	Male	Master's level	English
	N09	21-25	Female	Master's level	Assamese
Tinsukia	T04	36-40	Male	PhD level	English
	T05	21-25	Male	Bachelor's level	English
	T06	36-40	Male	Master's level	Assamese
	T07	31-35	Female	PhD level	English
	T08	21-25	Male	Master's level	Assamese
	T09	31-35	Male	PhD level	English
	T10	31-35	Male	PhD level	Assamese

Appendix B

Complete datalist

B.1 Word list

Table B.1: Complete list of Assamese words elicited

Sl.no.	Assamese word (IPA)	English meaning	Sl.no.	Assamese word (IPA)	English meaning
1	bal	a male child	38	k^hur	razor
2	bɛl	woodapple	39	k^hʊl	drum
3	bel	bell	40	mas	fish
4	beli	sun	41	mɛz	table
5	b^hal	good	42	mɔn	mind
6	b^haz	fry	43	mɔr	die
7	b^hɛl	raft	44	mɔs	wipe
8	b^hiz	to get wet	45	mɔz	enjoy
9	b^hɔr	fill	46	mud	close eyes
10	b^hul	wrong	47	mur	head
11	b^hur	raft	48	mʊn	unit of weight
12	b^hʊl	vegetable	49	mʊr	mine
13	b^hʊz	party	50	mʊs	moustache
14	bil	pond	51	pɔn	oath
15	bol	let's go	52	pʊn	straight
16	bɔl	strength	53	sɔk	fright
17	bɔr	big	54	tal	cymbal
18	bul	name	55	tap	heat
19	bʊl	color	56	tɛl	oil
20	gal	cheek	57	til	sesame
21	gat	pit	58	tɔl	below
22	g^hɔr	house	59	tɔp	meditate
23	g^hur	go round	60	tup	bait
24	g^hʊr	very	61	tʊl	pick
25	gol	gone	62	tʊp	drop
26	gɔl	throat	63	xat	seven
27	gɔr	rampart	64	xaz	meal
28	gɔt	sort	65	xol	type of fish
29	gul[1]	mix	66	xɔt	honest
30	gʊl	round	67	xul	thorn
31	gʊt	group	68	xut	loan
32	hol	done	69	zak	group
33	hul	thorn	70	zap	jump
34	kah	cough	71	zɔg	jug
35	k^hal	ditch	72	zug	time scale
36	k^hɛl	game	73	zʊg	addition
37	k^hɛr	hay			

B.2 Passage for reading

uttɔra bɔtah aru belir mazɔt karnʊ besi bɔl buli kazija lagil. ɛnɛtɛ ʊmal sʊla pind^{h}a batoruwa ɛzɔn xeik^{h}ini palehi. uttɔra bɔtah aru belije thik korilɛ zɛ teulʊkɔr mazɔt zizonei batoruwazɔnɔr gar pɔra ʊmal sʊlatʊ atɔrabɔ paribɔ, teukei besi boli bʊla hobɔ. takei buli uttɔra bɔtahɛ zʊrɛrɛ boliboloi arɔmb^{h}ɔ korilɛ. bɔtah zimanei besi hol, batoruwazɔnɛ ximanei zʊrɛrɛ sʊlatʊ k^{h}amusi d^{h}ori t^{h}akil. ɔbɔxɛxɔt bɔtahɛ nizɔr pɔrazɔi mani lole. tar pasɔr bɛla, belije nizɔr tap bɔrhaboloi lolɛ. tɛtija belir tap aru d^{h}uli balit batoruwazɔnɔr bilai nʊhʊwa hol. teu gɔrɔmɔtɛ gar ʊmal sʊlatʊ k^{h}uli pɛlalɛ. takɛ dek^{h}i uttɔra bɔtahɛ mani lolɛ zɛ dujʊre mazɔt belihɛ besi boli.

Appendix C

Miscellaneous

C.1 Results of Bonferroni post-hoc tests for F3

Table C.1: Results of Bonferroni post-hoc tests for F3 conducted on vowels produced by Tinsukia speakers.

Contrasts	Isolation					Sentence				
	estimate	SE	df	t-ratio	p-value	estimate	SE	df	t-ratio	p-value
a-e	-0.79	0.22	869	-3.6	0.0096	-0.32	0.21	778	-1.53	1
a-ɛ	-0.6	0.11	867	-5.29	< .0001	-0.41	0.11	764	-3.75	0.0054
a-i	-1.02	0.17	995	-6.09	< .0001	-0.56	0.16	945	-3.5	0.0137
a-o	0.77	0.12	840	6.28	< .0001	0.85	0.12	732	7.06	< .0001
a-ɔ	0.97	0.08	990	11.89	< .0001	0.99	0.08	828	12.39	< .0001
a-u	0.72	0.08	1075	8.56	< .0001	0.87	0.09	766	10.14	< .0001
a-ʊ	0.78	0.08	1026	9.42	< .0001	0.81	0.08	956	10.06	< .0001
e-ɛ	0.19	0.22	1035	0.84	1	-0.09	0.21	931	-0.45	1
e-i	-0.24	0.25	1183	-0.97	1	-0.25	0.23	1169	-1.06	1
e-o	1.55	0.23	1044	6.83	< .0001	1.16	0.21	872	5.41	< .0001
e-ɔ	1.75	0.21	927	8.19	< .0001	1.31	0.2	927	6.45	< .0001
e-u	1.51	0.21	947	7.02	< .0001	1.18	0.2	888	5.81	< .0001
e-ʊ	1.57	0.21	945	7.32	< .0001	1.13	0.2	885	5.57	< .0001
ɛ-i	-0.43	0.18	1121	-2.4	0.4629	-0.15	0.17	1092	-0.92	1
ɛ-o	1.37	0.14	759	9.61	< .0001	1.26	0.14	466	9.21	< .0001
ɛ-ɔ	1.56	0.11	793	14.4	< .0001	1.4	0.11	727	13.38	< .0001
ɛ-u	1.32	0.11	921	12	< .0001	1.28	0.11	637	11.87	< .0001
ɛ-ʊ	1.38	0.11	956	12.78	< .0001	1.23	0.1	862	11.79	< .0001
i-o	1.79	0.19	1057	9.69	< .0001	1.41	0.18	788	8.04	< .0001
i-ɔ	1.99	0.16	1041	12.15	< .0001	1.56	0.16	1102	9.92	< .0001
i-u	1.75	0.17	1040	10.53	< .0001	1.43	0.16	972	9	< .0001
i-ʊ	1.81	0.16	1067	11.02	< .0001	1.38	0.16	1063	8.8	< .0001
o-ɔ	0.2	0.12	812	1.69	1	0.15	0.12	813	1.27	1
o-u	-0.04	0.12	1063	-0.38	1	0.02	0.11	1081	0.17	1
o-ʊ	0.02	0.12	760	0.14	1	-0.03	0.12	627	-0.26	1
ɔ-u	-0.24	0.08	1047	-3.11	0.0543	-0.13	0.08	956	-1.68	1
ɔ-ʊ	-0.18	0.07	1171	-2.44	0.4185	-0.18	0.07	1174	-2.52	0.3304
u-ʊ	0.06	0.08	1120	0.77	1	-0.05	0.08	940	-0.64	1

Table C.2: Results of Bonferroni post-hoc tests for F3 conducted on vowels produced byJorhat speakers.

Contrasts	Isolation					Sentence				
	estimate	SE	df	t-ratio	p-value	estimate	SE	df	t-ratio	p-value
a-e	-0.56	0.2	1119	-2.77	0.1617	-0.19	0.15	1211	-1.27	1
a-ɛ	0.56	0.1	1120	5.47	< .0001	-0.27	0.09	1186	-2.82	0.1389
a-i	-1.15	0.16	1119	-7.44	< .0001	-0.54	0.11	1374	-4.77	0.0001
a-o	0.73	0.12	1119	6.23	< .0001	0.3	0.1	1201	2.9	0.1082
a-ɔ	-1.06	0.08	1120	-13.28	< .0001	0.67	0.07	1273	8.92	< .0001
a-u	0.69	0.08	1119	8.18	< .0001	0.27	0.08	1518	3.55	0.0112
a-ʊ	-0.75	0.08	1120	-9.15	< .0001	0.31	0.08	1375	4.06	0.0015
e-ɛ	0	0.21	1119	0	1	-0.08	0.16	1387	-0.49	1
e-i	-0.6	0.24	1119	-2.49	0.358	-0.36	0.17	1592	-2.14	0.9048
e-o	1.29	0.22	1119	5.97	< .0001	0.48	0.16	1436	3.03	0.0689
e-ɔ	-1.62	0.2	1119	-8.19	< .0001	0.85	0.15	1152	5.77	< .0001
e-u	1.25	0.2	1119	6.24	< .0001	0.46	0.15	1264	3.11	0.0541
e-ʊ	-1.31	0.2	1119	-6.59	< .0001	0.5	0.15	1223	3.36	0.0224
ɛ-i	-0.6	0.16	1120	-3.65	0.0076	-0.28	0.12	1492	-2.23	0.7219
ɛ-o	1.29	0.13	1119	10.13	< .0001	0.56	0.12	916	4.82	< .0001
ɛ-ɔ	1.62	0.09	1120	17.24	< .0001	0.93	0.09	936	9.98	< .0001
ɛ-u	1.25	0.1	1120	12.78	< .0001	0.54	0.09	1282	5.71	< .0001
ɛ-ʊ	1.31	0.1	1119	13.67	< .0001	0.58	0.09	1335	6.17	< .0001
i-o	1.89	0.17	1119	10.91	< .0001	0.84	0.13	1275	6.38	< .0001
i-ɔ	-2.22	0.15	1120	-14.78	< .0001	1.21	0.11	1228	10.67	< .0001
i-u	1.84	0.15	1119	12.1	< .0001	0.82	0.11	1388	7.15	< .0001
i-ʊ	-1.91	0.15	1120	-12.61	< .0001	0.85	0.11	1408	7.52	< .0001
o-ɔ	-0.33	0.11	1119	-3.02	0.072	0.37	0.1	1075	3.69	0.0067
o-u	-0.04	0.11	1119	-0.39	1	-0.02	0.1	1371	-0.23	1
o-ʊ	-0.02	0.11	1119	-0.2	1	0.01	0.1	985	0.14	1
ɔ-u	-0.38	0.07	1120	-5.09	< .0001	-0.39	0.07	1368	-5.28	< .0001
ɔ-ʊ	-0.31	0.07	1119	-4.33	0.0005	-0.36	0.07	1589	-4.9	< .0001
u-ʊ	-0.07	0.08	1119	-0.86	1	0.04	0.08	1476	0.49	1

Table C.3: Results of Bonferroni post-hoc tests for F3 conducted on vowels produced by Nagaon speakers.

Contrasts	Isolation					Sentence				
	estimate	SE	df	t-ratio	p-value	estimate	SE	df	t-ratio	p-value
a-e	-0.26	0.19	1306	-1.37	1	-0.17	0.21	1017	-0.84	1
a-ɛ	-0.17	0.09	1306	-1.93	1	-0.18	0.1	1012	-1.82	1
a-i	-0.61	0.14	1306	-4.52	0.0002	-0.39	0.15	1149	-2.52	0.3316
a-o	1.23	0.1	1305	11.91	< .0001	0.89	0.12	927	7.6	< .0001
a-ɔ	1.15	0.07	1305	16.82	< .0001	1.18	0.08	1070	15.69	< .0001
a-u	1.08	0.07	1305	15.08	< .0001	0.91	0.08	1199	11.56	< .0001
a-ʊ	1.03	0.07	1306	14.66	< .0001	0.91	0.08	1131	11.73	< .0001
e-ɛ	0.09	0.2	1306	0.47	1	-0.01	0.21	1099	-0.04	1
e-i	-0.35	0.23	1306	-1.55	1	-0.21	0.24	1286	-0.9	1
e-o	1.5	0.21	1306	7.24	< .0001	1.07	0.22	1153	4.85	< .0001
e-ɔ	1.41	0.19	1306	7.36	< .0001	1.35	0.21	1005	6.61	< .0001
e-u	1.35	0.19	1306	6.97	< .0001	1.09	0.21	1041	5.27	< .0001
e-ʊ	1.29	0.19	1306	6.72	< .0001	1.08	0.21	1025	5.26	< .0001
ɛ-i	-0.44	0.15	1306	-3.04	0.0667	-0.2	0.16	1195	-1.25	1
ɛ-o	1.41	0.12	1305	12.12	< .0001	1.07	0.13	775	8.03	< .0001
ɛ-ɔ	1.32	0.09	1305	15.36	< .0001	1.36	0.1	839	13.97	< .0001
ɛ-u	1.25	0.09	1305	14.13	< .0001	1.09	0.1	1085	11.06	< .0001
ɛ-ʊ	1.2	0.09	1305	13.73	< .0001	1.09	0.1	1115	11.23	< .0001
i-o	1.85	0.15	1305	11.94	< .0001	1.28	0.17	1117	7.32	< .0001
i-ɔ	1.76	0.13	1305	13.17	< .0001	1.57	0.15	1120	10.37	< .0001
i-u	1.7	0.14	1306	12.5	< .0001	1.3	0.15	1123	8.46	< .0001
i-ʊ	1.64	0.13	1305	12.19	< .0001	1.29	0.15	1158	8.52	< .0001
o-ɔ	-0.09	0.1	1305	-0.87	1	0.29	0.11	852	2.52	0.3377
o-u	-0.15	0.1	1305	-1.49	1	0.02	0.12	1129	0.18	1
o-ʊ	-0.21	0.1	1305	-2.01	1	0.02	0.12	792	0.13	1
ɔ-u	-0.07	0.07	1305	-0.98	1	-0.27	0.07	1124	-3.58	0.0101
ɔ-ʊ	-0.12	0.07	1305	-1.79	1	-0.27	0.07	1276	-3.85	0.0035
u-ʊ	-0.05	0.07	1306	-0.74	1	-0.01	0.08	1203	-0.08	1

Table C.4: Results of Bonferroni post-hoc tests for F3 conducted on vowels produced by Nalbari speakers.

Contrasts	Isolation					Sentence				
	estimate	SE	df	t-ratio	p-value	estimate	SE	df	t-ratio	p-value
a-e	-0.39	0.12	1831	-3.32	0.0258	-0.3	0.12	1585	-2.45	0.4013
a-ɛ	-0.61	0.08	1822	-7.58	< .0001	-0.39	0.08	1717	-4.98	< .0001
a-i	-1.09	0.09	1845	-11.9	< .0001	-0.65	0.09	1840	-7.06	< .0001
a-o	0.66	0.09	1749	7.54	< .0001	0.93	0.09	1362	10.76	< .0001
a-ɔ	0.92	0.07	1773	13.91	< .0001	0.88	0.06	1727	13.81	< .0001
a-u	0.62	0.07	1842	9.04	< .0001	0.68	0.07	1970	10.07	< .0001
a-ʊ	0.73	0.07	1811	11.05	< .0001	0.75	0.06	1822	11.57	< .0001
e-ɛ	-0.22	0.12	1853	-1.84	1	-0.09	0.13	1849	-0.72	1
e-i	-0.71	0.13	1851	-5.61	< .0001	-0.35	0.13	2026	-2.62	0.2455
e-o	1.05	0.12	1849	8.45	< .0001	1.24	0.13	1842	9.46	< .0001
e-ɔ	1.31	0.11	1837	11.36	< .0001	1.18	0.12	1636	9.77	< .0001
e-u	1.01	0.12	1842	8.59	< .0001	0.98	0.12	1604	7.9	< .0001
e-ʊ	1.12	0.12	1843	9.76	< .0001	1.05	0.12	1681	8.66	< .0001
ɛ-i	-0.49	0.1	1854	-4.9	< .0001	-0.26	0.1	1970	-2.59	0.2676
ɛ-o	1.27	0.1	1768	13.04	< .0001	1.33	0.1	1281	13.73	< .0001
ɛ-ɔ	1.53	0.08	1796	19.33	< .0001	1.28	0.08	1518	16.54	< .0001
ɛ-u	1.23	0.08	1841	15.16	< .0001	1.07	0.08	1754	13.39	< .0001
ɛ-ʊ	1.34	0.08	1844	17.15	< .0001	1.14	0.08	1825	14.85	< .0001
i-o	1.76	0.11	1833	16.63	< .0001	1.59	0.11	1682	14.91	< .0001
i-ɔ	2.01	0.09	1843	22.22	< .0001	1.53	0.09	1796	16.93	< .0001
i-u	1.71	0.09	1850	18.42	< .0001	1.33	0.09	1785	14.17	< .0001
i-ʊ	1.83	0.09	1853	20.25	< .0001	1.4	0.09	1918	15.46	< .0001
o-ɔ	0.25	0.09	1759	2.95	0.0914	-0.05	0.08	1389	-0.62	1
o-u	-0.05	0.09	1841	-0.51	1	-0.25	0.09	1748	-2.93	0.0973
o-ʊ	0.07	0.09	1729	0.8	1	-0.18	0.09	1212	-2.14	0.9182
ɔ-u	-0.3	0.07	1835	-4.48	0.0002	-0.2	0.07	1859	-3.11	0.054
ɔ-ʊ	-0.18	0.06	1854	-2.92	0.0998	-0.13	0.06	2038	-2.17	0.8445
u-ʊ	0.12	0.07	1844	1.71	1	0.07	0.07	1869	1.06	1

Table C.5: Results of Bonferroni post-hoc tests for F3 conducted on vowels produced by Barpeta speakers.

Contrasts	Isolation					Sentence				
	estimate	SE	df	t-ratio	p-value	estimate	SE	df	t-ratio	p-value
a-e	-0.56	0.2	1119	-2.77	0.1617	-0.73	0.21	949	-3.47	0.0155
a-ɛ	0.56	0.1	1120	5.47	< .0001	-0.61	0.1	1067	-5.85	< .0001
a-i	-1.15	0.16	1119	-7.44	< .0001	-1	0.16	1067	-6.3	< .0001
a-o	0.73	0.12	1119	6.23	< .0001	1.03	0.12	873	8.65	< .0001
a-ɔ	-1.06	0.08	1120	-13.28	< .0001	1.04	0.08	1073	12.89	< .0001
a-u	0.69	0.08	1119	8.18	< .0001	0.46	0.08	1136	5.57	< .0001
a-ʊ	-0.75	0.08	1120	-9.15	< .0001	0.85	0.08	1100	10.26	< .0001
e-ɛ	0	0.21	1119	0	1	0.12	0.21	1044	0.58	1
e-i	-0.6	0.24	1119	-2.49	0.358	-0.27	0.24	1146	-1.13	1
e-o	1.29	0.22	1119	5.97	< .0001	1.76	0.22	1052	8.07	< .0001
e-ɔ	-1.62	0.2	1119	-8.19	< .0001	1.77	0.21	972	8.61	< .0001
e-u	1.25	0.2	1119	6.24	< .0001	1.19	0.21	999	5.78	< .0001
e-ʊ	-1.31	0.2	1119	-6.59	< .0001	1.58	0.21	986	7.66	< .0001
ɛ-i	-0.6	0.16	1120	-3.65	0.0076	-0.39	0.16	1110	-2.37	0.5086
ɛ-o	1.29	0.13	1119	10.13	< .0001	1.64	0.13	773	12.4	< .0001
ɛ-ɔ	1.62	0.09	1120	17.24	< .0001	1.65	0.1	894	16.85	< .0001
ɛ-u	1.25	0.1	1120	12.78	< .0001	1.07	0.1	1041	10.87	< .0001
ɛ-ʊ	1.31	0.1	1119	13.67	< .0001	1.45	0.1	1057	14.98	< .0001
i-o	1.89	0.17	1119	10.91	< .0001	2.03	0.17	1037	11.62	< .0001
i-ɔ	-2.22	0.15	1120	-14.78	< .0001	2.03	0.15	1059	13.3	< .0001
i-u	1.84	0.15	1119	12.1	< .0001	1.45	0.15	1068	9.42	< .0001
i-ʊ	-1.91	0.15	1120	-12.61	< .0001	1.84	0.15	1077	12	< .0001
o-ɔ	-0.33	0.11	1119	-3.02	0.072	0.01	0.11	760	0.06	1
o-u	-0.04	0.11	1119	-0.39	1	-0.57	0.11	1011	-5.1	< .0001
o-ʊ	-0.02	0.11	1119	-0.2	1	-0.19	0.12	730	-1.61	1
ɔ-u	-0.38	0.07	1120	-5.09	< .0001	-0.58	0.07	1039	-7.93	< .0001
ɔ-ʊ	-0.31	0.07	1119	-4.33	0.0005	-0.19	0.07	1139	-2.73	0.181
u-ʊ	-0.07	0.08	1119	-0.86	1	0.39	0.07	1102	5.19	< .0001

C.2 Responses for identification tests

Table C.6: Responses for identification of front vowels

Stimuli	Region	Responses					
		/i/	/e/	/e, ɛ/	/ɛ/	None	Total
/i/	Barpeta	66					66
	Jorhat	72					72
	Nagaon	54					54
	Nalbari	54					54
	Tinsukia	60					60
/e/	Barpeta		31	16	17	2	66
	Jorhat		57	4	11	2	72
	Nagaon		39	9	6		54
	Nalbari		32	19	3		54
	Tinsukia	1	42	6	11		60
/ɛ/	Barpeta		5	13	47	1	66
	Jorhat		4	3	62	3	72
	Nagaon		7	5	42		54
	Nalbari			11	43		54
	Tinsukia		17	7	35	1	60

Table C.7: Responses for identification of back vowels

Stimuli	Region	Responses											
		/u/	/u, ʊ/	/ʊ/	/ʊ, o/	/o/	/o, ɔ/	/ɔ/	/u, o/	/u, ɔ/	/ʊ, o, ɔ/	none	Total
/u/	Barpeta	92	10	21	1	4						4	132
	Jorhat	51	35	53		1			1			3	144
	Nagaon	15	46	46								1	108
	Nalbari	87	9	12									108
	Tinsukia	40	15	58	2	1		1	1			2	120
/ʊ/	Barpeta	20	3	65	16	24					1	3	132
	Jorhat	25	28	71		16						4	144
	Nagaon	19	25	45		19							108
	Nalbari	13		61	27	6		1					108
	Tinsukia	30	8	46	1	30		1	3	1			120
/o/	Barpeta	3		19	30	76	2	2					132
	Jorhat	7		12		108		6	8			3	144
	Nagaon	1	1	1		84	1	11	3			6	108
	Nalbari			24	32	48		4					108
	Tinsukia	7	1	13		89		9	1				120
/ɔ/	Barpeta	3		1		10	14	94		1	1	8	132
	Jorhat					5	10	124				5	144
	Nagaon	1				5	11	86	1			4	108
	Nalbari					4	20	84					108
	Tinsukia	2			1	15	1	101					120

References

Abercrombie, D. (1967). *Elements of general phonetics* (Vol. 203). Edinburgh University Press Edinburgh.

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

Adank, P., Van Hout, R., & Smits, R. (2004). An acoustic description of the 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 *Ninth European Conference on Speech Communication and Technology.*

Archangeli, D. B., & Yip, J. (2020). Assamese vowels and vowel harmony. *Journal of South Asian Languages and Linguistics*, *6*(2), 151–183.

Arvaniti, A. (2012). The usefulness of metrics in the quantification of speech rhythm. *Journal of Phonetics*, *40*(3), 351–373.

Baranowski, M. A. (2006). Phonological variation and change in the dialect of charleston, south carolina. *Publication of the American Dialect Society*, *92*.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48. doi: 10.18637/jss.v067.i01

Baxter, L. (2010). Lexical diffusion in the early stages of the merry-marry merger. *University of Pennsylvania Working Papers in Linguistics*, *16*(2), 3.

Biadsy, F., Hirschberg, J., & Habash, N. (2009). Spoken Arabic dialect identification using phonotactic modeling. In *Proceedings of the EACL 2009 workshop on computational approaches to Semitic languages* (pp. 53–61).

Biadsy, F., & Hirschberg, J. B. (2009). Using prosody and phonotactics in Arabic dialect identification. In *Proceedings of Interspeech'09.*

Boersma, P., & Weenink, D. (2014). *Praat: doing phonetics by computer (version 5.4).[logiciel, en ligne].*

Bradlow, A. R. (1995). A comparative acoustic study of English and Spanish vowels. *The Journal of the Acoustical Society of America*, *97*(3), 1916–1924.

Brubaker, R. S., & Altshuler, M. W. (1959). Vowel overlap as a function of fundamental frequency and dialect. *The Journal of the Acoustical Society of America*, *31*(10), 1362–1365.

Carré, R. (1996). Prediction of vowel systems using a deductive approach. In *Proceeding of Fourth International Conference on Spoken Language Processing. ICSLP'96* (Vol. 3, pp. 1593–1596).

Chambers, J. K., & Trudgill, P. (1998). *Dialectology*. Cambridge University Press.

Chatterji, S. K. (1970). *The origin and development of the Bengali language* (Vol. 3). London:George Allen and Unwin.

Chiba, T., & Kajiyama, M. (1958). *The vowel: Its nature and structure* (Vol. 652). Phonetic society of Japan Tokyo.

Chittaragi, N. B., & Koolagudi, S. G. (2019). Acoustic-phonetic feature based Kannada dialect identification from vowel sounds. *International Journal of Speech Technology*, *22*(4), 1099–1113.

Choi, H. (2002). Acoustic cues for the Korean stop contrast-dialectal variation. *ZAS papers in Linguistics*, *28*, 1–12.

Choi, J. D. (1992). *Phonetic underspecification and target interpolation: An acoustic study of Marshallese vowel allophony* (Unpublished doctoral dissertation). Phonetics Laboratory, Department of Linguistics, UCLA.

Clopper, C. G., Pisoni, D. B., & De Jong, K. (2005). Acoustic characteristics of the vowel systems of six regional varieties of American English. *The Journal of the Acoustical Society of America*, *118*(3), 1661–1676.

Commons, W. (2020). *File:indoarische sprachen gruppen.png — wikimedia commons, the free media repository.* Retrieved from `https://commons.wikimedia.org/w/index.php?title=File:Indoarische_Sprachen_Gruppen.png&oldid=456291916` ([Online; accessed 16-March-2021])

Cunningham-Andersson, U., & Engstrand, O. (1988). IRIS- A data base for cross-linguistic phonetic research. *Unpublished manuscript, Department of linguistics, Uppsala Universitet*.

Daniel McCloy. (2020). *R Code to find the overlap area of the convex hulls of vowel measurements.* `https://gist.github.com/drammock/db05be8e03a4f09e5a14973a7a6a5c2c`. GitHub.

Das, B. (2010). Asamiya bhasha. In B. Das & P. C. Basumatary (Eds.), *Axamiya aru Axamar Bhasa* (pp. 13–33). Guwahati.

Dauer, R. M. (1983). Stress-timing and syllable-timing reanalyzed. *Journal of phonetics*.

Deka, K. (2007). Kamrupi upabhaxa. In D. P. Patgiri (Ed.), *Asamiya bhasar upabhasa.* University Publication Department, Gauhati University, Guwahati.

Dellwo, V. (2006). Rhythm and speech rate: A variation coefficient for δC. *Language and language-processing*, 231–241.

Dellwo, V., & Wagner, P. (2003). Relations between language rhythm and speech rate. In *Proceedings of the 15th International Congress of Phonetics Sciences* (pp. 471–474).

Deo, A. (2018). Dialects in the Indo-Aryan landscape. *The Handbook of Dialectology*, 535–546.

Deterding, D. (1997). The formants of monophthong vowels in Standard Southern British English pronunciation. *Journal of the International Phonetic Association*, *27*(1-2), 47–55.

Devi, T. C., & Thaoroijam, K. (2020). Vowel-based Meeteilon dialect identification using a Random Forest classifier. In *Proceedings of O-COCOSDA 2020.*

Dihingia, L., & Sarmah, P. (2017a). Vowel perception in two Assamese varieties. In *Proceedings of the Seoul International Conference of Speech Sciences.*

Dihingia, L., & Sarmah, P. (2017b). *Vowel space in Assamese dialects* [Paper presented at the 23rd Himalayan Languages Symposium, Tezpur].

Dimitriadou, E., Hornik, K., Leisch, F., Meyer, D., & Weingessel, A. (2008). Misc functions of the department of statistics (e1071), tu wien. *R package*, *1*, 5–24.

Dinkin, A. J. (2009). *Dialect boundaries and phonological change in Upstate New York* (Unpublished doctoral dissertation). University of Pennsylvania.

Dinkin, A. J. (2016). Phonological transfer as a forerunner of merger in upstate new york. *Journal of English Linguistics*, *44*(2), 162–188.

Disner, S. F. (1983). *Vowel quality: The relation between universal and language specific factors* (Vol. 58). Phonetics Laboratory, Department of Linguistics, UCLA.

Disner, S. F. (1984). Insights on vowel spacing. *Patterns of sounds*, 136–155.

Donegan, P. J., & Stampe, D. (1983). *Rhythm and the holistic organization of language structure.* Chicago Linguistic Society, University of Chicago.

Escudero, P., Boersma, P., Rauber, A. S., & Bion, R. A. (2009). A cross-dialect acoustic description of vowels: Brazilian and European Portuguese. *The Journal of the Acoustical Society of America*, *126*(3), 1379–1393.

Fant, G. (1970). *Acoustic theory of speech production* (No. 2). Walter de Gruyter.

Ferragne, E., & Pellegrino, F. (2007). Automatic dialect identification: A study of British English. In *Speaker classification II* (pp. 243–257). Springer.

Fox, J., & Weisberg, S. (2019). *An R companion to applied regression* (Third ed.). Thousand Oaks CA: Sage. Retrieved from `https://socialsciences.mcmaster.ca/jfox/Books/Companion/`

Fox, R. A., & Jacewicz, E. (2009). Cross-dialectal variation in formant dynamics of American English vowels. *The Journal of the Acoustical Society of America*, *126*(5), 2603–2618.

Fridland, V., Kendall, T., & Farrington, C. (2014). Durational and spectral differences in American English vowels: Dialect variation within and across regions. *The Journal of the Acoustical Society of America*, *136*(1), 341–349.

Gendrot, C., & Adda-Decker, M. (2007). Impact of duration and vowel inventory size on formant values of oral vowels: an automated formant analysis from eight languages. In *Proceedings of the 16th International Congress of Phonetic Sciences* (pp. 1417–1420).

Ghazali, S., Hamdi, R., & Barkat, M. (2002). Speech rhythm variation in Arabic dialects. In *Speech Prosody 2002, International Conference.*

Goswami, G. C. (1966). *An introduction to Assamese phonology.* Deccan College Postgraduate and Research Institute.

Goswami, G. C. (1982). *Structure of Assamese.* Gauhati University.

Goswami, G. C., & Tamuli, J. (2003). Asamiya. In D. Jain & G. Cardona (Eds.), *The Indo-Aryan languages* (pp. 391–443). Routledge.

Goswami, U. (1970a). *A study on kamrupi: a dialect of assamese.* Department of Historical and Antiquarian Studies, Assam.

Goswami, U. (1970b). *A study on Kamrupi, a variety of Assamese.* Gauhati University.

Goswami, U. (1988). Banikanta Kakati's contribution to dialectological studies in Assamese. In T. Taid & R. D. Goswami (Eds.), *Banikanta Kakati, the man and his works* (pp. 113–120). Publication Board Assam.

Goswami, U. (1991). *Asamiya bhashar udbhav samriddhi aru vikash.* Gauhati University.

Grabe, E., & Low, E. L. (2002). Durational variability in speech and the rhythm class hypothesis. *Papers in laboratory phonology*, *7*(515-546).

Grenon, I., & White, L. (2008). Acquiring rhythm: A comparison of L1 and L2 speakers of Canadian English and Japanese. In *Proceedings of the 32nd Boston University Conference on Language Development* (pp. 155–166).

Grierson, G. A. (1903). *Linguistic Survey of India: Vol. v. Indo-Aryan Family, Eastern Group; Pt. I. Specimens of the Bengali and Assamese languages*. Office of the superintendent of government printing, India, Reprinted 1967 by Motilal Banarsidass.

Grierson, G. A. (1903-1928). *Linguistic survey of India* (Vol. I-XI). Office of the superintendent of government printing, India, Culcutta, Reprinted 1967 by Motilal Banarsidass.

Guy, G. R. (1990). The sociolinguistic types of language change. *Diachronica*, *7*(1), 47–67.

Hagiwara, R. (1997). Dialect variation and formant frequency: The American English vowels revisited. *The Journal of the Acoustical Society of America*, *102*(1), 655–658.

Hakacham, U. R. (2000). *Asamiya aru Asamar Tibbot-Bormiya bhasa*. Manjula Rabha Hakacham.

Hall-Lew, L. (2009). *Ethnicity and phonetic variation in a San Francisco neighborhood* (Unpublished doctoral dissertation). Stanford University.

Hall-Lew, L. (2010). Improved representation of variance in measures of vowel merger. In *Proceedings of Meetings on Acoustics 159ASA* (Vol. 9, p. 060002).

Hamdi, R., Barkat-Defradas, M., Ferragne, E., & Pellegrino, F. (2004). Speech timing and rhythmic structure in Arabic dialects: a comparison of two approaches. In *Proceedings of Interspeech* (Vol. 4).

Han, M. (1962). The feature of duration in Japanese. *Onsei no kenkyu*, *10*, 65-80.

Haugen, E. (1966). *Dialect, language, nation* [reprinted in Pride and Holmes, J.(eds.) Sociolinguistics]. Penguin, London.

Hay, J., Drager, K., & Warren, P. (2010). Short-term exposure to one dialect affects processing of another. *Language and speech*, *53*(4), 447–471.

Hay, J., Warren, P., & Drager, K. (2006). Factors influencing speech perception in the context of a merger-in-progress. *Journal of phonetics*, *34*(4), 458–484.

Haynes, E. F., & Taylor, M. (2014). An assessment of acoustic contrast between long and short vowels using convex hulls. *The Journal of the Acoustical Society of America*, *136*(2), 883–891.

Herold, R. (1990). *Mechanisms of merger: The implementation and distribution of the low back merger in Eastern Pennsylvania* (Unpublished doctoral dissertation). University of Pennsylvania dissertation.

Herold, R. (1997). Solving the actuation problem: Merger and immigration in eastern Pennsylvania. *Language variation and change*, *9*(2), 165–189.

Hewitt, B. G. (1979). *Lingua Descriptive Studies 2: Abkhaz.* Amsterdam: North-Holland Publishing Company.

Higgins, A., Benson, P., Li, K., & Porter, J. (1999). *Dialect identification* (Tech. Rep.). Clifton, New Jersey, USA: ITT Aerospace/Comminications Division.

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), 3099–3111.

Ho, T. K. (1995). Random decision forests. In *Proceedings of 3rd International conference on document analysis and recognition* (Vol. 1, pp. 278–282).

Hockett, C. F. (1955). *A manual of phonology* (No. 11). Waverly Press.

Hoenigswald, H. M. (1965). *Language change and linguistic reconstruction.* University of Chicago Press.

Jacewicz, E., Fox, R. A., & Lyle, S. (2009). Variation in stop consonant voicing in two regional varieties of American English. *Journal of the International Phonetic Association*, *39*(3), 313–334.

Jacewicz, E., Fox, R. A., O'Neill, C., & Salmons, J. (2009). Articulation rate across dialect, age, and gender. *Language variation and change*, *21*(2), 233–256.

Jacewicz, E., Fox, R. A., & Salmons, J. (2007). Vowel space areas across dialects and gender. In *International congress of phonetic sciences* (Vol. 16, pp. 1465–1468).

Jacewicz, E., Fox, R. A., & Wei, L. (2010). Between-speaker and within-speaker variation in speech tempo of American English. *The Journal of the Acoustical Society of America*, *128*(2), 839–850.

Jaeger, J. J. (1978). Speech aerodynamics and phonological universals. In *Annual Meeting of the Berkeley Linguistics Society* (Vol. 4, pp. 312–329).

Johnson, D. E. (2007). *Stability and change along a dialect boundary: The low vowels of southeastern New England* (Unpublished doctoral dissertation). University of Pennsylvania.

Jongman, A., Fourakis, M., & Sereno, J. A. (1989). The acoustic vowel space of Modern Greek and German. *Language and speech*, *32*(3), 221–248.

Joseph, J. E. (1987). *Eloquence and power: The rise of language standards and standard languages*. Burns & Oates.

Kakati, B. K. (1941). Assamese its formation and development. *Guwahati: Department of Historical and Antiquarian Studies*.

Keating, P. (2005). *D-prime (signal detection) analysis.* Retrieved 2019-05-07, from `http://phonetics.linguistics.ucla.edu/facilities/statistics/dprime.htm`

Kelley, M. C., & Tucker, B. V. (2020). A comparison of four vowel overlap measures. *The Journal of the Acoustical Society of America*, *147*(1), 137–145.

Kennedy, M. (2006). *Variation in the pronunciation of English by New Zealand school children* [Master's thesis]. Victoria University of Wellington.

King, R. D. (1967). Functional load and sound change. *Language*, *43*(4), 831–852.

Kristoffersen, G. (2007). Dialect variation in East Norwegian tone. *Tones and tunes*, *1*, 91–111.

Kuhn, M. (2015). Caret: classification and regression training. *Astrophysics Source Code Library*, ascl–1505.

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software*, *82*(13), 1–26. doi: 10.18637/jss.v082.i13

Labov, W. (1991). The three dialects of English. *New ways of analyzing sound change*, *5*, 1–44.

Labov, W. (1994). Principles of linguistic change: Volume 1: Internal factors. vol. 20. *Language in society*.

Labov, W. (2006). *The social stratification of English in New York city*. Cambridge University Press.

Labov, W., Ash, S., & Boberg, C. (2008). *The atlas of North American English: Phonetics, phonology and sound change*. Walter de Gruyter.

Labov, W., Karen, M., & Miller, C. (1991). Near-mergers and the suspension of phonemic contrast. *Language variation and change*, *3*(1), 33–74.

Labov, W., Yaeger, M., & Steiner, R. (1972). *A quantitative study of sound change in progress* (Vol. 1). US Regional Survey.

Ladefoged, P. (2003). *Phonetic data analysis: An introduction to fieldwork and instrumental techniques*. Wiley-Blackwell.

Ladefoged, P., & Disner, S. F. (2012). *Vowels and consonants*. John Wiley & Sons.

Ladefoged, P., & Maddieson, I. (1990). Vowels of the world's languages. *Journal of Phonetics*, *18*(2), 93–122.

Ladefoged, P., & Maddieson, I. (1996). *The sounds of the world's languages* (Vol. 1012). Blackwell Oxford.

Lenth, R. (2019). emmeans: Estimated marginal means, aka least-squares means [Computer software manual]. Retrieved from `https://CRAN.R-project.org/package=emmeans` (R package version 1.4.1)

Liaw, A., & Wiener, M. (2002). Classification and regression by randomforest. *R news*, *2*(3), 18–22.

Liljencrants, J., & Lindblom, B. (1972). Numerical simulation of vowel quality systems: The role of perceptual contrast. *Language*, 839–862.

Lindau, M., & Wood, P. (1977). Acoustic vowel spaces. *UCLA Working papers in Phonetics*, *38*, 41–48.

Lindblom, B. (1986). Phonetic universals in vowel systems. *Experimental phonology*, 13–44.

Livijn, P. (2000). Acoustic distribution of vowels in differently sized inventories–hot spots or adaptive dispersion. *Phonetic Experimental Research, Institute of Linguistics, University of Stockholm (PERILUS)*, *11*.

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

Macmillan, N. A., & Creelman, C. D. (2004). *Detection theory: A user's guide*. Psychology press.

Maddieson, I., & Disner, S. F. (1984). *Patterns of sounds*. Cambridge University Press.

Maguire, W., Clark, L., & Watson, K. (2013). Introduction: what are mergers and can they be reversed? *English Language & Linguistics*, *17*(2), 229–239.

Maguire, W. N. (2008). *What is a merger, and can it be reversed?: the origin, status and reversal of the 'NURSE-NORTH merger' in Tyneside English* (Unpublished doctoral dissertation). Newcastle University.

Mahanta, S. (2001). *Some aspects of prominence in Assamese and Assamese English*. (Unpublished MPhil Dissertation) Dissertation submitted to CIEFL, Hyderabad.

Mahanta, S. (2008). *Directionality and locality in vowel harmony: With special reference to vowel harmony in Assamese*. Netherlands Graduate School of Linguistics.

Mahanta, S. (2012). Assamese. *Journal of the International Phonetic Association*, *42*(2), 217–224.

Martinet, A. (1952). Function, structure, and sound change. *Word*, *8*(1), 1–32.

Martinet, A. (1957). *Économie des changements phonétiques*. Bern, Francke.

Masica, C. P. (1993). *The Indo-Aryan languages*. Cambridge University Press.

Moral, D. (1991). *Upabhasa-bijnan*. Banalata.

Moral, D. (1992). *A phonology of Asamiya dialects: Contemporary Standard and Mayong* (Unpublished doctoral dissertation). PhD Thesis ed. Pune: PhD Thesis, Deccan College.

More, A. S., & Rana, D. P. (2017). Review of random forest classification techniques to resolve data imbalance. In *2017 1st International Conference on Intelligent Systems and Information Management (ICISIM)* (p. 72-78). doi: 10.1109/ICISIM .2017.8122151

Morrison, G. S. (2008). Comment on "a geometric representation of spectral and temporal vowel features: Quantification of vowel overlap in three linguistic varieties"[j. acoust. soc. am. 119, 2334–2350 (2006)]. *The Journal of the Acoustical Society of America*, *123*(1), 37–40.

Nearey, T. M., & Assmann, P. F. (1986). Modeling the role of inherent spectral change in vowel identification. *The Journal of the Acoustical Society of America*, *80*(5), 1297–1308.

Nespor, M. (1990). On the rhythm parameter in phonology. *Logical issues in language acquisition*, 157–175.

Nycz, J., & Hall-Lew, L. (2013). Best practices in measuring vowel merger. In *Proceedings of Meetings on Acoustics 166ASA* (Vol. 20, p. 060008).

Oshiro, T. M., Perez, P. S., & Baranauskas, J. A. (2012). How many trees in a random forest? In *International workshop on machine learning and data mining in pattern recognition* (pp. 154–168).

Pan, H.-h., Chen, M.-h., Lyu, S.-r., & Ke, Y.-c. (2011). Dialectal variation of voice quality in Taiwan Min: An EGG and Acoustical Study. In *ICPhS* (pp. 1554–1557).

Patgiri, B. (2011). *A comparative phonological study of two varieties of the Assamese language*. (Unpublished MPhil Dissertation) Dissertation submitted at the Center for Linguistics, JNU, New Delhi.

Pellegrino, F., Coupé, C., & Marsico, E. (2011). A cross-language perspective on speech information rate. *Language*, 539–558.

Penny, R. J. (2000). *Variation and change in Spanish*. Cambridge University Press.

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*(2), 175–184.

Pike, K. L. (1945). *The intonation of American English.* ERIC.

Port, R. F., Al-Ani, S., & Maeda, S. (1980). Temporal compensation and universal phonetics. *Phonetica*, *37*(4), 235–252.

Port, R. F., Dalby, J., & O'Dell, M. (1987). Evidence for mora timing in Japanese. *The Journal of the Acoustical Society of America*, *81*(5), 1574–1585.

R Core Team. (2019). R: A language and environment for statistical computing [Computer software manual]. Vienna, Austria. Retrieved from `https://www.R-project.org/`

Ramus, F., Nespor, M., & Mehler, J. (1999). Correlates of linguistic rhythm in the speech signal. *Cognition*, *73*(3), 265–292.

Ray, P. S. (1966). *A reference grammar of Bengali.* (Tech. Rep.). Washington DC: Center for Applied Linguistics.

Recasens, D., & Espinosa, A. (2006). Dispersion and variability of Catalan vowels. *Speech communication*, *48*(6), 645–666.

Rickford, J. R. (2002). *How linguists approach the study of language and dialect.* Retrieved 2019-02-03, from `http://www.stanford.edu/~rickford/papers/l73_reading_1.doc.` (Lecture Notes)

Roach, P. (1982). On the distinction between 'stress-timed' and 'syllable-timed' languages. *Linguistic controversies*, 73–79.

Robb, M. P., & Gillon, G. T. (2007). Speech rates of New Zealand English and American English speaking children. *Advances in Speech Language Pathology*, *9*(2), 173–180.

Sarma, M., & Sarma, K. K. (2016). Dialect identification from assamese speech using prosodic features and a neuro fuzzy classifier. In *2016 3rd International Conference on Signal Processing and Integrated Networks (SPIN)* (p. 127-132). doi: 10.1109/SPIN.2016.7566675

Savithri, S. R. (2009). *Speech rhythm in Indian languages.* [Lecture delivered at the Dr. N. Rathna Oration , 41st ISHACON, Pune.].

Schwartz, G. (2010). Rhythm and vowel quality in accents of English. *Research in Language*.

Schwartz, J.-L., Boë, L.-J., Vallée, N., & Abry, C. (1997). Major trends in vowel system inventories. *Journal of phonetics*, *25*(3), 233–253.

Scornet, E. (2017). Tuning parameters in random forests. *ESAIM: Proceedings and Surveys*, *60*, 144–162.

Sircar, D. (1990). Pragjyotisha-Kamarupa. *Barpujari, H K, The Comprehensive History of Assam I, Guwahati: Publication Board, Assam*, 59–78.

Spa, J. J. (2015). *Economie des changements phonétiques. traité de phonologie diachronique [the economy of phonetic changes; treatise of diachronic phonology]* (Vol. 61) (No. 3).

Stevens, K. N. (1972). The quantal nature of speech: Evidence from articulatory-acoustic data. In E. E. David & P. B. Denes (Eds.), *Human communication: A unified view* (pp. 51–66). New York: McGraw-Hill.

Stevens, K. N. (1989). On the quantal nature of speech. *Journal of phonetics*, *17*(1), 3–45.

Stevens, S. S., & Volkmann, J. (1940). The relation of pitch to frequency: A revised scale. *The American Journal of Psychology*, *53*(3), 329–353.

Stevens, S. S., Volkmann, J., & Newman, E. B. (1937). A scale for the measurement of the psychological magnitude pitch. *The Journal of the Acoustical Society of America*, *8*(3), 185–190.

Taid, T. (1988). A note on the synchronic description of Assamese sounds in Kakati's AFD. In T. Taid & R. D. Goswami (Eds.), *Banikanta Kakati, the man and his works* (pp. 70–83). Publication Board Assam.

Thomas, E. R. (2000). Applying phonetic methods to language variation. *American Speech*, *75*(4), 368–370.

Trudgill, P., & Foxcroft, T. (1978). On the sociolinguistics of vocalic mergers: Transfer and approximation in East Anglia. *Sociolinguistic Patterns in British English*, 69–79.

Twaha, A. I. (2017). *Intonational phonology and focus in two varieties of Assamese* (Unpublished doctoral dissertation). Dissertation submitted at the Department of HSS, IIT Guwahati.

Vance, T. J. (1987). *An introduction to Japanese phonology*. Albany: State University of New York Press.

Vaux, B., & Samuels, B. (2015). Explaining vowel systems: dispersion theory vs natural selection. *The Linguistic Review*, *32*(3), 573–599.

Verhoeven, J., De Pauw, G., & Kloots, H. (2004). Speech rate in a pluricentric language: A comparison between Dutch in Belgium and the Netherlands. *Language and Speech*, *47*(3), 297–308.

Wassink, A. B. (1999). *A sociophonetic analysis of Jamaican vowels* (Unpublished doctoral dissertation). University of Michigan.

Wassink, A. B. (2006). A geometric representation of spectral and temporal vowel features: Quantification of vowel overlap in three linguistic varieties. *The Journal of the Acoustical Society of America*, *119*(4), 2334–2350.

White, L., & Mattys, S. L. (2007). Rhythmic typology and variation in first and second languages. *Amsterdam Studies in the Theory and History of Linguistic Science Series 4*, *282*, 237.

Wickham, H. (2016). *ggplot2: Elegant graphics for data analysis*. Springer-Verlag New York. Retrieved from `https://ggplot2.tidyverse.org`

Williams, D., & Escudero, P. (2014). A cross-dialectal acoustic comparison of vowels in Northern and Southern British English. *The Journal of the acoustical society of America*, *136*(5), 2751–2761.

Wong, A. W.-m., & Hall-Lew, L. (2014). Regional variability and ethnic identity: Chinese Americans in New York City and San Francisco. *Language & Communication*, *35*, 27–42.

Wood, S. A. (1991). Vertical, monovocalic and other "impossible" vowel systems: a review of the articulation of the Kabardian vowels. *Studia Linguistica*, *45*(1-2), 49–70.

Yang, B. (1996). A comparative study of American English and Korean vowels produced by male and female speakers. *Journal of phonetics*, *24*(2), 245–261.

Zhang, J. (2014). *A sociophonetic study on tonal variation of the Wuxi and Shanghai dialects*. Netherlands Graduate School of Linguistics.