

Phonetics of Angami



Viyazonuo Terhija

A Dissertation

Submitted in Partial Fulfillment of the Requirements
For the Degree of Doctor of Philosophy in Linguistics

Department of Humanities and Social Sciences
Indian Institute of Technology Guwahati
Guwahati, Assam-781039
September-2024



© 2024 Viyazonuo Terhiija

Declaration

I declare that, the dissertation titled, “**PHONETICS OF ANGAMI**”, 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 Prof. 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.



Viyazonuo Terhiija

Dated: 13-09-2024

Guwahati.

Dept. of Humanities and Social Sciences

Indian Institute of Technology Guwahati

Guwahati - 781039, Assam, India.

Certificate

This is to certify that the thesis entitled “**PHONETICS OF ANGAMI**”, submitted by **VIYAZONUO TERHIJA** (Registration Number: 156141013), a research scholar in the *Department of Humanities and Social Sciences, Indian Institute of Technology Guwahati*, for the award of the degree of **Doctor of Philosophy**, is a record of an original research work carried out by her under my supervision and guidance. The thesis has fulfilled all requirements as per the regulations of the institute and in our opinion has reached the standard needed for submission. The results embodied in this thesis have not been submitted to any other University or Institute for the award of any degree or diploma.

Dated: 13-09-2024
Guwahati.



Prof. Priyankoo Sarmah
Professor
Dept. of Humanities and Social Sciences
Indian Institute of Technology Guwahati
Guwahati - 781039, Assam, India.

This thesis is dedicated to my dear husband

Mr. Japunglie Chase

dzapunglie t^hase



Acknowledgements

Given the extensive nature of the research work spanning over many years, I am deeply indebted to numerous individuals and organizations for their unwavering support, without which completion of this thesis would not have been possible. Firstly, I would like to express my appreciation and gratitude to the Indian Institute of Technology Guwahati for allowing me to carry out this research and providing financial support. Also, I want to thank IIT Guwahati for providing state-of-the-art laboratories, computational resources, and other logistical supports.

I consider it a privilege and honor to be under the tutelage and supervision of a good teacher and equally good human, Professor Priyankoo Sarmah. He has given so much time and energy to learning and teaching to make the research studies par excellence with the new trends and findings. For all the conferences, workshops, and seminars, the progress meeting held at 8 a.m. every Monday, I would like to thank him for guiding me on this journey. I owe a debt of gratitude to several people, foremost among them Prof. Bidisha Som, the chairperson of my Doctoral Committee, for encouraging and providing valuable suggestions on my work throughout the years; Prof. Rohit Sinha and Dr. Bodhisattva Sengupta, member of my Doctoral Committee, for their insightful suggestions and critical feedback that helped to improve my thesis work; Professor K Samudra Vijaya for introducing and encouraging me to carry scientific study and research; and Professor Arupjyoti Saikia for the valuable comments and insights.

I am deeply grateful to the outstanding linguists who contributed to my Ph.D. journey. I express my profound gratitude to Dr. Alexander R. Coupe of Nanyang Technological University, Singapore; Dr. Seunghun J. Lee of International Christian University, Tokyo, Japan & Dr. Juraž Šimko, University of Helsinki, Finland. The meetings and discussions in their areas of expertise have significantly influenced the

development of this thesis. I appreciate their constant availability, both in person and through email communication.

At the Department of Humanities and Social Sciences, IIT Guwahati, I am incredibly grateful to the head of the department for being supportive of the academic process. I would also like to express my sincere gratitude to all non-teaching staff from the HSS (past and present), Bandana Khataniar, Khanthai Mala Basumatary, Parag Jyoti Kalita, Durga Sharma, and Rubul Gogoi for their unconditional support and prompt response in dealing with technical and administrative issues.

I am also grateful for all the provisions provided at the Phonetics and Phonology Laboratory (PP lab.) during my Ph.D. stay. I am grateful to the head of Research and Development for allowing me to use the facilities at the Pathological Speech Processing Laboratory of North East Centre for Biological Sciences and Healthcare Engineering (NECBH). I am grateful for the medical provisions on campus, and my special gratitude to Chief Medical Officer, Dr. Anuj Kr. Baruah.

This work would have not be possible without the love and support of the people of Kigwema, Viswema, and Kohima villages. I express my gratitude to each individual who participated and graciously allowed me to collect data during the numerous field visits. While this thesis ultimately focuses on the study of Kohima village data, its foundation was laid during the initial years of experimentation through my study of Kigwema and Viswema. As such, this study was only possible with permission and help extended by the village council and student council in the village. I express my deepest gratitude to the Kigwema Student Council for generously allowing me to utilize the Panchayat Hall for recordings. I extend my sincere thanks to Metsi Yhokha, my friend and guide, who facilitated introductions and networking within the village. My stay in Viswema village would not have been possible without the hospitality of Mr. Vito Mekro and his family, who graciously hosted me time and again. I am indebted to Ms. Akhruvi Mekro for her introduction to the natives of

Viswema. Additionally, I am thankful to Joy, Meza, and Vilol for their invaluable assistance. A special note of appreciation goes to the many octogenarians whom I had the privilege of meeting during the early years of my research, although few are no longer with us today.

In Kohima village, my heartfelt gratitude goes to Mr. and Mrs. Vizotuonyü Suokhrie and family for their unwavering assistance throughout my data collection. I extend special thanks to Kenei, Meme, Bubu, Rügüo, Nuoluo, and Dziesesatuo for their invaluable contributions. A sincere ‘Thank you’ to my dear friend and senior, Dr. Kelhouvinuo Suokhrie, for the support extended over the years. I am also grateful to all the people who participated in the data collection, even during the challenges posed by the pandemic and its aftermath.

I express my gratitude to Mr. and Mrs. Neikralie Khezhe and Mr. and Mrs. Ketholelie Terhijja for their generous support and hospitality during the numerous visits for data collection in the Kohima town area. I extend heartfelt thanks to Mr. and Mrs. Pekrusetuo Angami and their family for graciously hosting me in Dimapur each time I traveled for work and beyond. The love and support I received from them will forever be cherished and remembered.

I wish to further thank my friends and language teachers, especially Ms. Asano Zünyü, for readily being there for me all the time, and Dr. Metsei Mor, Dr. Rünguso, and Mr. Vilaneizo Punyü for the academic input and help.

I sincerely thank my seniors who have engaged in meaningful academic discourse and encouraged and helped me in my initial years, Dr. Luke Horo, Dr. Leena Dihingia, and Dr. Wendy Lalhminghlui. I am also grateful to all my Phonetics and Phonology laboratory colleagues.

I am always grateful to my colleagues from the campus, who have always been supportive and made my stay there beautiful. I earnestly thank Dr. Moakala Tzudir for the friendship and the constant support. I would also like to thank Dr. Sangay

Tamang, Dr. Pongli Sangtam, Dr. Chan, Ato, Imnanaro, Nini, Zhonei, Kesilenuo and all the Nagas at IITG. I am also grateful to Saidul Da for all the help extended during my stay on campus.

I am indeed fortunate to have friends outside the IIT Guwahati Campus who have always been there throughout this journey: Dr. Kanato Chophi, Dr. Azhoni Kricheni, Samuel Vaiphei, Neithono Yhokha, Menuse Sophie, Chunthailiu, Pretty Brahma, Awheno Medom, Amen Tzudir and Anguno Casavi.

I express my love and gratitude to my parents, Mr. Dzüziengulie and Mrs. Vikhole Terhiija; my siblings, Khries, Jess, and Asono; and all my in-laws who always believed and supported me in the completion of this thesis. No words are sufficient to express my gratitude to you all. I am grateful for the prayers of my grandmother, Late Atsa Kuosie-ü Pienyü who have been a source of sustenance and strength for me throughout this journey.

This work would not have seen the light of day without the love and support of my dear husband, Japungulie Chase. Thank you for believing in me, for being patient when I am away on campus, and for all the prayers.

Above all, I thank God for all His blessings!

Viyazonuo Terhiija

Abstract

This thesis provides a detailed acoustic phonetic analysis of the Angami language (also known as Tenyidie) speech sounds, a lesser-studied Tibeto-Burman language spoken in northeast India. Nine significant contributions are made in this thesis. The primary contribution of this thesis involves the development of a speech corpus, which is manually annotated and segmented, aimed at advancing research on the Angami language. The second contribution of the thesis is a detailed acoustic study of voice onset time (VOT) in Angami plosives. The third contribution of the thesis is the study of acoustic characteristics of voiced and voiceless nasals along with their effect on vowel nasality. Furthermore, a fine-grained study of voiceless nasals was performed using Nasometer II. The fourth contribution of the thesis is the acoustic characterization of affricates and fricatives in Angami, using features such as centre of gravity (CoG), central moment, intensity, and percentage of voicing. The fifth contribution of the thesis is the acoustic characterization of lateral and approximants in Angami using spectral features such as the first four formants, HNR, CoG, intensity, and percentage of voicing. The result showed a clear distinction between the voiced and voiceless laterals and approximants; the formants (F1-F4) further showed a clear pattern to differentiate the approximants based on PoA. The thesis's sixth contribution is an intricate examination of vowels. It utilizes the first three formant frequencies F1, F2, and F3 and the duration to analyze monophthongs. This study also employs features, such as, the Discrete Cosine Transforms (DCT) of F1 and F2 and duration to investigate the diphthongs. The findings clearly identified six distinct monophthongs and two diphthongs in the language. The seventh contribution of the thesis addresses issues in Angami tones and studies acoustic characteristics of tones using F_0 and duration. The results showed that Angami has four distinct level tones. The F_0 of the tones T2 and T3 overlap, and the tones are statistically not significantly different in terms of F_0 or durations. The study also investigates

the interaction of tones with segments, such as tones interaction with consonants and tone interaction with vowels. The finding of VOT interactions with F_0 shows that the extent of consonantal perturbation depended on the tonal categories. The correlation of vowel height and intrinsic F_0 (IF_0) was also explored. The findings show that in Angami, high vowels induce higher F_0 , and low vowels induce F_0 lowering. Additionally, Angami has bidirectional tonal co-articulation, showing both anticipatory and carryover effects. For its eighth contribution, this thesis delves deeper into the acoustic correlates and classification of tones. Using random forest (RF) to classify the five-tone categories resulted in an accuracy of 66.4%. Interestingly, combining T2 and T3 led to a significant increase in accuracy, reaching 80.8%. Among the various acoustic features studied, the most prominent correlates of tones in Angami were found to be F_0 mean, F_0 height, F_0 max, and F_0 min. The thesis's ninth contribution involves studying native speakers' perception of tones through two separate tests. The first test, using male and female stimuli, revealed higher accuracy in identifying the high (T1) and low (T5) tones than the intermediary tones (T2, T3, & T4). Confusion mainly emerged between tones T2 and T3. In the second perception test with synthesized stimuli, findings indicated that the identification of tones relies on two primary cues: F_0 height and duration. Lower tones (T5 and T4) were predominantly identified based on duration, while higher tones (T1 and T2) were identified with F_0 height.

Contents

Declaration	iii
Acknowledgements	vi
List of Figures	xviii
List of Tables	xxvi
List of Acronyms	xxxvi
1 Angami language and the current study	1
1.1 Introduction	1
1.2 Background	4
1.2.1 Linguistic phylogeny of Angami	4
1.2.2 Nomenclatures of Angami	7
1.2.3 Previous works on the Angami language	9
1.2.4 Overview of Angami phonological inventory	11
1.3 Methodology	15
1.3.1 Area of study	16
1.3.2 Speech data and participants	17
1.3.3 Data recording and annotation	18
1.3.4 Acoustic and temporal measurements	21

1.3.5	F0 Extraction Using Zero-frequency filtering (ZFF)	24
1.3.6	Statistical measurements	25
1.4	Organization of the dissertation	26
2	Acoustics of Angami plosives	29
2.1	Introduction and literature review	29
2.1.1	Literature review of Angami plosives	33
2.2	Methodology	34
2.2.1	Speakers	34
2.2.2	Materials	35
2.2.3	Recording of data	36
2.2.4	VOT annotation and extraction	36
2.2.5	Statistical measurement	37
2.3	Results	38
2.3.1	VOT in Angami stops	40
2.4	Discussion	47
3	Acoustics of Angami nasals	48
3.1	Introduction and literature review	49
3.2	Literature review of Angami nasals	52
3.3	Methodology	53
3.3.1	Speakers	53
3.3.2	Materials	53
3.3.3	Recording of Data	54
3.3.4	Acoustic and statistical analysis	54
3.4	Results	55
3.4.1	Acoustic properties of voiced and voiceless nasals	56
3.4.2	Vowel nasality	59

3.4.3	Duration of voiced and voiceless nasals	65
3.5	Discussion	67
3.6	Nasals in Angami using nasometer II	71
3.7	Methodology	72
3.7.1	Speakers	72
3.7.2	Materials	73
3.7.3	Recording of data using Nasometer II	73
3.7.4	Data extraction and Analysis	74
3.8	Results	75
3.8.1	Oral and Nasal waveforms of Angami nasals	75
3.8.2	Voicing in Angami Nasals	80
3.8.3	Nasalance of voiced and voiceless nasals	83
3.8.4	Duration of voiced and voiceless nasals	84
3.9	Conclusion	85
4	Acoustics of Angami Fricatives & Affricates	89
4.1	Introduction and Literature review	89
4.2	Literature review of Fricatives & Affricates in Angami	94
4.3	Methodology	97
4.3.1	Speakers	98
4.3.2	Materials	98
4.3.3	Extraction of acoustic features	98
4.4	Results	99
4.4.1	Acoustics features of fricatives in Angami	99
4.4.2	Acoustic features of Affricates in Angami	118
4.5	Discussion & Conclusion	133

5	Acoustics of Angami Laterals and Approximants	135
5.1	Introduction and literature review	136
5.2	Literature review of Angami laterals & approximants	138
5.3	Data and Methodology	142
5.3.1	Speakers	142
5.3.2	Materials	142
5.3.3	Acoustic features and statistical analyses	143
5.4	Results	144
5.4.1	Acoustics of laterals in Angami	144
5.4.2	Acoustics of approximants in Angami	153
5.5	Conclusion	162
6	Acoustics of Angami Vowels	170
6.1	Introduction and literature review	170
6.1.1	Literature Review of Angami monophthong Vowels	172
6.1.2	Literature Review of Angami Diphthong Vowels	173
6.1.3	Representation of Angami vowels in orthography	175
6.2	Methodology	177
6.3	Results	181
6.3.1	Acoustic Features of Monophthong Vowels in Angami	181
6.3.2	Acoustic Features of Diphthong Vowel in Angami	197
6.4	Discussion	203
7	Acoustics of suprasegmentals in Angami	205
7.1	Introduction and literature review	206
7.1.1	Tone notation	208
7.1.2	The physiology of tone production	208
7.1.3	Theories on the vocal-fold tension (pitch perturbation)	209

7.2	Literature Review of Angami tones	212
7.3	Tones in Angami	214
7.3.1	Methodology	214
7.3.2	Results	216
7.3.3	Conclusions	226
7.4	Interaction of tone and segments	228
7.4.1	Tone and Angami consonants	228
7.4.2	Tone and Angami vowels	240
7.4.3	Discussion and conclusions	247
7.4.4	Tonal coarticulatory effects in Angami	247
8	Acoustic Correlates and Classification of Tones in Angami	259
8.1	Introduction	259
8.2	Methodology	261
8.2.1	F0 Extraction Using Zero-frequency filtering (ZFF)	262
8.2.2	Random Forest	263
8.3	Results	264
8.3.1	Acoustic analysis of tones	264
8.3.2	Feature Importance using Random Forest	265
8.3.3	Classification of tones using Random Forest	266
8.3.4	Statistical Analysis	267
8.4	Discussion and Conclusion	269
9	Perceptual correlates of tones in Angami	272
9.1	Introduction	273
9.2	Perception of Angami tones using natural stimuli (E1)	277
9.2.1	Methodology	277
9.2.2	Results of the perceptual study	280

9.2.3	Discussion	284
9.3	Perception of Angami tones using natural and synthetic stimuli (E2)	285
9.3.1	Methodology	286
9.3.2	Results of the perceptual study	290
9.3.3	Discussion	295
10	Summary and Conclusion	297
10.1	General Conclusions	297
10.1.1	VOT in Angami laryngeal contrasts	299
10.1.2	Acoustic correlates of Angami nasals	299
10.1.3	Acoustic correlates of Angami fricatives and affricates	300
10.1.4	Acoustic properties of Angami Approximants	300
10.1.5	Summary of Angami consonants	301
10.1.6	Acoustic properties of Angami vowels	302
10.1.7	Acoustics properties of Angami tones	303
10.1.8	Interaction of tones and segments	303
10.1.9	Classification of tones in Angami	304
10.2	Directions for future research	305
	Bibliography	322

List of Figures

1-1	Map of India with Nagaland highlighted.	3
1-2	Map of the Tengima area (in red), adapted from Hutton (1969) issued by John Bartholomew & Son Ltd. Edinburgh.	10
1-3	Area of study (Kohima)	17
1-4	Sample and example of recording of speech data using Nasometer II 6450 (KayPENTAX medical)	20
1-5	An example of segmentation and annotation of the sound files	21
2-1	An example of the segmentation of the VOT and vowel boundary for /t ^h u/.	37
2-2	An example of the VOT and vowel boundary in three laryngeal contrast of Angami.	39
2-3	Kernel density estimation plots of VOT in Angami stops.	41
2-4	Skewness and Kurtosis value of the VOT density across the nine Angami stops.	41
2-5	The average VOT duration in milliseconds for the nine Angami plosives with standard deviation.	43
2-6	The average VOT duration in milliseconds of the 10 Angami speakers.	43
2-7	Variable interaction plot generated from an LME model with VOT as the dependent variable and laryngeal contrast and PoA as fixed effects.	46

3-1	An example of the nasal and vowel boundary for the word /ma/ ‘price’ and /m ^h a/ ‘things’ as produced by a female speaker.	61
3-2	Formant contours (N1-N4) in Hertz for the Angami nasal consonants.	62
3-3	Bar plots of (a) the mean of nasal formants bandwidths (BW1-BW4), (b) SD and CoG calculated from the nasal spectrum.	63
3-4	Nasality measures for the vowels following Angami nasals.	64
3-5	Variable interaction plot generated from an LME model with the duration of nasal as the dependent variable and voicing types, contexts, PoA, and their interactions as fixed effects. The statistically significant interactions are indicated with asterisks.	68
3-6	Oral and nasal channel waveforms extracted from Nasometer for a plain /me/.	76
3-7	Oral and nasal channel waveforms extracted from Nasometer for /m ^h e/ in isolation.	76
3-8	Oral and nasal channel waveforms extracted from Nasometer for /m ^h e/ in sentence.	76
3-9	Voiceless nasals in Angami in three places of articulation, bilabial, alveolar and palatal, produced by a female speaker.	79
3-10	Bars showing the average percentage of voicing with standard errors.	80
3-11	Nasalance of nasals produced in isolation with standard errors as colored ribbons.	84
3-12	Nasalance of nasals produced in sentence frames with standard errors as colored ribbons.	84
4-1	Spectrograph of (a) labio-dental voiceless /f/ and (b) voiced unaspirated /v/ followed by vowel /ə/ as produced by a female speaker in isolation.	101

4-2	Spectrograph voiceless alveolar /s/ & voiceless palatal /ʃ/ followed by /a/ vowel as produced by a female speaker in isolation.	101
4-3	Kernel density estimation plots for the six Angami fricatives.	102
4-4	Violin plot of the average duration of the six Angami fricatives.	103
4-5	Violin plot of the average duration of the six Angami fricatives across the three contexts.	104
4-6	Violin plot of the average center of gravity of the six Angami fricatives.	106
4-7	Violin plot of the average center of gravity of the six Angami fricatives across the three contexts.	106
4-8	Violin plot of the average percentage of voicing of the six Angami fricatives.	107
4-9	Violin plot of the average percentage of voicing of the six Angami fricatives across the three contexts.	107
4-10	Violin plot of the average intensity of six fricatives in Angami.	109
4-11	Violin plot of the average intensity of six fricatives across the three contexts in Angami.	110
4-12	Spectrograph of palatal voiceless unaspirated /tʃ/ (a), voiceless aspirated /tʃ ^h / (b) and voiced aspirated /dʒ/ (c) followed by vowel /a/ as produced by a female speaker in isolation.	119
4-13	Kernel density estimation plots for the six Angami Affricates.	120
4-14	Violin plot of the average duration of the six Affricates in Angami.	121
4-15	Violin plot of the average duration of the six Affricates across the three contexts in Angami.	122
4-16	Violin plot of the average center of gravity of the six Angami Affricates.	124
4-17	Violin plot of the average center of gravity of the six Angami Affricates across the three contexts in Angami.	124

4-18	Violin plot of the average percentage of voicing of the six Angami Affricates.	125
4-19	Violin plot of the average percentage of voicing of the six Angami Affricates across the three contexts.	126
4-20	Violin plot of the average intensity of the six Angami Affricates.	126
4-21	Violin plot of the average intensity of the six Angami Affricates across the three contexts.	127
5-1	Spectrograms of English approximants for the word <i>wet, yet, let</i> and <i>ret(ch)</i> in Ladefoged and Disner (2012)	138
5-2	An example of the lateral and vowel boundary for the words /li/ 'pot' and /l ^h i/ 'business' as produced by a female Angami speaker in isolation.	146
5-3	Voiceless lateral /l ^h i/ 'business' as produced by a female speaker in three contexts.	147
5-4	Violin plots showing the duration (ms) in voiced and voiceless laterals across the three contexts.	148
5-5	Formants (F1-F4) in voiced and voiceless laterals across the three contexts.	150
5-6	HNR in voiced and voiceless laterals across the three contexts.	151
5-7	Intensity in voiced and voiceless laterals across the three contexts.	151
5-8	Voiceless aspirated approximants /w ^h e/ 'cover', /ɽ ^h a/ 'redeem', and /j ^h a/ 'lift' as produced by a female speaker in isolation.	154
5-9	Violin plots showing the duration (ms) in voiced and voiceless approximants	155
5-10	Violin plots showing the duration (ms) in voiced and voiceless approximants across the three contexts	155
5-11	Violin plots showing the percentage of voicing in the voiced and voiceless approximants.	156

5-12	Violin plots showing the percentage of voicing in the voiced and voiceless approximants across the three contexts.	156
5-13	Violin plots showing the center of gravity in voiced and voiceless approximants	157
5-14	Violin plots showing the center of gravity in voiced and voiceless approximants across the three contexts	158
5-15	Formants (F1-F4) in voiced and voiceless bilabial (top), alveolar (mid) and palatal (bottom) approximants across the three contexts in Angami.	160
5-16	Harmonic to noise ratio in voiced and voiceless approximants across the three contexts.	161
5-17	Intensity in voiced and voiceless approximants across the three contexts.	161
6-1	Annotation example showing the vowel boundary of the word /mi/ ‘fire’ as produced by a female speaker	180
6-2	Skewness and Kurtosis graph of normalized F1 (top) and F2 (bottom) in the six vowels of Angami	182
6-3	Mean F1 and F2 plots for the six Angami vowels (top); Normalized average F1 and F2 plots of the six Angami vowels (bottom)	185
6-4	Normalized F1 and F2 with one standard deviation ellipses for 7 vowels (top) and 6 vowels (bottom)	186
6-5	Normalized F1 and F2 with one standard deviation ellipses for Male (top) and Female (bottom) speakers)	188
6-6	Interaction plot for the LME models where normalized F1 (top), F2 (mid), and F3 (bottom) are dependable variables, whereas the interactions of gender, tones, and vowels are the fixed effects.	191
6-7	Violin plots of the average duration of vowels where they can appear in a lexical word without an onset or a coda.	193
6-8	Violin plot of the average duration of the six Angami vowels.	194

6-9	Violin plots of the average duration of the six Angami vowels by contexts and gender	195
6-10	The formant values of F1 and F2 for monophthong and diphthong showing the space diphthong vowel occupies.	198
6-11	Inverse DCT contours of F1 and F2 calculated from the mid-80% of the total duration of the two diphthongs in Angami.	199
6-12	Inverse DCT contours of F1 and F2 calculated from the mid 80% of the total duration of the two diphthongs for male and female Angami speakers.	201
6-13	Violin plots of the average diphthongs duration across male and female speakers.	202
6-14	Violin plots of the average diphthongs duration by contexts and gender.	203
7-1	Spectrographs of the word ‘le’ across the five tones as produced by a female speaker.	216
7-2	Kernel density estimation plots for the F_0 extracted from five Angami tones.	218
7-3	F_0 contour plot tones in Angami by male and female speakers.	218
7-4	F_0 contours across the five tones in Angami, (top) is the average raw F_0 values while, (bottom) is the average normalized F_0 values.	220
7-5	Normalized F_0 contour plot of the tones in Angami across the three context.	221
7-6	Violin plots of the average duration of the five Angami tones	224
7-7	Violin plots of the average duration of the five Angami tones across the three contexts.	224
7-8	Normalized F_0 contour plot of the four tones in Angami	227
7-9	F_0 contour tones of the five level tones in Black Miao as produced by male speaker (Kuang, 2013)	227

7-10	Normalized mean F_0 tone contours of five Angami tones with standard errors	231
7-11	Variable interaction plot generated from an LME model with VOT as the dependent variable and laryngeal contrast and tones as fixed effects.	232
7-12	Normalized F_0 contours of the five Angami tones in the nine Angami stops.	234
7-13	Normalized F_0 contours for Angami tones grouped by three laryngeal features.	235
7-14	Average F_0 contours of five Angami tones.	244
7-15	F_0 contour plot of T1 with all the six Angami vowels.	245
7-16	F_0 contours of T2/T3 with all Angami vowels.	245
7-17	F_0 contour plot of T4 with all the six Angami vowels.	246
7-18	F_0 contour plot of T5 with all the six Angami vowels.	246
7-19	Normalized mean F_0 of the disyllable tone contours showing anticipatory coarticulatory effects.	253
7-20	Normalized mean F_0 of the disyllable tone contours, showing carryover coarticulatory effects.	254
8-1	Overall framework of the Angami tone detection system.	262
8-2	Mean (mid 80%) of normalized F_0 contours of Angami tones.	265
8-3	Variable importance of top 10 features for tones in Angami.	266
8-4	Confusion matrix in %. (top) Five level tones, (bottom) Tones T2 and T3 merged as one unit.	268
9-1	F_0 contour of the five tones stimuli produced by a female and male Angami speaker	278
9-2	Duration of the five tones stimuli as produced by a female and male Angami speaker	279

9-3	An example of the options displayed on the screen for perception study (E1)	288
9-4	F ₀ contour of the five tones stimuli for the word 'le' as produced by a female Angami speaker	288
9-5	Duration of the five tones stimuli for the word 'le' as produced by a female Angami speaker	289
9-6	An example of the options displayed on the screen for perception study (E2)	290
9-7	Average responses of the synthesized stimuli.	293
9-8	Average responses of the synthesized stimuli across short (top), mid (middle) and long (bottom) durations.	294
9-9	Average reaction time to the synthesized stimuli.	294
-1	Map of migration of the Tengima (Tenyimia) adapted from Hutton (1921)	314

List of Tables

1.1	Classification of the Tibeto Burman language with special reference to Angami language	8
1.2	Consonant inventory in Khonoma Angami (Blankenship et al. 1993) .	12
1.3	Consonant inventory in standard Angami (Kuolie, 2006)	12
1.4	Syllable structure in Angami (adapted from Kuolie, 2003)	13
1.5	Distribution of consonants and vowels in onset position (CV) in Angami.	14
1.6	Vowel inventory Angami (Khonoma) based on Blankenship et al. (1993)	15
1.7	Vowel inventory Angami (standard) based on Kuolie (2006)	15
1.8	The IPA labels of consonants of Angami along with the corresponding labels used in this work.	16
1.9	Data distributed in the dissertation. Here, the details column includes the number of unique tokens, the syllable structure and number of speakers	18
1.10	Example of the target word /l ^h a/ in three contexts.	19
2.1	Example of stop minimal set in Angami.	34
2.2	Distribution of consonants in the five tone categories included in this study.	35
2.3	The average (avg) VOT duration in milliseconds with their standard deviation(sd).	42

2.4	The average (avg) VOT duration in milliseconds with their standard deviation(sd) across speakers.	42
2.5	Results of an ANOVA conducted on an LME model with VOT as the dependent variable and place of articulation (POA), laryngeal contrast, tones, vowels and their interactions as fixed effects. Only statistically significant interactions are shown in this table.	45
2.6	Results of a Bonferroni post-hoc test conduct on an LME model with VOT as dependent variable and POA and laryngeal contrast as fixed factors (df=5297).	46
2.7	Average VOT in milliseconds with their standard deviations across the three laryngeal categories in Angami.	47
3.1	Example of nasal minimal pairs in Angami	52
3.2	Distribution of Nasals and vowels in the Data	54
3.3	Results of Analysis of Deviance (ANOVA) tests on the LME models for estimating the effect of nasals consonants and estimates of post-hoc Bonferroni tests for pair-wise contrasts.	58
3.4	Analysis of Deviance (ANOVA) tests on the LME models with average nasality values as dependent variables and pair-wise contrasts estimated from post-hoc Bonferroni tests.	65
3.5	Significance of nasality values from Bonferroni post-hoc pairwise comparisons conducted on LME models for 11 equidistant points of vowels following nasals.	66
3.6	The mean duration and standard deviation (SD) in milliseconds (ms) of the six nasals in Angami by contexts.	66
3.7	Results of an Analysis of Deviance (ANOVA) test on the LME model for estimating the effect of voicing, PoA, and contexts on nasal duration.	67

3.8	Estimates of the duration of nasal pairs obtained from a Bonferroni post-hoc pairwise comparison conducted on an LME model, with significant estimates ($p < 0.0001$) marked with an asterisk.	67
3.9	Average percentage of voicing in nasal consonants with standard deviations in parentheses.	81
3.10	Results of Analysis of Deviance Tests on the three models.	82
3.11	Significance of difference in voicing in isolation and in sentence frames for each nasal in three models.	83
3.12	Average duration of nasal segments with standard deviations.	85
3.13	Results of Analysis of Deviance Tests with duration.	85
3.14	Results of Analysis of Deviance Tests with duration.	86
4.1	Fricatives and Affricates described in Angami by various authors. Here, ‘Fct’ and ‘Afr’ represent fricatives and Affricates respectively.	95
4.2	Example of fricatives minimal set in Angami.	96
4.3	Example of affricates minimal set in Angami.	97
4.4	Distribution of fricatives and affricates included in this study.	99
4.5	Average (in bold) and standard errors (in reduced font) of the acoustic features in Angami voiced and voiceless fricatives.	108
4.6	Average (in bold) and standard errors (in reduced font) of the acoustic features in Angami voiced and voiceless fricatives across the three contexts.	111
4.6	Average (in bold) and standard errors (in reduced font) of the acoustic features in Angami voiced and voiceless fricatives across the three contexts.	112

4.7	Results of an ANOVA conducted on an LME model with the acoustic-phonetic features such as duration, CoG, Intensity and voicing as the dependent variable and PoA, phonation, context and their interactions as fixed effects for fricatives in Angami.	115
4.8	Results of a Bonferroni post-hoc test conduct on an LME model with skewness, kurtosis, CoG, duration as dependent variable and PoA and phonation as fixed factors.	116
4.9	Results of a Bonferroni post-hoc test conduct on an LME model with intensity and percentage of voicing as dependent variable and PoA and phonation as fixed factors.	117
4.10	Average (in bold) and standard errors (in reduced font) of the acoustic features in Angami voiceless unaspirated, voiceless aspirated and voiced affricates.	128
4.11	Average (in bold) and standard errors (in reduced font) of the acoustic features in Angami voiceless unaspirated, voiceless aspirated and voiced affricates.	128
4.11	Average (in bold) and standard errors (in reduced font) of the acoustic features in Angami voiceless unaspirated, voiceless aspirated and voiced affricates.	129
4.12	Results of an ANOVA conducted on an LME model with the six acoustic-phonetic features as the dependent variable and PoA, laryngeal contrasts, context and their interactions as fixed effects for Affricates in Angami.	131
4.13	Results of a Bonferroni post-hoc test conduct on an LME model with skewness, kurtosis, CoG, duration as dependent variable and PoA and Laryngeal contrast as fixed factors for Affricates.	132

5.1	Approximants described in Angami by various authors. Here, Approx. stands for approximants and L-velar for labial velar. Original transcription are retained for the cited references.	141
5.2	Laterals and Approximants minimal pairs in Angami.	142
5.3	Average (in bold) and standard errors of the acoustic features in Angami voiced and voiceless laterals.	150
5.4	Results of Bonferroni post-hoc pairwise comparison conducted on LME models with acoustic features as the dependent variable and voicing contrasts, contexts and their interactions as fixed effects.	152
5.5	Analysis of Deviance (ANOVA) tests on the LME models with acoustic features as the dependent variable and voicing contrasts, contexts and their interactions as fixed effects, pair-wise contrasts estimated from post-hoc Bonferroni tests. The abbreviation ISO, ST and CR represent isolation, sentence and carrier phrase, respectively.	152
5.6	Average (in bold) and standard errors of the acoustic features in Angami voiced and voiceless approximants	162
5.7	Average (in bold) and standard errors of the acoustic features in Angami voiced and voiceless approximants across the three contexts. . .	165
5.7	Contd. . . Average (in bold) and standard errors (in reduced font) of the acoustic features in Angami voiced and voiceless approximants across the three contexts.	166
5.8	Analysis of Deviance (ANOVA) tests on the LME models with acoustic features of the approximants as the dependent variable and voicing contrasts, contexts and their interactions as fixed effects, pair-wise contrasts estimated from post-hoc Bonferroni tests. The abbreviation ISO, ST and CR represent isolation, sentence and carrier phrase, respectively.	167
5.8	Contd.	168

5.8	Contd...	169
6.1	Description of monophthong vowels in Angami by various authors . . .	174
6.2	Description of Diphthong vowels in Angami by various authors	174
6.3	Orthography vs. IPA of the vowels in Angami	176
6.4	Example of monophthong vowel minimal pair in Angami	176
6.5	Example of diphthong vowels minimal pairs in Angami	177
6.6	Distribution of six monophthong vowels in Angami across the three contexts in the data.	178
6.7	Distribution of the two diphthong vowels in Angami	179
6.8	The mean values and standard deviation (SD) of F1, F2, and F3 of the six vowels in Angami.	187
6.9	The mean values and standard deviation (SD) of F1, F2, and F3 of the six vowels in Angami as produced by Male and Female speakers.	189
6.10	Results of an ANOVA conducted on an LME model with normalized F1 as the dependent variable and vowel, context, gender, tone, and their interactions as fixed effects. Only statistically significant interactions are shown in this table.	189
6.11	Results of an ANOVA conducted on an LME model with normalized F2 as the dependent variable and vowel, context, gender, tone, and their interactions as fixed effects. Only statistically significant interactions are shown in this table.	190
6.12	Results of an ANOVA conducted on an LME model with normalized F3 as the dependent variable and vowel, context, gender, tone, and their interactions as fixed effects. Only statistically significant interactions are shown in this table.	190
6.13	Example of a minimal pair of vowels where they can appear in a lexical word without an onset or a coda.	192

6.14	Pairwise comparison of the vowels that appear as a lexical word without an onset or a coda.	192
6.15	Average vowel duration and standard deviation in six Angami vowels.	195
6.16	Average vowel duration and standard deviation in six Angami vowels by three context and gender.	195
6.17	Results of an ANOVA conducted on an LME model with vowel duration as the dependent variable and vowel, context, gender, tone, and their interactions as fixed effects. Only statistically significant interactions are shown in this table.	196
6.18	Results of a Bonferroni post hoc test conducted on an LME model with vowel duration as the dependent variable and vowel as fixed factors (df=Inf).	197
6.19	The mean values and standard deviation (SD) of F1 and F2 of the two diphthong vowels in Angami.	200
6.20	The mean values and standard deviation (SD) of F1 and F2 of the two diphthong vowels in Angami as produced by Male and Female speakers.	200
6.21	Results of an ANOVA conducted on an LME model with C0 (F1 & F2) as the dependent variable and vowel, context, gender, tone, and their interactions as fixed effects. Only statistically significant interactions are shown in this table.	200
6.22	Results of an ANOVA conducted on an LME model with C1 (F1 & F2) as the dependent variable and vowel, context, gender, tone, and their interactions as fixed effects. Only statistically significant interactions are shown in this table.	200

6.23	Results of an ANOVA conducted on an LME model with C2 (F1 & F2) as the dependent variable and vowel, context, gender, tone, and their interactions as fixed effects. Only statistically significant interactions are shown in this table.	201
6.24	Average duration of the diphthongs in Angami.	202
6.25	Average duration of the two diphthongs with SD in three contextual frames divided by speakers in Angami.	203
7.1	Example of tonal minimal set in Angami (Liezietsu et al., 2019) . . .	213
7.2	Tones described in Angami by various authors	214
7.3	Distribution of tone tokens in Angami.	215
7.4	Results of an Analysis of Deviance (ANOVA) test on the LME model for estimating the effect of tones, contexts and their interactions on the mid-points (40%, 50% & 60%) of the normalized F_0 contour and estimates of post-hoc Bonferroni tests for pair-wise contrasts.	222
7.5	Mean and standard deviation of the duration (ms) of the five Angami tones	223
7.6	Mean and standard deviation of the duration (ms) of the five Angami tones across the three contexts	223
7.7	Results of an Analysis of Deviance (ANOVA) test on the LME model for estimating the effect of duration of tones, contexts and their interactions on the mid-points (40%, 50% & 60%) of the normalized F_0 contour and estimates of post-hoc Bonferroni tests for pair-wise contrasts.	225
7.8	Results of post-hoc Bonferroni tests for pair-wise contrasts across the three context.	225
7.9	Mean VOT and standard deviation for Angami stops across all tones.	231
7.10	Mean and standard deviation of the duration of the five Angami tones	232

7.11	Result of Bonferroni post-hoc pairwise comparison conducted on an LME model with normalized VOT as the dependent variable and tone and laryngeal contrast as fixed effects (VD= voiced, VLA= voiceless aspirated, VLU = voiceless unaspirated)	238
7.12	Result of Bonferroni post-hoc pairwise comparison conducted on an LME model with tone duration as dependent variable and tone as fixed effects.	239
7.13	Distribution of vowel tokens in Angami.	242
7.14	Result of LME test in Angami tones	247
7.15	Example of the Noun-Noun combination and the Verb - Grammatical category (GC) combination used in the study.	250
7.16	Results of Analysis of Deviance Tests on LME models for estimating anticipatory effects at three middle points, showing χ^2 values ($df=4$) with statistical significance indicated by an asterisk *	256
7.17	Results of Analysis of Deviance Tests on LME models for estimating carryover effects at three middle points, showing χ^2 values ($df=4$) with statistical significance indicated by an asterisk *	256
8.1	Tonal space of the five tones in Angami between male and female speakers in Hertz.	266
8.2	Distribution of the number tones in the tone classification study.	267
8.3	Analysis of Deviance (ANOVA) tests on the LME models, with the top six F_0 features as dependent variables and pair-wise contrasts estimated from post-hoc Bonferroni tests.	271
9.1	Response of the perception test in percentage % values and average reaction time	281
9.2	Reaction Time for every stimulus and responses in the experiment (E1)	281

9.3	Result of post-hoc Bonferroni pairwise comparison conducted on an LME model with reaction time (RT) as the dependent variable and the responses as fixed effects.	282
9.4	Responses of the perception test based on stimuli produced by female speaker and male speaker in percentage % values	283
9.5	Reaction time based on stimuli produced by female speaker and male speaker	283
9.6	Tonal minimal pair for the word /le/ in Standard Angami as per MKS Dieda (dictionary)	286
9.7	Response of the perception test (natural stimuli) in percentage (%) .	291
9.8	Reaction Time of the perception test (natural stimuli) in seconds (sec)	291
10.1	Phonemic inventory of consonants in Angami.	302
10.2	Phonemic inventory of vowels in Angami.	303
3	List of Nasal datalist	309
4	List of fricatives, affricates, laterals and approximants datalist.	310
4	List of fricatives, affricates, laterals and approximants datalist.	311
5	List of lexical tones datalist.	312
6	List of vowels that have the same phoneme and tone but different orthography.	312
7	List of the Angami villages adapted from District Census Handbook Kohima, 2011. The sub-groupings and the division of the villages is done based on common knowledge and cross-checking with native Angami. (subjected to correction)	315
8	Information of participants in this thesis (2018-2022)	317
8	Information of participants in this thesis (2018-2022) (continued) . . .	318



List of Acronyms

ANOVA	Analysis of variance
CC	Coordinating conjunction
CP	Categorical perception
CoG	Center of Gravity
CV	Consonant Vowel
DCT	Discrete Cosine transform
DF	Degrees of Freedom
F ₀	Fundamental Frequency
GCI _s	Glottal Closure Instants
HNR	Harmonic to Noise Ratio
IMP-ASP	Imperative Aspect
IPA	International Phonetic Alphabet
JND	Just Noticeable Difference
ISO	International Organization for Standardization
L1	First language (native language)
LME	Linear Mixed Effects
MIL	Modern Indian Languages
NA	Not Available
NHS	Nagaland Heritage Studies
NEP	New Education Policy

PoA	Place of Articulation
PrPROG	Present Progressive
RF	Random Forest
SD	Standard Deviation
SE	Standard Errors
TBU	Tone Bearing Unit
VD	Voiced
VLU	Voiceless unaspirated
VLA	Voiceless aspirated
UCLA	University of California, Los Angeles
UNESCO	The United Nations Educational, Scientific and Cultural Organization
VOT	Voice Onset Time
WALS	World Atlas of Language Structure
ZFF	Zero Frequency Filtering

Chapter 1

Angami language and the current study

1.1 Introduction

The Angami language, also known as Tenyidie (ISO 639–3: njm) is an under-described Tibeto-Burman language, primarily spoken in the state of Nagaland in India. This dissertation provides an acoustic analysis of the speech sounds of the standard variety of Angami language as spoken in the Kohima district. The work is motivated by the lack of comprehensive acoustic-phonetic study on the speech sounds of the Angami language. This dissertation provides acoustic analysis of the consonants, vowels and tones, supported by statistical evidence and further substantiated by the application of the findings using machine learning.

While a small body of work on the phonetics of Angami consonants is available (Blankenship et al., 1993; Bhaskararao and Ladefoged, 1991), there have been descriptive works on the grammar and phonology of the language (Burling, 1960; Giridhar, 1980; Ravindran, 1974; Chase, 1992; Kuolie, 2006; Suokhrie, 2015; Meyase, 2016, 2021). From the previous studies, we identify the following research gaps in the

study of Angami language:

- Lack of phonetic analysis of the standard variety of Angami.
- Lack of large speech database for conducting reliable quantitative analysis.
- Lack of appropriate methodologies for acoustic-phonetic analysis, considering the typologically rare speech sounds in Angami.
- Considering the register tone system in Angami, there was no previous effort to study their phonetic characteristics and perceptual correlates of the register tones.

As seen above, over the span of the last 23 years, no other acoustic-phonetic analysis of Angami is reported. Apart from that, in the existing literature on the language, there are discrepancies in the reporting of the inventory size and types of consonants, vowels and tones. Hence, in this work we are motivated to study the phonetics of Angami and to provide a comprehensive account of the segmental and suprasegmental features of the language.

Nagaland, the northeastern state in India (see Figure 1-1), where Angami is spoken, comprises seventeen distinct indigenous tribes. Among these, fourteen, namely, Angami, Ao, Chakesang, Chang, Kachari, Khiamniungan, Konyak, Kuki, Lotha, Phom, Pochury, Rengma, Sangtam, Sumi, Tikhir, Yimkhiung, Zeliang, belong to the Naga community, while the remaining two tribes are the Kukis and Kacharis. Nagaland is a multilingual society where the inhabitants commonly speak several languages. Apart from their mother tongue (L1), which may be one of the primary languages of the community, English is used as an official language. Nagamese, which emerged as a contact language between the Nagas and the Assamese, functions as a *lingua franca*. Additionally, Hindi or one of the Modern Indian Languages (MIL) is included in school curriculum as a subject.

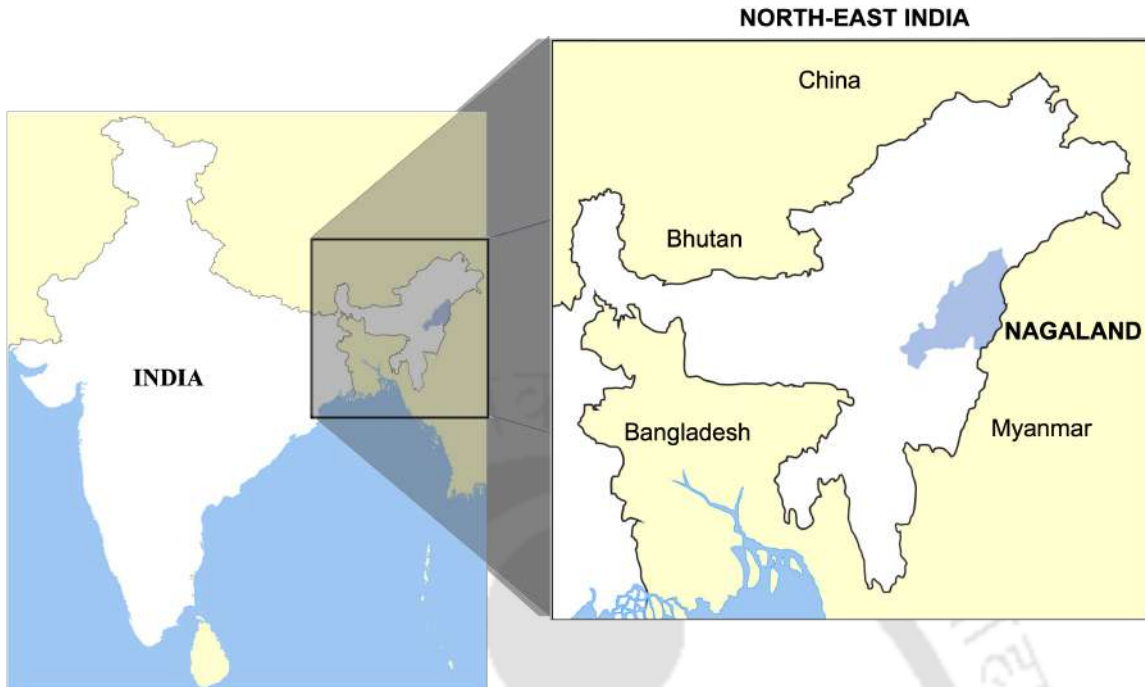


Figure 1-1: Map of India with Nagaland highlighted.

Angami is one of the major communities of the Nagas in Nagaland and is primarily found in the Kohima district of Southern Nagaland, bordering with Manipur and the Dimapur district of Nagaland, spoken by about 152,796 speakers (Office of the Registrar General & Census Commissioner of India, 2011). Angami is considered to be a language at a vulnerable stage by UNESCO (Eberhard et al., 2024). While the term Angami is considered an exonym, the language is also referred to by its endonym, Tenyidie. However, the term Tenyidie subsumes Angami and several linguistic groups in the vicinity of the Angami speaking areas (Kuolie, 2006)¹. In the current study, we focus on the spoken form used by the standard Angami speakers.

The Angamis are broadly divided into four subgroups depending on their geographic location, namely, the Northern, Southern, Western, and Central (commonly known as Chakroma²). Each group consists of several villages, each of which is said to

¹For a discussion on the term Tenyidie, please refer to Section 1.2.2

²The word means people who reside below the road. As mentioned in Hutton (1921), the term ‘chā-krō-ma’ translates literally to ‘Road-below men,’ a name derived from the geographical positioning of the villages in relation to the old Samguting-Kohima bridle path.

have its own variety. The current study investigates the Northern Angami subgroup, as spoken in the Kohima village. This variety, spoken around Kohima village, is considered to be the standard form of the Angami language (Mahapatra and McConnell Grant D., 1989; Ravindran, 1974).

Considering the complex organization of the Angami linguistic communities, the following sections provide brief descriptions of the Angami linguistic group and its linguistic phylogeny. This is followed by discussion on the previous works on the language, and then a brief overview of the Angami phonology.

1.2 Background

1.2.1 Linguistic phylogeny of Angami

Nathan Brown, a pioneer American Baptist missionary in Assam, is recognized for being among the earliest individuals to categorize the Naga languages based on lexical evidence (Brown, 1851). Grierson (Grierson, 1903), in the linguistic survey of India, divides Naga languages into five subgroups; western subgroup, Central subgroup, Eastern Sub-group, Naga- Bodo subgroup and Naga-Kuki Sub-group. The five subgroups and the languages associated with are as follows:

- (i) Western Naga: Angami, Sema, Rengma & Kezhama
- (ii) Central Naga: Ao, Lhotas, Tengsa, Thukumi & Yachumi
- (iii) Eastern Naga: Tableng, Tamlu, Banpara, Mutonia, Chang, Mohongia Namsangia, Assiringia, Moshang & Shangge
- (iv) Naga-Bodo: Mikhir, Kachcha Naga, Kabui & Khoirao
- (v) Naga-Kuki: Mao, Maram, Miyangkhang, Kwoireng, Luppa & Maring

Much of Grierson's research drew upon the earlier works and manuscripts of the British administrator R. B. McCabe. However, Grierson's linguistic analysis did not delve into the specifics of the dialects within individual Naga languages. Rather than dividing languages based on geography, Grierson employed a morphological approach, particularly focusing on the placement of negative marking. The western subgroup, for instance, includes languages where the negative follows the root of the word it modifies. Angami is one of the four languages that fall into this category. In his documentation of the Angami language, Grierson noted that it is spoken by approximately 35,410 individuals residing in the southern region of the Naga Hills. Furthermore, he classified Angami as belonging to the Naga subgroup of Tibeto-Burman languages (Grierson, 1903). The classification of the languages provided by Grierson has been questioned, especially in regard to groupings under of the Naga-Bodo and Naga-Kuki, and in regard to the geographical location of the languages (Burling, 2003; Marrison, 1967).

Shafer (1955) attempted the classification of the Naga languages through phonological analysis, with special references to the syllable finals. Shafer classified Angami along with the Rengma and Sema languages under the Eastern branch of Burmic division of the Sino-Tibetan family. Marrison (1967) classified the Naga group of languages based on typological comparisons at the level of phonology, morphology, and syntax. At the phonological level, he subdivided the syllabic patterns into word-initial, word-medial and word-final positions. He grouped the Naga language family into seven subgroups and arranged them from North-East to South-West. Angami falls under the subgroup Type C-1 along with Sema, Chokri, Kezhama & Mao. The details of the subgroup provided by the author are listed below:

(i) Type A -1: Tangsa (Yongli), Tangsa (Moshang), Nocte & Wancho

(ii) Type A - 2: Konyak, Phom & Chang

- (iii) Type B - 1: Yacham, Ao (Chungli), Ao (Mongsen) & Sangtam
- (iv) Type B -2: Lotha, Yimchurgru, Ntenyi & Meluri
- (v) Type B - 3: Tangkhul & Maring
- (vi) Type C - 1: Sema, Angami (Kohima), Angami (Khonoma), Chokri, Kezhama & Mao
- (vii) Type C - 2: Rengma, Maram, Khoirao, Mzieme, Zeme, Liangmai, Puiron, Nruanghmei

Burling (2003) categorized the Tibeto-Burman languages spoken in the Northeast of India into three broad groups, the central area which consist of the Bodo-Konyak-Jighpaw, the northern area which includes the languages spoken in Arunachal area, and the Eastern border area. Within these groups, he conducted a comparative lexical study to further classify the languages based on potential relationships among them. The eastern group listed below is often considered forming a linguistic subgroup of the Tibeto-Burman family more closely related to each other than other languages. Angami falls under the Eastern border area under the subgroup of Angami-Pochuri. He reported that Chokri (present-day Chakesang variety) could even be counted as Angami variety because of its similarities with Angami.

- (i) Ao group: Yacham-Tengsa, Ao-chungli, Ao-Mongsen, Sangtam, Yimchugrü, Lotha
- (ii) Angami-Pochuri group: Rengma, Pochuri, Rengma, Simi, Angami, Chokri, Kheza, Mao
- (iii) Zeme group: Nruanghmei, Puiron, Khoirao, Zeme, Mzieme, Liangmai, Maram
- (iv) Tangkhul group : Tangkhul, Maring

(v) Karbi (Mikir)

(vi) Meitei

(vii) Mizo-Kuki-Chin (Kukish)

The classifications proposed by Grierson (1903), Shafer (1955) and Marrison (1967) are based on geographical location or on phonological similarities. Burling (2003) offers a much different perspective, as he classifies the languages to a greater extent exclusively on lexical comparisons. He identifies his classification to be similar to Shafer than Marrison. Table 1.1 shows the classification of the Tibeto–Burman languages with special reference to Angami language in the literature. Given the variety of classification for the Naga languages, in this study we are sympathetic to the classification provided by Burling as it captures the linguistic similarities along with historical affiliations among the Tibeto-Burman languages.

1.2.2 Nomenclatures of Angami

As mentioned earlier, ‘Angami’ serves both as the name of the community and as the name of the language spoken by the community. The origin of the term ‘Angami’ can be traced back to the way their neighboring communities addressed them. According to Hutton (1969), the name ‘Angami’ is believed to be a distortion of the word ‘Gnamei’ given by the neighboring Manipuris. Grierson (1903) offers insights into the Angami people through the account provided by Mr. Davis, stating that while the people refer to themselves as ‘Tengima’³, neighboring tribes such as the Kezhamas, Semas, and Lhotas knew them as ‘Tsoghami’, ‘Tsongumi’, and ‘Tsangho’, respectively. Sanyu (1996) suggests that the term ‘Angami’ was coined by the Zemi people, which is derived from the word ‘gami’, meaning ‘invaders’. Despite these

³This is possibly the genesis of the term ‘Tenyimia’ which found currency to describe Angami and the related languages in the latter part of the twentieth century.

Table 1.1: Classification of the Tibeto Burman language with special reference to Angami language

Source	Tibeto - Burman groupings	Sub-Group (not based on hierarchical order)	Linguistics feature
Grierson (1903)	Western group	Rengma Sema Angami	morphology - negative marking that follows the verb
Shafer (1955)	Eastern Branch	Rengma Sema Angami Tsakrima Kezama Spovoma Kabui	phonology - special reference to syllable finals
Marrison (1967)	South-East Type C-1	Sema Angami Chokri Khezama Mao	phonology, morphology, syntax and geographical proximity
Burling (2003)	Eastern Branch	Rengma N Pochuri Rengma Simi Angami Chokri Kheza Mao	lexical comparison, geographical proximity, and language affinities

various accounts, the term ‘Angami’ had been embraced by the community itself for a considerable period, serving as their self-identifier.

Hutton, recognized as one of the pioneering British anthropologists and administrators who studied the Angamis extensively, provided valuable insights into the history of settlement, customs, and traditions (Hutton, 1969) of the Angami people. In the account of the Angami Nagas, Hutton (1969) gives a detailed map of the ‘Angami Naga country’ which encompasses four main groups: Tengima, Chakrima, Kezama, and Memi (see Figure 1-2). The map indicates that the ‘Tengima group’ corresponds to the present-day Kohima, Chumukedima, and Dimapur districts, the

primary settlements of the Angamis. The group known as ‘Chăkrîma’ (previously referred to as the ‘Eastern Nagas’) is situated in present-day Phek district, inhabited by the Chakesang ethnic group. The ‘Kezama’ group’s territory is currently shared between Nagaland and Manipur, while the ‘Memî’ group is located within the Manipur state. This grouping is based on the similarity of language, customs and shared ancestry. Kuolie (2006) gives an account of the genesis of Tenyidie where he points out that Tenyidie (meaning the Tenyi-language) is the name of the language spoken by the Tenyimia ⁴ (meaning, the Tenyi-people) which includes the following Naga groups: Angami, Chakesang, Rengma, Zeliangrong, Mao, Poumai, Memai, and Pochuri. He further asserts that these groups have a common ancestry. Consequently, from the recent past, the Tenyidie language is considered to include all the varieties spoken by each group under the Tenyimia community. The language varieties spoken in the area are identified as dialects of the Tenyidie language (Kuolie, 2006). Hence, the term Tenyimia serves as an umbrella term for the many communities, mentioned above, and it has been used to bring the communities together asserting that they all share the same ancestry. Tenyidie is taught in schools, as well as used in the church services, radio broadcasts and in the print media. The Ura Academy, established in 1939 (previously known as Angami literature committee under Rev.J.E Tanquist) is the organization responsible for the development and popularization of the Tenyidie Angami language. The organization has produced a number of literature and textbooks for school and college level education.

1.2.3 Previous works on the Angami language

While linguistic analysis of Angami is scarce, there are works since the beginning of the 20th century that primarily describe the grammar of the language. Angami was

⁴Nuh (2002) describes Tenyimia as a combination of three root words: ‘te’ meaning ‘set out’ or ‘go’, ‘nyi’ meaning ‘willing’, ‘mi’ meaning ‘people’, in Chakesang Kezha dialect. In short, ‘Tenyimia’ means ‘the one who is willing to set out’.

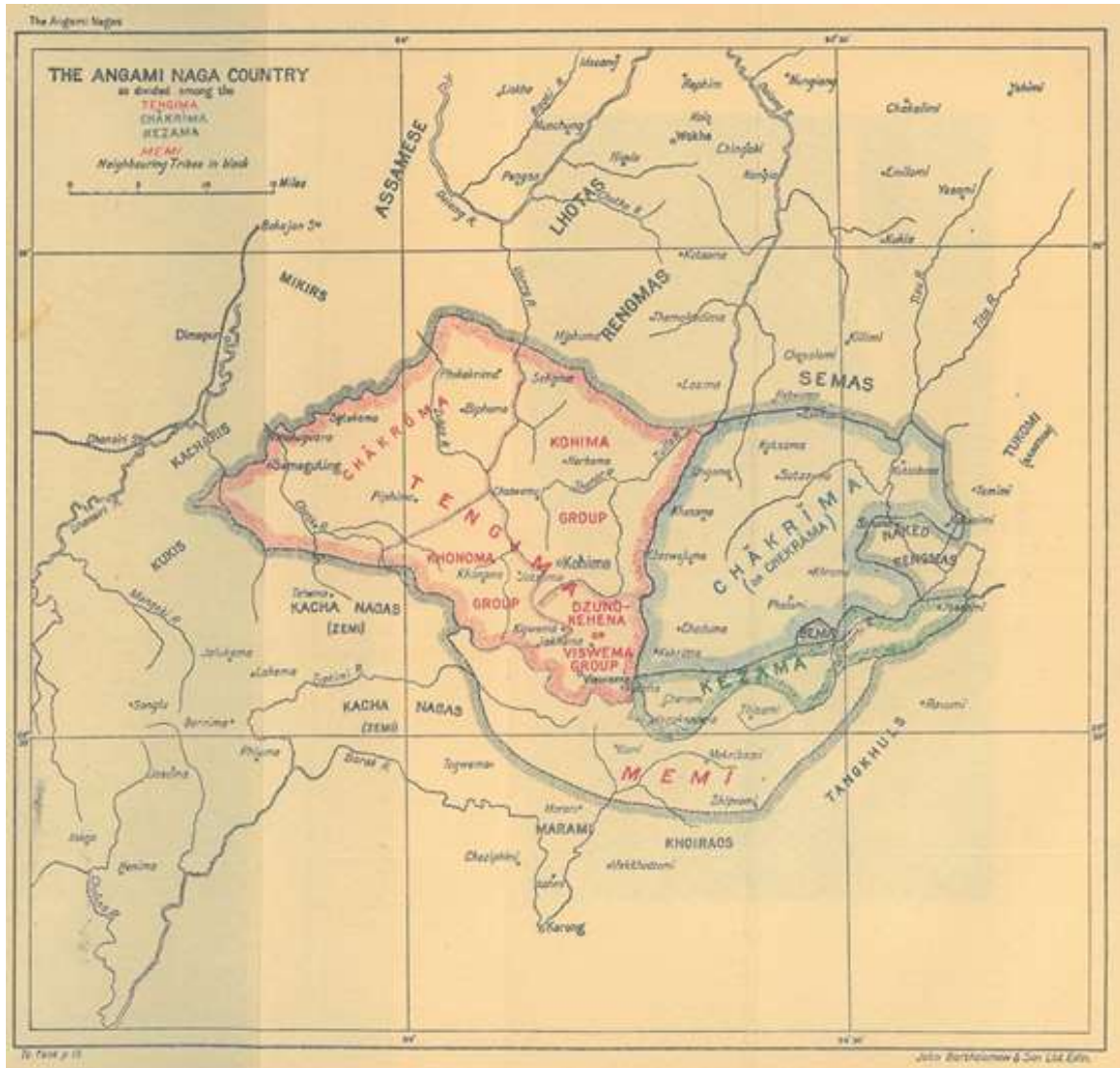


Figure 1-2: Map of the Tengima area (in red), adapted from Hutton (1969) issued by John Bartholomew & Son Ltd. Edinburgh.

first reported in McCabe's field report, where he had given a brief description of the grammar and provided a list of words translated from English to Angami (McCabe, 1887). General descriptions of the grammar of standard Angami are reported in the works of Burling (1960), Girdhar (1980), Kuolie (2006) and Ravindran (1974). In their structural description of the language, Kuolie (2006) thoroughly examined the phonology, morphology, numeral system, and the syntax of the language. Typological overview and morphosyntactic study of the language is also reported in the

works of Ezung (2014, 2018). One of the earliest preliminary acoustic studies of the language was on the study of voiceless nasals in Khonoma Angami, as reported by Bhaskararao and Ladefoged (1991). Another early study provides a phonetic description of Khonoma Angami vowels, tones, and voiceless nasals (Blankenship et al., 1993). A morphological analysis of Tenyidie tonal phonology and tone characteristics has also been provided by Meyase (2016, 2021, 2019). Besides the works in standard Angami, there are works done in dialects of the language as well. Dialectal variations and internal variations based on kinship ties and geography have also been reported (Suokhrie, 2015, 2017).

While resources for the Angami language are limited, two online sources have documented the language. The World Atlas of Language Structure (WALS) contains descriptive works on the language, while the UCLA Phonetic Lab archives contain recorded speech data in the Khonoma dialect of Angami (Giridhar, 1980; Grierson, 1903; Subbarao and Kevichüsa, 1999). However, the available information and speech samples are primarily focused from the Northern Angami and the Khonoma dialects. Hence, they may not represent the standard variety of the Angami language which is considered to be the variety used in the Kohima village. Apart from that, the available data from Khonoma Angami speech data is limited with utterances from only a handful of speakers.

1.2.4 Overview of Angami phonological inventory

A notable feature of Angami is its extensive inventory of consonants (Marrison, 1967; Burling, 2003). The earliest records of Angami provide varying accounts of consonant inventories. In the Khonoma variety, the consonant inventory ranges from 33 to 34 (Marrison, 1967; Blankenship et al., 1993). Conversely, in the Kohima variety, the consonant inventory ranges from 39 to 40 (Ravindran, 1974; Kuolie, 2006). The consonant inventories for the Khonoma variety as described by Blankenship et al.

(1993) and the Kohima variety of Angami by Kuolie (2006) are presented in Table 1.2 and Table 1.3, respectively. Given these discrepancies in the existing studies, determining the exact number of consonants in the inventory of the language is one of the research objectives in this study.

One of the motivations to study the standard variety of Angami, namely, the Kohima variety, is to examine the large inventory of affricates in the variety, when compared to the other varieties such as the Khonoma variety (Kuolie, 2006; Blankenship et al., 1993). According to Marrison (1967), Angami, classified as type C-1 (see Section 1.2.1), displays a distinctive phonology characterized by the presence of voiceless nasals and liquids. This observation is reinforced by the study of the Khonoma variety of Angami as well, which also demonstrates the existence of voiceless nasals

Table 1.2: Consonant inventory in Khonoma Angami (Blankenship et al. 1993)

	Bilabial		Labio-dental		Alveolar		Post-alveolar	Retroflex	Palatal		Velar		Labial Velar	Glottal
Plosives	p	b	t	d					k	g	k ^ʷ	g ^w		
	p ^h		t ^h						k ^h		k ^w			
Nasals	m ^h	m	n ^h	n					ɲ ^h	ɲ	ŋ			
Fricative			v	s	z	ʃ	ʒ							h
Affricate				ts		tʃ								
Lateral			l ^h	l										
Approximant							ɹ	ɹ	j	j			ʍ	w

Table 1.3: Consonant inventory in standard Angami (Kuolie, 2006)

	Bilabial		Labio-dental		Dental alveolar	Alveolar	Palatal	Velar	Glottal	
Plosives	p	b	t	d			k	g		
	p ^h		t ^h				k ^h			
Nasals		m			n			ɲ		
		m ^h			n ^h			ɲ ^h	ŋ	
Fricative			f	v	s	z		ʃ	ʒ	h
Affricate			pf	bv			ts	dʒ	tʃ	dʒ
			pf ^h				ts ^h		tʃ ^h	
Trill							r			
							r ^h			
Lateral							l			
							l ^h			
Approximant		w						j		
		w ^h						j ^h		

and laterals (Blankenship et al., 1993). Angami primarily consists of open syllables with a preference for CV syllable structure, and rarely does it have a CVC syllable. The few CVC structures in the language occur with a sonorant [ɾ] as the coda (for example /t^hoɾ/ which is a superlative marker). Additionally, the language allows for consonant clusters at the onset position, typically consisting of a consonant followed by the phoneme /ɾ/. The syllable structure in Angami can be summarized as follows: a single vowel, a consonant followed by a vowel, or an initial consonant cluster followed by a vowel. Example of the syllable structures in Angami shown in Table 1.4. Further, the distribution of the consonants and vowels in the onset positions of CV syllables in Angami is shown in Table 1.5.

Table 1.4: Syllable structure in Angami (adapted from Kuolie, 2003)

V - pattern:	/a/	‘ɪ’
CV - pattern:	/ta/	‘run’
CCV - pattern:	/kɪa/	‘cry’

Existing literature on Angami vowels have led to disagreements regarding the phonemic inventory of the vowels. Previous studies have proposed varying numbers of phonemic vowels for Angami, ranging from six to seven (Burling, 1960; Ravindran, 1974; Giridhar, 1980; Kuolie, 2006; Suokhrie, 2015). While all studies acknowledge the presence of the five peripheral vowels in Angami (/i, e, u, o, a/), there is inconsistency in the description of the central vowel. Some studies describe it as a mid-central vowel, others as a high central vowel, and in some cases, both descriptions are provided. The vowels as described in the literature, are shown in Table 1.6 & Table 1.7. A detailed study of the Angami vowels, both monophthongs and diphthongs, will be discussed in Chapter 6 of this dissertation.

A common feature of the Tibeto-Burman languages under the eastern group of Burling’s phylogenetic classification is that the languages are tonal. Angami is a tone language which has four to five contrastive tones. The descriptions of Angami tones

Table 1.5: Distribution of consonants and vowels in onset position (CV) in Angami.

	p	p ^h	b	t	t ^h	d	k	k ^h	g	m	m ^h	n	n ^h	ɲ	ɲ ^h	ŋ
i	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	×	✓	×	✓
e	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	×	✓
a	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
ə	×	×	×	×	×	×	×	×	×	×	×	×	×	✓	✓	×
u	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	×	✓
o	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	×	✓
	f	v	s	z	ʃ	ʒ	h	pf	pf ^h	by	ts	ts ^h	dz	tʃ	tʃ ^h	dʒ
i	×	✓	✓	✓	✓	×	✓	✓	✓	×	✓	×	✓	✓	✓	✓
e	×	✓	✓	✓	✓	×	✓	✓	✓	×	✓	✓	✓	✓	×	×
a	×	✓	✓	✓	✓	✓	✓	×	×	×	✓	✓	✓	✓	✓	✓
ə	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
u	×	✓	✓	✓	✓	✓	✓	✓	×	×	✓	✓	×	×	✓	×
o	×	✓	✓	✓	✓	✓	✓	×	×	×	✓	✓	✓	✓	✓	✓
	l	l ^h	ɾ	ɾ ^h	w	w ^h	j	j ^h	mv							
i	✓	✓	✓	✓	✓	✓	×	×	×							
e	✓	✓	✓	✓	✓	✓	×	×	×							
a	✓	✓	✓	✓	✓	✓	✓	✓	×							
ə	×	×	✓	✓	×	×	×	×	✓							
u	✓	✓	✓	✓	×	×	✓	✓	×							
o	✓	✓	✓	✓	✓	✓	✓	✓	×							

are neither exhaustive nor uniform. For example, the number of tones in Angami has been reported to be either four or five in the literature (Burling, 1960; Giridhar, 1980; Ravindran, 1974; Kuolie, 2006; Chase, 1992; Suokhrie, 2017; Meyase, 2021). Apart from the number of tones, studies have also not agreed on the types of tones in the language. For example, while a group of authors has reported Angami to have both contour and level tones (Burling, 1960; Giridhar, 1980; Ravindran, 1974; Kuolie, 2006; Liezietsu et al., 2019), others have settled with only level tones in the language (Chase, 1992; Suokhrie, 2017; Lalminghlui et al., 2019; Meyase, 2021). The issues in tones are further discussed in Chapters 7 to 9.

While writing systems are imperfect representations of the languages they represent, some clues about the distinct phonemes in Angami can be obtained from the orthography. Since the 19th century, Angami has been using the Roman ortho-

Table 1.6: Vowel inventory Angami (Khonoma) based on Blankenship et al. (1993)

	Front	Central	Back
Close	i		u
Close-mid		ə	o
Open-mid	e		
Open		a	

Table 1.7: Vowel inventory Angami (standard) based on Kuolie (2006)

	Front	Central	Back
Close	i		u
Close-mid		ə	o
Open-mid	e	ü	
Open		a	

graphy. The orthography does not indicate tones, however, at the segmental level it has nearly one-to-one correspondence between graphemes and phonemes. Angami graphemes representing the consonants are provided in Table 1.8 with the corresponding International Phonetic Alphabet (IPA) symbols.

1.3 Methodology

The description of Angami speech sounds is based on speech data collected from native speakers through field studies conducted from 2019 to 2023. The area of study is discussed in Subsection 1.3.1, while the types of data elicited and the description of participants are discussed in Subsection 1.3.2. The procedure of recording speech data, as well as the procedures for annotation and segmentation of the speech data, are reported in Subsection 1.3.3. The speech sounds were subjected to acoustic analysis, with the features discussed in Subsection 1.3.4. The extracted features are statistically analyzed using linear mixed models, as discussed in Subsection 1.3.6.

Table 1.8: The IPA labels of consonants of Angami along with the corresponding labels used in this work.

Label	IPA symbol	Label	IPA symbol
p	p	ph	p ^h
b	b	t	t
th	t ^h	d	d
k	k	kh	k ^h
g	g	m	m
mh	m ^h	n	n
nh	n ^h	ny	ɲ
nyh	ɲ ^h	ng	ŋ
f	f	v	v
s	s	z	z
sh	ʃ	zh	ʒ
h	h	pf	pf
pfh	pf ^h	bv	by
ts	ts	tsh	ts ^h
dz	dz	c	tʃ
ch	tʃ ^h	j	dʒ
l	l	lh	l ^h
r	r	rh	r ^h
w	w	wh	w ^h
y	j	yh	j ^h

1.3.1 Area of study

Angami is geographically divided into four regions: Northern Angami, Southern Angami, Western Angami, and Central Angami (Kuolie, 2006). According to the District Census Handbook 2011, there are multiple hamlets and villages under each region, ranging from 6 to 28 villages (Directorate of Census Operations, Nagaland, 2011). The sub-groupings and division of the villages are detailed in Table A.1 in Appendix A. It is believed that each Angami village has its own dialect. In reality, the amount of mutual intelligibility differs depending on the proximity of the villages. For example, the Kohima district has villages from Northern, Southern and Western Angami regions and hence, in the Kohima district itself numerous varieties of the language are spoken. While the spoken forms of the language differ, all the varieties write the



Figure 1-3: Area of study (Kohima)

language in Tenyidie which is followed as a standard form of writing for the Tenyi languages. The written form of Tenyidie is closest to the form spoken in the Kohima village. Hence, the current study is conducted on the variety spoken in the Kohima village (Mahapatra et al., 1989; Ravindran, 1974; Ezung, 2018). The Kohima district is shown in Figure 1-3.

1.3.2 Speech data and participants

As the goal of this work is to provide a detailed phonetic analysis of Angami speech sounds, speech databases were created as part of the work. The speech databases used for the investigation comprise recordings from native Angami speakers. A short

description of the databases created for the current study is provided in Table 1.9. Besides the production studies, two perception test were also conducted aimed at finding out the distinctiveness and perceptual correlates of the Angami tones.

The data collection spanned over five years (2019-2023) and involved numerous field visits. Several speakers have participated in data collection throughout the years, and some have participated in more than one data collection study. Seventy-five unique native Angami speakers contributed to the study (see Table A.2 in Appendix B). The age of the participants in the studies ranged from 21 to 66 years, with an average age of 34 (SD=10). There were 36 male and 38 female participants; all were literate with a minimum of a high school diploma. Further details about the participants and their involvement are discussed in each chapter where the experiments are reported.

1.3.3 Data recording and annotation

All speech recordings were conducted in the field, and it was ensured that environmental noises were minimal. To ensure high-quality recordings, the data were collected using a Shure unidirectional head-worn microphone (model: Shure SM10A-CN) connected to a Tascam linear PCM recorder (model: TASCAM DR-100MKII)

Table 1.9: Data distributed in the dissertation. Here, the details column includes the number of unique tokens, the syllable structure and number of speakers

Database name	Database code	Details	Measures
Plosives	Avot	270, CV, 10	VOT, F0 of tone, classification of tones
Nasals	Anl	74, CV, 13	nasal formants
Nasals	Anaso	28, CV, 4	nasalance
Fricatives & Affricates	Aaf	72, CV, 10	spectral moments and features
Lateral & approximants	Apx	37, CV, 10	formants, spectral features
Vowel & Tone interactions	Avt	25, V, 6	vowel formants, F0
Tones	Amt	40, CV, 15	F0
Contextual tones	Act	75, CVCV, 24	F0

Table 1.10: Example of the target word /l^ha/ in three contexts.

Context	IPA	Meaning
a) Sentence	tʃet ^h ie a l ^h a vi se	this year, my grains are good
b) Carrier	a l ^h a puba	I said grains
c) Isolation	l ^h a	grains

through an XLR jack. The sampling frequency at the time of recording was 44.1 kHz, 24 bits in WAV format.

To collect the nasalance data, database Anaso in Table 1.9, Nasometer II 6450 (KayPENTAX medical) was used⁵. The sample and example of the speech recording using Nasometer 11 6450 is shown in Figure 1-4.

After recording of the data, the speech sounds were transferred to a computer and were saved in .wav format, which were subjected to further analysis. Since automatic segmentation was not possible for Angami, all speech sound files were segmented and annotated manually, by means of listening and visual examination of the waveforms and spectrograms using Praat 6.0.43 (Boersma, Paul and Weenink, David, 2020). The sound files were segmented and annotated marking the word, syllable, phoneme, tone boundaries. The annotation for each word accompanied a unique code, as shown in Figure 1-5. Each unique code embedded information about the speaker, speaker's gender, village, target word, token number, repetition number and the context in which the token was produced. An example of the annotation and segmentation of five level tiers is shown in Figure 1-5. Further, all acoustic measures were calculated using a Praat script written for each experiment.

For perception experiments, the experiments were designed using Praat, and they were run on an Apple laptop. The speech stimuli were played through a pair of professional headphones (Audio-Technica ATH-M50x) with frequency range of 15 –

⁵The Nasometer II 6450 (KayPENTAX medical) is located in the Pathological Speech Processing Laboratory of North East Centre for Biological Sciences and Healthcare Engineering (NECBH), IIT Guwahati, funded by the Department of Biotechnology, Government of India, project no. BT/COE/34/SP28408/2018



Figure 1-4: Sample and example of recording of speech data using Nasometer II 6450 (KayPENTAX medical)

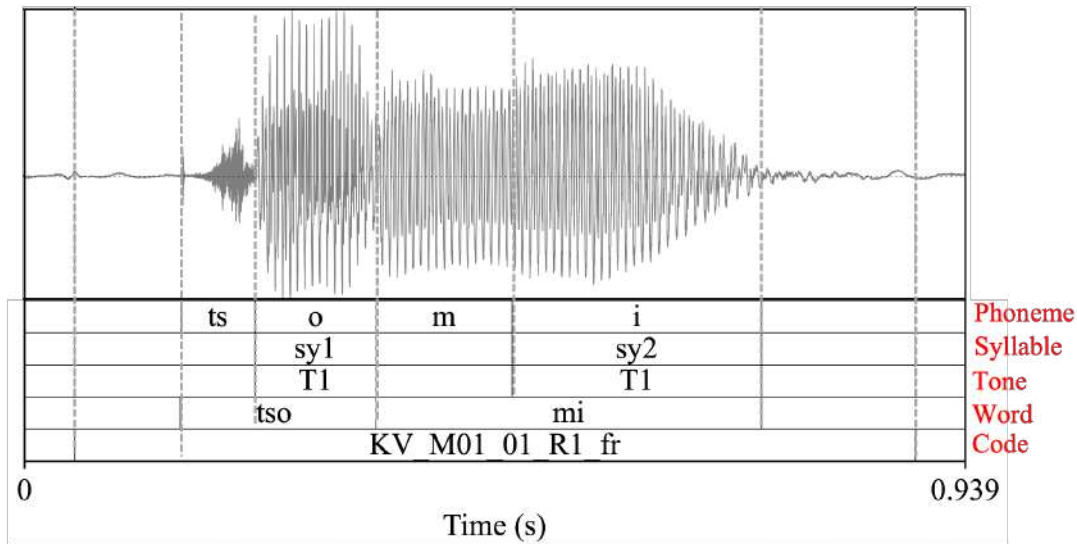


Figure 1-5: An example of segmentation and annotation of the sound files

28,000 Hz. There were two perception experiments where the first task was conducted using natural speech stimuli while for the second experiment, synthesized speech stimuli were used.

1.3.4 Acoustic and temporal measurements

For the acoustic analysis of plosives, the acoustic property focused on Voice Onset Time (VOT), which involves measuring the duration between the release of the oral stop consonant and the onset of the glottal pulse of the subsequent vowel. Meanwhile, the acoustic-phonetic properties of the nasal consonants are measured from the nasal segment and determined by the low intensity compared to the following vowels. The fraction of unvoiced frames in the nasal is measured using the Voice Report function of Praat (Boersma, Paul and Weenink, David, 2020) to determine the amount of voicing in the nasals. Fricatives were measured at the frication period of the sound segment, while affricate measurements include the VOT burst and the frication period. The acoustic-phonetic properties include spectral features (CoG, intensity, percentage of voicing, Harmonics to Noise Ratio) and their moments (skewness & kurtosis). In the study of approximants in Angami, the first four formant frequencies (F1-F4), spectral

features, and spectral moments were measured.

In order to conduct acoustic studies on Angami vowels, we extracted vowel-specific acoustic measurements. Vowel qualities of languages are best represented by the first three formants, where the first formant (F1) is inversely correlated with the height of the vowel. The second formant (F2) is directly proportionate to the frontness of the vowel. Finally, the third formant (F3) is inversely proportionate to the roundness of the vowel. Hence, to see the acoustics of Angami vowels, we have extracted the vowels' F1, F2, and F3 values from the midpoint of the vowels. Apart from the vowel quality acoustics (F1, F2, and F3), we also extracted vowel quantity information as expressed by the duration of the vowels. In case of the Angami tones study, we have extracted the fundamental frequency (F_0) values associated with each tone category at every 2% of the total duration. For each study, the temporal measurements were also measured. The temporal measurement for the study is measured from the onset of the sound segment till the termination. The average value of the duration is used for analysis.

As mentioned, Harmonic-to-Noise Ratio (HNR), Center of Gravity (CoG), skewness, kurtosis are some of the features extracted for the study using the functions in Praat (Boersma, Paul and Weenink, David, 2020). A brief overview of the features are as follows:

- **HNR:** HNR is a measure used in signal processing and speech analysis to quantify the ratio of harmonic components to noise in a signal. In terms of speech, the harmonic components correspond to the fundamental frequency (pitch), and the noise represents any non-harmonic or random components in the signal. HNR can be computed as the ratio of the harmonic power to the noise power, often expressed in decibels (dB).

$$HNR = 10 \log_{10} \frac{P_{harmonic}}{P_{noise}}$$

where:

- $P_{harmonic}$ is the power of the harmonic components in the speech signal.
- P_{noise} is the power of the noise components in the speech signal.
- **CoG:** CoG is a measure used to characterize the spectral properties of a speech signal by providing information about the distribution of energy across different frequency bands. The CoG represents the centroid or average frequency of the spectral energy distribution, indicating where most of the energy in the speech spectrum is concentrated. A higher CoG indicates that the energy is concentrated towards higher frequencies, while a lower CoG indicates that the energy is concentrated towards lower frequencies.
- **Kurtosis:** Kurtosis is a statistical measure used to quantify the “tailedness” or “peakedness” of a distribution compared to a normal distribution. In the context of speech analysis, kurtosis can be used to characterize the shape of the distribution of speech signal samples. It measures how much a distribution deviates from a normal distribution in terms of the distribution of its tails. A high kurtosis value indicates that a distribution has heavy tails and is more peaked than a normal distribution, while a low kurtosis value indicates that a distribution has lighter tails and is less peaked than a normal distribution.
- **Skewness:** Skewness is a statistical measure used to quantify the asymmetry in terms of the degree and direction of a distribution of data around its mean. A positive skewness value indicates that the distribution is skewed to the right, implying that the tail on the right side of the distribution is longer or heavier than the left side. Conversely, a negative skewness value indicates that the distribution is skewed to the left, with a longer or heavier tail on the left side.

1.3.5 F0 Extraction Using Zero-frequency filtering (ZFF)

In order to obtain tone information from the analyzed syllables, F_0 values need to be extracted from the vocalic parts. However, pitch estimation becomes challenging considering the onsets of the syllables contain voiceless aspirated and unaspirated stops. For example, auto-correlation-based Praat pitch tracking often returns null values from the pitch track regions contiguous to the voiceless stops. Often, the aspiration associated with the voiceless stop consonants results in breathiness in the following speech segment, which in turn leads to spurious pitch estimation. However, pitch estimation methods using Zero-Frequency Filtering (ZFF) are known to be robust against such consonantal effects.

Zero-frequency filtering (ZFF) is a technique proposed by Murty and Yegnanarayana (2008), which estimates the instantaneous fundamental frequency (F_0) from the location of the epochs, i.e., the instants of significant excitation (Murty and Yegnanarayana, 2008). Tzudir et al. (2021) reported that the ZFF method captures the F_0 contour with sufficient micro-prosodic information compared to the autocorrelation based Praat method (Tzudir et al., 2021). Therefore, the ZFF method is used in this work to extract the F_0 values of each monosyllabic tone utterance. The speech signal is initially pre-processed to sample at 8 kHz. In this process, the differenced speech signal is passed through a cascade of two ideal resonators in order to reduce the effects of high-frequency resonances. In addition, trend removal is performed by a local mean subtraction process. The resultant signal obtained is sinusoidal in nature, known as the ZFF signal, whose positive zero crossings are used to compute the location of the glottal closure instants (GCIs, henceforth). The inverse between two successive GCIs gives the resultant F_0 . In this study, the F_0 values are extracted using the ground truth from the TBU.

1.3.6 Statistical measurements

To eliminate speaker effects, the raw fundamental frequency and formants were normalized using the z-score normalization method (Rose, 1987). Speech data collected from multiple speakers usually vary across speakers due to anatomical, physiological, and talker-specific factors. These factors systematically affect the fundamental frequency and the formant frequencies of vowels and tones. This variation arising from speaker-specific features is considered an artifact, and speech data is usually processed with speaker normalization (Peterson and Barney, 1952). In a study by Adank et al. (2004), the z-score method of frequency normalization used by Lobanov (1971) for vowel normalization has been found to be the best vowel normalization metric. Lobanov’s z-score transform is a vowel extrinsic, talker intrinsic normalization procedure, reducing the anatomical variation of the formant frequencies but maintaining phonological and sociolinguistic variation. The formula for speaker normalization using the Lobanov normalization method can be written as in Eq. (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 .

$$F_{n[V]}N = \frac{F_{n[V]} - MEAN_n}{S_n} \quad (1.1)$$

Hence, in this study, before proceeding with the analysis of the formant frequencies, the data were normalized with the Lobanov method for speaker effects by using the `normLobanov` function from the *phonR* package on the open-source R platform (R Core Team, 2022).

For exploratory statistics, Linear Mixed Effects (LME) models were built for analyzing the quantitative data, using the `lme4` package (Bates et al., 2015) on the R platform (R Core Team, 2022). We built full models by including a maximum number

of fixed and random effects in the models. All models were subjected to backward reduction of effects using the *step* function of the *lmerTest* package (Kuznetsova et al., 2017b). The reduced models were subjected to a Type II Wald χ^2 test for analysis of deviance using the *car* package on R (Fox and Weisberg, 2019). The test results provide χ^2 values corresponding to each fixed effect, where a higher χ^2 value would imply more class separability. The results also provide a probability value that ascertains whether a fixed effect is significant. Each model was also subjected to a pairwise comparison test using the *emmeans* package (Lenth, 2019) with Bonferroni adjustment to explore the significance of difference among the levels in the fixed factor. Skewness and kurtosis were calculated to determine the density and distribution of the data.

All values extracted for each sound segment were plotted for visualization and produced on R with the *ggplot2* package (Wickham, 2016). For the classification of the tones, a Random Forest classification experiment was set up using the R software with the *randomForest*, *caret* and *e1071* packages (Liaw and Wiener, 2002; Kuhn, 2021; Meyer et al., 2021). The statistical analyses of the acoustic data are presented towards the end of each chapter of this dissertation.

1.4 Organization of the dissertation

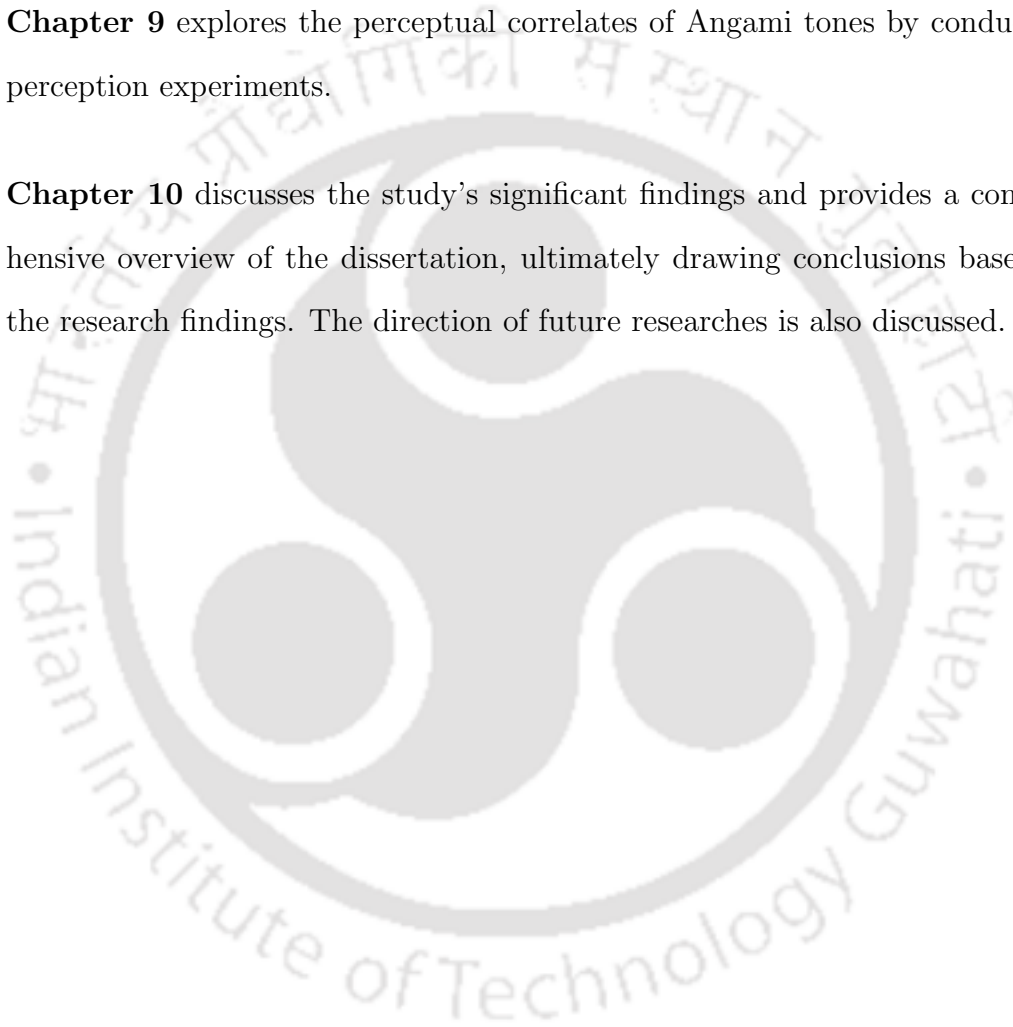
The dissertation is structured into ten chapters, organized as follows.

- **Chapter 1** comprises a comprehensive background of the Angami language, offering insights into its history and linguistic characteristics. Additionally, it provides an overview of the current study, outlining its objectives and scopes. Furthermore, it reviews the existing literature on the Angami language and briefly summarizes its phonological features. The chapter also presents a brief overview of the methodology employed in this dissertation.

- **Chapter 2** delves into the acoustic properties of plosives in Angami, with a specific focus on Voice Onset Time (VOT).
- **Chapter 3** investigates the acoustic properties of nasals in Angami, mainly focusing on the difference between voiced and voiceless nasals through various acoustic-phonetic parameters. Additionally, it conducts a detailed examination of Angami nasals using the Nasometer II.
- **Chapter 4** presents the acoustic characteristics of fricatives and affricates in Angami by analyzing parameters such as spectral moments and features, which include the skewness, kurtosis, duration, center of gravity (COG), intensity, and voicing. The study investigates each sound segment's phonation types and place of articulation to understand the fricatives and affricates in Angami comprehensively.
- **Chapter 5** delves into the acoustic properties of laterals and approximants in the Angami language, examining distinctions between voiced and voiceless variants and variations in place of articulation (POA) within these sound segments.
- **Chapter 6** provides an in-depth exploration of the vowel inventory in Angami. It examines the acoustic properties of monophthongs and diphthongs, focusing on parameters such as formant frequencies and duration to comprehensively analyze these vowel sounds.
- **Chapter 7** addresses the issues pertaining to Angami tones and conducts acoustic analyses utilizing fundamental frequency and duration measurements. In this chapter, the interaction of tones and segments are investigated. Additionally, it delves into the effects of contextual tones within the language and determines the types of effects. The chapter also explores the interaction of vowels and

tones, the intrinsic F_0 (IF_0).

- **Chapter 8** introduces a classification approach for Angami tones utilizing the random forest classifier. Multiple F_0 features are extracted and used for the classification of Angami tones.
- **Chapter 9** explores the perceptual correlates of Angami tones by conducting perception experiments.
- **Chapter 10** discusses the study's significant findings and provides a comprehensive overview of the dissertation, ultimately drawing conclusions based on the research findings. The direction of future researches is also discussed.



Chapter 2

Acoustics of Angami plosives

*Angami features a three-way contrast in laryngeal stops: voiced, voiceless unaspirated, and voiceless aspirated. Acoustic-phonetic analysis, mainly focusing on Voice Onset Time (VOT), is investigated for the study of the Angami stops. The findings indicate that voiced stops exhibit longer negative VOT, whereas voiceless unaspirated and voiceless aspirated stops display positive VOT. Additionally, the study underscores that laryngeal contrasts and place of articulation significantly affect VOT in Angami.*¹

2.1 Introduction and literature review

Since voice onset time (VOT) was proposed by Lisker and Abramson almost 50 years ago (Lisker and Abramson, 1964), numerous studies have been conducted on the major languages of the world, providing VOT characteristics in stop contrasts (Abramson and Whalen, 2017; Cho and Ladefoged, 1999; Flege and Eefting, 1986; Hussain, 2018; Lisker and Abramson, 1970). Of late, typological data for VOT of under-resourced languages have also started to emerge (Kirby and Hyslop, 2019; Lalhminghlui and

¹Part of this chapter has been published in a peer-reviewed journal, “Voice Onset Time in Angami”, in *Linguistics of the Tibeto-Burman Area*, volume 47, issue 1 (2024). John Benjamins Publishing Company.

Sarmah, 2020; Lee et al., 2019; Teo et al., 2014; Zakaria, 2020) as a result of that, the universals and variations in VOT can now be better characterized for both well-known and lesser-known languages. In this work, we add to the increasing typology on the subject by reporting the characteristics of VOT in Angami (Tenyidie), a lesser-described tone language of the Tibeto-Burman language family. Angami has nine stop consonants namely, /p, p^h, b, t, t^h, d, k, k^h, g/. To the best of our knowledge, not many studies have provided a detailed analysis of VOT in the Tibeto-Burman languages. Hence, this study attempts to fill the research gap.

While VOT was defined initially as the time between the release of a stop consonant and the onset of subsequent glottal pulsing, in recent times, it has also been extended to include the intervocalic and final position of the stop consonants (Abramson and Whalen, 2017; Lisker and Abramson, 1964). VOT captures the temporal characteristics associated with the phonation type of the stop consonants. Hence, voiced stops would have significant voicing lead with negative VOT values in relation to the point of stop release, which is considered the zero-point. A voiceless stop, however, would have voicing lags and positive VOT values. Unaspirated voiceless stops, however, have substantially lower VOT values (Abramson and Whalen, 2017; Lisker and Abramson, 1964). However, cross-linguistic studies on VOT have shown that while VOT characteristics for different phonation categories are comparable, there are language-specific differences.

Lisker and Abramson (1964), categorized 11 languages they studied into three groups based on their number of laryngeal contrasts. These groups comprised languages with two, three, and four-way laryngeal contrasts. It was shown that, within a language, each laryngeal contrast could be associated with a unique VOT value when produced in isolation. For example, for voiced stops, the VOT value is between -125 and -75 ms; for unaspirated voiceless stops, it is between zero and 25 ms; for aspirated voiceless stops, it is +75 ms. In the initial studies, VOT boundaries were

categorically defined; however, in the latter studies, it was found that VOT values may not be absolute.

Further, language-specific VOT differences were also reported in Mandarin and English studies (Cho and Ladefoged, 1999). For example, Cho and Ladefoged (1999) has shown that VOT values of the voiceless velar stops in different languages vary substantially. They showed that voiceless velar stops may be produced as unaspirated, slightly aspirated, aspirated, and highly aspirated. Similar findings are also reported for Mandarin and English, where voiceless aspirated stops in Mandarin and English are significantly different regarding their VOT values (Chao and Chen, 2008). While voiceless unaspirated stops (p, t, k) occupy the same range in the two languages, for voiceless aspirated stops (p^h, t^h, k^h), Mandarin falls in the highly aspirated category, in terms of Lisker and Abramson (1964).

Recently, Hussain (2018) reported VOT characteristics of ten lesser-studied Indo-Iranian languages. He reported VOT consistently differentiating languages with two or three laryngeal contrasts. However, as more than three laryngeal contrasts appeared, such as voiceless unaspirated vs. voiceless aspirated vs. voiced unaspirated vs. voiced aspirated vs. voiced implosives, in languages like Jangli, Urdu, Sindhi and Siraiki, the differences in VOT decreases as supplementary acoustic correlates such as voice quality or release characteristics distinguish the laryngeal categories and places of articulation (Hussain, 2018).

Languages that make three-way phonation distinctions in stops, namely, pre-voiced, voiceless unaspirated, and voiceless aspirated, are relatively rare across languages. PHOIBLE 2.0 reports that of 451 languages reported, only 38 have a three-way distinction in stop contrasts (Moran and McCloy, 2019). Nevertheless, such contrasts are common in the Sino-Tibetan languages. In North-East India, Tibeto-Burman languages widely attest to three-way distinction in the phonation of stops. Three-way contrast in stops is attested in Hyow spoken in Bangladesh, Mizo in

Mizoram, Sumi, and Angami in Nagaland, India (Kirby and Hyslop, 2019; Lalhminghlui and Sarmah, 2020; Lee et al., 2019; Teo et al., 2014; Zakaria, 2020), to name a few. Despite that, acoustic studies of phonation in these languages are sporadic. The two Tibeto-Burman languages we have come across with VOT studies are Drejongke (Bhutia), one of the official languages of the Sikkim state of India, and Diaspora Bhutanese spoken in Kathmandu valley, Nepal (Geissler, 2019; Lee et al., 2019). It is reported that there are four laryngeal contrasts, namely, aspirated, voiceless, voiced, and “devoiced” in Drejongke. While aspirated & and voiceless have positive VOT, voiced with negative VOT, the category “devoiced” is said to be a variable realized with either negative or positive VOT. It is seen that high F_0 is attested to voiceless unaspirated and voiceless aspirated stops, and low F_0 is attested to voiced and devoiced laryngeal category. In the study of diaspora Tibetans, it is reported that high tone is said to affect aspirated stops. VOT is said to be longer in aspirated stops with high tones (mean 72.7 ms) than in low tones (mean 48.1 ms) (Geissler, 2019). In General, voiceless and aspirated obstruents were associated with high F_0 . On the other hand, voiced obstruents are associated with low F_0 (Kirby and Hyslop, 2019; Lee et al., 2019).

Cross-linguistically, studies have reported that VOT is affected by place of articulation (Lisker and Abramson, 1964; Cho and Ladefoged, 1999), prosodic position (Lisker and Abramson, 1967; Van Dam, 2003; Simonet et al., 2014; Kirby and Ladd, 2016), rate of speech (Kessinger and Blumstein, 1997; Boucher, 2002) and the vowel and sonorants (Klatt, 1975). VOT is also affected by the F_0 of the tones, as reported in tonal language such as Thai (Gandour, 1974; Ewan and Kronen, 1974; Maddieson, 1978a), Yoruba (Hombert, 1978), Mandarin (Xu and Xu, 2003; Lai and Jongman, 2004), Hongkong Cantonese (Tse, 2005) and Mizo (Lalhminghlui and Sarmah, 2020). As Angami is a tone language, the interaction of VOT and tones are examined in detail in Chapter 7 (Section 7.4.1).

This chapter is organized in the following manner. Section 2.1 contains the introduction to VOT and the study's objective. Section 2.2 describes the study's methodology, and the study results are reported in section 2.3. The chapter is concluded with a discussion in section 7.4.1.4

2.1.1 Literature review of Angami plosives

Angami stops make a three-way laryngeal contrast, i.e., voiced, voiceless unaspirated, and voiceless aspirated stops, as seen in Table 2.1. Each of these contrasts has three places of articulation (POA), namely, bilabial, denti-alveolar, and velar. In this chapter, the VOT of stops in the three POA with the three laryngeal contrasts is explored. The only previous work in Angami VOT is based on the Khonoma variety (Blankenship et al., 1993). It is reported in this work that there are three voicing contrasts in Angami, namely, voiced, unaspirated voiceless, and aspirated voiceless. However, in terms of the places of articulation, the work provides three contrasts in Khonoma Angami, namely, bilabial, dental² and velar. VOT values are known to differ based on their POA, Cho and Ladefoged (1999) shows that cross-linguistically, the bilabial stops have the shortest VOT value, followed by alveolar stops, while velar stops have the longest VOT values. However, this universal pattern is not attested in the Khonoma Angami as reported in (Blankenship et al., 1993). For example, in the case of voiced stops, bilabials have higher VOT values than the velars in both isolation and sentence frames. In the case of voiceless aspirated stops produced in isolation, the bilabials are marginally shorter than the velars. The inconsistency in the VOT measurements, as reported in Blankenship et al. (1993), motivates us to look into the VOT of Angami stops more carefully and in detail.

²Blankenship et al. (1993) provided two types of place of articulation in Khonoma Angami stops in the same paper, while listing the consonant inventory of the language, sounds such as, /t/, /t^h/ & /d/ were placed under alveolar, the same were categorized as dental under the VOT section.

2.2 Methodology

This section describes the participants, the materials (database Avot), the recording procedure, and the annotation of this study.

2.2.1 Speakers

Ten Angami native speakers (5 Male and 5 Female) born and raised in the Kohima village, Nagaland, took part in this experiment. The speakers were chosen from Kohima village, as the variety spoken in this area is regarded the standard variety of the Angami language. The study area in the Kohima district of Nagaland, North-East India. The average speakers' age is 27.6 years (Standard Deviation = 2.8) at the recording time. All speakers were born and brought up in Kohima village, where Angami (Tenyidie) is the mother tongue (L1). All participants can also speak English (state official language) and Nagamese (lingua franca) besides Angami. All participants had a minimum education of a bachelor's degree, and some were working professionals. They were all proficient readers of Angami. The participants signed a consent form and received payment for their participation in the study.

Table 2.1: Example of stop minimal set in Angami.

Laryngeal contrasts	Stops	Word	Meaning
Voiceless unaspirated	p	pa	pick
	t	ta	run
	k	ka	serve
Voiceless aspirated	p ^h	p ^h a	tie
	t ^h	t ^h a	stand
	k ^h	k ^h a	stop
Voiced unaspirated	b	ba	seat
	d	da	cut
	g	ga	leaf

Table 2.2: Distribution of consonants in the five tone categories included in this study.

Tones	/p/	/t/	/k/	/p ^h /	/t ^h /	/k ^h /	/b/	/d/	/g/	Total
T1	118	119	119	119	118	119	115	118	115	1061
T2	118	118	120	119	120	120	117	119	119	1070
T3	119	118	120	119	120	120	114	118	118	1066
T4	118	117	120	119	120	118	114	117	118	1061
T5	117	115	119	118	119	117	112	117	117	1051
Grand Total										5309

2.2.2 Materials

The material was designed to capture the VOT in the nine stops of Angami and its interaction with the F_0 of the tones. Data used for this study consist of the three laryngeal contrast in Angami, i.e, voiced, voiceless and voiceless aspirated in three places of articulation, namely, bilabial (p, p^h & b), denti-alveolar (t, t^h & d) and velar (k, k^h & g). The laryngeal/consonant sounds that were of the target were monosyllabic (CV). All the nine consonants were followed by six vowels (i, e, a, o, u & ə) independently produced, which is combined with each of the five tones in Angami, namely, T1-T5 as reported in the standard Angami dictionary, MKS Dieda (Liezietsu et al., 2019). Here, T1 represents the high tone, T5 represents the low tone, whereas T2, T3 & T4 are the intermediate tones. The corresponding representation of tones in Chao's tone numerals are 55 (T1), 44 (T2), 33 (T3), 22 (T4) and 11 (T5). The tones in Angami are discussed in detailed in Chapter 7. The target tokens (270) were repeated twice by ten speakers. Altogether, 5400 tokens were recorded: 9 consonants x 6 vowels x 5 tones x 2 repetitions x 10 participants. Of the 270 unique tokens, 153 potentially related to an Angami lexical word, and the remaining 117 may not be associated with any lexical meaning.

After removing mispronounced tokens, the distribution of the consonants by the tone categories is as shown in Table 2.2.

2.2.3 Recording of data

Using a meaningful lexical word /pe/ with each of the five potential tones as a cue, the participants were requested to imitate the tone of the cue in producing the tones on the monosyllabic CV combinations. The word /pe/, which has five distinct meanings depending on the tone, was used as a cue for producing the tones on the CV monosyllables. The word /pe/ associated with T1 means ‘incline’; T2 means ‘bridge’, T3 means ‘march’, T4 means ‘shiver’ and T5 means ‘shoot’. While recording the data, /pe/ with one of its five meanings was displayed on a computer screen for the speakers to receive the cue for the tone. The screen displayed the target syllable on which the speakers were supposed to implement the tone of the cued word. Even though most of the syllables produced were not associated with lexical meaning, the speakers performed the tone-mimicking task effortlessly.

A total of 5400 in tokens were elicited, of which 5309 were used for analysis because the tokens for 91 were mispronounced and were therefore excluded. Table 2.2 displays the distribution of tokens taken into consideration for analysis in this study. The recordings were conducted in the speaker’s home in a noise controlled environment. Each person took about an hour for the recording session. The data were collected at two different times, first in October 2019 and later in March 2021. The data were collected using following the recording procedure discussed in Chapter 1 (Section 1.3.3).

2.2.4 VOT annotation and extraction

For analysis, the speech recordings were uploaded to a computer. The sound files were automatically segmented using Dr.VOT (Shrem et al., 2019), which is designed to indicate the boundaries for VOT. However, while Dr.VOT gave correct boundaries for VOT in voiceless unaspirated stop consonants, its accuracy was lower for voiceless aspirated stops and voiced stops. Therefore, the aspirated voiceless and the voiced

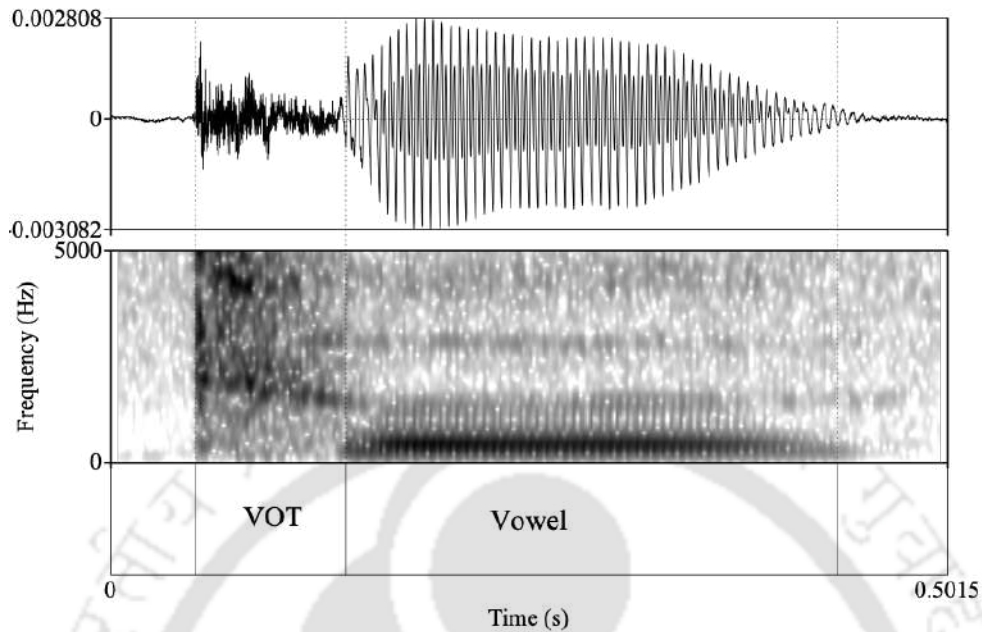


Figure 2-1: An example of the segmentation of the VOT and vowel boundary for /t^hu/.

stops were manually segmented for the VOT boundaries using Praat 6.0.43 to listen and visually inspect the waveforms and spectrograms (Boersma, Paul and Weenink, David, 2020). Additionally, the boundaries for voiceless unaspirated stops, determined by Dr.VOT, were checked for errors. As seen in Figure 2-1, the VOT boundaries, along with the vowel boundaries, are annotated. A Praat script was used to extract VOT durations and transfer them to an Excel for further analysis.

2.2.5 Statistical measurement

The statistics analysis were conducted following the standard procedure as discussed in Chapter 1 (Section 1.3.6).

2.3 Results

Angami comprises three laryngeal contrasts, as was previously mentioned in Section 1. The waveforms and spectrograms are extracted from Praat in order to confirm the VOT boundaries of the contrasting Angami stops. The waveforms of the syllables /pa/, /p^ha/ & /ba/ are seen in Figure 2-2. The area between the two dashed lines in the figures indicates the VOT of the stops. The figure illustrates that the bilabial voiceless unaspirated stop, /p/, has the least VOT value since it contains only a burst, whereas the voiceless aspirated stop, /p^h/, is longer as it has a following aspirated part. On the other hand, the bilabial voiced stop, /b/, has a significant voicing lead, characteristic of such stops. In order to analyze the VOT characteristics of Angami stops, we extracted VOT values derived from 5309 tokens containing the nine Angami stops in CV syllables. The following subsection presents density plots of the VOT values to visualize the categorical differences among the Angami stops. Following that, the average VOT values are presented for all the subjects in the study. As Angami is a tone language, in the subsequent subsection, the interaction of tones and VOT in the language is discussed. The density of the VOT across the nine Angami stops is shown in Figure 2-3.

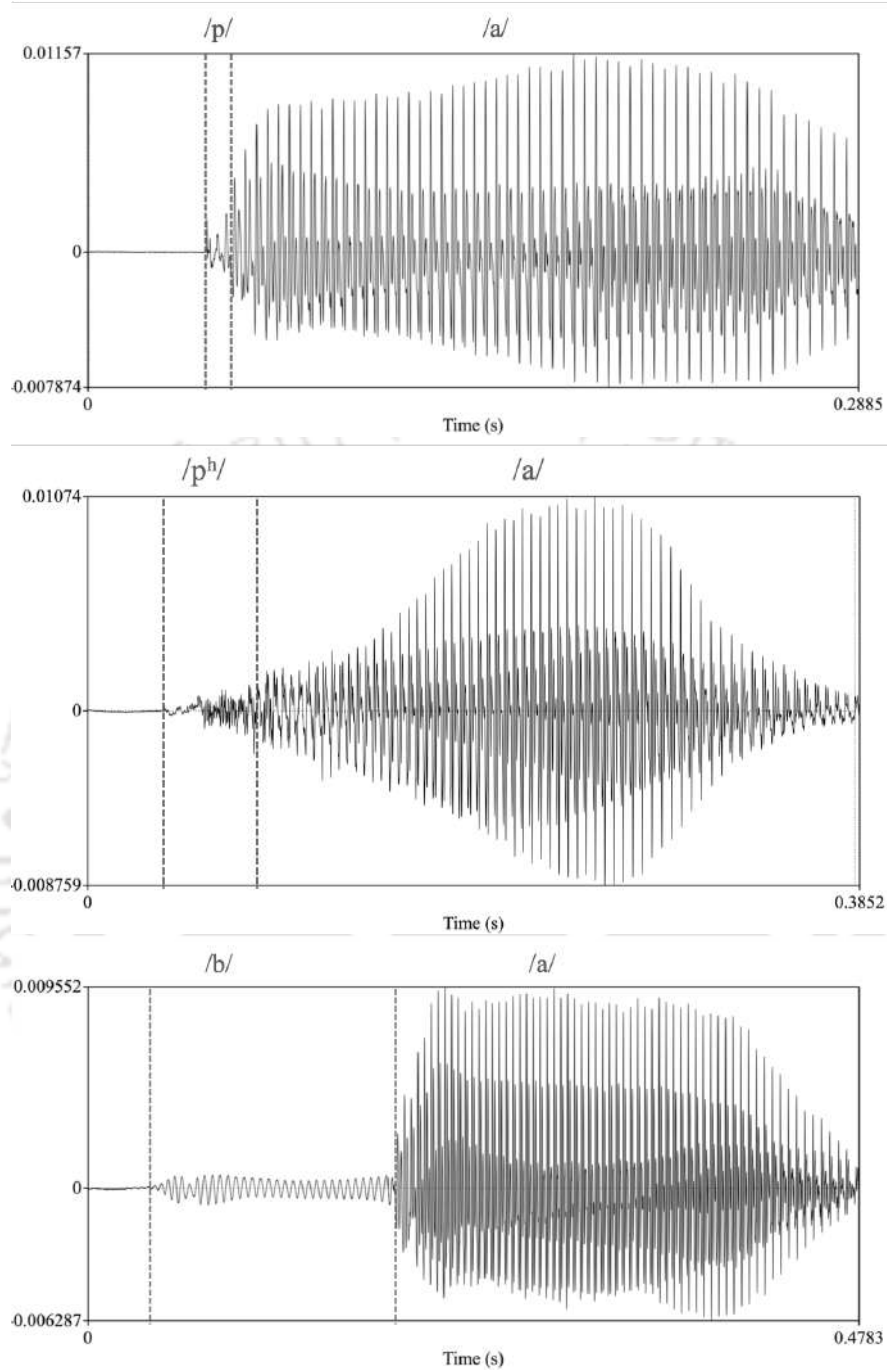


Figure 2-2: An example of the VOT and vowel boundary in three laryngeal contrast of Angami.

2.3.1 VOT in Angami stops

In order to visualize the VOT difference in the nine Angami stops, density plots for VOT are presented in Figure 2-4. The figures also accompany the kurtosis and skewness values for each stop, representing the normalcy of the distribution of the VOT values. Skewness indicates symmetry of the data, whereas kurtosis indicates the magnitude of the tails. A symmetrically distributed data has skewness between -0.5 and 0.5. While kurtosis values around 3 indicate a normal distribution, values more than 3 and less than 3 are short-tailed and heavy-tailed distribution indicators, respectively. As seen in the figure, voiced /b/, /d/ and /g/ have distributions peaking at -104, -112 and -115 ms, respectively, which are symmetrical considering their skewness values. As indicated by the kurtosis values, /b/ and /g/ have normal VOT distribution, while /d/ is slightly heavy-tailed. Similarly, the voiceless unaspirated stops, /p/, /t/, and /k/, have VOT distributions peaking at 11, 13, and 36 ms, respectively. The VOT values for /t/ are moderately skewed, and for /p/ and /k/, they are highly skewed. The high kurtosis values indicate that the VOT of the voiceless unaspirated stops have heavy-tailed distribution. Aspirated stops /p^h/, /t^h/, and /k^h/ have VOT distributions with peaks at 76, 66, and 105 ms, respectively. While the distribution of VOT data for /k^h/ is symmetrical, data distributions for /p^h/ and /t^h/ are moderately skewed, however, the kurtosis values indicate that VOT values are normally distributed.

The average VOT values for all nine plosives are shown in Table 2.3. Voiced unaspirated stops have the highest negative VOT values, with -110, -116, and -127 ms for /b/, /d/, and /g/, respectively. As seen in the table, voiceless unaspirated stops have the shortest VOT values, with 15, 17 and 39 ms for /p/, /t/, and /k/; voiceless aspirated stops have a relatively higher VOT with 82, 77, 104 ms for /p^h/, /t^h & /k^h/ respectively.

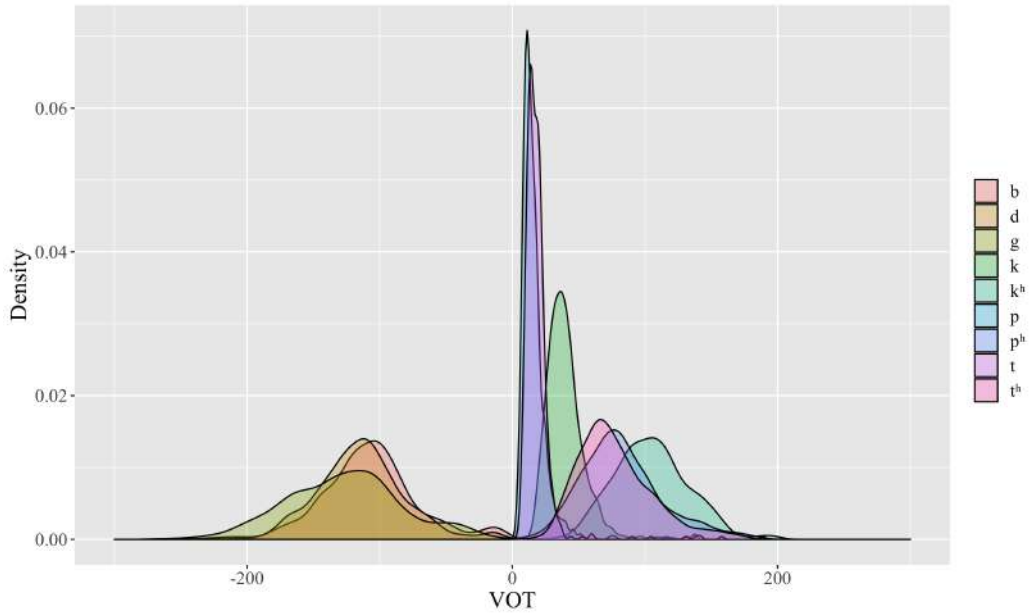


Figure 2-3: Kernel density estimation plots of VOT in Angami stops.

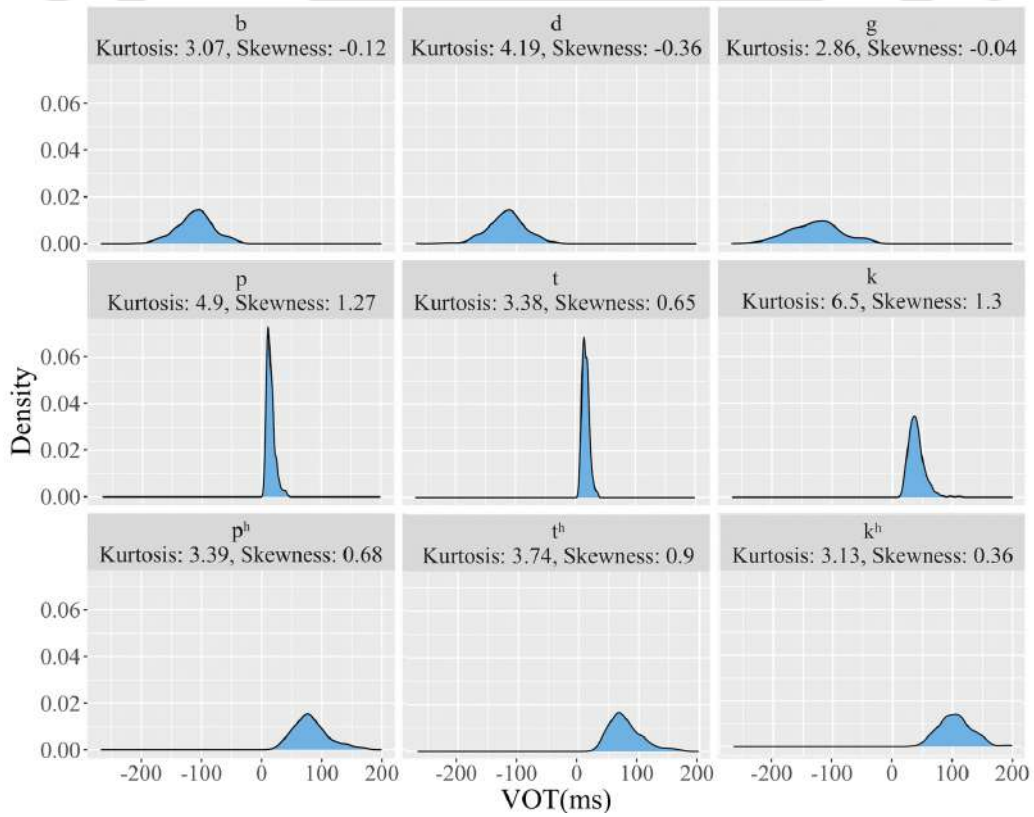


Figure 2-4: Skewness and Kurtosis value of the VOT density across the nine Angami stops.

Table 2.3: The average (avg) VOT duration in milliseconds with their standard deviation(sd).

	/p/	/t/	/k/	/p ^h /	/t ^h /	/k ^h /	/b/	/d/	/g/
avg	15	17	39	82	77	104	-110	-116	-127
sd	(7)	(6)	(13)	(29)	(27)	(28)	(29)	(31)	(41)

Table 2.4: The average (avg) VOT duration in milliseconds with their standard deviation(sd) across speakers.

Speakers		/p/	/t/	/k/	/p ^h /	/t ^h /	/k ^h /	/b/	/d/	/g/
F01	avg	19	17	46	72	70	109	-148	-147	-168
	sd	8	4	17	30	28	25	21	31	34
F02	avg	12	11	30	76	65	90	-113	-119	-128
	sd	4	2	10	16	15	21	34	37	33
F03	avg	13	15	35	82	71	103	-105	-107	-123
	sd	6	6	9	18	18	20	25	23	35
F04	avg	17	16	42	132	122	138	-122	-123	-137
	sd	7	4	13	22	27	26	21	21	30
F05	avg	16	17	39	69	65	95	-105	-113	-123
	sd	6	3	13	16	15	20	20	21	33
M01	avg	11	13	37	65	63	89	-110	-120	-131
	sd	3	4	11	18	17	20	28	20	33
M02	avg	18	23	44	83	75	108	-105	-118	-115
	sd	5	4	10	15	16	16	28	27	48
M03	avg	9	15	36	99	91	114	-119	-135	-158
	sd	3	4	11	28	25	25	16	27	39
M04	avg	18	20	37	53	55	77	-75	-75	-78
	sd	7	5	13	16	14	19	27	21	32
M05	avg	14	19	43	89	93	122	-98	-105	-107
	sd	5	5	9	22	20	23	15	18	21

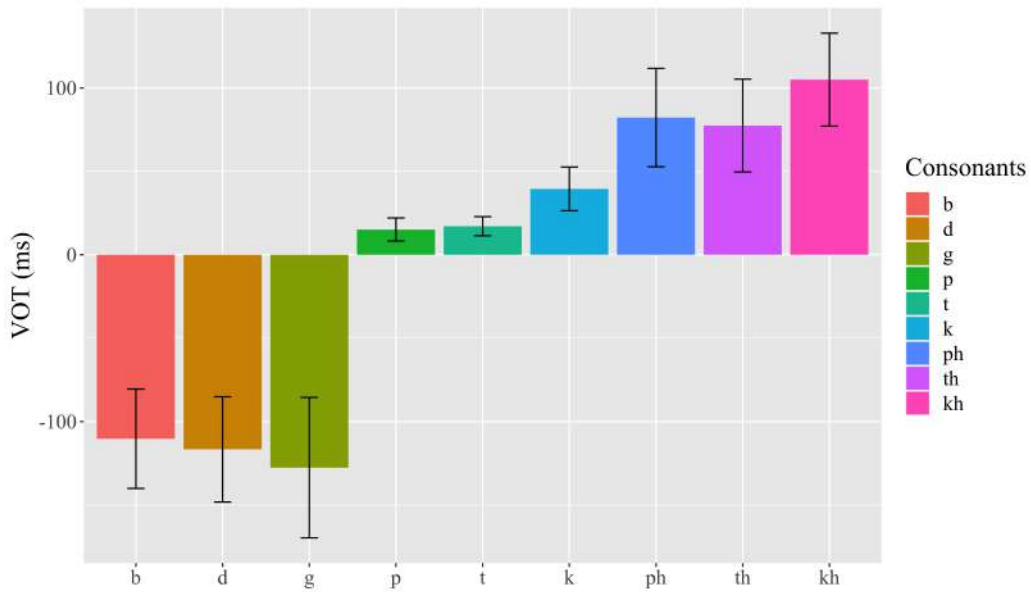


Figure 2-5: The average VOT duration in milliseconds for the nine Angami plosives with standard deviation.

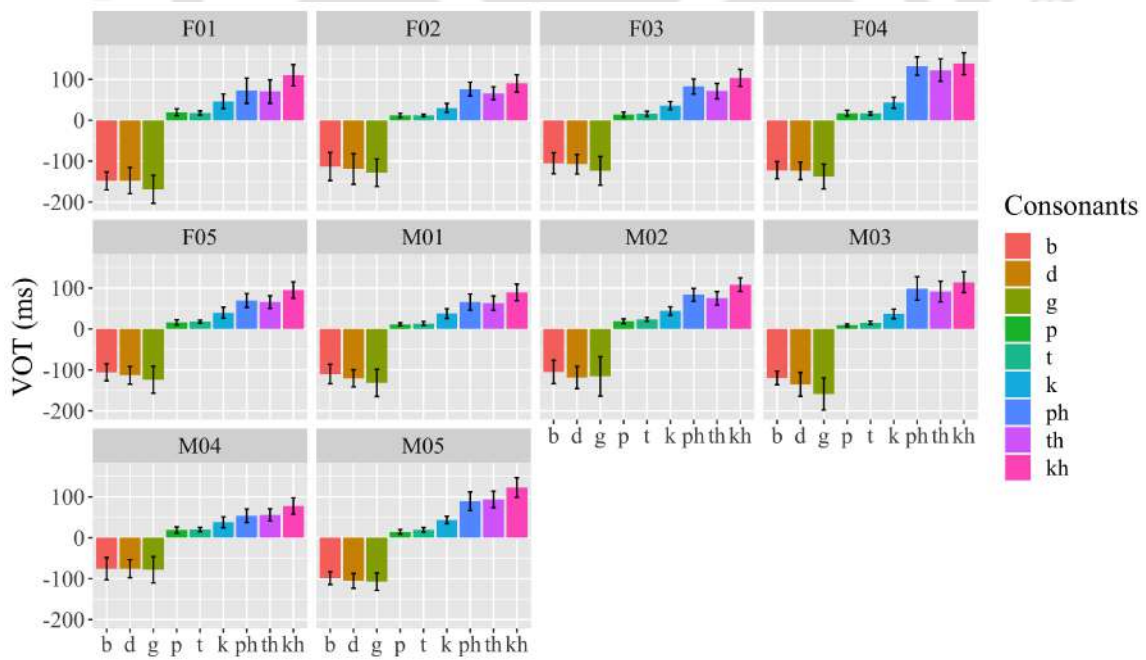


Figure 2-6: The average VOT duration in milliseconds of the 10 Angami speakers.

Figure 2-5 shows the average VOT for the nine Angami plosives and their standard deviations. As observed, voiced stops have significant voicing leads in negative values, while voiceless unaspirated stops and voiceless aspirated stops have voicing lags with positive VOT values. Additionally, it was seen that in all categories, the velar stops are consistently longer than the bilabial and denti-alveolar stops. Figure 2-6 shows subject-wise VOT values for Angami stops and the values provided in Table 2.4. The figure confirms that VOT patterns observed from the average data are indeed attested by individual speakers. While the absolute VOT values are not similar, the trend shown by the VOT values is largely similar.

In order to confirm the statistical significance of VOT differences, we constructed a linear mixed-effect (LME) model with VOT values as dependent variable. To explore the interactions, tones (T), gender, place of articulation, vowels (V), laryngeal contrasts (LC), real/nonce lexeme were considered as fixed effects. Word (W), speaker (Sp), and iteration (I) were considered random effects. Backward reduction using the step function yielded the final model as in 2.1 (Kuznetsova et al., 2017a). Fixed effects such as gender, real/nonce lexemes and random effects such as word, speaker and iteration were eliminated in the backward reduction process.

$$\begin{aligned}
 VOT \sim & L.contrasts + POA + tones + vowels + L.contrasts * POA \\
 & + L.contrasts * Tones + L.contrasts * vowels + POA * vowels + \\
 & L.contrasts * POA * Vowel + (1|Sp) + (1|word) + (1|iteration)
 \end{aligned} \tag{2.1}$$

The results of the Type II Wald chi square tests for analysis of deviance conducted on the LME model is presented in Table 2.5. The three laryngeal contrasts, POA, tones, vowels and their interaction show statistically significant effects on VOT. The results confirm that while the VOT of the three laryngeal contrasts are distinct, they may be different depending on the POA, vowel and tones. Hence, looking into the

pairwise comparisons of VOT by laryngeal contrasts and place of articulation (POA) is pertinent. The variable interaction plot generated from the model is shown in Figure 2-7 which shows the significant effect. In order to see the pairwise comparisons, the LME model was subjected to a Bonferroni post-hoc test. As seen in the Table 2.6, in case of voiced stops, denti-alveolar, bilabial, and velar are significantly different from each other. In the case of voiceless unaspirated stops, except the denti-alveolar and bilabial stops, POA and laryngeal contrasts are significantly different in terms of VOT. In case of voiceless aspirated stops, denti-alveolars, bilabials, and velars are significantly different from each other in terms of VOT. This explains why POA effects on VOT are statistically significant.

Table 2.5: Results of an ANOVA conducted on an LME model with VOT as the dependent variable and place of articulation (POA), laryngeal contrast, tones, vowels and their interactions as fixed effects. Only statistically significant interactions are shown in this table.

Fixed Effects	<i>df</i>	χ^2	<i>p</i> -value
Contrasts	2	66534.16	0.001
POA	2	285.27	0.001
Tones	4	11.42	0.05
Vowels	5	55.76	0.001
Contrasts x POA	4	667.10	0.001
Contrasts x Tones	8	88.14	0.001
Contrasts x Vowels	10	107.64	0.001
Contrasts x POA x Vowels	20	41.42	0.01
Conditional $R^2 = .93$, Marginal $R^2 = .92$			

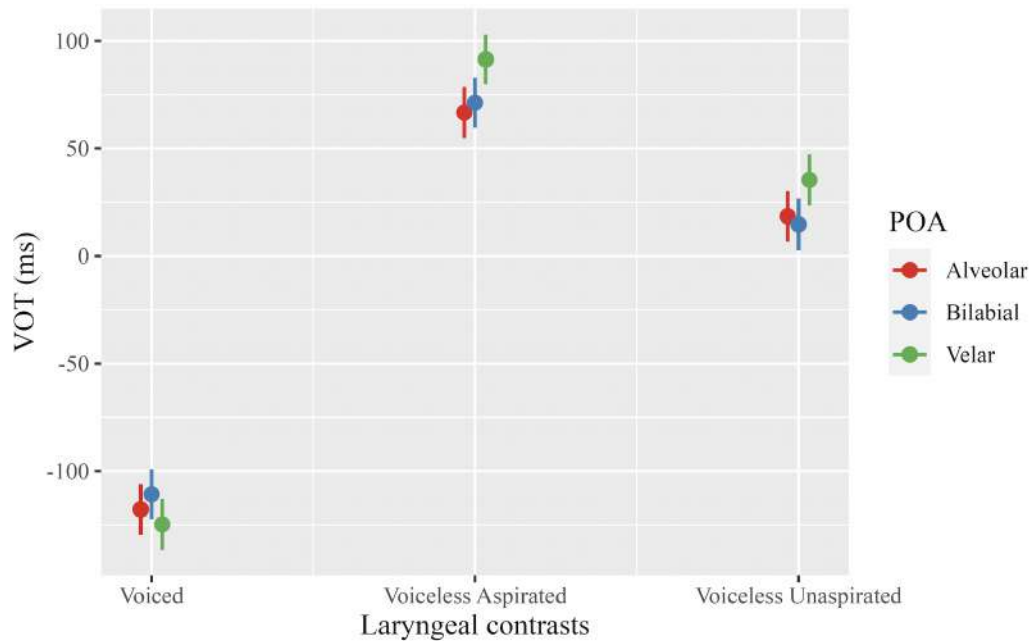


Figure 2-7: Variable interaction plot generated from an LME model with VOT as the dependent variable and laryngeal contrast and PoA as fixed effects.

Table 2.6: Results of a Bonferroni post-hoc test conduct on an LME model with VOT as dependent variable and POA and laryngeal contrast as fixed factors (df=5297).

Contrasts	Estimate	SE	<i>t</i> -ratio	<i>p</i> -value
Voiced				
Denti-alveolar-Bilabial	-6.35	1.49	-4.27	< 0.001
Denti-alveolar-Velar	10.95	1.48	7.41	< 0.0001
Bilabial-Velar	17.30	1.49	11.62	< 0.0001
Voiceless Unaspirated				
Denti-alveolar-Bilabial	1.99	1.48	1.35	n.s.
Denti-alveolar-Velar	-22.41	1.47	-15.23	< 0.0001
Bilabial-Velar	-24.41	1.47	-16.61	< 0.0001
Voiceless Aspirated				
Denti-alveolar-Bilabial	-4.81	1.47	-3.28	< 0.05
Denti-alveolar-Velar	-27.49	1.47	-18.73	< 0.0001
Bilabial-Velar	-22.67	1.47	-15.43	< 0.0001

Table 2.7: Average VOT in milliseconds with their standard deviations across the three laryngeal categories in Angami.

Unaspirated Voiced /b, d, g/	Unaspirated voiceless /p, t, k/	Aspirated voiceless /p ^h , t ^h , k ^h /
-118 (34)	24 (9)	88 (28)

2.4 Discussion

In this chapter, the voice onset time (VOT) in the nine Angami stops have been analyzed. These nine stops have three places of articulation, namely, bilabial, denti–alveolar, and velar, and three laryngeal types, namely, voiced, voiceless aspirated, and voiceless unaspirated. Using both lexical and non-lexical tokens, this study is an exhaustive account of VOT in Angami stops.

The overall patterns noticed for Angami in this study are similar to the patterns noticed cross-linguistically. The voiced stop has a large negative VOT (–118 ms), while voiceless unaspirated and aspirated stops have positive VOT (24 ms and 88 ms, respectively). The average VOT values of the laryngeal contrasts and the standard deviation is shown in Table 2.7. These numbers are slightly different compared to the last known quantitative study on Angami VOT. For example, in Blankenship et al. (1993), the VOT values for voiced, voiceless unaspirated, and voiceless aspirated, produced in isolation, are -140 ms, 13 ms, and 77 ms, respectively. This variation may occur due to the difference in the number of speakers and amount of data in the two studies; for instance, Blankenship et al. (1993), analyzed 175 tokens in isolation for Angami VOT. Compared to that, in the current study, we have reported VOT data from 5309 tokens collected from 10 speakers. The difference may also arise due to the dialectal differences between Khonoma Angami and Kohima Angami (this study). It is also possible that there have been changes in the VOT characteristics in the Angami speech in the last 30 years. However, confirming any of these hypotheses would require further investigation.

Chapter 3

Acoustics of Angami nasals

The voiced and voiceless contrasts in Angami nasals is discussed in this chapter. Voiceless nasals are relatively rare and less studied cross-linguistically. This chapter attempts to study the acoustic properties of voiced and voiceless nasals in Angami, produced in three places of articulation: bilabial, alveolar, and palatal. The study also investigates the nasalization of vowels following voiced and voiceless nasals. Acoustic properties such as the first four nasal formants (N1-N4), their bandwidths (BW1-BW4), Centre of Gravity (CoG), and standard deviation (SD) of the nasal spectrum were used to characterize the Angami voiced and voiceless nasals. Nasality was measured in the vowels following the nasals, using A1-P0, A1-P1, and A3-P0 features. Results showed that all four nasal formants demonstrated significant voicing-specific effects. Vowel nasality measures also showed significant differences depending on the voicing of the preceding nasal. Furthermore, in this chapter, Angami voiceless nasals were studied in detail using nasometric measurement. The results showed that the voiceless nasals produced in isolation were mostly voiceless, with oral aspiration appearing around the middle of the articulation. When produced in sentence frames, an additional voiced nasal portion was noticed at the onset of the voiceless nasal. In terms of nasality, determined by nasalance scores, effect of the place of articu-

lation was observed. It was also noticed that the onset of oral aspiration resulted in reduced nasalance. The voiceless nasals were consistently longer than their voiced counterparts. Generally, when produced in sentence frames, all the nasals were longer, regardless of voicing.¹

3.1 Introduction and literature review

Voiceless nasals are less studied (cross-linguistically), and their distribution is localized—for instance, they are more prevalent in Tibeto-Burman languages. The focus language of the current study, Angami, is a language that phonemically contrasts voiced and voiceless nasals in three places of articulation. In the chapter, we examine the acoustic properties of the voiceless nasals and compare them with their voiced counterparts. Secondly, we also investigate the nasalization-related acoustic properties of the vowels following the Angami nasals. Finally, we also explore the Angami nasal characteristics using nasometric data.

The Tibeto-Burman languages have a prevalence of voicing contrasts in nasals, as in Angami. Five different subgroups of the Tibeto-Burman language family — Himalayish, Qiangic, Lolo-Burmese, Nungish, and Kamarupan are reported to have voicing contrasts in nasals (Matisoff, 2003). In addition to the Tibeto-Burman languages, Welsh and Ikema Ryukyuan have also reported voicing contrasts in nasals (Ford, 2016; Hammond, 2019). Among the Tibeto-Burman languages, Burmese is relatively well-studied in terms of nasal voicing contrasts. Several instrumental studies in Burmese confirmed that a voiceless nasal consists of a voiceless nasal friction and a voiced nasal region (Dantsuji, 1984, 1986). The Burmese voiceless nasals are initially voiceless but voiced towards the termination of the nasals (Bhaskararao and Lade-

¹This chapter has been published in a peer-reviewed journal and in the proceedings of a peer-reviewed conference - “Acoustics of Voiceless Nasals in Angami”, Journal of Asian and African Studies, Supplement (in press, June 2024); “Aspiration in voiceless nasals in Angami”. In Proceedings of Meetings on Acoustics 179 ASA (Vol. 42, No. 1, p. 060008). Acoustical Society of America.

foged, 1991). Burmese nasals differ from Angami nasals, as the latter has continuous nasal airflow with no voicing throughout the production of the voiceless nasal. As mentioned before, investigation into the voiceless nasals in Burmese, Mizo and Angami revealed that there are two distinct types of voiceless nasals (Bhaskararao and Ladefoged, 1991). Recently, with the addition of data from Kham Tibetan and Xumi, the two types of nasals are summarized as follows (Chirkova et al., 2019). The first type are the preaspirated nasals, which contain a voiceless portion with nasal and oral airflow, followed by a voiced portion with only nasal airflow. The second type are the voiceless nasals with continuous nasal airflow throughout the duration of the nasal. However, towards the end of such nasals, oral airflow is added and characterized by partial voicing (Chirkova et al., 2019). Note that as far as the final partial voicing of the second type of voiceless nasals is concerned, Bhaskararao and Ladefoged differ by reporting that the final portion also is devoid of any voicing.

Burmese is one of the few languages where the acoustic properties of the voiceless nasals have been studied (Dantsuji, 1984). In his study of Burmese nasals, Dantsuji (1984) has analyzed the spectral and temporal characteristics of the voiced and voiceless nasals. He reports several prominent energy peaks in the spectrum of the voiceless nasal friction portion. The intensity of the voiceless portion of the Burmese voiceless nasal is higher than the intensity of the voiced portion of the voiceless nasal. It is also noticed that the voiced portions of the voiceless nasals are shorter than the voiced nasal stops (Dantsuji, 1984). In a subsequent study on the Burmese voiceless nasals by Dantsuji, acoustic properties of the nasal murmur spectra (nasal part of the voiceless nasal) such as the first nasal formants (N1), second nasal formants (N2), third nasal formants (N3) along bandwidths (B1, B2, B3) were used to find cues for place of articulation for voiceless nasals (Dantsuji, 1986). It is reported that cues of places of articulation could be distinguished by the nasal formants and their bandwidths using the nasal portion of the voiceless nasals (Dantsuji, 1986).

Besides the nasal formants and nasal formant bandwidths, previous works have reported measurements of the voicing rate for voiceless nasals. In one such study, Xumi and Burmese showed significant differences in the rate of voicing for voiced and voiceless nasals (Chirkova et al., 2019). In the case of Mizo, instrumental studies using the Nasometer II were also conducted, which showed fine grained characteristics of Mizo voiceless nasals (Lalhminghlui and Sarmah, 2021). Nevertheless, nasal formants and their bandwidths have been shown to be reliable measures in characterizing nasals. For instance, in an acoustic study of the voiced nasal consonants in three central Australian languages, it was reported that nasal formants, their bandwidths along with the center of gravity (CoG), and standard deviations (SD) calculated from the nasal spectrum provided reliable characterization of nasals with different places of articulation (PoA) (Tabain et al., 2016b).

This chapter also looks into the effect of voiced and voiceless nasals on following vowels in terms of vowel nasality. Vowel nasality occurs when the velum is lowered during vowel production, which opens the velopharyngeal port and allows air to flow through the nose and mouth (Styler, 2015; Chen, 1997). While nasalization in vowels may be phonemic, as in French or Hindi, vowels may also be nasalized due to coarticulatory effects of the preceding nasal (Maeda, 1993; Delvaux et al., 2002; Ohala and Ohala, 1975; Malécot, 1960). Hence, we assume that vowels in Angami that are preceded by nasals will be nasalized, and the nasalization will differ for voiced and voiceless nasals.

Considering the discussion in this section, we present the results of acoustic studies conducted on Angami voiced and voiceless nasals. In the current study, we consider acoustic properties directly extracted from the nasal consonants, such as four nasal formants (N1, N2, N3, N4) and their bandwidths, CoG, and SD of the nasal spectrum. As we expect the effect of the nasals on the following vowels, we also investigate the vowel nasality using A1-P0, compensated A1-P0, A1-P1, and A3-P0 (Styler, 2013).

A detailed discussion on these features is provided in Section 3.3.4.

3.2 Literature review of Angami nasals

In Angami, nasals occur in four places of articulation (PoA), namely, bilabial (/m/), alveolar (/n/), velar(/ŋ/), and palatal (/ɲ/). While the bilabial, alveolar and palatal nasals have voiceless counterparts, velar nasals don't. Angami also has a limited number of words with the voiced labio-dental nasal, [m̥] (mv), considered an allophonic variation of the bilabial nasal /m/, arising due to the presence of a following /ə/ (Kuolie, 2006; Hajek, 2009). As seen here, among the nasals in Angami, the velar nasal does not have a voiceless counterpart. An example of nasal minimal pairs in Angami is shown in Table 3.1. For the acoustic analyses, we took only the nasals that have both voiced and voiceless counterparts.

Existing literature on the Angami nasals does not make a distinction between voiced and voiceless nasals, instead referring to them as voiced and aspirated voiced nasals (Ravindran, 1974; Kuolie, 2006; Burling, 1960), a few studies have noted the presence of voicing in their descriptions (Suokhrie, 2015; Meyase, 2016). Blankenship et al. (1993) study on the Khonoma Angami was the pioneering work that reported the existence of voiceless nasals. The description is discussed in the previous Section 3.1, where the authors reported that voiceless nasals in Angami is characterized by entirely voiceless nasal portions followed by aspiration (Blankenship et al., 1993).

The first part of the chapter is organized as follows: Section 3.3 describes the

Table 3.1: Example of nasal minimal pairs in Angami

Voiced		Voiceless		
IPA	Meaning	IPA	Meaning	Tone
ma	price	m̥ ^h a	something	T4
na	budding	n̥ ^h a	plant	T1
ɲa	crazy	ɲ̥ ^h a	messy	T1
ɲa	shame			

methodology, which include description of speech corpus and the acoustic and statistical analyses. Section 3.4 discusses the study's results, and discussion in Section 3.5. The second part of the study entails nasometric study of the voiceless nasals, which is discussed in Section 3.6.

3.3 Methodology

The participants, materials, the recording procedures, and the acoustic and statistical analyses are all addressed in this section.

3.3.1 Speakers

13 Native Angami speakers (6M, 7F) participated in the study. The participants are all residents of Kohima village, except for one female, the variety spoken there is considered the standard variety of the Angami language (Ezung, 2018). The average speakers' age is 40 years (Standard Deviation = 15.5) at the recording time. In addition to Angami, all participants are also fluent in English (the state's official language) and Nagamese (*lingua franca*). All the participants had at least a bachelor's degree, and some were working professionals.

3.3.2 Materials

The material was designed in a way to capture the voiced and voiceless nasals (/m, m^h, n, n^h, ɲ & ɲ^h/) in the context of six vowels (/i, e, a, o, u & ə/) whenever possible. The combination of nasals and vowels were meaningful words in the language. The nasals occur in three places of articulation, namely, bilabial, alveolar, and palatal. The target nasal sounds were monosyllabic (CV) in structure. There are two sets of data: the first set produced by six speakers includes 74 unique nasal tokens produced in three environments, namely, sentence, isolation, and carrier phrase. The second data

set is a subset of the first, where 28 nasal tokens were produced only in isolation and repeated three times by seven speakers. The number of tokens elicited in the first and second data sets is, 1332 (n = 6 speakers) and 588 (n = 7 speakers), respectively. Due to mispronunciation, 43 tokens were removed. A total of 1,877 tokens are considered for analysis, combining both data sets. The distribution of the six nasals with all possible vowel combinations in the data set is shown in Table 3.2.

Table 3.2: Distribution of Nasals and vowels in the Data

Vowels	/m/	/m ^h /	/n/	/n ^h /	/ɲ/	/ɲ ^h /	Total
i	73	39	75	21	75	19	339
e	76	39	76	89	16	0	310
a	76	87	72	89	74	89	392
ə	0	0	0	0	73	111	307
o	76	21	75	87	65	0	343
u	38	124	94	21	40	20	186
Grand Total							1877

3.3.3 Recording of Data

The recordings were conducted in a noise-controlled environment at the speaker's home. Each person took about an hour for the recording session. The data were collected at two different times, first in September 2018 and later in October 2019. The recording was collected following the recording procedure as described in Section 1.3.3.

3.3.4 Acoustic and statistical analysis

The speech recordings were transferred to a computer for analysis. The sound files were segmented and annotated by means of listening and visual examination of the waveforms and spectrograms using Praat 6.0.43 (Boersma, Paul and Weenink, David, 2020). An example of the nasal and vowel boundary for the word /ma/ and /m^ha/ produced by a female speaker is shown in Figure 3-1. A Praat script was used to

extract the first four nasal formants, namely, N1, N2, N3, and N4 in hertz (HZ), along with their bandwidths (BW1, BW2, BW3, BW4) were calculated at the temporal midpoint of the nasals. Further, Centre of Gravity (CoG) and standard deviation (SD) were also extracted. In order to visualize the nasal formant contours, nasal formants were extracted at every 10% of the total duration.

To study vowel nasality, the nasality auto-measure Praat script developed by Styler was used (Styler, 2013). The spectral measurements and vowel duration were taken at 11 points (0-100%) with 10% intervals. The four spectral measurements taken as adapted for the current study are A1-P0, compensated A1-P0, A1-P1, and A3-P0. A1-P0 is the amplitude of the first formant (A1) minus the amplitude of the first nasal peak (P0). Usually, P0 occurs between the first two harmonics (H1, H2). A1-P1 is calculated by subtracting the amplitude of the second nasal peak that occurs in the 850-1050 Hz range from A1. A3-P0 is the difference in amplitude between the third formant and P0. A3-P0 captures the spectral tilt that may help characterize nasality. As seen in the study of French (nasalized vowels) and English (phonetic nasality), the values in features such as A1-P0, A1-P1, and A3-P0 decreased with increased nasality (Chen, 1997; Styler, 2017). A detailed discussion on the effectiveness of these features in measuring vowel nasality can be found in Styler (Styler, 2013, 2017). and Chen (1997). R's statistical packages were used for exploratory statistics, the analyses are conducted as described in Section 1.3.6.

3.4 Results

The acoustic properties of voiced and voiceless nasals in Angami and the nasalization-related acoustic properties will be discussed in the following section.

3.4.1 Acoustic properties of voiced and voiceless nasals

The nasal formant contours have been plotted for visualization in Figure 3-2. As mentioned in section 3.3.4, the nasal contours (N1-N4) were calculated at every 10% of the total duration. Figure 3-2 clearly distinguishes between voiced and voiceless nasals for the first nasal formant (N1). The voiced nasals have low N1 values (< 400 Hz), while the voiceless have high N1 (> 600 Hz). We observed that in the voiceless nasal, the N1 value is highest for bilabial, followed by alveolar and palatal ($/m^h/ > /n^h/ > /ɲ^h/$). Whereas, amongst voiced nasal, bilabial nasal has higher N1 values than alveolar and palatal ($/m/ > /n/ \& /ɲ/$). In figure 3-2, similar to N1, the N2 values of voiced are lower than voiceless nasals, but the differences between the two are relatively less. Considering the mid 50% of the total duration, the N2 values for voiceless nasals are about 1800-2400, while the voiced is about 1600-1800. In The N2 values for voiceless nasals, the palatal is highest, followed by alveolar and bilabial ($/ɲ^h/ > /n^h/ > /m^h/$). The voiced bilabial $/m/$ is the lowest, followed by palatal nasal $/ɲ/$ and alveolar nasal $/n/$. Figure 3-2 shows that for the N3 values, there is a clear distinction between voiceless nasals and voiced nasals. The range of voiceless nasal is 2850 Hz and above, while voiced nasal is 2750 Hz and below. The PoA for the voiceless nasals for N3 is not obvious, as all the nasals have merged or overlapped. However, for voiced nasals, we see similar patterns with N2.

There is a marginal difference between voiceless and voiced nasals, as seen in Figure 3-2. The N4 values for voiced nasal ranges from 3680 Hz and below, while voiceless nasal ranges from 3850 and above. There is a clear PoA pattern in voiced nasals ($/m/ > /n/, /ɲ/$). However, in voiceless nasals, there is an overlapping.

Furthermore, the mean and standard error of the frequencies of nasal formant bandwidths, along with CoG and SD, is displayed in Figure 3-3. The figure shows that the frequency values of BW1 and BW3 are lower than BW2 and BW4. The distinction between voiced nasal and voiceless nasal is seen in the values of BW1 and

BW3. Furthermore, the CoG and SD mean values are displayed in the figure, where it is observed that the CoG values of voiced nasals are relatively lesser than voiceless nasals. CoG depends on PoA (bilabial > alveolar > palatal) in both the contrasting nasals. As for SD, there is no significant difference in the nasals.

In order to confirm the statistical significance of the nasal formant values and bandwidths among the six nasals, we constructed LME models. The dependent variables were the nasal features, which included the four nasal formant values at midpoint and their bandwidths. Additionally, the average Centre of Gravity (COG) and standard deviation (SD) of the nasal spectrum were considered as dependent variables. Hence, nasals, context, and their interaction were the fixed effects, while the speaker was the random effect, as shown in 3.1.

$$NasalFeatures \sim Nasals + Context + Nasals * Context + (1|Speaker) \quad (3.1)$$

The summary of the LME model is presented in Table 3.3. The constructed models were subjected to Type II Wald Chi-square tests for analysis of deviance on LME models separately, where it showed that the effect of nasals consonants on nasal formants were significant. It is also observed that the interaction of nasals and context also have significant effect. The χ^2 value, degrees of freedom (df), and significance for each feature are shown in Table 3.3 (2nd column). Hence, looking into the pairwise comparisons of nasal features by voicing contrasts of the nasals is pertinent. The LME models were further subjected to post-hoc Bonferroni analysis to see the significant differences between voiced and voiceless nasals. The result of the test is reported in Table 3.3. The estimate, standard error (SE), degrees of freedom (df), t-ratio, and significance of the voiced and voiceless nasals under each model is reported. As seen in the table, the voicing contrasts in the three places of articulation, that is, bilabial,

alveolar, and palatal, significantly effect nasal features such as N1, N2, N3, and BW1. As for N4, significant effect is observed in bilabial and alveolar voicing contrasts. It is also noticed that voicing contrast do not have any significant effect on standard deviation (SD) and Centre of Gravity (COG).

Table 3.3: Results of Analysis of Deviance (ANOVA) tests on the LME models for estimating the effect of nasals consonants and estimates of post-hoc Bonferroni tests for pair-wise contrasts.

Feature	χ^2	Contrasts	Estimate	SE	df	<i>t</i> -ratio	<i>p</i> -value
N1	Nasals: 1119.6 ^{***}	$m - m^h$	-519.4	26.3	1899	-19.7	< 0.0001
	Context: 13.5 [*]	$n - n^h$	-543.5	26.1	1899	-20.8	< 0.0001
	Nasals x Context: 81.5 ^{***}	$\eta - \eta^h$	-463.3	30.3	1899	-15.3	< 0.0001
N2	Nasals: 379.1 ^{***}	$m - m^h$	-426.0	32.8	1899	-13.0	< 0.0001
	Context: 5.2	$n - n^h$	-299.5	32.6	1899	-9.1	< 0.0001
	Nasals x Context: 74.1 ^{***}	$\eta - \eta^h$	-438.4	37.8	1899	-11.5	< 0.0001
N3	Nasals: 204.8 ^{***}	$m - m^h$	-276.0	25.3	1899	-11.0	< 0.0001
	Context: 8.0	$n - n^h$	-175.0	25.2	1899	-7.0	< 0.0001
	Nasals x Context: 61.1 ^{***}	$\eta - \eta^h$	-225.4	29.2	1899	-7.7	< 0.0001
N4	Nasals: 73.1 ^{***}	$m - m^h$	-249.0	40.1	1871	-6.1	< 0.0001
	Context: 3.2	$n - n^h$	-174.4	40.2	1871	-4.4	< 0.0001
	Nasals x Context: 30.4	$\eta - \eta^h$	-132.5	47.0	1871	-3.0	0.0708
BW1	Nasals: 605.1 ^{***}	$m - m^h$	-262.0	27.1	1899	-9.6	< 0.0001
	Context: 7.3	$n - n^h$	-417.6	27.0	1899	-15.5	< 0.0001
	Nasals x Context: 46.9 ^{**}	$\eta - \eta^h$	-424.3	31.2	1899	-13.6	< 0.0001
BW2	Nasals: 49.5 ^{***}	$m - m^h$	-40.2	48.5	1899	-0.8	1.0000
	Context: 2.5	$n - n^h$	128.3	48.2	1899	2.7	0.1185
	Nasals x Context: 23.8	$\eta - \eta^h$	255.0	56.0	1899	4.5	< 0.0001
BW3	Nasals: 32.1 ^{***}	$m - m^h$	-123.6	44.2	1899	-2.8	0.0779
	Context: 1.3	$n - n^h$	-157.6	44.0	1899	-3.6	0.05
	Nasals x Context: 41.4 [*]	$\eta - \eta^h$	-98.5	51.0	1899	-1.9	0.8052
BW4	Nasals: 31.3 ^{***}	$m - m^h$	91.4	51.7	1870	1.8	1.0000
	Context: 2.0	$n - n^h$	202.3	51.7	1871	3.9	0.05
	Nasals x Context: 26.1	$\eta - \eta^h$	200.1	60.3	1871	3.3	0.0139
CoG	Nasals: 237.5 ^{***}	$m - m^h$	-197.0	31.5	1899	-6.2	< 0.0001
	Context: 2.2	$n - n^h$	-53.9	31.4	1899	-1.7	1.0000
	Nasals x Context: 96.7 ^{***}	$\eta - \eta^h$	-102.3	36.4	1899	-2.8	0.0752
SD	Nasals: 16.5 ^{**}	$m - m^h$	-28.6	14.4	1899	-1.9	0.6997
	Context: 4.3	$n - n^h$	24.0	14.3	1899	1.7	1.0000
	Nasals x Context: 24.6	$\eta - \eta^h$	-45.0	16.6	1899	-2.7	0.1032

3.4.2 Vowel nasality

Figure 3-4 provides a visual representation of the nasal effects on the vowel in each context. In the top two Figures 3-4, the nasals have been plotted on A1-P0 and compensated A1-P0 values. An upward trajectory of the voiceless nasals is seen from the initial to 25% of the nasal formant contour. There is a distinction between voiced and voiceless nasals, as observed. In carrier and sentence context, the values correlate with PoA in voiceless nasals ($m^h > n^h > p^h$). A significant effect on A1-P1 is seen in figure, where the voiced nasal tends to differentiate from the voiceless nasal across the whole nasal contour. Voiceless alveolar nasal $/n^h/$ have the lowest A1-P1, followed by bilabial and palatal. Further, as seen in Figure 3-4 (bottom), the A3-P0 values overlap and do not show any clear pattern. To see the effect of voicing contrasts on nasality, we conducted exploratory modeling where the average four nasality features, A1-P0, compensated A1-P0, A1-P1, and A3-P0, were the dependent variables, Nasals, context, and their interactions were the fixed effects, and the speaker was the random effects as shown in 3.2.

$$\text{Nasality features} \sim \text{Nasals} + \text{context} + \text{Nasals} * \text{context} + (1|\text{speaker}) \quad (3.2)$$

Type II Wald Chi-square tests for analysis of deviance test conducted on the four LME models separately showed that voicing contrasts have a significant effect on all four nasality features, as seen in Table 3.4. There is no interaction of nasal types and contexts, indicating no change in nasality depending on context. The LME models were further subjected to Bonferroni post-hoc test to see the differences between voiced and voiceless nasals in three places of articulation (PoA). Table 3.4 shows the estimates, standard errors (SE), degrees of freedom (df), t-ratios, and p-values of homorganic voiced and voiceless nasals. The pair-wise comparison of the voiced and

voiceless nasals showed that nasality features such as A1-P0, compensated A1-P0, and A1-P1 have a significant effect on voicing. A3-P0 does not show any systematic difference between voiced and voiceless nasals and hence, it can be concluded that A1-P0, compensated A1-P0, and A1-P1 are the salient features that manifest the difference of voicing in nasals in the case of Angami.

While the effect of voicing on nasals is manifested on the following vowels in terms of vowel nasality, it is not clear to what extent into the following vowel is the effect significant. To investigate that, 11 separate LME models were constructed with nasality measures (A1-P0, compensated A1-P0, and A1-P1) at every 10% of the total duration of the vowel following nasals. Only A1-P0 compensated A1-P0, and A1-P1 were considered as they showed a systematic effect of the preceding nasal voicing as discussed previously. All 11 models were subjected to post-hoc Bonferroni tests to obtain pair-wise comparisons. The results of the comparisons are provided in Table 3.5. As seen in the table, the difference between voiced and voiceless nasals in terms of nasality is noticed in varying degrees for different nasality correlates. As far as A1-P0 is concerned, the effect of nasal voicing is significant up to 50% of the total duration of the following vowel. The effect of nasal voicing permeates more into the following vowel if the preceding nasal is palatal. On the other hand, in the case of bilabial nasals, the effect is the least. In the case of A1-P1, the effect is seen in the entire vowel duration.

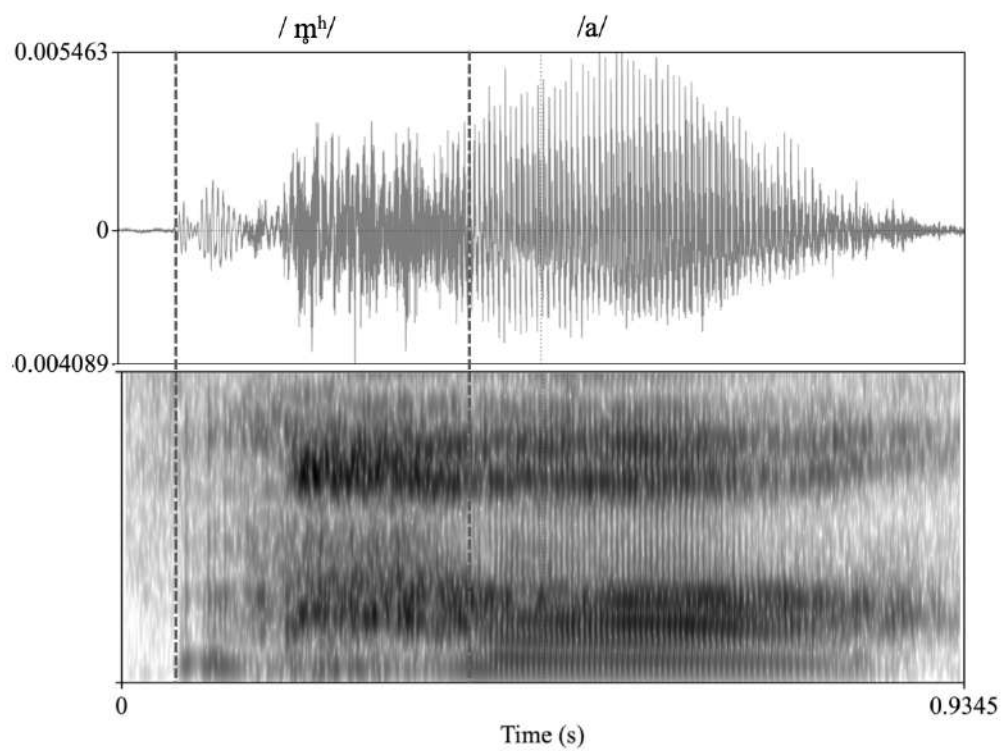
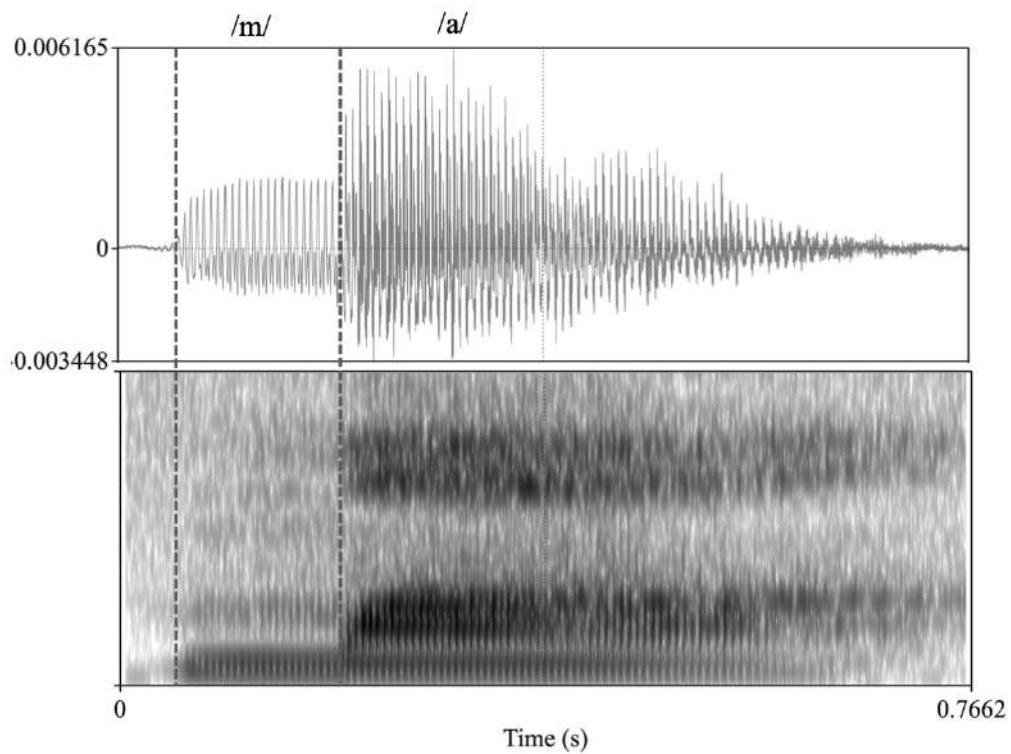


Figure 3-1: An example of the nasal and vowel boundary for the word /ma/ ‘price’ and /m^ha/ ‘things’ as produced by a female speaker.

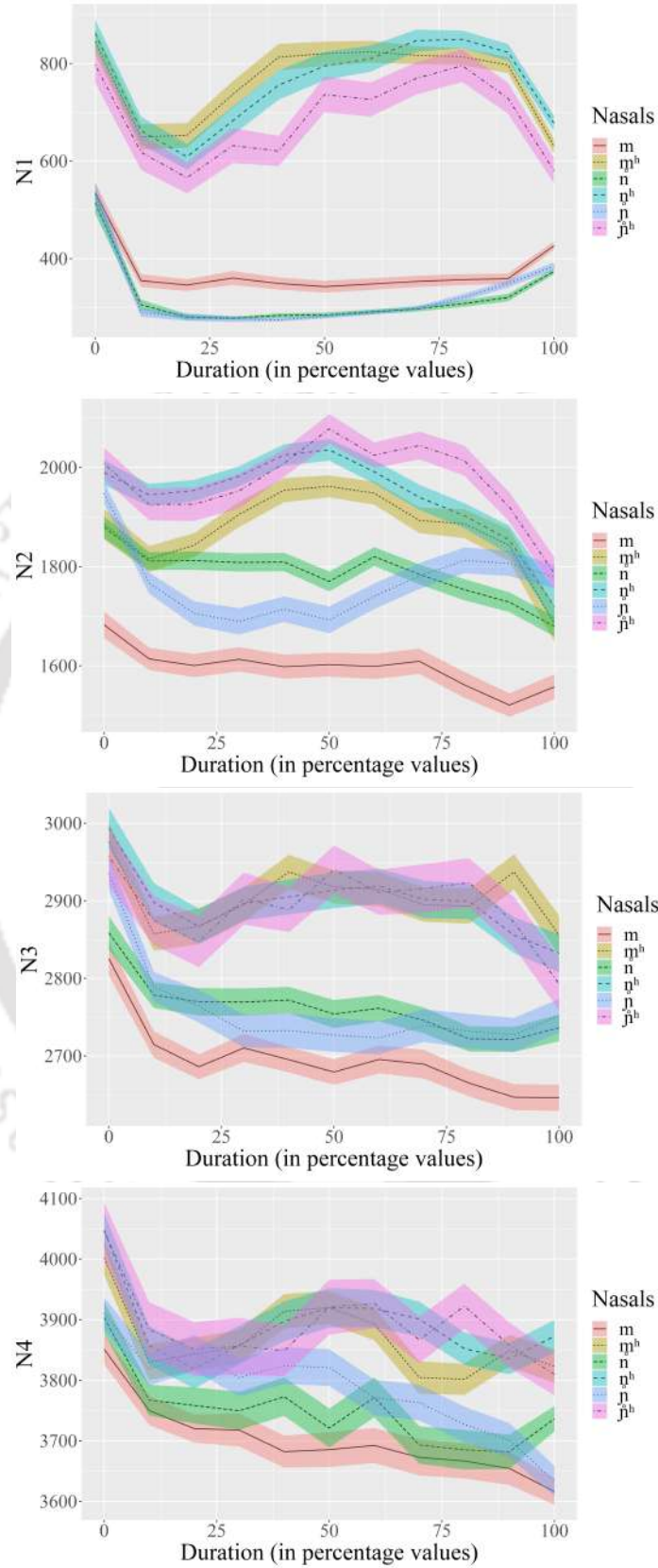


Figure 3-2: Formant contours (N1-N4) in Hertz for the Angami nasal consonants.

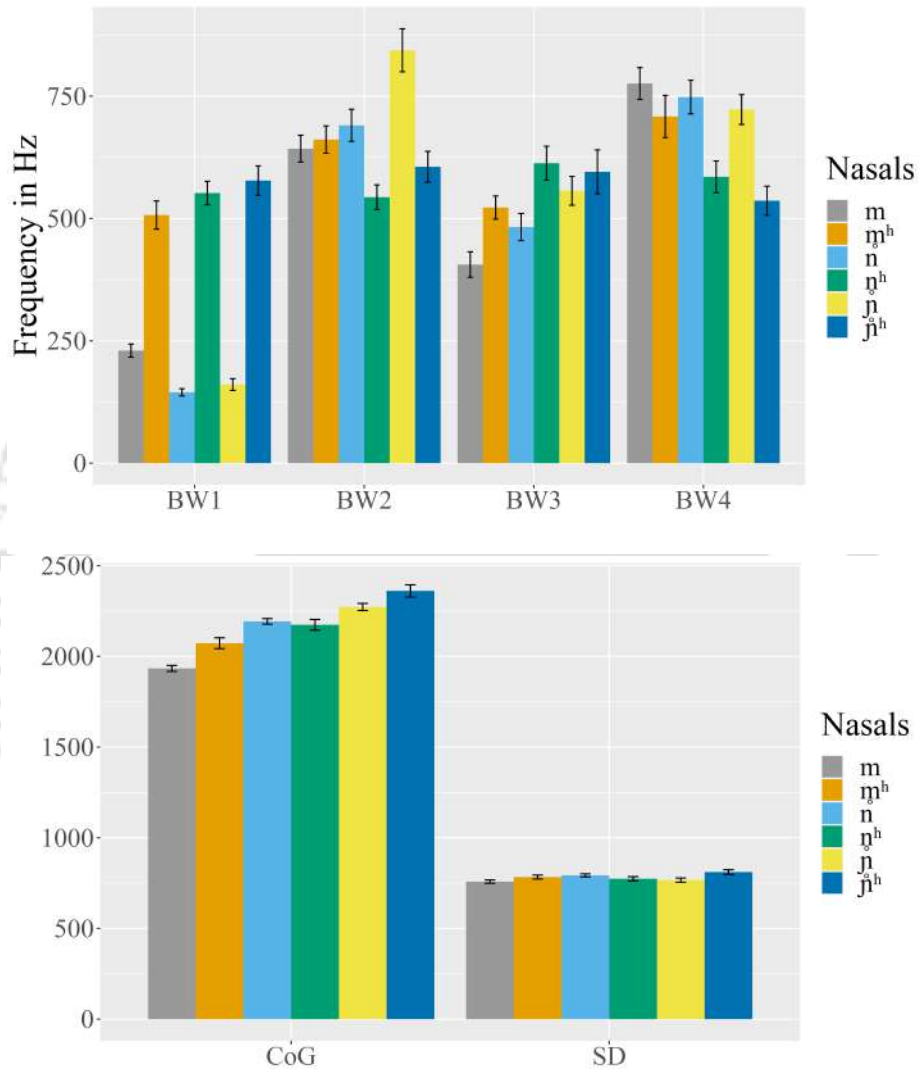


Figure 3-3: Bar plots of (a) the mean of nasal formants bandwidths (BW1-BW4), (b) SD and CoG calculated from the nasal spectrum.

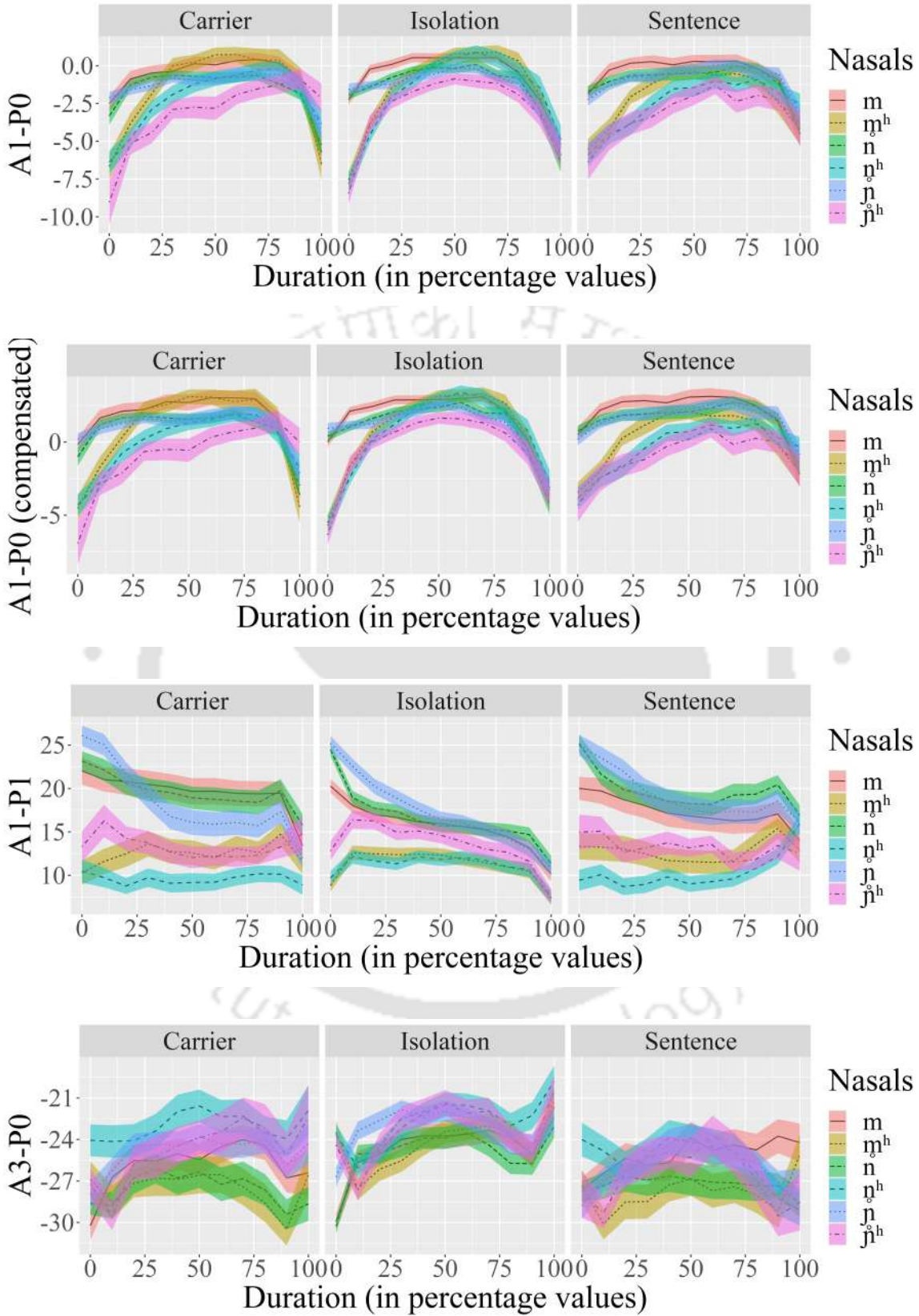


Figure 3-4: Nasality measures for the vowels following Angami nasals.

Table 3.4: Analysis of Deviance (ANOVA) tests on the LME models with average nasality values as dependent variables and pair-wise contrasts estimated from post-hoc Bonferroni tests.

Model*	Contrasts	Estimate	SE	df	t-ratio	
A1-P0 $\chi^2(5)=68.1^*$	m - m ^h	0.9	0.3	1511	3.5	< 0.01
	n - n ^h	0.8	0.3	1511	3.0	< 0.05
	ɲ- ɲ ^h	1.9	0.3	1511	6.0	< 0.0001
com. A1-P0 $\chi^2(5)=84.4^*$	m - m ^h	1.7	0.3	1511	4.2	< 0.001
	n - n ^h	1.1	0.3	1511	4.0	< 0.001
	ɲ- ɲ ^h	2.1	0.3	1512	6.5	< 0.0001
A1-P1 $\chi^2(4)=193.9^*$	m - m ^h	5.8	0.9	1513	6.7	< 0.0001
	n - n ^h	8.9	0.8	1513	10.6	< 0.0001
	ɲ- ɲ ^h	5.0	1.0	1515	5.0	< 0.0001
A3-P0 $\chi^2(5)=30.2^*$	m - m ^h	1.6	0.7	1511	2.2	0.3871
	n - n ^h	-3.1	0.6	1511	-4.6	< 0.001
	ɲ- ɲ ^h	0.1	0.8	1512	0.2	1.000

p < 0.001 is indicated by an asterisk

3.4.3 Duration of voiced and voiceless nasals

Furthermore, the six nasals' average duration across the contexts was examined, showing that voiceless nasals are longer than their voiced counterparts as shown in Table 3.6.

To confirm the statistical significance of the nasal duration, we constructed an LME model with nasal duration as the dependent variable. Assuming the contextual effects on nasal duration, voicing contrasts, place of articulation (PoA), context, and their interactions were considered fixed effects, whereas, the speaker was considered a random effect as shown in 3.3.

$$Duration \sim voicing * PoA + contexts + voicing * PoA * context + (1|speaker) \quad (3.3)$$

Results of the Type II Wald chi-square tests for analysis of deviance conducted

Table 3.5: Significance of nasality values from Bonferroni post-hoc pairwise comparisons conducted on LME models for 11 equidistant points of vowels following nasals.

	0%	10%	20%	30%	40%	50%	60%	70%	80%	90%	100%
A1-P0											
m - m ^h	****	****	****								
n - n ^h	****	****	****	***							
ɲ- ɲ ^h	****	****	****	****	***	**					
com.A1-P0											
m - m ^h	****	****	****								
n - n ^h	****	****	****	****							
ɲ- ɲ ^h	****	****	****	****	****	****	*	*			
A1-P1											
m - m ^h	****	****	****	****	****	****	****	****	****	*	
n - n ^h	****	****	****	****	****	****	****	****	****	****	****
ɲ- ɲ ^h	****	****	****	****	**	*	**	*	*		**

**** p < 0.0001 , *** p < 0.001, ** p < 0.01, * p < 0.05

Table 3.6: The mean duration and standard deviation (SD) in milliseconds (ms) of the six nasals in Angami by contexts.

Nasals	Sentence	Carrier	Isolation
m	149.7 (76)	166.2 (45)	156.2 (59)
m ^h	198.6 (72)	223.6 (73)	190.0 (72)
n	153.0 (57)	198.2 (64)	149.3 (56)
n ^h	212.5 (90)	219.3 (81)	190.1 (67)
ɲ	187.5 (74)	221.8 (78)	169.6 (54)
ɲ ^h	251.4 (105)	234.0 (79)	207.2 (75)

on the LME model show that voicing contrast, PoA, contexts, and their interactions have a significant effect on the nasal duration as seen in Table 3.7. Hence, it can be concluded that the voicing of the nasals has significant duration differences, however, the durational differences are also affected by the place of articulation of the nasals and the contexts in which they are produced. The interaction plot shown in Figure 3-5 confirms that in all cases, the voiceless nasals are longer than their voiced counterparts. However, in the carrier frame context, the difference between the duration of voiced and voiceless nasals with palatal and alveolar PoA seems to be less. Hence, to explore these interactions in detail, the LME model was subjected to a pairwise

comparison using a Bonferroni post-hoc test. The results of the post-hoc test as shown in Table 3.8 confirm that the bilabial nasals maintain statistically significant durational differences across all contexts. However, the alveolar and palatal nasals do not show any significant durational difference in the carrier frame context. In all other contexts, the durational difference between voiced, and voiceless nasal is systematic and statistically significant. These results also concur with the findings in Burmese (Dantsuji, 1984; Chirkova et al., 2019), Icelandic (Jessen and Pétursson, 1998), and Mizo (Lalhminghlui and Sarmah, 2021) where it was shown that voiceless nasals are longer than their voiced counterpart.

Table 3.7: Results of an Analysis of Deviance (ANOVA) test on the LME model for estimating the effect of voicing, PoA, and contexts on nasal duration.

Fixed effects	χ^2	<i>df</i>	<i>p</i> -value
Voicing	267.0	1	< 0.001
PoA	86.8	2	< 0.001
Context	136.7	2	< 0.001
Voicing x PoA	1.0	2	n.s
Voicing x Context	13.0	2	< 0.01
PoA x Context	19.1	4	< 0.001
Voicing x PoA x Context	13.7	4	< 0.01

Table 3.8: Estimates of the duration of nasal pairs obtained from a Bonferroni post-hoc pairwise comparison conducted on an LME model, with significant estimates ($p < 0.0001$) marked with an asterisk.

Contrast	Sentence	Carrier	Isolation
m-m ^h	-48.1*	-55.8*	-34.1*
n-n ^h	-59.6*	-22.0*	-40.7*
ɲ-ɲ ^h	-61.6*	-13.0*	-34.7*

3.5 Discussion

In this chapter, we look into Angami’s acoustic features of voiced and voiceless nasals. Regarding the nasal consonants, we observed that the voiced and the voiceless nasals

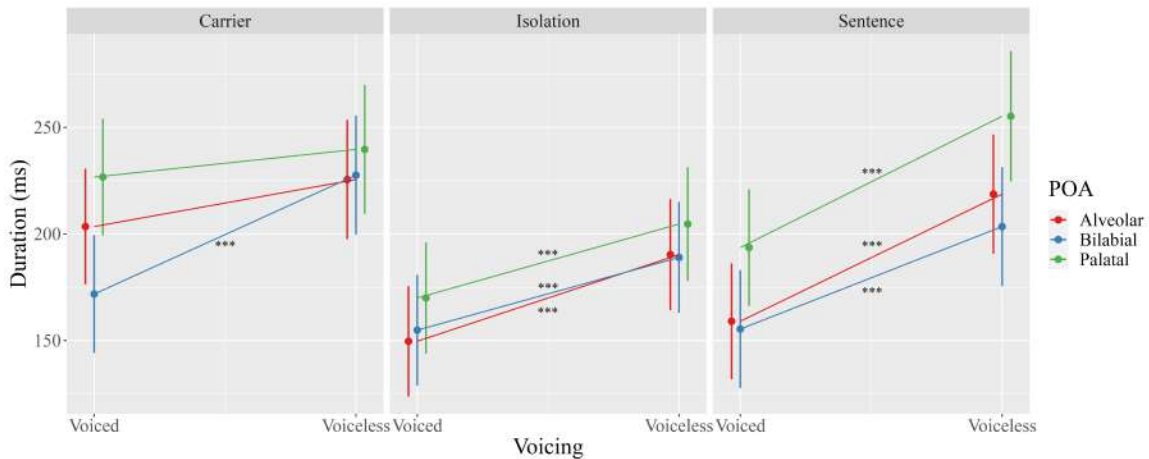


Figure 3-5: Variable interaction plot generated from an LME model with the duration of nasal as the dependent variable and voicing types, contexts, PoA, and their interactions as fixed effects. The statistically significant interactions are indicated with asterisks.

have significantly different patterns. As shown in the nasal formant tracks (3-2), the voiced nasals consistently have lower formant values than their voiceless counterparts. The difference between voiced, and voiceless nasals is visually prominent in the case of N1 and N3. Another cursory observation here is that the nasal formants have PoA-specific effects in both voiced and voiceless nasal contexts. The PoA effect of the voiceless nasal is pronounced in N1 and N2. However, the PoA effect is asymmetric in N1 and N2. Regarding N1, the voiceless bilabial nasal has the highest, and the palatal nasal has the lowest values.

On the other hand, in terms of N2, the reverse is observed. In the case of voiced nasals, a similar pattern is observed. While the voiced bilabial nasal has higher N1 than the voiced alveolar and palatal nasals, in the case of N2, the reverse is observed. In terms of statistical models, all four nasal formants showed significant P-values when all six types of nasals were used as fixed factors. Post-hoc analyses of the LME models confirmed that there was a significant difference between the voiced and the voiceless nasals in each pair of nasals (bilabial, alveolar, palatal). Similar post-hoc analyses on CoG and SD failed to yield any statistically significant for voicing in

nasals.

As voiceless nasals are relatively rare in the world's languages, there are no studies that have tried exploring the physiology in the production of such phonemes. The closest comparison that can be drawn is from the observations on voiceless or whispered vowels. In the case of whispered vowels, they are observed to be produced with partially open glottis. Whispered vowels are characterized by higher formant frequencies than their voiced counterparts (Peterson, 1961; Smith, 1973; Lehiste, 1964). In the study of English vowels, Lehiste (1964, 1970) noted that F1 was approximately 200–250 Hz higher, whereas F2 and F3 were 100–150 Hz higher, in whispered vowels than in their voiced counterparts. Lehiste (1964, 1970) postulated that the differences in formant frequencies are due to the degree of glottal openness. During voiced vowel production, the glottis is closed while, during whispered vowel production, the glottis is open, resulting in formants that are higher in frequency. This formant raising is due to the fact that the resonant frequencies of an acoustic tube open at both ends are higher than that of a tube closed at one end. Considering this, we argue that voiceless nasals also function similarly. The open glottis in the production of voiceless nasals contributes to the raising of the formant frequencies.

It has been noted that open glottis have wider formant bandwidths than closed glottis (Smith, 1973; Fujimura and Lindqvist, 1971). In the case of whispered vowels, the coupling of the trachea to the supraglottic spaces resulting in the damping of formant frequencies leads to the higher first formant bandwidth (BW1) values (Peterson, 1961; Fujimura and Lindqvist, 1971). Hence, BW1 is larger in voiceless vowels than in voiced vowels. Additionally, it has also been shown that the value of formant frequency is positively correlated with their bandwidths, beyond a certain threshold (Hawks and Miller, 1995).

As for vowel nasality, average A1-P0, compensated A1-P0, and A1-P1 showed effects of nasal voicing. A detailed investigation showed that the effect of the preceding

nasal voicing contrast is manifested in at least 30% of the following vowels. However, in the case of A1-P0, the effect seems to be dependent on the place of articulation of the nasals. We postulate that the varying changes in nasality by PoA are due to the size of oral occlusion in the production of the nasals. As the oral cavity is larger in the case of the bilabial nasals, more energy is lost in the oral cavity, resulting in less nasal energy. This results in shorter nasality in vowels followed by bilabial nasals. On the other hand, the palatal nasals, which have the shortest oral cavity, result in the least energy lost in the oral cavity resulting in higher nasal energy leading to higher nasality. Hence, oral cavity size in the production of voiceless nasals is inversely proportionate to the vowel nasality.

Nasality is higher after the voiceless nasals than after the voiced nasals. As studies on voiceless nasals are extremely rare, not many studies have looked into the relationship between the voicing of nasals and nasality in the following vowels. In the current study, it was noticed that nasality in vowels following voiceless nasals is higher than in voiced nasals. It is to be noted that the Angami voiceless nasals are produced with aspiration towards the end of the voiceless nasal. This breathy aspiration increases the airflow, mimicking the production of a nasal. It has been seen that the increase associated with a sound segment mimics the production of a contextual vowel with nasalization. Ohala (1993) noticed that when breathiness is associated with a sound segment, the contextual vowel is nasalized as seen in the Hindi word /ãnk^h/ which is derived from the Sanskrit word /akṣii/. Intuitively, we hear more nasality in vowels following voiceless nasals in Angami than after voiced nasals. We postulate that the higher nasalization in the vowels following Angami voiceless nasals may be due to the aspiration followed by the voiceless nasals, which may be phonologized by the Angami speakers. However, at this point in time, we would like to conduct more experimental studies to come to a conclusion regarding the pathway that prompted higher nasality in vowels following voiceless nasals in Angami.

3.6 Nasals in Angami using nasometer II

In the previous section, acoustic correlates of Angami voiceless nasals were studied. In continuation of the study, this section explores the voiceless nasals with fine-grained measurements using nasometer II. As discussed in Section 3.1, among the languages that have voicing contrasts in nasals, Burmese is relatively well studied and a number of instrumental studies are conducted on the nasals of the language (Dantsuji, 1984, 1986; Bhaskararao and Ladefoged, 1991; Chirkova et al., 2019). It is reported that Burmese has voiced nasals, / m, n, ɲ, ŋ /, contrasting with / m̥, n̥, ɲ̥ & /ŋ̥ /, in terms of voicing. Early acoustic studies on Burmese showed that the voiceless nasals consist of a portion of voiceless nasal friction and a voiced nasal portion (Dantsuji, 1984). The energy of the voiceless nasal friction portion is much lower than that of the voiced nasal portion. Additionally, the duration of the voiced portion is found to be shorter than that of the nasal friction. Among the three voiceless nasals in Burmese, the voiced portion of the voiceless nasal /hŋ/ is reported to be shorter than the other voiceless nasals (Dantsuji, 1984). Later it was reported that the Burmese voiceless nasals are voiceless in the initial portion, however voiced towards their terminal portions (Bhaskararao and Ladefoged, 1991). This makes Bhaskararao and Ladefoged to claim the existence of two types of voiceless nasals in the world, one the Burmese type and the other the Angami type which has continuous nasal airflow with no voicing throughout the production of the voiceless nasal (Bhaskararao and Ladefoged, 1991).

Aerodynamic and electroglottographic (EGG) studies were conducted on the voiceless nasals in three Tibeto-Burman languages, namely, Burmese, Kham Tibetan and Xumi (Chirkova et al., 2019). The results of the study indicate that the Xumi and Kham Tibetan voiceless nasals are similar as both have momentary closure of apertures characterized by nasal airflow and complete absence of oral airflow. In terms of Burmese, their results support the findings of the previous studies that the Burmese

voiceless nasals tend to be voiced towards the end. In sentence frames, Xumi voiceless nasals are found to be ‘devoiced’ while Tibetan voiceless nasals are completely voiced (Chirkova et al., 2019). In this section, the acoustic characteristics of Angami nasals are studied using nasometric data. The Nasometer allows the capture of nasal and oral acoustic signals separately, enabling one to examine nasal segments with more detail. The general aim of the study is to provide more fine-grained measurements of the voiced and voiceless nasals in Angami with the analyses of the oral and nasal channel data obtained from the Nasometer. In the following section, i.e. Section 3.7, we provide the methodology used in the work. In Section 3.8, we report the results obtained by examining the data from the nasal and oral channels of the Angami nasals. We conclude that the voiceless nasals produced in isolation and in sentence frames are significantly different. Moreover, we also report the evidence of oral airflow in the voiceless nasals. The results also show that the voiceless nasals have voiced portions when preceded by vowels. While the voicing difference in isolation vs. sentences is significant for all the nasals in the initial 50% of the nasals, it is not significant in the final 50% of the nasal. In terms of nasalance scores measured at every 10% of the total duration of the nasal, we noticed an overall difference between the voiced and the voiceless nasals. Apart from that, nasalance differences were significant, depending on the place of articulation of the nasals. It is also reported that the voiceless nasals are significantly longer than their voiced counterparts in both isolation and sentence contexts. Finally, in Section 3.9, we conclude the study.

3.7 Methodology

3.7.1 Speakers

Four Angami native speakers (2 Male and 2 Female) residing in Guwahati, Assam participated in this experiment. The mean speakers’ age is 29.5 years (SD= 6.7). The

recordings were made in a sound-attenuated booth at the Phonetics and Phonology Laboratory of the Indian Institute of Technology Guwahati. While the total duration of recorded sound segments was 15 minutes, the speakers spent about 1 hour in the lab as the equipment was able to capture only a minute of speech data at one sitting. Secondly, since the headset is in direct contact with the informant, the moisture from the nostrils built up on the separation plate of the Nasometer. Hence, after a few recording sessions, the device was removed for cleaning to avoid any effect of the moisture on the recordings. The participants were compensated for their time and travel expenses, and they signed a consent form agreeing to participate in the study.

3.7.2 Materials

Data used for the experiment consist of voiced and voiceless nasals in three places of articulation, namely, bilabial, alveolar, and palatal (m , m^h , n , n^h , η , η^h). The target nasal sounds were monosyllabic (CV) in nature. The bilabial and alveolar nasal sounds were followed by five vowels (i , ε , a , o , u), whereas the palatal nasal was followed by three vowels (i , a , ε) which makes meaningful words in the language. As mentioned, Angami is a tone language, however, tones are not marked here. There are 26 unique tokens produced in two environments, i.e. isolation and embedded in a sentence frame, /a 'X' puba/ (I said 'X'). The tokens were displayed on a computer screen with their meanings in English. Each token was repeated three times, resulting in a total of 156 tokens for each speaker. In total 624 tokens were elicited out of which 620 were subjected to analysis as 4 tokens were mispronounced.

3.7.3 Recording of data using Nasometer II

Using Nasometer II 6450 (KayPENTAX medical), the nasal segments produced by speakers were recorded². The headset was calibrated with a value range of 0.9. Fol-

²The authors used the facilities at the Pathological Speech Processing Laboratory of North East Centre for Biological Sciences and Healthcare Engineering (NECBH), IIT Guwahati, funded by the

lowing that, the headset was positioned on the informants' head carefully. The informants were instructed to read the list of words, displayed one after another on a computer screen, at a normal pace. Their speech was recorded using the Nasometer II software that came with the Nasometer equipment.

3.7.4 Data extraction and Analysis

The data captured with the Nasometer was saved in the .wav format at a sampling rate of 11 kHz at 16 bits per sample. By default, this data is stored as stereo files with oral and nasal channels separated. Apart from that, the Nasometer software also provides nasalance scores in .nsp format, with nasalance calculated at every 8 ms. However, for this study, the software provided nasalance values were not used. Rather, we extracted the nasal and the oral channel separately from the stereo files and segmented and annotated them manually. In order to obtain data at a controlled temporal resolution, we calculated nasalance from the amplitude of the extracted channels at every 10% of the total duration of the nasal segments.

Nasalance was calculated using the formula in (3.4), where, N is the amplitude of the nasal signal and O is the amplitude of the oral signal. The stereo data extraction and calculation of nasalance were performed using a Praat script written for the purpose (Boersma, Paul and Weenink, David, 2020).

$$Nasalance = \left(\frac{N}{O + N} \right) \times 100 \quad (3.4)$$

In order to determine the amount of voicing in the nasals, the fraction of unvoiced frames is measured using the Voice Report function of Praat. Nasalance values and percentage of voicing were exported to a spreadsheet for visualization and statistical analyses. Descriptive and exploratory statistic were calculated using statistical packages on R. While descriptive statistics such as, mean, standard deviation and standard

errors were calculated using core functions of R, exploratory statistical modelling such as the Linear Mixed Effects (LME) models were performed using the *lme4* package on R (Bates et al., 2015; R Core Team, 2022). In order to see the significance of the fixed effects, Type II Wald chisquare tests were conducted using the *Anova* function in the *car* package on R (Fox and Weisberg, 2019). Finally, to estimate the contrasts between the levels in the fixed factors, Bonferroni post-hoc tests were conducted using the *emmeans* package on R (Lenth, 2019). The plots in the results section were generated using the *ggplot2* package on R (Wickham, 2016).

3.8 Results

3.8.1 Oral and Nasal waveforms of Angami nasals

In order to see the amplitude associated with the nasal and oral segments separately, we plotted the waveforms of the oral and nasal channels as shown in Figure 3-6. The waveforms associated with the nasal and oral channels are seen to be different for voiced and voiceless nasals, as seen in Figure 3-6 & Figure 3-7. Figure 3-6 shows the oral and nasal waveforms for the voiced nasal /m/ when followed by an /e/. The figure shows that the voiced /m/, when followed by the vowel /e/, has continuous energy in the nasal channel with higher amplitude in the nasal segment /m/. As the following vowel segment /e/ is oral, the amplitude in the nasal channel is lesser when the vowel is produced. An opposite trend is noticed in the oral channel. In Figure 3-6, in the oral channel, the nasal segment remains void of any significant amplitude, whereas the vowel segment shows high amplitude.

On the other hand, when we look at the voiceless nasal /m^he/ in Figure 3-7, the nasal channel shows continuous amplitude from the beginning of the production till the termination of the syllable. As seen from the nasal channel, the beginning of the nasal has higher amplitude, and it eventually decreases at the onset of the aspiration.

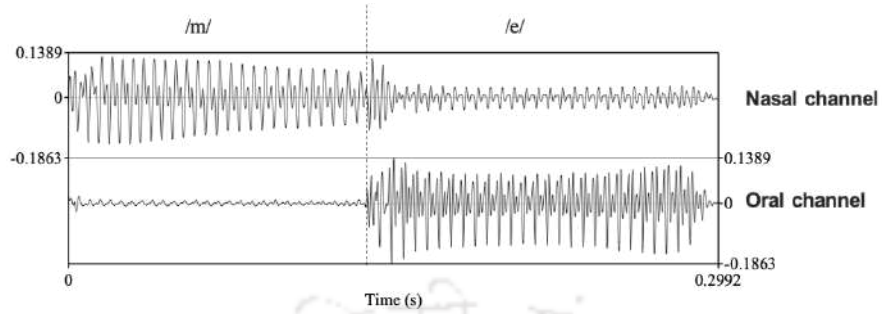


Figure 3-6: Oral and nasal channel waveforms extracted from Nasometer for a plain /me/.

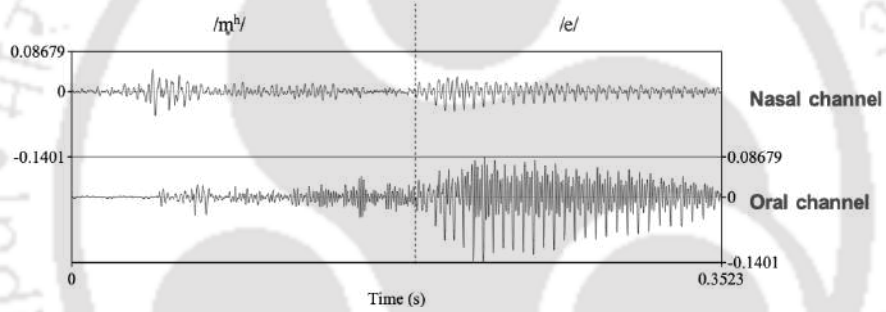


Figure 3-7: Oral and nasal channel waveforms extracted from Nasometer for /m^he/ in isolation.

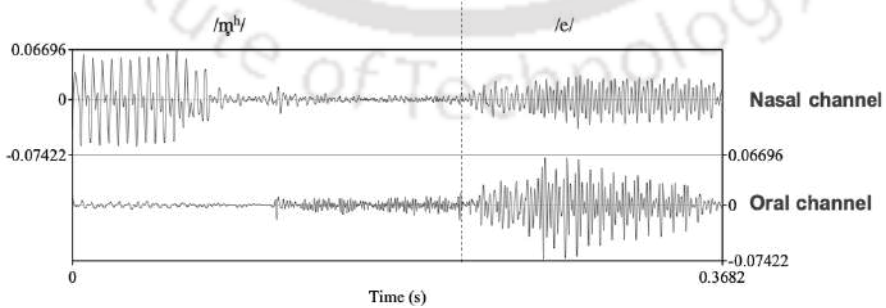


Figure 3-8: Oral and nasal channel waveforms extracted from Nasometer for /m^he/ in sentence.

The amplitude begins rising at the onset of the vowel and continues throughout the production of the vowel. On the contrary, in the oral channel, there is an absence of amplitude at the beginning of the nasal segment. However, the oral amplitude starts rising with the onset of aspiration in the nasal. Unsurprisingly, the amplitude begins to rise further with the onset of the vowel.

Figure 3-8 shows the oral and nasal waveforms of the voiceless nasal /m^he/ embedded in the sentence frame, / a m^he puba / meaning 'I said blow'. In the nasal channel, the initial segment of the voiceless nasal /m^h/ has high amplitude with periodic waveforms as it is preceded by a vowel of the carrier phrase. It indicates that the nasal /m^h/ is initially produced as a voiced nasal, and eventually it becomes voiceless. The initial voicing of the /m^h/ is the carryover effect from the preceding vowel in the carrier phrase. However, in case of the nasal channel for /m^h/ in Figure 3-8, we noticed that with the onset of the aspiration, the periodicity and the amplitude subside. Meanwhile, in the oral channel, the onset of the voiceless nasal is devoid of any amplitude as the apertures are close. At the onset of the aspiration, we see a rise in the amplitude, which eventually increases when the oral vowel begins.

From the waveforms of nasals contrasting in voice (such as, /m/ vs. /m^h/), produced in isolation and in sentence frames, we make the following conclusions. Firstly, in case of the word /me/ in Figure 3-6, the vowel is nasalized as indicated by the increase of amplitude of vowel /e/ in the nasal channel. The oral amplitude remains near-zero in the production of the nasal consonant /m/. On the other hand, in case of the voiceless aspirated nasal /m^h/, when produced in isolation as seen in Figure 3-7, we see two distinct parts of the voiceless nasal. It has a nasal only part initially, which is followed by aspiration. We see higher nasal energy from the beginning to the end of the /m^h/ as evident from the increased amplitude in the nasal channel. The amplitude of the nasal channel is also relatively high during the aspiration period. In the oral channel, the part corresponding to the initial nasal

only part does not show any significant energy as confirmed by the low amplitude of the waveform, however with the onset of the aspiration, we see the onset of higher amplitude values. Similar trends are noticed in all the voice and voiceless nasals we recorded for the study. Finally, we noticed that when the aspirated voiceless nasals are produced in sentence frames, a carryover effect of voicing on the initial part of the voiceless nasal is noticed, rendering a voiced portion at the onset of the voiceless nasal. This observation is consistent with the findings reported for Angami by Blankenship et al. (1993).

Figure 3-9 shows the aspirated voiceless nasals in Angami produced in three places of articulation (PoA). While the bilabial and alveolar nasals have similar waveform patterns for the nasal and oral channels, the palatal nasal shows a slightly different pattern. In case of the voiceless palatal nasal, the amplitude is relatively low in the oral channel. Moreover, while the nasal amplitude is lowered towards the end of the voiceless bilabial and alveolar nasals, such abrupt lowering is not noticed in case of the palatal nasal.

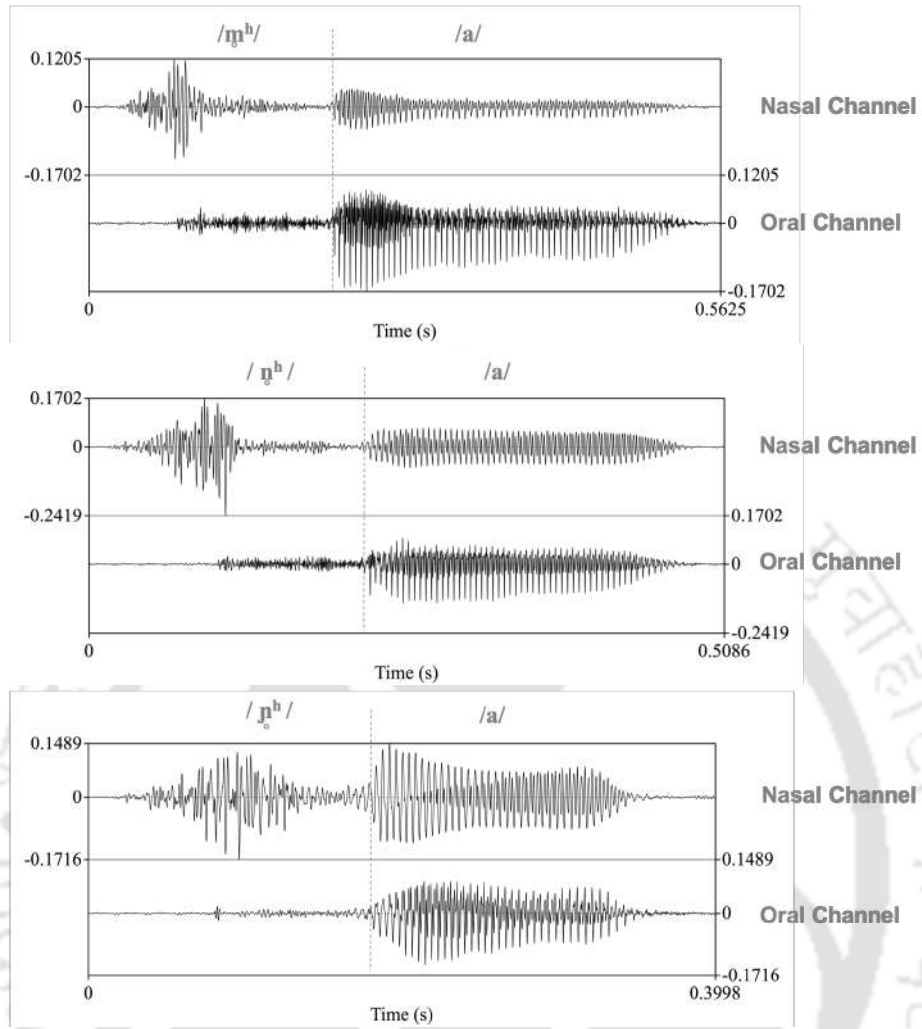


Figure 3-9: Voiceless nasals in Angami in three places of articulation, bilabial, alveolar and palatal, produced by a female speaker.

3.8.2 Voicing in Angami Nasals

It was noticed in the previous section that the voiceless nasals in Angami may not be completely voiceless. The results in these study and in the previous studies have also shown that the voiceless nasals in Angami may have voiced parts. Moreover, it was also noticed that voiced portions are longer in connected speech, when there is a preceding vowel to the voiceless nasal. Hence, we measure the amount of voicing in the first and the second half of each pair of homorganic nasals contrasting in voicing and report the findings in this section.

For all the nasals, the fraction of unvoiced frames is measured using the Voice Report function of Praat. The average percentage of voicing in the nasals for the initial 50% portion and final 50% were extracted using Praat script. Figure 3-10 shows the average percentage of voicing with standard errors for the initial and final portions of the voiced and voiceless nasals. The overall amount of voicing in each nasal is shown in Table 3.9.

From Table 3.9, we observe that the amount of voicing in the voiced nasals is high when they are in isolation as compared to when they appear in sentence frames.

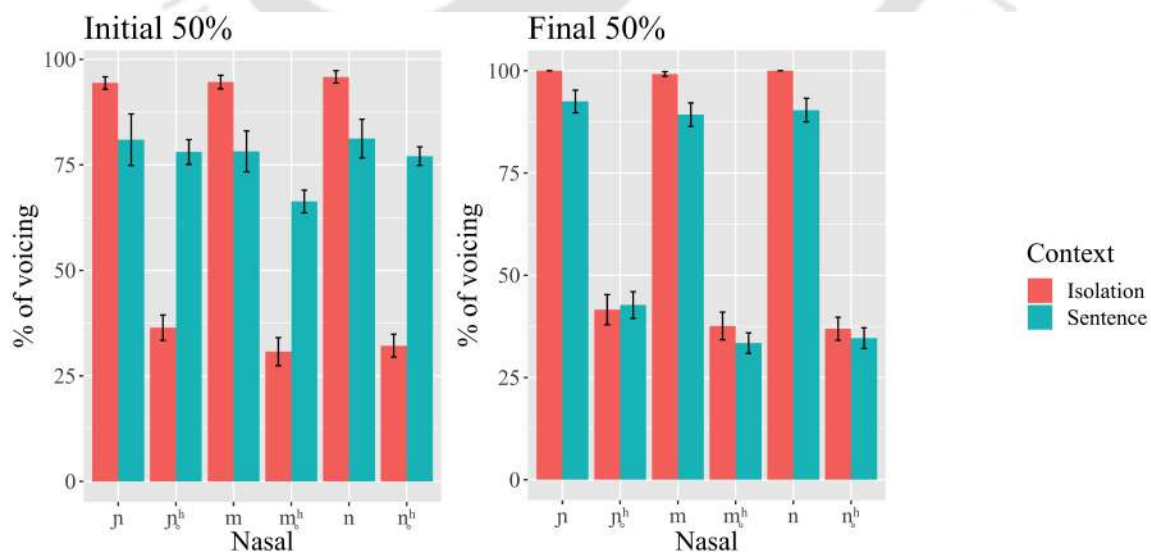


Figure 3-10: Bars showing the average percentage of voicing with standard errors.

Table 3.9: Average percentage of voicing in nasal consonants with standard deviations in parentheses.

Nasals	Isolation	Sentence
/m/	96.80 (6.51)	83.69 (29.07)
/m ^h /	34.21 (18.18)	49.85 (14.05)
/n/	97.97 (5.35)	85.80 (27.14)
/n ^h /	34.64 (15.76)	55.74 (11.96)
/ɲ/	97.23 (4.42)	86.74 (26.31)
/ɲ ^h /	39.07 (13.65)	60.30 (13.48)

Interestingly, the reverse trend is noticed for the voiceless nasals. In case of the voiceless nasals, the amount of voicing increases in sentence frames. When we looked at the voicing in the initial 50% and the final 50% separately, a slightly different picture emerged. While the voiced nasals had consistently higher amount of voicing in isolation, the voiceless nasals had substantially higher amount of voicing when produced in the sentence frames, as noticed in Figure 3-10. Finally, voicing in the final 50% of the voiced nasals was lower when produced in the sentence frames. The voiceless nasals also seemed to follow the same trend however, the difference in voicing in the two contexts was quite small.

In order to see the differences in the amount of voicing in the nasals across the contexts, we modelled the data on voicing using LME. Three LME models were built with overall voicing in the nasal, voicing in the initial 50% of the nasal and voicing in the final 50% of the nasal as dependent variables (%VD) where, two contexts and two voicing types in nasals were considered fixed effects and speaker, iteration, token number, previous segment and following segments were considered as random effects. Backward reduction of the random variables of all the three models yielded the final model as shown in (3.5).

$$\%VD \sim context + voicing + context * voicing + (1|speaker) \quad (3.5)$$

The three models built for amount of voicing are subjected to a type II, Wald-

Table 3.10: Results of Analysis of Deviance Tests on the three models.

Fixed Effects (<i>df</i>)	Overall Voicing	Initial 50%	Final 50%
Context (1)	6.4 *	55.35 ***	14.9 ***
Voicing(1)	1204.7 ***	411.2 ***	1512.2 ***
Context*Voicing(1)	137.9 ***	270.8 ***	5.5 *

n.s.: not significant, *: $p < 0.05$, **: $p < 0.01$, ***: $p < 0.001$

Chisquare analysis of deviance test. The results of the test are presented in Table 3.10. In the table, the interaction of context and voicing is significant, indicating the amount of voicing to be significantly different depending on whether the nasals are voiced or voiceless and if they are in isolation or in sentence frames. However, the χ^2 values confirm that such interactions are stronger in the initial 50% than the final 50%.

In order to see nasal consonant-wise differences in the amount of voicing, we built another three LME models with percentage of voicing in the entire nasal segment, percentage of voicing in the initial 50% and percentage of voicing in the final 50% of the nasal as dependent variables and the two contexts and six nasal consonant types as fixed effects. Speaker, iteration number, token number, previous segment and following segments were considered as random effects, however, backward reduction of the random variables resulted in a reduced model only with speaker as a random effect. The three models showed significant interaction of context and nasal types in terms of the percentage of voicing in the entire nasal segment and the percentage of voicing in the initial 50% of the nasal segments. However, no interaction was noticed in terms of the percentage of voicing in the final 50% of the nasals. The three models were further subjected to a pairwise variability test using the *emmeans* function to see the difference in voicing in each of the nasals in terms of the contexts they were produced in. Table 3.11 summarizes the results of the pairwise comparison using Bonferroni post-hoc tests. The table shows significant differences in the amount of voicing in the initial 50% of the nasals, in the case of five out of six Angami nasal

Table 3.11: Significance of difference in voicing in isolation and in sentence frames for each nasal in three models.

Nasals (<i>df</i>)	Overall Voicing	Initial 50%	Final 50%
/m/	**	**	n.S.
/m ^h /	***	***	n.S.
/n/	*	!	n.S.
/n ^h /	***	***	n.S.
/ɲ/	n.S.	n.S.	n.S.
/ɲ ^h /	***	***	n.S.

n.s.: not significant, !: $p < 0.05$, *: $p < 0.01$, **: $p < 0.001$, ***: $p < 0.0001$

consonants. Only in case of /ɲ/, there is no difference in the amount of voicing in the two contexts. The statistical results shown here confirms the trends that are seen in Figure 3-10.

3.8.3 Nasalance of voiced and voiceless nasals

We looked into the change in nasalance scores across the production of the voiced and voiceless nasals. The nasalance scores are calculated using Eq. (3.4). Nasalance score is calculated throughout the duration of the nasal at eleven equidistant points, that is at every 10% of the total duration of the nasal. Figure 3-11 & Figure 3-12 show the nasalance in the nasals produced in isolation and sentence frames, respectively. In case of the voiced nasals, /m/, /n/ and /ɲ/, it is noticed that the nasalance remain high up to 80% - 90% of the total duration of the nasals, before dropping sharply. On the other hand, the nasalance of the voiceless nasals (/m^h/, /n^h/ and /ɲ^h/) drop gradually from about 20% onwards when produced in isolation. On the other hand, when produced in a sentence frame, the nasalance of voiceless nasals reduces sharply from about 60% of the total duration.

Figure 3-11 and Figure 3-12 also show a systematic correlation between the nasalance scores and the places of articulation (PoA) of the nasals. In terms of nasalance, in both voiced and voiceless nasals, the palatal nasals seems to be the highest followed by the alveolar nasals. The bilabial nasals seem to have the least amount of

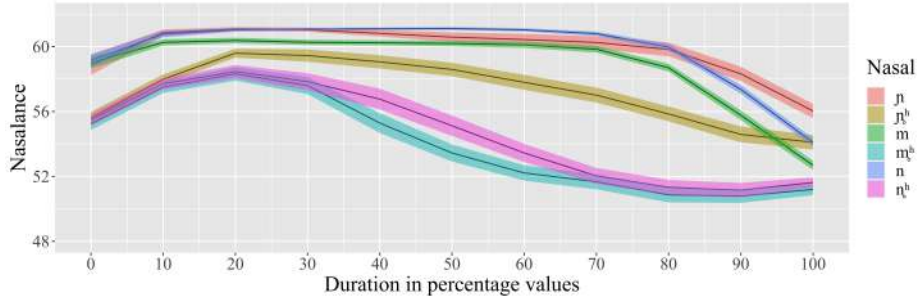


Figure 3-11: Nasalance of nasals produced in isolation with standard errors as colored ribbons.

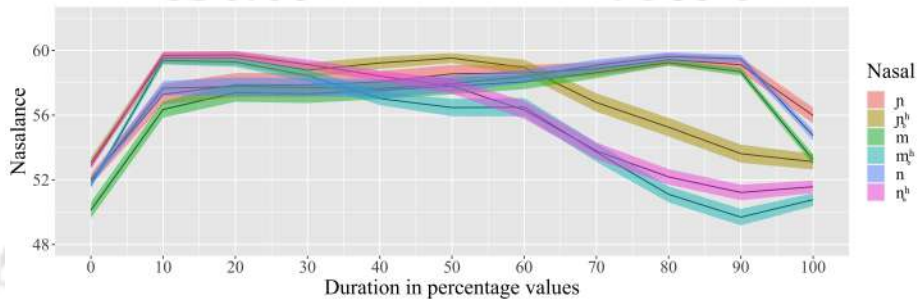


Figure 3-12: Nasalance of nasals produced in sentence frames with standard errors as colored ribbons.

nasalance. However, the PoA-wise distinction is more noticeable towards the end of the nasals.

3.8.4 Duration of voiced and voiceless nasals

The duration of the six nasals in Angami shows that voiceless nasals are longer than their voiced counterparts (see Table 3.12). Both in isolation and in sentence frames, trends are similar. In order to confirm the effect of voicing on duration, we constructed a linear mixed-effect (LME) model. Initially, a maximal model was constructed with duration as dependent variable and the two contexts and six nasal types as fixed effects. Speaker, iteration number, token number, previous segment and following segments were considered random effects. Backward reduction of the random variables of all the three models yielded the final model as shown in 3.6.

Table 3.12: Average duration of nasal segments with standard deviations.

Nasals	Isolation	Sentence
/m/	97.45 (35.9)	199.12 (62.02)
/m ^h /	149.34 (48.87)	253.45 (38.61)
/n/	95.32 (26.05)	215.34 (61.82)
/n ^h /	147.12 (54.64)	254.90 (66.47)
/ɲ/	109.91 (28.07)	212.15 (57.46)
/ɲ ^h /	151.03 (45.48)	263.21 (57.30)

Table 3.13: Results of Analysis of Deviance Tests with duration.

Fixed Effects (<i>df</i>)	χ^2	Significance
Context (1)	1.6	<i>not significant</i>
Nasals(5)	194.5	$p < 0.001$
Context*Nasals(5)	3.5	<i>not significant</i>

$$Duration \sim context + nasals + context * nasals + (1|speaker) + (1|following\ segment) \quad (3.6)$$

Table 3.13 shows the results of the Type II Wald chi square tests for analysis of deviance. As seen in the table, only the nasal types have a significant effect on the duration. Whereas, the interaction of context and nasal types was not significant. We also subjected the duration model to a pairwise comparison using a Bonferroni post-hoc test. The pairwise comparison showed that each voiceless nasal consonant is significantly longer than its voiced counterpart. The summary of the comparisons are provided in Table 3.14. Table 3.14 shows that regardless of the context of the nasal, the voiceless nasals are systematically longer than the voiced nasals.

3.9 Conclusion

In the current work, the effectiveness of dual channel recording obtained with Nasometer II is evident. The separate recordings of the oral and nasal channels in

Table 3.14: Results of Analysis of Deviance Tests with duration.

Contrasts	Context	Estimate	SE	df	t-ratio	p-value
/m/ - /m̥ ^h /	Isolation	-52.03	8.24	626	-6.312	< .0001
/n/ - /n̥ ^h /	Isolation	-51.80	8.03	626	-6.448	< .0001
/ɲ/ - /ɲ̥ ^h /	Isolation	-41.11	10.37	626	-3.964	< 0.01
/m/ - /m̥ ^h /	Sentence	-52.99	8.01	626	-6.617	< .0001
/n/ - /n̥ ^h /	Sentence	-39.414	8.04	626	-4.904	< .0001
/ɲ/ - /ɲ̥ ^h /	Sentence	-51.14	10.30	626	-4.963	< .0001

producing Angami nasals have provided us with some important insights into the acoustic characteristics of the nasals produced in Angami. Firstly, it was noticed that the voiced nasals, as seen in Section 3.8.1, have complete oral closure in their production. On the other hand, voiceless nasals, produced in isolation, after the onset of the nasal, oral aspiration begins to appear. As soon as oral energy appears, the nasal energy is decreased, as seen in Figure 3-8. When produced in isolation, on average, only the initial 33% of the voiceless nasals are voiced. However, when produced in sentence frames, as the one used in this study (see Section 3.8.2), about 74% of the initial portion of the nasal is voiced. This is also noticed in Figure 3-8. In such cases, the voiced portion of the voiceless nasal will have no energy from the oral channel. That will be followed by a completely voiceless nasal portion without any oral energy. Finally, oral aspiration will begin and continue with the nasal until the termination of the voiceless nasal. The voicing of the voiceless nasals in the sentence context is probably due to the re-syllabification of the voiceless nasal, where the part of the voiceless nasal becomes the nasal coda for the preceding vowel, as seen in 3.9.

$$(i) a \# m̥^h a... \rightarrow am \# m̥^h a...$$

In terms of nasality, as seen from the nasalance measures, it is noticed that both the voiced and voiceless nasals start peaking from about 10-20% of the total duration of the nasal segment. In case of voiced nasals produced in isolation, nasality continues to be high with gradual downward slope until the 80% of the total duration of the nasal

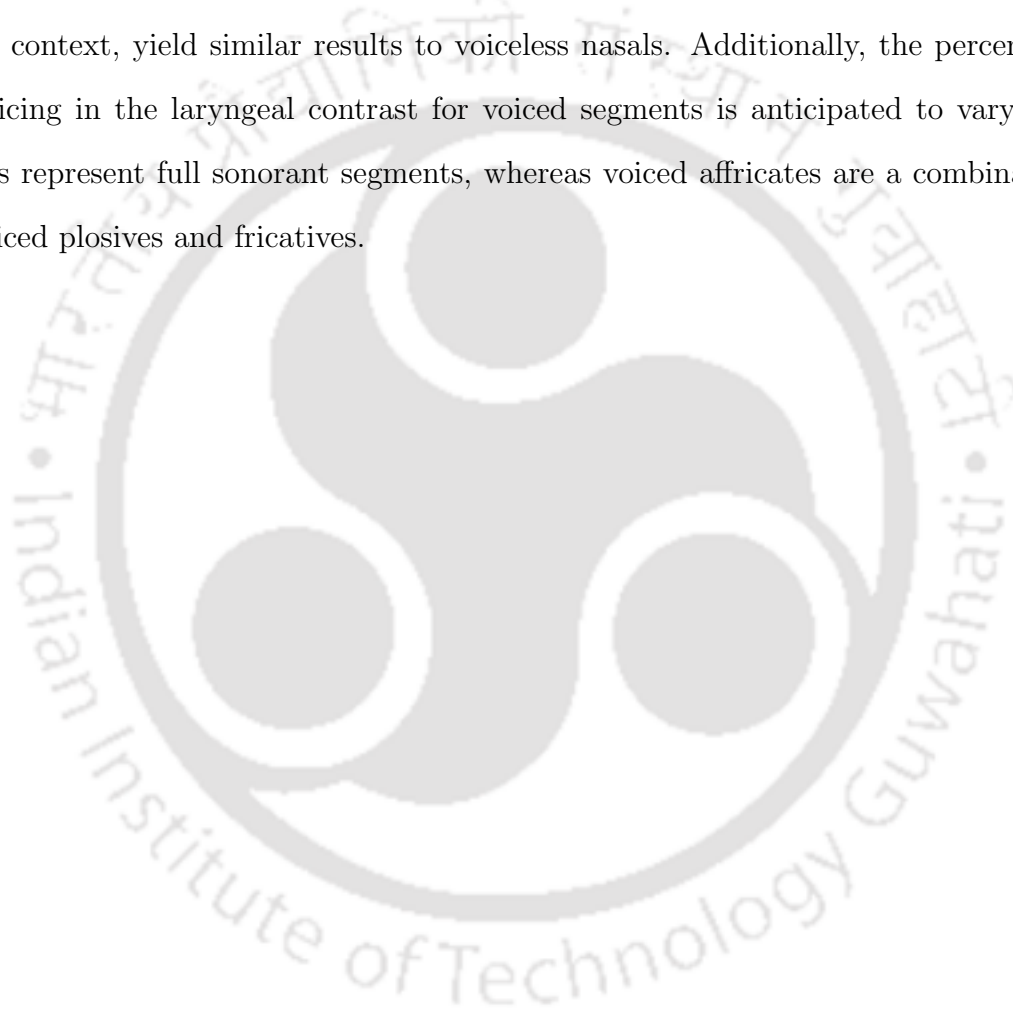
and then nasality abruptly drops. On the other hand, when produced in sentence frames, nasalance values are high with an upward slope until the 90% of the total duration of the nasal, when nasality abruptly drops. In case of the voiceless nasals in isolation, nasality starts decreasing from about 30% of the total duration of the nasal. When produced in sentence frames, nasality voiceless nasals continue to be high until the 60% of the total duration of the nasals. The reason behind the drop in nasalance can be associated with the emergence of the oral aspiration in the production of voiceless nasals in Angami. As the oral aspiration appears late in the nasals produced in sentence frames, the drop in nasality is also delayed in the voiceless nasals produced in sentence frames. We also see a strong correlation between the places of articulation of the nasals and nasalance. In both isolation and sentence frames, it is noticed that the palatal nasals induce the highest nasalance, while the bilabial nasals induce the lowest nasalance. This may be correlated to the energy loss to the side branch created by the oral cavity in the production of alveolar and bilabial nasals. The difference of nasalance is more noticeable as the side branch gradually opens for the voiceless nasals.

In terms of duration of the nasals, we notice that the homorganic nasals differing in voicing also differs significantly in terms of their duration. In case of a voiceless nasal in isolation, articulation involves nasal airflow followed by oral aspiration whereas, in case of the same nasals in sentence frames, articulation additionally involves a voiced nasal portion at the beginning of the nasal. The complex nature of the voiceless nasal articulation requires more duration, making it longer than its voiced counterpart.

The analysis of the data in this chapter has shown that the voiceless nasals in Angami have a significant oral aspiration associated with them. As mentioned in the previous studies, Angami voiceless nasals are entirely voiceless when produced in isolation. However, when produced in sentence frames, a voiced nasal portion may appear at the initial part of the nasal. However, unlike reported in Chirkova

et al. (2019), we did not find any evidence of significant voicing towards the terminal boundary of the voiceless nasals.

In the following chapter, we will investigate the acoustic characterization of fricatives and affricates. Similar to nasals and plosives, affricates also exhibit three laryngeal contrasts, while fricatives contrast only in voiced and voiceless qualities. The study aims to explore whether voiceless segments, when integrated into a sentence context, yield similar results to voiceless nasals. Additionally, the percentage of voicing in the laryngeal contrast for voiced segments is anticipated to vary. As nasals represent full sonorant segments, whereas voiced affricates are a combination of voiced plosives and fricatives.



Chapter 4

Acoustics of Angami Fricatives & Affricates

*Angami possesses a substantial inventory of fricatives and affricates. Fricatives exhibit two distinctions in voicing and are produced in three places of articulation, while affricates demonstrate two to three distinctions in voicing and are articulated in three places. This chapter addresses the issue in the inventory sizes of fricatives and affricates. Additionally, it examines acoustic-phonetic correlates such as duration, center of gravity (CoG), intensity, and voicing of these sound segments. The result shows that skewness, kurtosis and CoG are significant features in distinguishing fricatives, while, only skewness and kurtosis were significant for affricates.*¹

4.1 Introduction and Literature review

Fricatives are defined as sounds that are produced with narrow constriction of the articulatory, which results in turbulence due to the high-velocity jet of air formed in the vocal tract (Ladefoged and Maddieson, 1996). Fricatives are commonly divided into

¹Part of this chapter has been presented as an oral presentation- “Affricates in Angami” at the 27th Himalayan Languages Symposium (June, 2024)

two categories: non-sibilants and sibilants. A hissing or buzzing sound distinguishes sibilant fricatives. The sounds are produced by directing airflow with the tongue towards the dental or alveolar ridge, which strikes the teeth, causing turbulence. The acoustic property of sibilant sound is one which has strong energy at high frequencies. Sibilant fricatives include [s], [z], [ʃ], and [ʒ], whereas, the non-sibilant fricatives are [f], [v], [θ], and [ð]. Sibilant and non-sibilant fricatives differ in laryngeal contrast as well as the place of articulation. Ewe, spoken in West Africa has a four non-sibilant fricatives which contrasts in bilabial and labio-dental fricatives that have both voiced and voiceless distinction namely, [ɸ],[β], [f] and [v] (Maddieson, 2005)

Cross-linguistically, the inventory size of fricatives varies from as few as 1 to as many as 12 or more. For example, Margi, a language belonging to the Afroasiatic family spoken in Nigeria, is reported to have 12 fricatives with voicing contrast, while Kabardian, a Northwest Caucasian language, boasts 22 fricatives. However, the average number of fricatives across languages is 2, representing the modal size of fricative inventories. It is estimated that 93.4% of languages worldwide feature at least one fricative sound (Maddieson, 1984a). In PHOIBLE 2.0, among 2186 languages surveyed, the most prevalent fricative is [s], present in 2020 languages, accounting for 67% of the sample. Following [s], [f] is the next most common fricative, occurring in 44% of languages (Moran and McCloy, 2019). According to the World Atlas of Languages (WALs), 34.5% of the languages surveyed (196 out of 347) exhibit a voicing contrast between at least one pair of fricatives. This contrast may exist independently or in conjunction with any contrast in voicing among plosives (Maddieson, 2013b). In Tibeto-Burman languages, particularly among the Naga languages, the average frequency of fricatives is six, with voicing contrast. Specifically, Mao and Chokri languages are documented to possess six fricatives, whereas Rengma, Sema, and Lotha languages exhibit seven fricatives. Additionally, the Khezha language is reported to have eight fricatives (Bielenberg and Nienu, 2001; Sreedhar, 1976; Marrison, 1967).

In Angami, there are seven fricatives in its inventory (Kuolie, 2006).

A stop onset and a brief friction period characterize an affricate sound. This friction component is regarded as part of the release of the stop. According to Ladefoged (1996), “Affricates are an intermediate category between simple stops and a sequence of a stop and a fricative” (Ladefoged and Maddieson, 1996). In PHOIBLE 2.0, the most common affricates among the 2186 languages are voiceless palato-alveolar [tʃ] with 40% of the world’s languages (1218), followed by voiced palato-alveolar [dʒ] and voiceless alveolar [ts] with 27% and 22%, respectively (Moran and McCloy, 2019). Affricates are typically found in three primary places of articulation: dental/alveolar, retroflex, and alveopalatal. The most common voicing contrast observed in affricates is between voiceless and voiced affricates. However, some languages exhibit four laryngeal contrasts in their affricate inventory. For instance, Kalasha, spoken in Pakistan, features four laryngeal contrasts: voiceless unaspirated, voiceless aspirated, non-breathy voiced, and breathy voiced (Kochetov and Arsenault, 2020). It is not common, but rare, to have more than two laryngeal contrasts in affricates in the language. In the Naga languages, the inventory size of affricates varies from 2-9. While Rengma is reported to have two affricates, Sema and Chokri have five affricates. Khezha and Angami, with nine affricates, have the largest affricate inventory (Marrison, 1967; Giridhar, 1980; Kuolie, 2006). The detailed discussion on the inventory of the fricatives and affricates in Angami will be discussed in detail in Section 4.2.

Acoustic investigations of fricatives in different languages have employed diverse features such as spectral characteristics, amplitude (intensity), temporal information, and formant transition of the following vowel to differentiate between sibilant and non-sibilant fricatives, as well as to distinguish the voicing contrasts and place of articulation (Svantesson, 1986; Gordon et al., 2002; Jongman et al., 2000; Nirgianaki, 2014). Gordon et al. (2002) conducted a cross-linguistic study examining voiceless

fricatives across seven languages (Aleut, Apache, Chickasaw, Scottish Gaelic, Hupa, Montana Salish, and Toda); the authors analyzed various measurements, including duration, center of gravity (CoG), overall spectral shape, and formant transitions from adjacent vowels. The findings indicate that formant transition and CoG were effective in distinguishing fricatives. Notably, the fricative /s/ exhibited the highest CoG values (Gordon et al., 2002). In a study of Chinese fricatives, it is reported that voiceless sibilants have low dispersion of critical band spectra and high center of gravity. The same study also reported that the intensity is lower in alveolar than palatal fricatives (Svantesson, 1986). The acoustic parameters correlated to determine the place of articulation (PoA) in fricatives in both English and Greek include spectral mean, spectral variance, kurtosis, skewness, and amplitude, as reported by Jongman et al. (2000) and Nirgianaki (2014). In the study of English fricatives, differences between alveolar and palatal fricatives were observed, where alveolar fricatives exhibited a spectral peak at higher frequencies compared to palatal fricatives (Jongman et al., 2000). Alveolar fricatives have notably large positive kurtosis in contrast to palatal fricatives. Additionally, sibilants were found to have higher amplitudes than non-sibilants. In the study of Greek fricatives, F2 onset, and spectral mean were identified as distinguishing factors for the five places of articulation: labio-dental, dental, alveolar, palatal & velar (Nirgianaki, 2014). Furthermore, duration and amplitude were highlighted as distinguishing features between voiced and voiceless fricatives in Greek.

In the acoustic analysis of affricates from the existing literature, some of the most salient acoustic features in studying affricates are Center of Gravity, standard deviation (SD), kurtosis, skewness, duration, and F1 and F2 (Thurgood, 2009; Kochetov and Arsenault, 2020; Mays and Beckman, 2008). In Anong, a Tibeto-Burman language spoken in Yunnan Province, China, the center of gravity (CoG) is identified as a significant feature for distinguishing affricates. Thurgood (2009) conducted a

study on this language and reported that the voiceless aspirated alveolar affricate /ts^h/ exhibited the highest CoG, followed by the voiceless aspirated retroflex affricate /tʂ^h/ and the voiceless aspirated alveolo-palatal affricate /tʃ^h/. This pattern was also observed for the voiceless unaspirated affricates in the language. These findings suggest that CoG is crucial in distinguishing between different affricates in Anong. In a similar study, in Kalasha (a language spoken in Pakistan), the place of articulation for affricates was distinguishable from features such as COG, standard deviation (SD), kurtosis, and skewness (Kochetov and Arsenault, 2020). The study reports that COG and SD were significantly lower for retroflexes, while kurtosis and skewness were higher than dental and alveolopalatal affricates. In the same study, the distinction between voicing and aspiration affricates was distinct based on the duration of the affricate release. Unaspirated voiced affricates were characterized by the shortest release durations among the examined categories, while aspirated and breathy affricates demonstrated the longest release durations (Kochetov and Arsenault, 2020). Based on the discussion above, this chapter will focus on the acoustic characterization of fricatives and affricates in Angami. Specifically, spectral moments (skewness, kurtosis), intensity, Center of Gravity (CoG), percentage of voicing, and duration will be analyzed. The objective is to identify the features that most effectively distinguish fricatives and affricates in Angami based on laryngeal contrasts and place of articulation. This chapter is organized as follows: Section 4.2 looks into the existing literature and addresses the variation within affricates. Section 4.3 discusses the methods used in this chapter, whereas Section 4.4 discusses the results. Finally, Section 4.5 concludes the chapter.

4.2 Literature review of Fricatives & Affricates in Angami

The fricative and affricate sound inventories in the existing literature are descriptive. The number of inventories varies in both fricatives and affricates. In the study of Khonoma Angami (western variety), the inventory of fricatives varies from 4 to 7, and only two affricates are reported in this variety (McCabe, 1887; Blankenship et al., 1993; Marrison, 1967). On the other hand, in Kohima Angami² variety, it is reported that the fricatives vary from 5 to 7, while affricates vary from 4 to 9 (Burling, 1960; Ravindran, 1974; Giridhar, 1980; Kuolie, 2006; Suokhrie, 2017). The discrepancies in the variation for fricatives are seen when authors do not account for the palatal fricatives (Burling, 1960; Marrison, 1967). The summary of various authors' descriptions of Fricatives and Affricates in Angami is shown in Table 4.1. As seen in the table, the description is inconsistent, and there seems to be a gap between Khonoma and Kohima Angami (standard). While Khonoma has a lesser sound inventory, especially in affricates, Kohima Angami has more affricates. Aspiration of affricates is missing in the description of the Khonoma dialect.

Dialectal differences were observed by McCabe (1887), who noted the interchangeability of alveolar affricates /ts/ and palatal affricate /ch/ based on dialects. Furthermore, Suokhrie (2015) conducted a comprehensive study of affricates within the Kohima village, focusing on the variation among four clans. The author reported that clans Lhisemia (L) and Tsütunuomia (T) predominantly use alveolar affricates, while clans Pfuchatsumia (P) and Dapfhütsumia (D) prefer palatal affricates (Suokhrie, 2015).

²Giridhar notes that the Angami speakers were from T.Khel and L.Khel of Kohima village.

Table 4.1: Fricatives and Affricates described in Angami by various authors. Here, ‘Fct’ and ‘Afr’ represent fricatives and Affricates respectively.

Author	Variety	Type	Bilabial	Labio-dental	Alveolar	Palatal	Glottal
McCabe 1887	Khonoma	Fct Afr		f v	s z	ʃ ʒ tʃ tʃ ^h	h
Blankenship 1993	Khonoma	Fct Afr		v	s z ts	ʃ ʒ tʃ	h
Marrison 1967	Khonoma	Fct Afr		f v	s z	tʃ dʒ	
Burling 1960	Kohima	Fct Afr	pf	f v	s z ts* dz* ts ^h *	tʃ dʒ tʃ ^h	h
Marrison 1967	Kohima	Fct Afr	pf by	f v	s z	tʃ dʒ tʃ ^h	h
Ravindran 1974	Kohima	Fct Afr	pf by	f v	s z ts dz	ʃ ʒ tʃ dʒ tʃ ^h	h
Giridhar 1980	Kohima	Fct Afr	pf by p ^h	f v	s z ts dz ts ^h	ʃ ʒ tʃ dʒ tʃ ^h	h
Kuolie 2006	Kohima	Fct Afr	pf by p ^h	f v	s z ts dz ts ^h	ʃ ʒ tʃ dʒ tʃ ^h	h
Suokhrie 2015	Kohima	Fct Afr	pf p ^h	f v	s z ts dz ts ^h	ʃ ʒ tʃ dʒ tʃ ^h	h

Given that there are dialectal differences in the production of fricatives and affricates, a careful study of the phonemes was conducted. Angami language comprises seven fricatives, each distinguished by four different places of articulation: labio-dental, alveolar, palatal, and glottal. Furthermore, the fricatives have two laryngeal contrasts: voiceless and voiced. The sound /f/ is classified as a voiceless labiodental fricative, while /v/ is identified as a voiced labiodental fricative. In Angami, the /f/ sound is consistently followed by the vowel /ə/. The voiceless labio-dental /f/ is in a contrastive environment with labio-dental affricate /pf/ as it contrasts in words such as /fə/ ‘charred’ and /pfə/ ‘carry.’ On the other hand, /s/ represents a voiceless alveolar palatal fricative, while /z/ corresponds to a voiced alveolar fricative. Additionally, /ʃ/ is a voiceless palatal fricative, and /ʒ/ is its voiced counterpart. Lastly, /h/ denotes a voiceless glottal fricative. The minimal pair of the fricatives are shown in Table 4.2.

In the Angami sound inventory, there are nine affricates, each differing in three places of articulation: labio-dental, alveolar, and palatal. Additionally, these affricates exhibit three voicing distinctions: voiceless, voiceless aspirated, and voiced. For instance, /pf/ represents a voiceless unaspirated labio-dental affricate, which may alternate with /f/ in certain instances, as observed in words like /tepfə/ or /tefə/, meaning ‘dog’. The sound /p^h/ is classified as a voiceless aspirated labio-dental affricate, while [by] is identified as a voiced labio-dental affricate. Like the nasal

Table 4.2: Example of fricatives minimal set in Angami.

Laryngeal contrasts	Fricatives	Word	Meaning
Voiceless unaspirated	f	fə	blow
	s	sə	bottle
	ʃ	ʃə	wings
	h	hə	encourage
Voiced unaspirated	v	və	hit
	z	zə	top
	ʒ	ʒə	sleep

Table 4.3: Example of affricates minimal set in Angami.

Laryngeal contrasts	Affricates	Word	Meaning
Voiceless unaspirated	$\underline{p}f$	$\underline{p}f\text{ə}$	carry
	$\underline{t}s$	$\underline{t}s\text{ə}$	sprout
	$\underline{t}ʃ$	$\underline{t}ʃ\text{ə}$	small
Voiceless aspirated	\underline{p}^{h}	$\underline{p}^{h}\text{ə}$	search
	$\underline{t}s^h$	$\underline{t}s^h\text{ə}$	corn cobb
	$\underline{t}ʃ^h$	$\underline{t}ʃ^h\text{ə}$	pain
Voiced unaspirated	$\underline{d}z$	$\underline{d}z\text{ə}$	water
	$\underline{d}ʒ$	$\underline{d}ʒ\text{ə}$	complacent

labio-dental /mv/, an allophone of the /m/, [bv] is considered an allophone of /b/. As voiced labio-dental fricative [bv] in Angami occurs in a conditional environment, specifically when it is followed by the vowel /ə/. The sound / $\underline{t}s$ / represents a voiceless alveolar unaspirated affricate in Angami, whereas / $\underline{t}s^h$ / represents a voiceless aspirated affricate. The voiceless unaspirated palatal affricate is represented by / $\underline{t}ʃ$ /, while the voiceless aspirated palatal affricate is represented by / $\underline{t}ʃ^h$ /. $\underline{d}z$ denotes the voiced unaspirated alveolar affricate, while $\underline{d}ʒ$ signifies the voiced palatal plosive. The minimal pair of the affricates are shown in Table 4.3.

The acoustic analysis of fricatives and affricates focused specifically on sound segments with distinctions in voicing. In examining fricatives acoustically, attention was given to voiceless unaspirated /f/, /s/ and /ʃ/, as well as voiced unaspirated /v/, /z/ and /ʒ/. As for the affricates, sound segments which makes three laryngeal contrast were taken into study, namely, voiceless unaspirated, voiceless aspirated and voiced affricates (/ $\underline{t}s$ /, / $\underline{t}s^h$ /, / $\underline{d}z$ /, / $\underline{t}ʃ$ /, / $\underline{t}ʃ^h$ /, / $\underline{d}ʒ$ /).

4.3 Methodology

This section outlines the methodology which includes the speech corpus, the participant, and the process of extracting acoustic-phonetic correlates.

4.3.1 Speakers

Ten Angami native speakers (5 Male and 5 Female) residing in the Kohima village, Nagaland, participated in this experiment. The mean speakers' age is 31.4 years (SD= 6.3) at the recording time. All speakers were born and raised in Kohima village, where Angami (Tenyidie) is the mother tongue (L1). All participants can also speak English (state official language) and Nagamese (*lingua franca*) besides Angami. One participant is proficient in Hindi as well. All participants had a minimum education of a bachelor's degree, and some were working professionals. They were all proficient readers of Angami.

4.3.2 Materials

The material consists of six fricatives and six affricates on the onset position of the CV syllable in meaningful lexical words. The target sounds were followed by all the six vowels (i, e, a, o, u & ə). It is important to note that certain sounds may not occur with specific vowels in the inventory. The target word was embedded in three contexts: sentence frames, carrier, and isolation, as discussed in Section 1.3.2. The total time required for each speaker for recording ranges from 30 to 50 minutes, depending on the pauses taken by each participant in between the recordings. In total, 1855 tokens were elicited and analyzed for fricatives and affricates. The distribution of tokens considered for analysis in this study is presented in Table 4.4. The recordings were conducted in the speaker's home in a noise controlled environment. The data were collected in October 2022. The standard recording procedure is followed as in Section 1.3.3.

4.3.3 Extraction of acoustic features

Five acoustic-phonetic features were derived for both fricatives and affricates. Among these features are spectral moments, encompassing skewness and kurtosis, as elabor-

Table 4.4: Distribution of fricatives and affricates included in this study.

f	v	s	z	ʃ	ʒ	Total
30	180	202	215	158	113	898
<u>ts</u>	<u>ts^h</u>	<u>dz</u>	<u>tʃ</u>	<u>tʃ^h</u>	<u>dʒ</u>	
211	155	175	154	170	94	959

ated in Section 1.3.6. Additionally, temporal measurements of acoustic features were obtained from the entire duration of the segments. Other acoustic measurements include the center of gravity (CoG), intensity, and percentage of voicing, all assessed at the midpoint of the sound segment. All measures were calculated using a Praat script written for the purpose, while descriptive and exploratory statistics were computed using R, as discussed in Section 1.3.6.

4.4 Results

This section presents the acoustic-phonetic correlates of fricatives and affricates. The discussion begins with an examination of spectral moments, followed by temporal measurements and the center of gravity. Furthermore, the analysis encompasses the percentage of voicing and intensity for each sound segment.

4.4.1 Acoustics features of fricatives in Angami

The spectrograph of voiceless and voiced fricative /f/ & /v/ followed by vowel /ə/ is shown in Figure 4-1. The boundary of the fricatives and the following vowel are marked with a horizontal line. The figures indicate that the voiced fricative exhibits a darker band compared to the voiceless sound. The voiceless fricative is distinguished by a frication period preceding the onset of the vowel. Within the vowel, distinct vocal fold pulses occur at regular intervals, while the fricatives display irregular pulses. In order to see the differences in place of articulation in the production of fricatives, the spectrographs of voiceless alveolar /s/ & voiceless palatal /ʃ/ followed by /a/

vowel are plotted in Figure 4-2. It is observed that the palatal has a higher energy of frication, as represented by the darker band in the upper band of the spectrogram. A region with greater intensity for voiceless palatal /ʃ/, which is between 3,000 and 4,000 Hz, is observed. In both cases, it is also observed that the frication period is carried over to the onset of the following vowel, which is visible in the higher formants (F3 & F4).

4.4.1.1 Spectral moments of Fricatives

In order to visualize the data distribution of the fricatives, density plots are plotted and presented in Figure 4-3. The average kurtosis and skewness values for each fricative, representing the normalcy of the data distribution, are provided in Table 4.5. As seen in the Figure, labio-dental fricatives /f/ and /v/ peak at 2.4 and 8.0, respectively, which indicates that the distribution is positively skewed, and the kurtosis values show that the fricatives are heavy-tailed as their values are at 17, and 102, respectively. Overall, /v/ does not seem to have a definite peak as it has a flat distribution, as seen in the Figure. In the case of alveolar fricatives, while the voiceless alveolar /s/ value is at -0.4, which indicates a symmetrical distribution of data based on the skewness values (which is between -0.5 and 0.5), the voiced alveolar value is at 1.2 which indicates moderate positive skewness. The kurtosis indicates a short-tailed distribution, as the values are less than 3. Regarding palatal fricatives, voiceless alveolar /ʃ/ and voiced alveolar /ʒ/ both have high positive skewness as the values are at 5.7 and 9.8. Moreover, like the alveolar fricatives, the kurtosis is a short-tailed distribution.

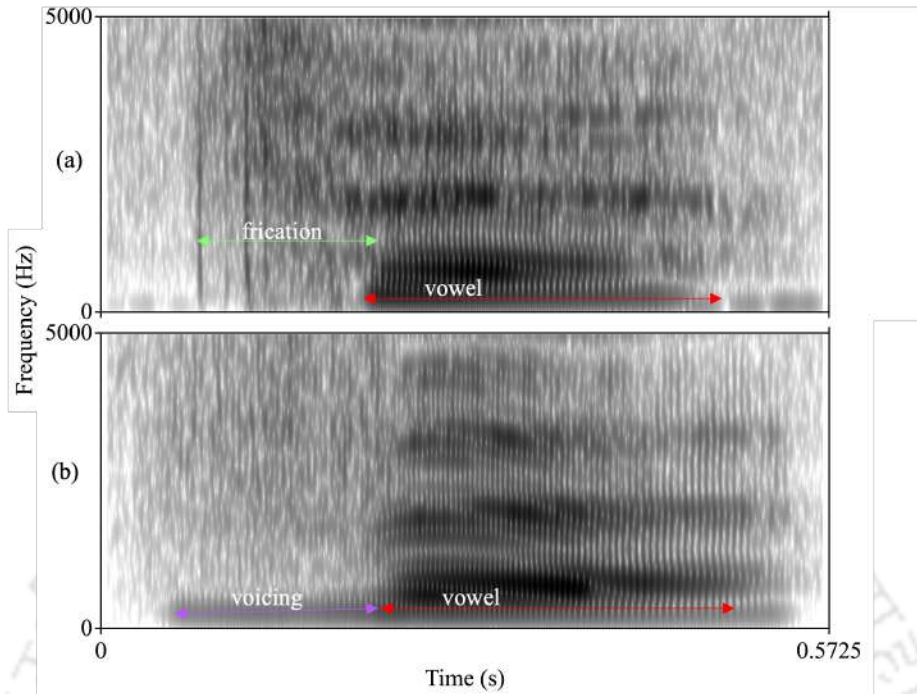


Figure 4-1: Spectrograph of (a) labio-dental voiceless /f/ and (b) voiced unaspirated /v/ followed by vowel /ə/ as produced by a female speaker in isolation.

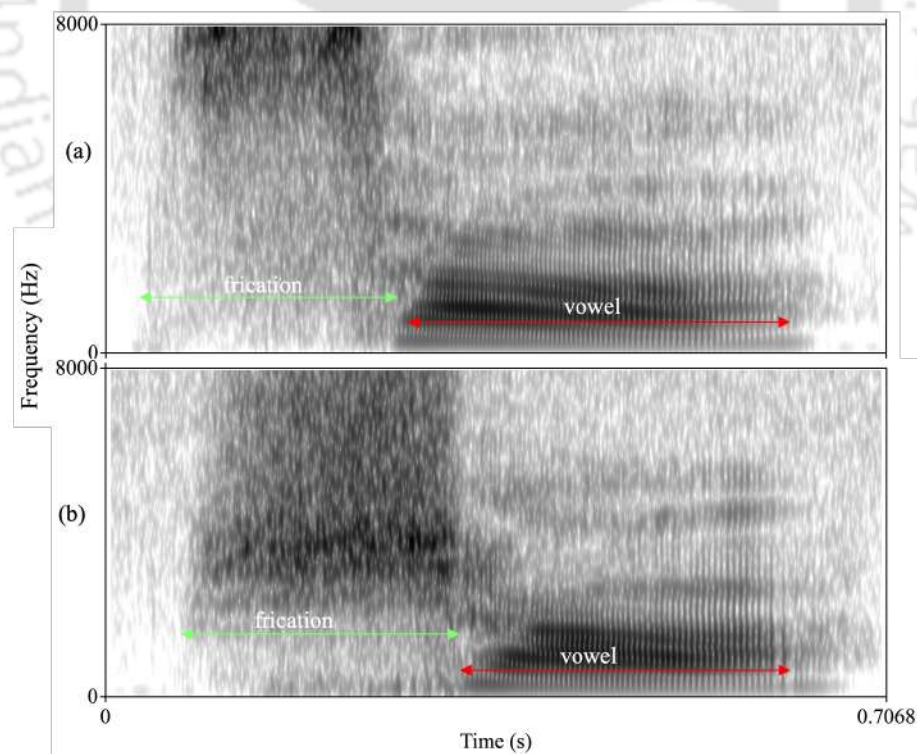


Figure 4-2: Spectrograph voiceless alveolar /s/ & voiceless palatal /ʃ/ followed by /a/ vowel as produced by a female speaker in isolation.

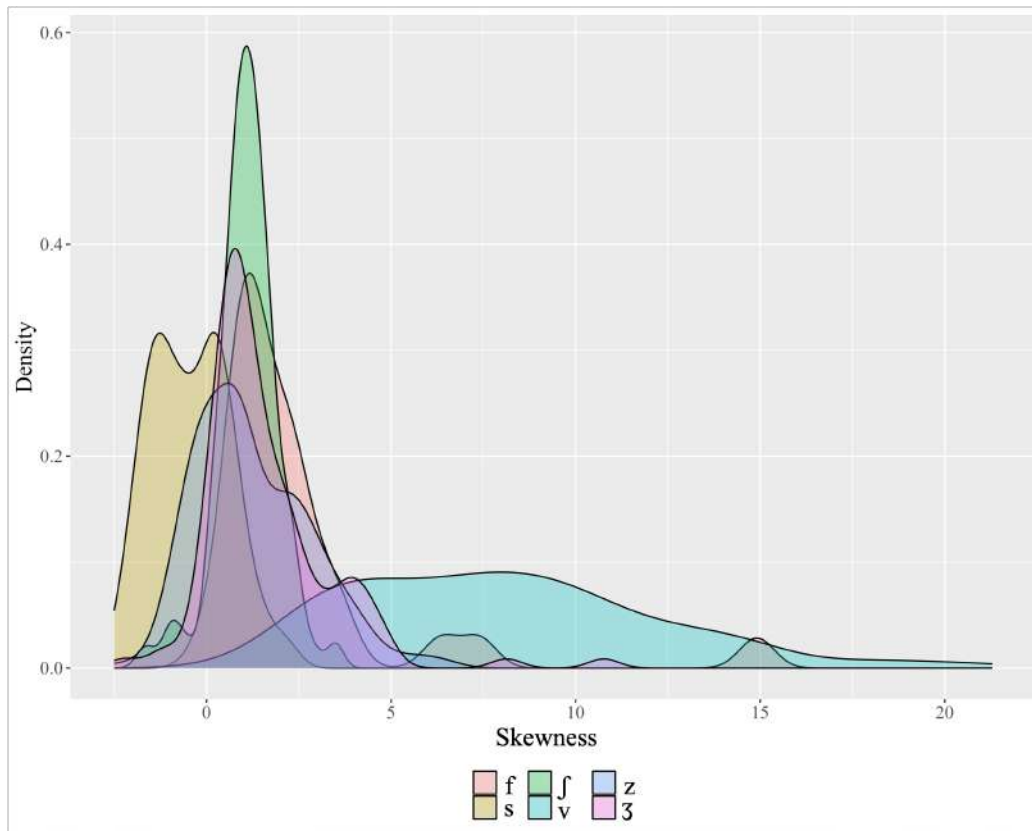


Figure 4-3: Kernel density estimation plots for the six Angami fricatives.

The skewness and kurtosis values for each fricative across the three contexts are provided in Table 4.6. Labio-dental and palatal fricatives produced in sentence frames have slightly higher skewness and kurtosis values than those produced in carrier and isolation. The kurtosis value of the /z/ fricative in the sentence is marginally higher.

4.4.1.2 Temporal features of Fricatives

Figure 4-4 displays the violin plot illustrating the average duration of the six fricatives in Angami. The mean durations are detailed in Table 4.5. As seen in the table, the duration of fricatives in labio-dental fricatives /f/ & /v/ at 207 ms and 156 ms; alveolar /s/ & /z/ at 236 and 182 ms; palatal /j/ & /ʒ/ at 216 and 183 ms, respectively. Notably, the voiceless alveolar /s/ exhibits the longest duration at 236 ms, while the voiced labio-dental fricative /v/ registers the shortest duration at 156 ms.

ms. In general, voiceless fricatives demonstrate longer durations compared to voiced fricatives. The duration of the fricatives is further analyzed based on context. The violin plot of the average duration across the three contexts is shown in Figure 4-5. The detailed values are listed in Table 4.6. As observed in both the figure and the table, the voiceless labio-dental /f/ exhibits a longer duration (213 ms) in sentence frames compared to carrier and isolation contexts. Conversely, when produced in isolation, the voiced alveolar /v/ demonstrates a longer duration (176 ms). Notably, within the alveolar section, fricatives produced in isolation display longer durations, as evidenced by /s/ at 244 ms and /z/ at 195 ms, respectively. A similar finding is seen in palatal fricatives when produced in an isolation context. /ʃ/ and /ʒ/ have a duration of 227 ms and 200 ms, which is relatively greater than the one produced in carrier and sentence. Overall, fricatives produced in isolation are longer than those produced in carrier and sentence. This may be because sentence and carrier frames have co-articulatory effects, which restrict longer duration than those produced in isolation.

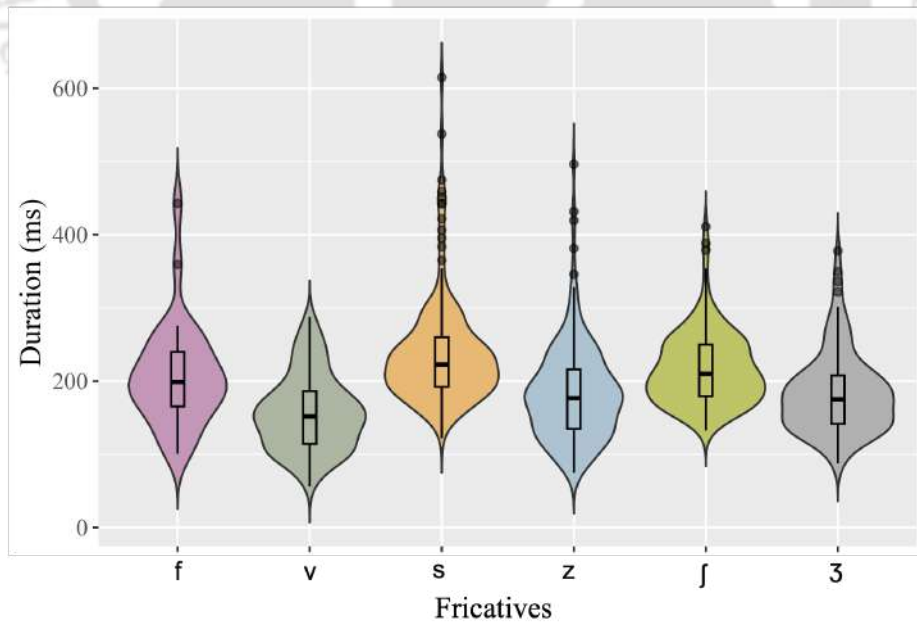


Figure 4-4: Violin plot of the average duration of the six Angami fricatives.

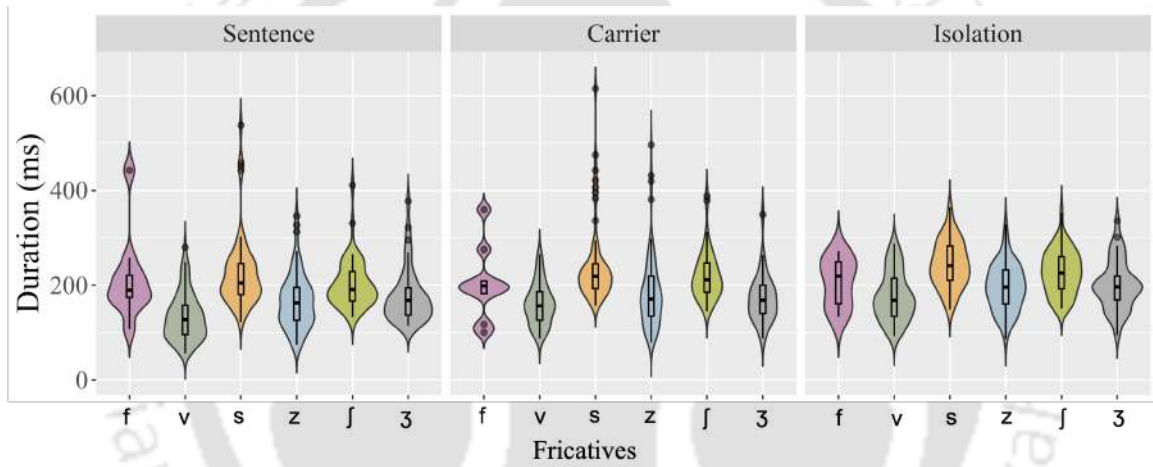


Figure 4-5: Violin plot of the average duration of the six Angami fricatives across the three contexts.

4.4.1.3 Center of Gravity for Fricatives

Figure 4-6 displays the violin plot for the average center of gravity calculated at the mid-point of the fricatives. The values are further provided in Table 4.5. As seen in the figure and table, the CoG values in labio-dental fricatives /f/ & /v/ at 2733 Hz and 638 Hz; alveolar /s/ & /z/ at 6070 Hz and 3060 Hz; palatal /ʃ/ & /ʒ/ at 3664 and 2100 Hz. The voiced labio-dental /v/ exhibits the lowest center of gravity (CoG) values, whereas the highest CoG is observed in the voiceless alveolar /s/. The values of the fricatives are further analyzed based on the contexts; the violin plots are displayed in Figure 4-7 and values in Table 4.6. Across all fricatives, including labio-dental, alveolar, and palatal, the center of gravity (CoG) values are highest when the fricatives are produced in isolation.

4.4.1.4 Voicing in Fricatives

The average value has been plotted for visualization in Figure 4-8 to determine the percentage of voicing in fricatives. The values are provided in Table 4.5. As anticipated, the voiceless fricatives have less (< 5) percentage of voicing. While the voiced fricatives /v/, /z/, and /ʒ/ have an average of 29, 39, and 34 % of voicing. The voiced and voiceless fricatives, when produced in isolation, have lower voicing than carrier and sentence frames as seen in Figure 4-7 and the values in Table 4.6. In all cases, alveolar voices have higher voicing % than bilabial and palatal. The voicing segment can range from high to low depending on the contexts: isolation < sentence < carrier.

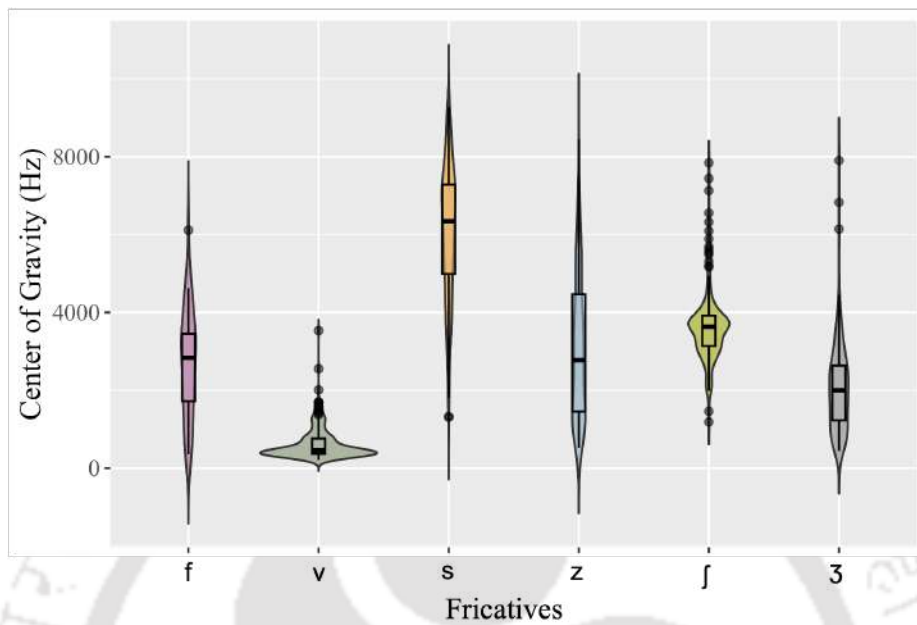


Figure 4-6: Violin plot of the average center of gravity of the six Angami fricatives.

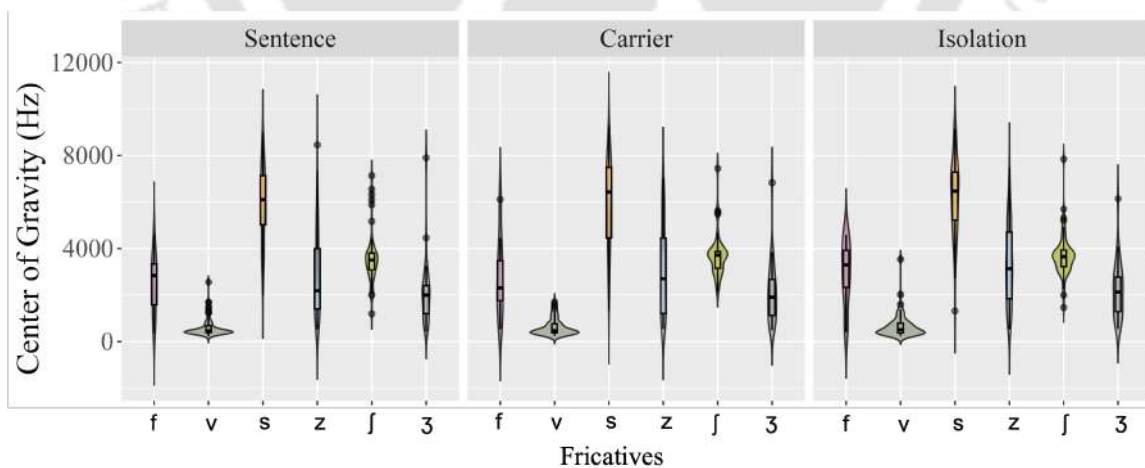


Figure 4-7: Violin plot of the average center of gravity of the six Angami fricatives across the three contexts.

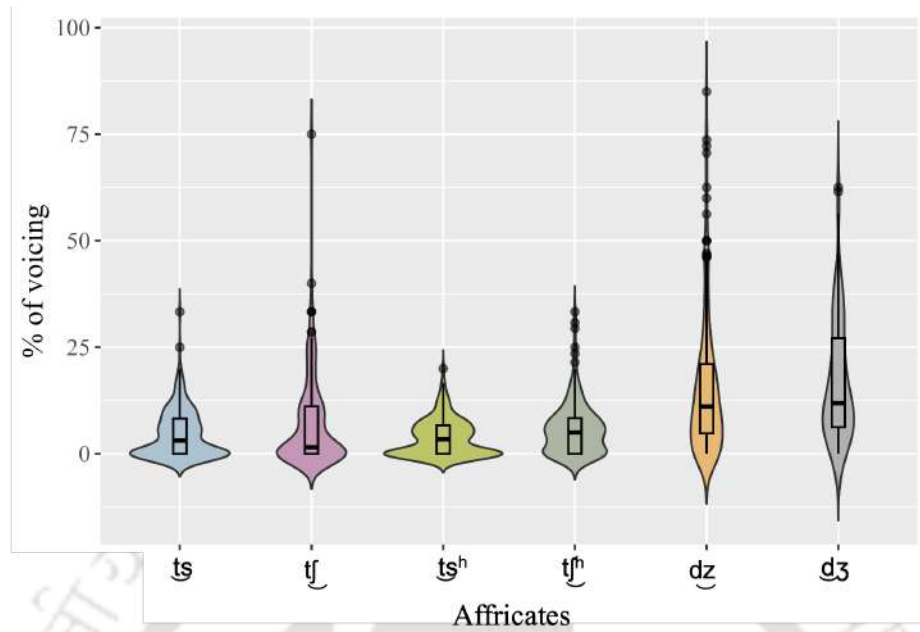


Figure 4-8: Violin plot of the average percentage of voicing of the six Angami fricatives.

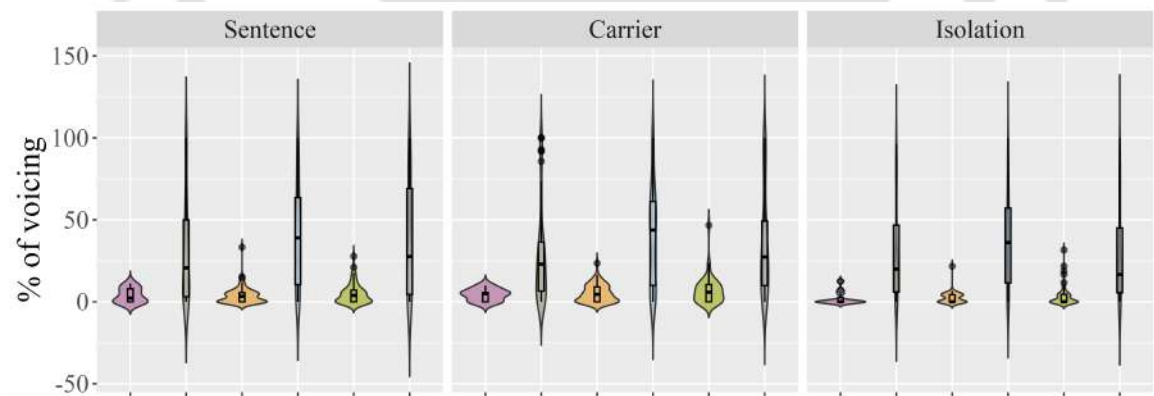


Figure 4-9: Violin plot of the average percentage of voicing of the six Angami fricatives across the three contexts.

4.4.1.5 Intensity in Fricatives

The intensity of each of the fricatives is measured and plotted as seen in Figure 4-10. The mean intensity (dB) remained consistent across all six fricatives. Specifically, the intensity values for labio-dental fricatives /f/ and /v/ were recorded at 27 dB and 30 dB, respectively. Similarly, both alveolar fricatives /s/ and /z/ displayed an intensity of 33 dB. Meanwhile, the palatal fricatives /ʃ/ and /ʒ/ exhibited intensity values of 34 dB and 33 dB, respectively. Consistent values observed across all three contexts suggest that intensity may not be a reliable acoustic correlate for distinguishing fricatives.

Table 4.5: Average (in bold) and standard errors (in reduced font) of the acoustic features in Angami voiced and voiceless fricatives.

Features	/f/	/v/	/s/	/z/	/ʃ/	/ʒ/
Duration	207	156	236	182	216	183
	13	3	4	4	3	5
Voicing	3	29	3	39	5	34
	0	2	0	2	0	3
CoG	2733	638	6070	3060	3664	2100
	262	33	123	124	79	116
Intensity	27	30	33	33	34	33
	1	0	0	0	0	0
Skewness	2.4	8.0	-0.4	1.2	1.1	1.5
	0	0	0	0	0	0
Kurtosis	17	102	1.1	3.1	5.7	9.8
	11	8	0	0	0	2

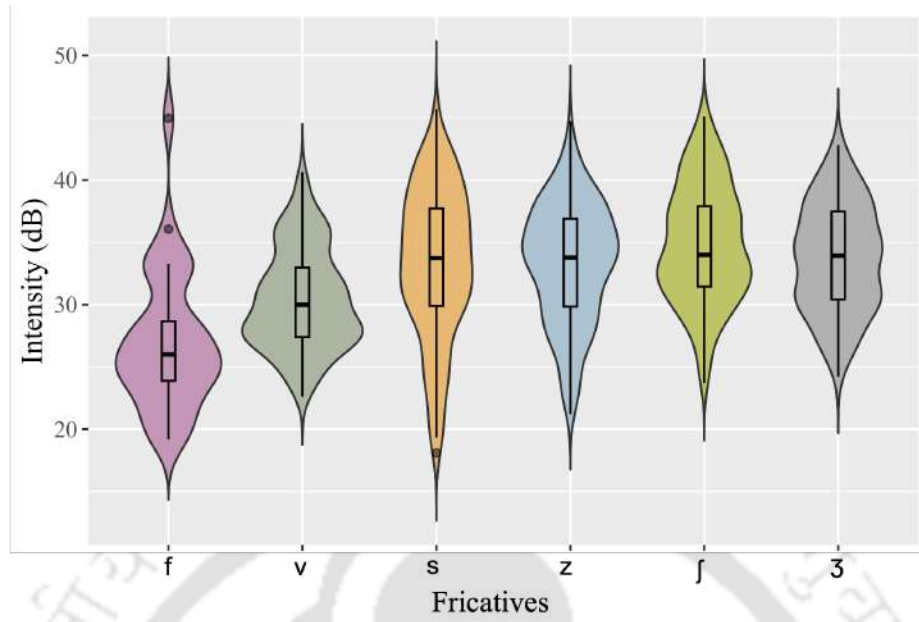
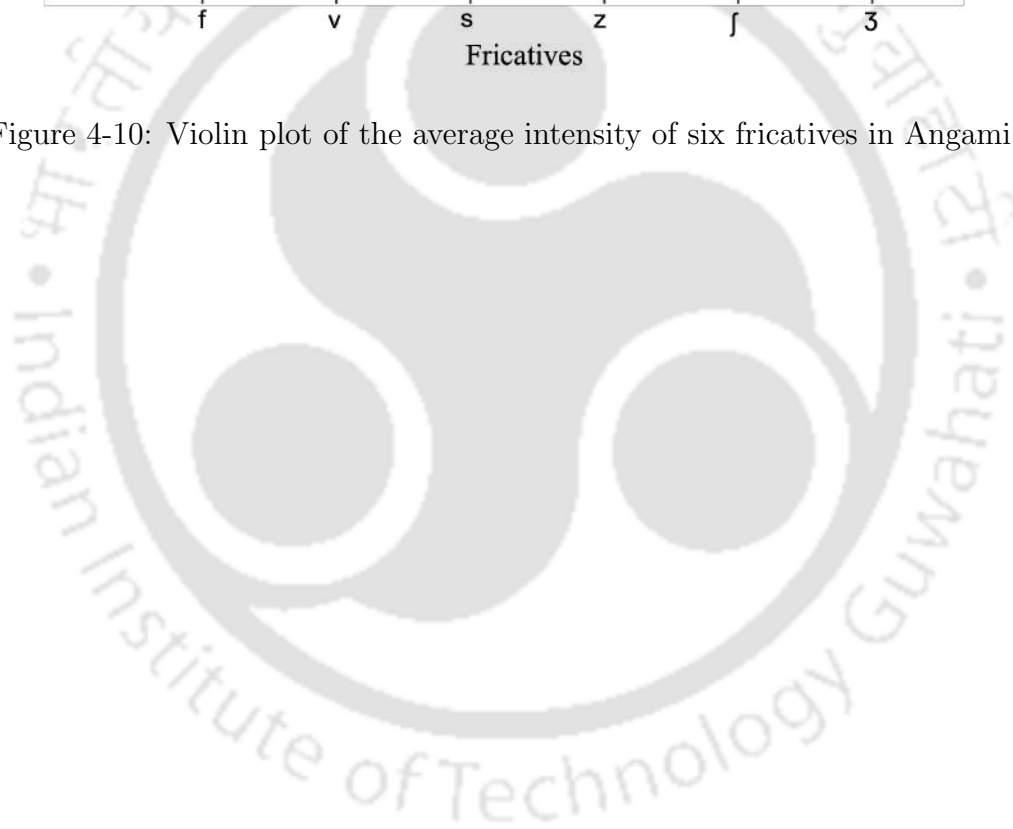


Figure 4-10: Violin plot of the average intensity of six fricatives in Angami.



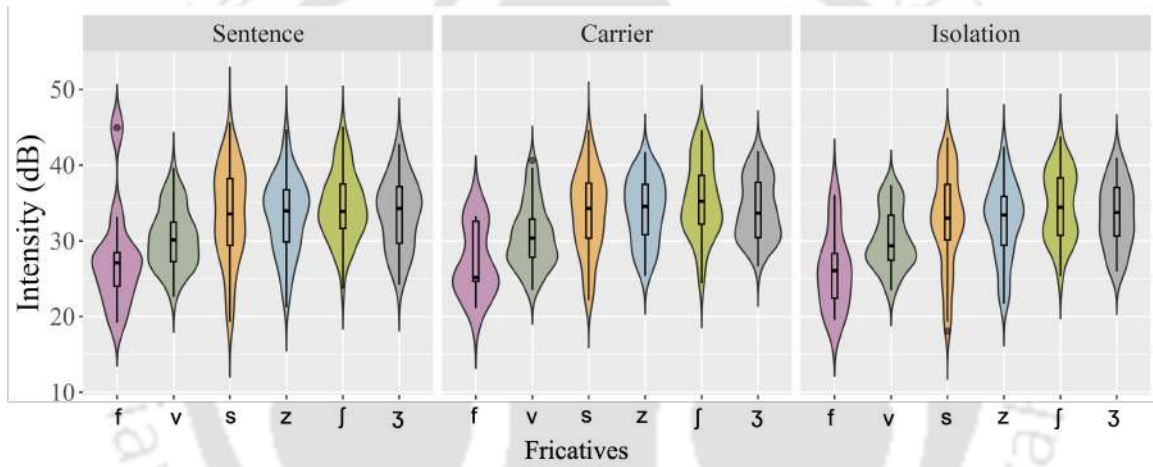


Figure 4-11: Violin plot of the average intensity of six fricatives across the three contexts in Angami.

Table 4.6: Average (in bold) and standard errors (in reduced font) of the acoustic features in Angami voiced and voiceless fricatives across the three contexts.

Features	Labio-dental					
	Carrier		Sentence		Isolation	
	/f/	/v/	/f/	/v/	f/	/v/
Duration	203	160	213	133	205	176
	25	5	28	6	15	6
Voicing	4	28	3	30	2	30
	1	3	1	4	1	3
CoG	2778	620	2491	629	2916	665
	583	47	414	56	415	67
Intensity	27	30	27	30	26	30
	1	0	2	0	1	0
Skewness	2	7	3	8	2	8
	0	0	1	0	0	0
Kurtosis	7	92	37	110	8	105
	4	11	33	15	5	15
Alveolar						
	Carrier		Sentence		Isolation	
	/s/	/z/	/s/	/z/	/s/	/z/
Duration	240	186	225	166	244	195
	9	9	9	6	6	5
Voicing	5	39	3	39	2	37
	0	3	0	3	0	3
CoG	5970	3044	5989	2841	6256	3295
	240	224	200	220	196	198
Intensity	33	33	33	33	32	32
	0	0	0	0	0	0
Skewness	-0.4	1.3	-0.4	1.4	-0.4	0.9
	0	0	0	0	0	0
Kurtosis	1	3	1	3	1	1
	0	1	0	0	0	0

Table 4.6: Average (in bold) and standard errors (in reduced font) of the acoustic features in Angami voiced and voiceless fricatives across the three contexts.

Features	Palatal					
	Carrier		Sentence		Isolation	
	/ʃ/	/ʒ/	/ʃ/	/ʒ/	/ʃ/	/ʒ/
Duration	219	171	202	178	227	200
	6	8	6	9	6	8
Voicing	6	35	4	37	3	31
	0	5	0	5	0	5
CoG	3677	2064	3607	2045	3707	2192
	122	202	159	213	132	194
Intensity	35	34	34	33	34	33
	0	0	0	0	0	0
Skewness	0.8	1.6	1	1.7	1.3	1.4
	0	0	0	0	0	0
Kurtosis	4	8	5	12	6	7
	0	3	1	6	0	1

Separate linear mixed-effect models were created to investigate the statistical significance of the acoustic-phonetic features across the fricatives. The LME models consist of the mean values of the acoustic features such as duration, CoG, intensity, and voicing as the dependent variable. Laryngeal contrasts (LC), PoA, three contexts, and their interactions were the fixed effects, while the speaker and types of words were random effects. The model is shown in 4.1. The summary of the LME model is presented in Table 4.7. The constructed models underwent Type II Wald Chi-square tests for analysis of deviance on Linear Mixed Effects (LME) models separately. These tests revealed that the two laryngeal contrasts, place of articulation (PoA) and context, demonstrate statistically significant effects on the duration of the fricatives. However, it was also noted that the interaction of laryngeal contrasts, PoA, and context has no significant effect. In order to see the pairwise comparisons, the LME model was subjected to a Bonferroni post-hoc test. The summary of the results is reported in Table 4.9.

$$Features \sim LC + PoA + Context + LC * PoA * Context + (1|Speaker) + (1|word) \quad (4.1)$$

The result indicates that duration in fricatives of Angami across laryngeal contrasts and PoA are not significant. On the other hand, we see that phonation and PoA have a significant effect on CoG ($\chi^2(1) = 6.3, p < 0.05$). The result of the post-hoc Bonferroni shows that in voiced fricatives, labio-dental and alveolar, labio-dental and palatal are significantly different. However, alveolar and palatal are not significant from each other. In the case of voiceless unaspirated fricatives, labio-dental, alveolar, and palatal are significantly different regarding CoG. As for intensity, though the fricatives are distinct in terms of PoA and context, the interaction of laryngeal contrast and PoA does not have an effect. In voiced labio-dental and alveolar, labio-dental

and palatal are distinct in voicing. In the case of voiceless fricatives, labio-dental, alveolar, and alveolar-palatal are distinct in intensity. The two laryngeal contrasts and PoA show statistically significant effects on voicing. However, their interaction is not significant.

Furthermore, as seen in Table 4.7, the two laryngeal contrasts, PoA and their interaction, show statistically significant effects on skewness and kurtosis. The results confirm that while the skewness and kurtosis of the two laryngeal contrasts are distinct, they may differ depending on the PoA. Pairwise comparison, as shown in Table 4.9, shows evidence that for skewness values, significant differences are seen between labio-dental and alveolar, as well as labio-dental and palatal fricatives in the context of voiced fricatives. For voiceless unaspirated fricatives, the only significant difference in skewness is observed between labio-dental and alveolar pairs. Similar post-hoc results in kurtosis values as well.

Table 4.7: Results of an ANOVA conducted on an LME model with the acoustic-phonetic features such as duration, CoG, Intensity and voicing as the dependent variable and PoA, phonation, context and their interactions as fixed effects for fricatives in Angami.

Features	Fixed Effects	<i>df</i>	χ^2	p-value
Skewness	Laryngeal contrast	1	47.9338	< 0.001
	PoA	2	383.9807	< 0.001
	context	2	2.0293	n.s
	Laryngeal contrast x PoA	2	36.8854	< 0.001
	Laryngeal contrast x PoA x context	4	1.8138	n.s
Kurtosis	Laryngeal contrast	1	8.2200	< 0.01
	PoA	2	223.2323	< 0.001
	context	2	1.7218	n.s
	Laryngeal contrast x PoA	2	335.7878	< 0.001
	Laryngeal contrast x PoA x context	4	1.2828	n.s
Duration	Laryngeal contrast	1	50.1841	< 0.001
	PoA	2	13.2771	< 0.01
	context	2	61.4931	< 0.001
	Laryngeal contrast x PoA	2	0.8645	n.s
	Laryngeal contrast x PoA x context	4	8.7173	n.s
CoG	Laryngeal contrast	1	66.7398	< 0.001
	PoA	2	57.6894	< 0.001
	context	2	6.2676	< 0.05
	Laryngeal contrast x PoA	2	6.3273	< 0.05
	Laryngeal contrast x PoA x context	4	1.2507	n.s
Intensity	Laryngeal contrast	1	0.0015	n.s
	PoA	2	27.4	< 0.001
	context	2	11.4305	< 0.01
	Laryngeal contrast x PoA	2	4.4307	n.s
	Laryngeal contrast x PoA x context	4	1.8420	n.s
Voicing	Laryngeal contrast	1	345.6842	< 0.001
	PoA	2	11.9950	< 0.01
	context	2	1.7255	n.s
	Laryngeal contrast x PoA	2	3.2124	n.s
	Laryngeal contrast x PoA x context	4	0.6829	n.s

Table 4.8: Results of a Bonferroni post-hoc test conduct on an LME model with skewness, kurtosis, CoG, duration as dependent variable and PoA and phonation as fixed factors.

	Contrasts	Estimate	SE	df	t-ratio	p-value
Skewness	Voiced					
	Labio-dental – Alveolar	-6.7	0.3	39.4	-17.3	< .0001
	Labio-dental – Palatal	6.5	0.4	48.4	14.9	< .0001
	Alveolar – Palatal	-0.1	0.3	115.8	-0.3	n.s
	Voiceless Unaspirated					
	Labio-dental – Alveolar	-2.6	0.7	36.7	-3.4	0.05
	Labio-dental – Palatal	1.6	0.7	37.4	2.1	n.s
	Alveolar – Palatal	-0.9	0.3	85.6	-2.6	n.s
	Kurtosis	Voiced				
Labio-dental – Alveolar		-99.0	7.1	37.6	-13.8	< .0001
Labio-dental – Palatal		94.1	8.3	43.9	11.3	< .0001
Alveolar – Palatal		-4.9	7.6	72.9	-0.6	n.s
Voiceless Unaspirated						
Labio-dental – Alveolar		-99.9	7.2	37.6	-13.7	< .0001
Labio-dental – Palatal		14.4	14.2	35.8	1.0	n.s
Alveolar – Palatal		-0.6	7.1	58.9	-0.0	n.s
Duration		Voiced				
	Labio-dental – Alveolar	27.4	9.6	37.0	2.8	n.s
	Labio-dental – Palatal	-23.4	10.7	46.2	-2.1	n.s
	Alveolar – Palatal	3.9	8.5	155	0.4	n.s
	Voiceless Unaspirated					
	Labio-dental – Alveolar	26.7	19.1	34.4	1.3	n.s
	Labio-dental – Palatal	-12.6	19.3	35.1	-0.6	n.s
	Alveolar – Palatal	14.0	8.3	108	1.6	n.s
	CoG	Voiced				
Labio-dental – Alveolar		2203	417	40.2	5.2	0.0001
Labio-dental – Palatal		-1893	443	49.1	-4.2	0.05
Alveolar – Palatal		310	271	426.2	1.1	n.s
Voiceless Unaspirated						
Labio-dental – Alveolar		2828	842	38.1	3.3	0.05
Labio-dental – Palatal		-3689	429	45.2	-8.5	.0001
Alveolar – Palatal		1229	277	288.6	4.4	0.001

Table 4.9: Results of a Bonferroni post-hoc test conduct on an LME model with intensity and percentage of voicing as dependent variable and PoA and phonation as fixed factors.

Intensity	Voiced					
	Labio-dental – Alveolar	2203	417	40.2	5.2	0.0001
	Labio-dental – Palatal	-1893	443	49.1	-4.2	0.01
	Alveolar – Palatal	310	271	426.2	1.1	n.s
	Voiceless Unaspirated					
	Labio-dental – Alveolar	2828	842	38.1	3.3	0.05
	Labio-dental – Palatal	-1599	847	38.9	-1.8	n.s
	Alveolar – Palatal	1229	277	288.6	4.4	0.001
Voicing	Voiced					
	Labio-dental – Alveolar	9.4	2.6	37.0	3.5	0.05
	Labio-dental – Palatal	-5.5	3.1	42.6	-1.7	n.s
	Alveolar – Palatal	3.9	2.9	65.7	1.3	n.s
	Voiceless Unaspirated					
	Labio-dental – Alveolar	0.7	5.1	34.7	0.1	n.s
	Labio-dental – Palatal	-1.1	5.2	35.2	-0.2	n.s
	Alveolar – Palatal	-0.3	2.7	54.3	-0.1	n.s

4.4.2 Acoustic features of Affricates in Angami

Figure 4-12 illustrates the spectrograph of affricates in three laryngeal contrasts: voiceless palatal affricates / tʃ/, voiceless aspirated affricate / tʃ^h/, and voiced palatal affricates / dʒ/. In the spectrograph of the voiceless affricate, marked by a burst and frication period, as depicted in Figure 4-12(a), the burst indicates the release of the closure, followed by a period of frication characterized by increased energy in higher frequencies. The spectrograph displays a burst, frication period, and subsequent aspiration for the voiceless aspirated affricate, shown in Figure 4-12(b). The aspiration is visible as a noise component characterized by low-intensity energy. Finally, in the spectrograph of the voiced affricate, seen in Figure 4-12(c), the voiced period (-VOT), burst, and friction is evident. The voiced period indicates the presence of vocal fold vibration during the closure, followed by the burst and frication, similar to the voiceless affricates, but with the additional presence of voicing.

4.4.2.1 Spectral moments of Affricates

Figure 4-13 presents the skewness distribution for each affricate, representing the normalcy of the data distribution. The skewness values of the alveolar affricates ts/, /tʃ^h/ & /dʒ/ are -0.2, 0.3 and 0.6, while the skewness values of the palatal affricates /tʃ/, /tʃ^h/ & /dʒ/ are at 0.9, 0.9 and 1.1. The alveolar affricates have a symmetrical distribution of data, except for the voiced alveolar, which has marginally positive skewness. The kurtosis values of the alveolar affricates indicate a short-tailed distribution (< 3). On the other hand, the skewness values of the palatal affricates indicate a positive skewness (> 0.5), while the kurtosis values indicate a normal distribution. The skewness and kurtosis values are also examined across the three contexts; the results are shown in Table 4.11. The table shows that the voiced alveolar produced in carrier frames has a higher value than those produced in isolation and sentence frames. The same result is reflected in the kurtosis values as well.

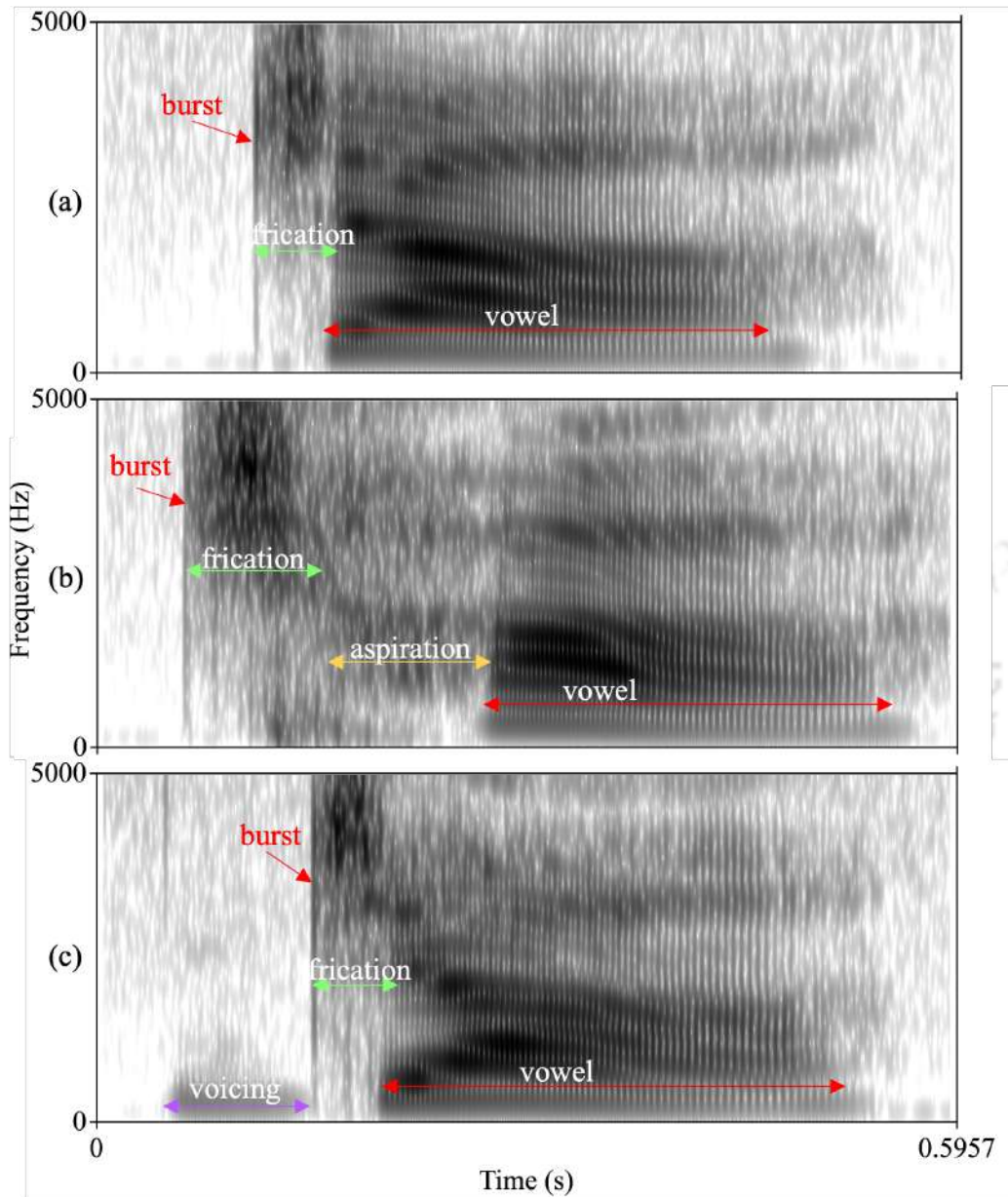


Figure 4-12: Spectrograph of palatal voiceless unaspirated /tʃ/ (a), voiceless aspirated /tʃ^h/ (b) and voiced aspirated /dʒ/ (c) followed by vowel /a/ as produced by a female speaker in isolation.

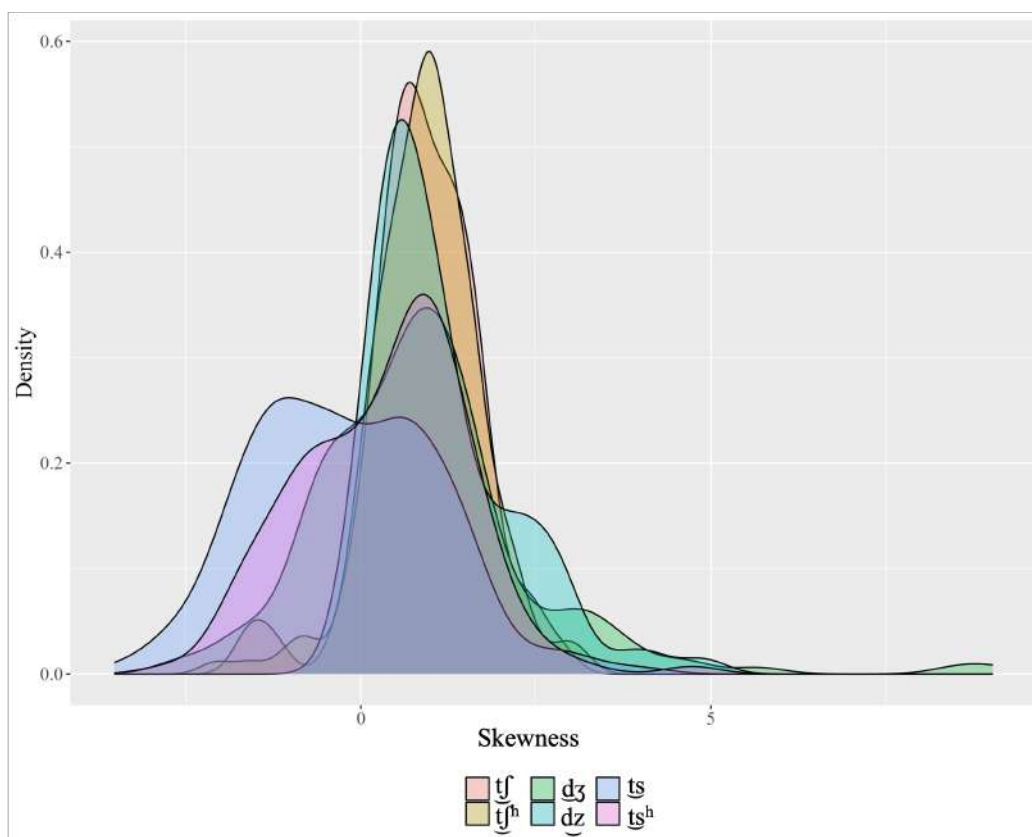


Figure 4-13: Kernel density estimation plots for the six Angami Affricates.

4.4.2.2 Temporal features of Affricates

The average duration of the six affricates is displayed in the violin plot in Figure 4-14. The detailed values are provided in Table 4.10. As seen in the table, the duration of the affricates in alveolar affricates $/\underline{ts}/$, $/\underline{tʃ}^h/$ & $/\underline{dʒ}/$ are at 149 ms, 205 ms and 178 ms. While the palatal affricates $/\underline{tʃ}/$, $/\underline{tʃ}^h/$ & $/\underline{dʒ}/$ are at 123 ms, 198 ms and 164 ms. The voiceless palatal affricate $/\underline{tʃ}/$ exhibit the shortest duration, while the voiceless aspirated alveolar $/\underline{tʃ}^h/$ has the longest duration. In general, voiceless unaspirated affricates have the shortest affricates, followed by voiced affricates. The pattern of affricate duration can be summarized as voiceless unaspirated < voiceless aspirated < voiced unaspirated. The duration of the affricates is further analyzed based on

the three contexts. Figure 4-15 displays the average violin plot, and the values are provided in Table 4.11. For the alveolar segments, affricates produced in sentence frames have a longer duration than the sentence and duration.

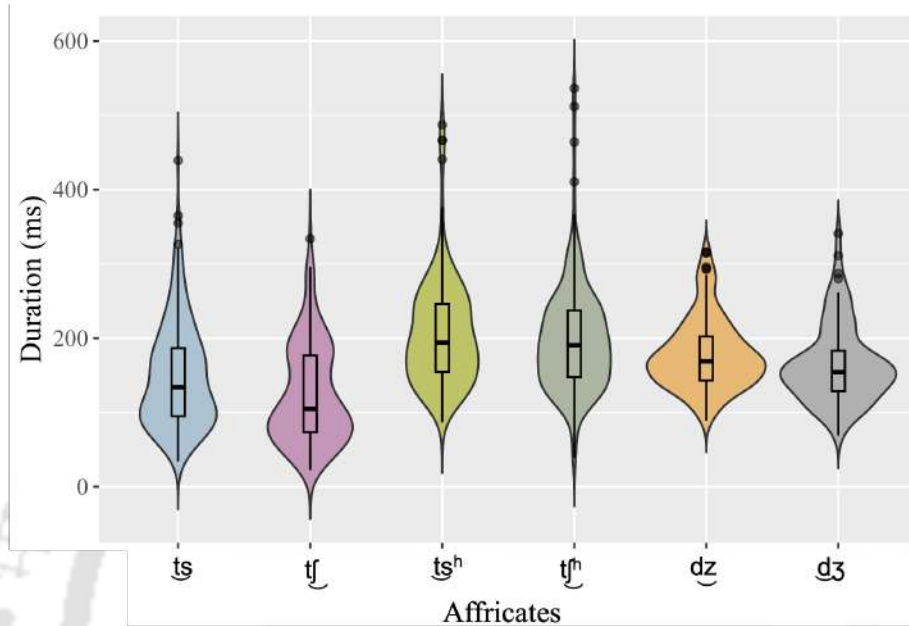


Figure 4-14: Violin plot of the average duration of the six Affricates in Angami.

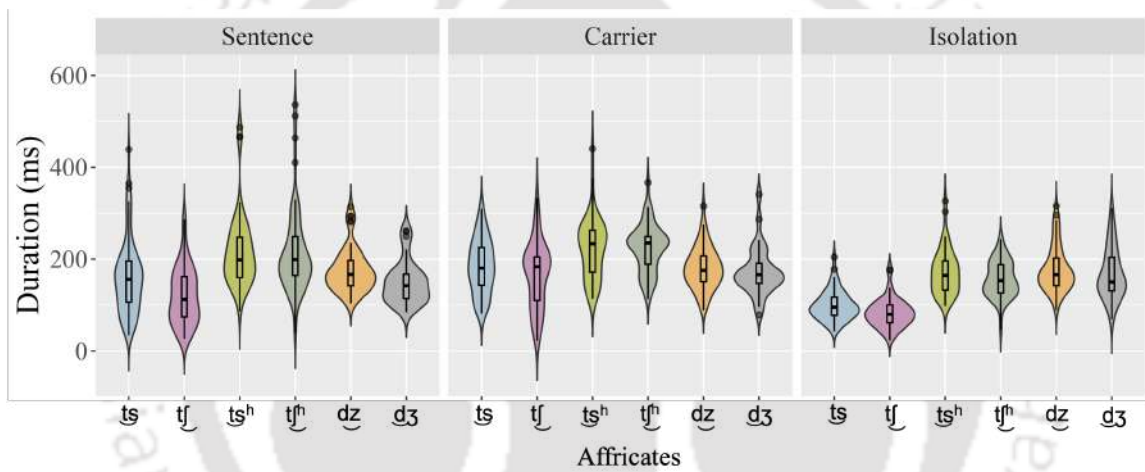


Figure 4-15: Violin plot of the average duration of the six Affricates across the three contexts in Angami.

4.4.2.3 Center of Gravity in Affricates

Figure 4-16 displays the violin plot for the average center of gravity calculated at the mid-point of the fricatives. The values are further provided in Table 4.10. The result shows that alveolar affricates /t͡s/, t͡ʃ^h & /d͡z/ are at 5545 Hz, 4473 Hz & 3499 HZ respectively; whereas the palatal affricates t͡ʃ, /t͡ʃ^h/ & /d͡ʒ/ are at 3507, 3355 and 2369 Hz, respectively. The voiced palatal affricates /d͡ʒ/ has the lowest CoG while the voiceless unaspirated /t͡s/ has the highest CoG. It is observed that the alveolar has higher CoG than the palatal affricates. The values of the affricates are further analyzed based on the contexts. In Figure 4-17 and Table 4.11, it is observed that during the production of alveolar affricates, the Center of Gravity (CoG) values tend to be higher when they are produced in carrier frames compared to when they are produced in sentence and isolation frames.

4.4.2.4 Voicing in Affricates

The average value has been plotted for visualization in Figure 4-18 to determine the percentage of voicing in fricatives. The values are provided in Table 4.10. Like the fricatives, the voiceless affricates have a lesser percentage of voicing, while the voiced affricates have a higher percentage of voicing.

While the voiceless unaspirated affricates have an average of 4 and 7 %, the voiceless aspirated affricates > 5%, and the voiced affricates are at 15 and 16%. The voiced affricates, when produced in carrier and sentence frame, have relatively higher voicing than isolation as seen in Figure 4-19 and the values in Table 4.11.

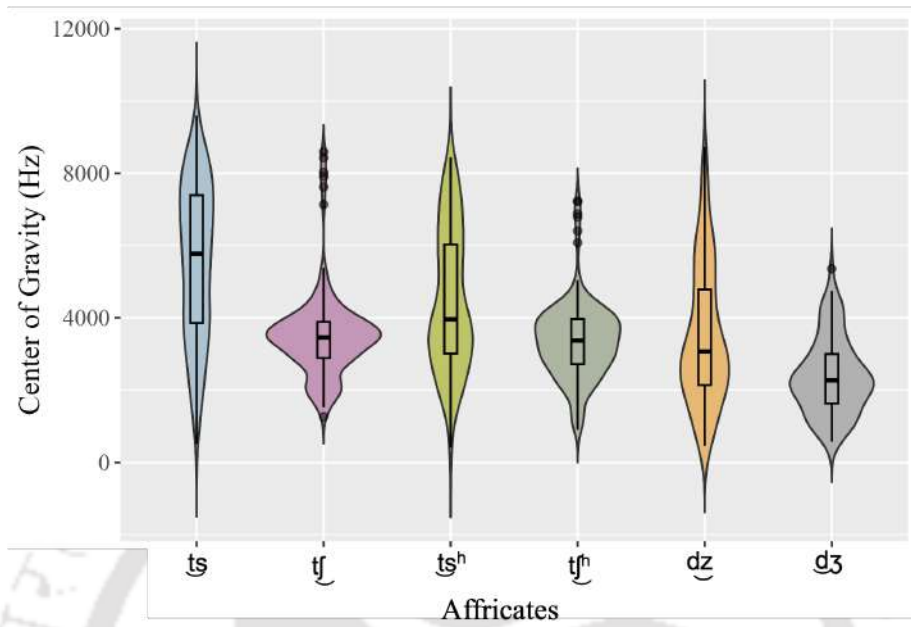


Figure 4-16: Violin plot of the average center of gravity of the six Angami Affricates.

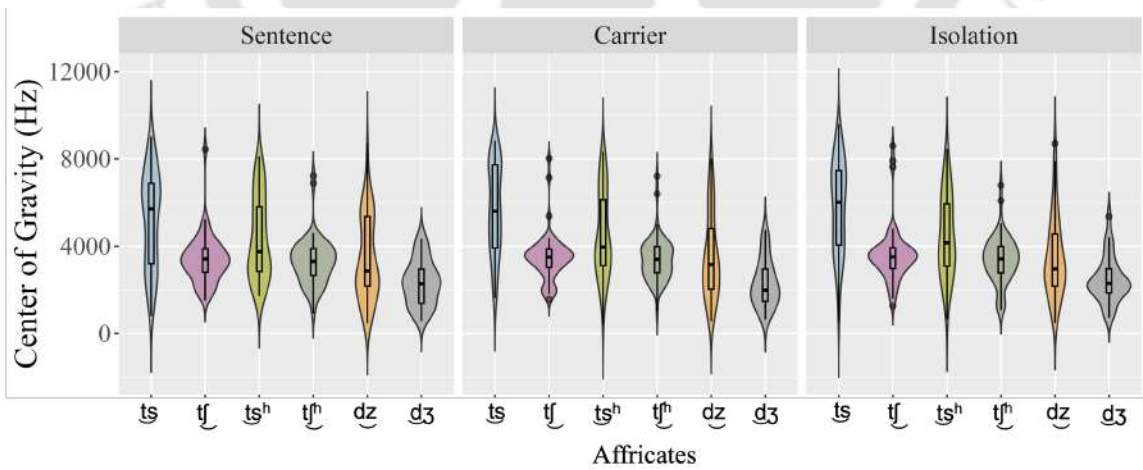


Figure 4-17: Violin plot of the average center of gravity of the six Angami Affricates across the three contexts in Angami.

4.4.2.5 Intensity in Affricates

The intensity of each of the affricates is measured and plotted as displayed in Figure 4-20. The mean intensity (dB) showed equal values for all the affricates, with palatal aspirated affricates at slightly higher values. Looking into the affricates produced in three contexts, the overall intensity (mean) of the affricates, voiceless aspirated affricates t_j^h produced in isolation has the highest intensity. It may have been induced by the aspiration feature of the sound.

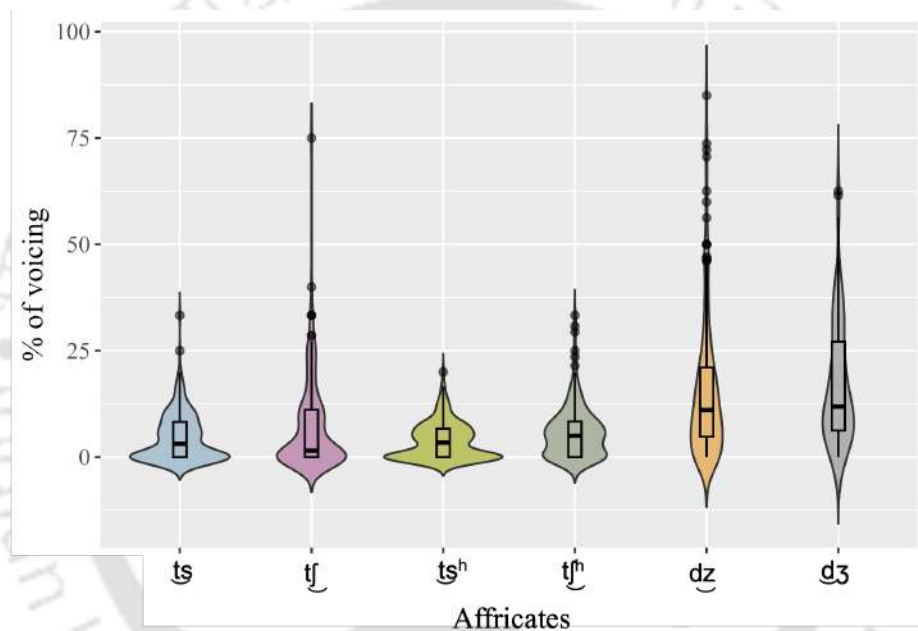


Figure 4-18: Violin plot of the average percentage of voicing of the six Angami Affricates.

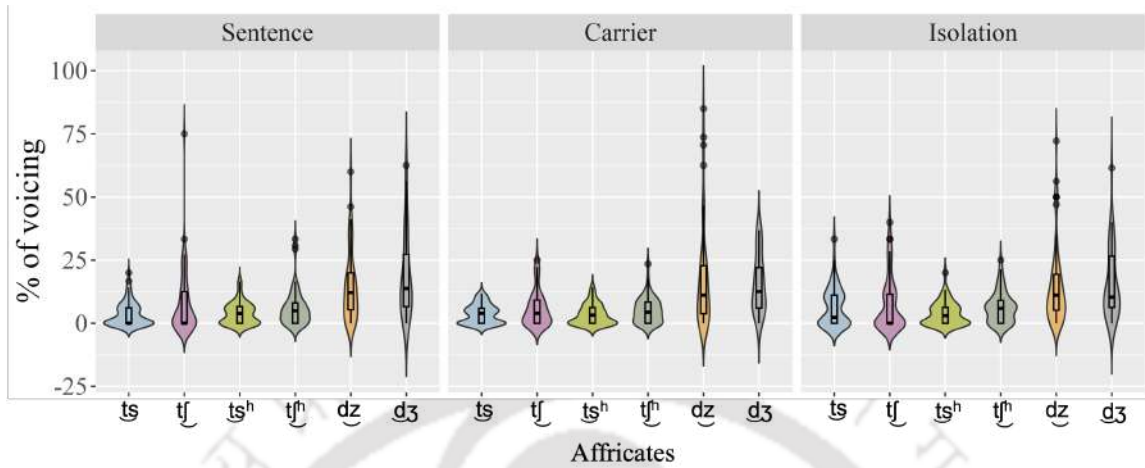


Figure 4-19: Violin plot of the average percentage of voicing of the six Angami Affricates across the three contexts.

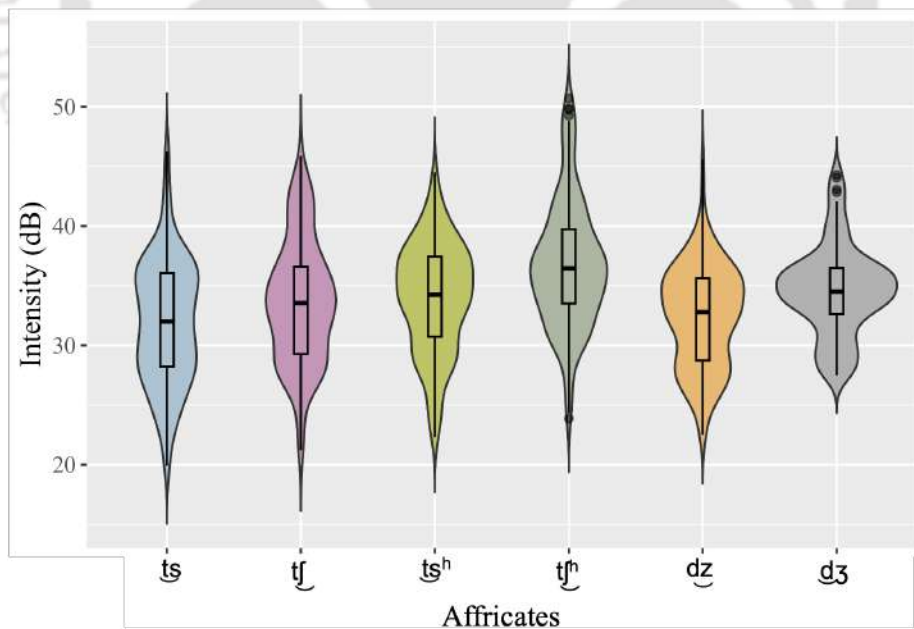


Figure 4-20: Violin plot of the average intensity of the six Angami Affricates.

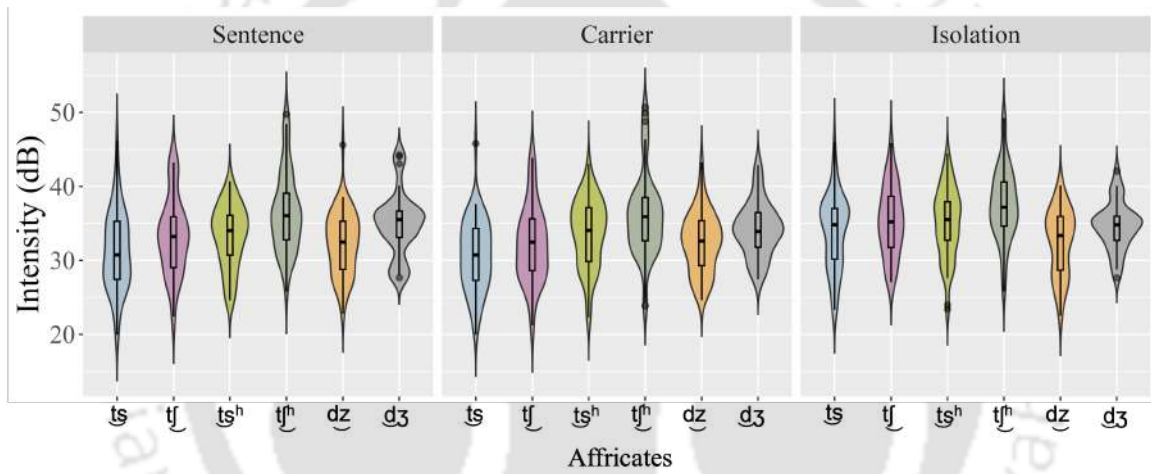


Figure 4-21: Violin plot of the average intensity of the six Angami Affricates across the three contexts.

Table 4.10: Average (in bold) and standard errors (in reduced font) of the acoustic features in Angami voiceless unaspirated, voiceless aspirated and voiced affricates.

Features	/ts/	/ts ^h /	/dz/	/tʃ/	/tʃ ^h /	/dʒ/
Skewness	-0.2	0.3	0.8	0.9	0.9	1.1
	0	0	0	0	0	0
Kurtosis	1.7	1.7	2.9	4.0	3.1	3.7
	0	0	0	0	0	0
Duration	149	205	178	123	198	164
	5	5	3	5	5	5
Voicing	4	3	15	7	5	16
	0	0	1	0	0	1
CoG	5545	4473	3499	3507	3355	2369
	153	160	145	96	83	110
Intensity	31	33	32	33	36	34
	0	0	0	0	0	0

Table 4.11: Average (in bold) and standard errors (in reduced font) of the acoustic features in Angami voiceless unaspirated, voiceless aspirated and voiced affricates.

Features	Alveolar								
	Carrier			Sentence			Isolation		
	/ts/	/ts ^h /	/dz/	/ts/	/ts ^h /	/dz/	/ts/	/ts ^h /	/dz/
Skewness	-0.3	0.2	0.8	-0.0	0.3	1.0	-0.3	0.3	0.7
	0	0	0	0	0	0	0	0	0
Kurtosis	1.4	2.1	2.2	1.5	1.1	4.7	1.7	2.3	1.7
	0	0	0	0	0	2	0	0	0
Duration	185	223	180	163	220	175	98	172	178
	7	9	6	10	12	6	3	7	6
Voicing	3	3	17	3	3	14	6	3	15
	0	0	2	0	0	1	0	0	2
CoG	5675	4552	3542	5196	4369	3463	5776	4495	3492
	257	285	262	270	276	256	269	279	239
Intensity	30	33	32	31	33	32	33	35	32
	0	0	0	0	0	0	0	0	0

In order to confirm the statistical significance of the six acoustic-phonetic features in affricates, we constructed a linear mixed-effect (LME) model similar to the model stated in 4.1. The fixed effects are the mean values of the skewness, kurtosis, duration, CoG, voicing, and intensity. Assuming PoA and laryngeal contrasts affect the features, the interaction of PoA and laryngeal contrasts was also explored (fixed

Table 4.11: Average (in bold) and standard errors (in reduced font) of the acoustic features in Angami voiceless unaspirated, voiceless aspirated and voiced affricates.

Features	Palatal								
	Carrier			Sentence			Isolation		
	$\underline{tj}/$	$\underline{/tj^h}/$	$\underline{/dʒ}/$	$\underline{tj}/$	$\underline{/tj^h}/$	$\underline{/dʒ}/$	$\underline{tj}/$	$\underline{/tj^h}/$	$\underline{/dʒ}/$
Skewness	0.8	0.7	1.2	1.0	1.0	1.2	0.8	1.0	0.9
	0	0	0	0	0	0	0	0	0
Kurtosis	4.0	2.6	4.0	4.7	3.6	4.4	3.3	3.1	2.7
	0	0	1	0	0	1	0	0	0
Duration	166	223	174	121	216	150	82	156	169
	9	7	9	8	12	8	4	5	10
Voicing	5	5	15	8	6	18	7	6	16
	0	0	2	1	1	3	1	0	2
CoG	3502	3397	2316	3387	3288	2316	3627	3383	2462
	156	138	199	158	147	199	186	148	192
Intensity	32	36	34	33	36	34	35	37	34
	0	0	0	0	0	0	0	0	0

effects). Speaker, context, and code were considered random effects. The summary of the LME model is presented in Table 4.12. Each LME model was subjected to a Bonferroni post-hoc test, and the result summary is shown in Table 4.13. The table provides the estimate, standard error (SE), degrees of freedom (df), t-ratio, and significance of the voiced, voiceless unaspirated, and voiceless aspirated affricates under each model are reported.

The result shows that PoA, laryngeal contrast, and their interactions show statistically significant effects on skewness. The pairwise comparison shows that alveolar and palatal are distinct in terms of the voiced, voiceless unaspirated, and voiceless aspirated in the skewness values. On the other hand, phonation and PoA do not have a significant effect on kurtosis values, but their interactions have a significant effect ($\chi^2(2) = 54, p < 0.001$). In examining the pairwise comparisons, it becomes evident that only voiced alveolar and palatal sounds exhibit distinction. Regarding the duration of affricate, laryngeal contrasts, place of articulation (PoA), and context all demonstrate significant effects. However, this significance is not evident in the

interaction between laryngeal contrasts and PoA. Furthermore, we also see that the laryngeal contrast and PoA do have a significant effect on CoG; however, the same is not observed in terms of their interactions. In the pairwise comparison, we see that only the voiceless unaspirated alveolar and palatal pair significantly affect the CoG. Regarding the intensity of affricates, phonation, and PoA have a significant effect. However, their interactions are not significant. The result of the pairwise comparison shows that voiceless unaspirated and voiceless aspirated do have a significant effect. Last but not least, laryngeal contrast and PoA do have a significant effect on voicing. However, their interactions do not. In summary, we see that the laryngeal contrast and PoA significantly affect only the spectral moments, i.e., skewness and kurtosis.

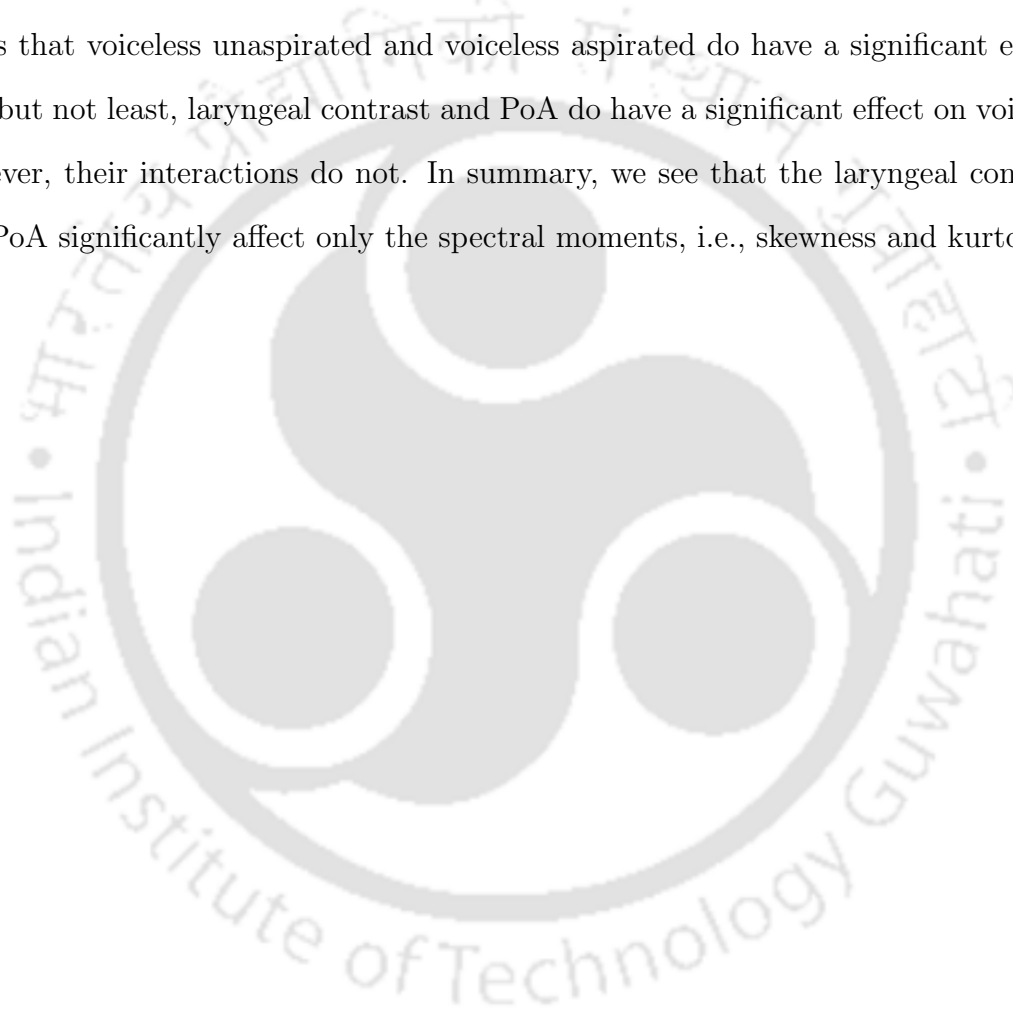


Table 4.12: Results of an ANOVA conducted on an LME model with the six acoustic-phonetic features as the dependent variable and PoA, laryngeal contrasts, context and their interactions as fixed effects for Affricates in Angami.

Features	Fixed Effects	df	χ^2	p-value
Skewness	Laryngeal contrast	2	4.2740	n.s
	PoA	1	4.3213	< 0.05
	context	2	12.2056	< 0.01
	Laryngeal contrast x PoA	2	27.3199	< 0.001
	Laryngeal contrast x PoA x context	4	1.6567	n.s
Kurtosis	Laryngeal contrast	2	1.2088	n.s
	PoA	1	1.4719	n.s
	context	2	3.5947	n.s
	Laryngeal contrast x PoA	2	54.5088	< 0.001
	Laryngeal contrast x PoA x context	4	4.2388	n.s
Duration	Laryngeal contrast	2	59.6284	< 0.001
	PoA	1	5.0331	< 0.05
	context	2	151.8281	< 0.001
	Laryngeal contrast x PoA	2	0.6904	n.s
	Laryngeal contrast x PoA x context	4	3.0438	n.s
CoG	Laryngeal contrast	2	10.5111	< 0.01
	PoA	1	30.3262	< 0.001
	context	2	5.5319	n.s
	Laryngeal contrast x PoA	2	0.8237	n.s
	Laryngeal contrast x PoA x context	4	1.7038	n.s
Intensity	Laryngeal contrast	2	12.0270	< 0.01
	PoA	1	25.0094	< 0.001
	context	2	58.8636	< 0.001
	Laryngeal contrast x PoA	2	0.8404	n.s
	Laryngeal contrast x PoA x context	4	0.3405	n.s
Voicing	Laryngeal contrast	2	111.0323	< 0.001
	PoA	1	5.0332	< 0.05
	context	2	1.2276	n.s
	Laryngeal contrast x PoA	2	0.0751	n.s
	Laryngeal contrast x PoA x context	4	0.80343	n.s

Table 4.13: Results of a Bonferroni post-hoc test conduct on an LME model with skewness, kurtosis, CoG, duration as dependent variable and PoA and Laryngeal contrast as fixed factors for Affricates.

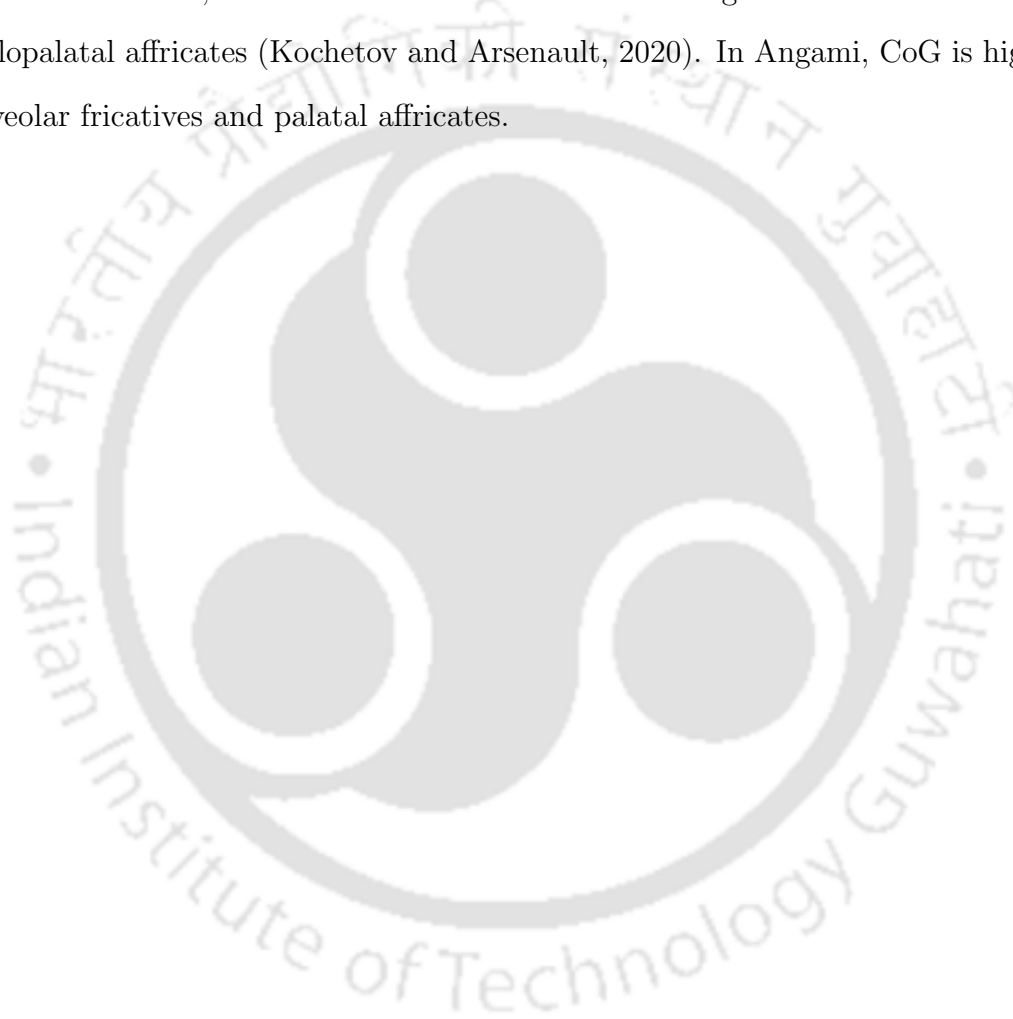
	Contrasts	Estimate	SE	df	t-ratio	p-value	
Skewness	Voiced						
	Alveolar – Palatal	0.783	0.252	618.0	3.109	0.0295	
	Voiceless Unaspirated						
	Alveolar – Palatal	-0.567	0.175	814.1	-3.247	0.0182	
Skewness	Voiceless aspirated						
	Alveolar – Palatal	-1.038	0.336	145.9	-3.089	0.0361	
	Kurtosis	Voiced					
		Alveolar – Palatal	10.373	1.56	495.8	6.635	< .0001
Voiceless Unaspirated							
Alveolar – Palatal		-2.269	1.09	725.2	-2.074	n.s	
Kurtosis	Voiceless aspirated						
	Alveolar – Palatal	-3.768	1.99	122.2	-1.893	n.s	
	Duration	Voiced					
		Alveolar – Palatal	13.5	12.02	114.7	1.123	n.s
Voiceless Unaspirated							
Alveolar – Palatal		17.7	9.09	193.3	1.944	n.s	
Duration	Voiceless aspirated						
	Alveolar – Palatal	6.0	12.09	52.6	0.497	n.s	
	CoG	Voiced					
		Alveolar – Palatal	776.5	348	634.1	2.229	n.s
Voiceless Unaspirated							
Alveolar – Palatal		1009.4	241	823.8	4.183	0.0005	
CoG	Voiceless aspirated						
	Alveolar – Palatal	1270.9	467	149.7	2.720	n.s	
	Intensity	Voiced					
		Alveolar – Palatal	-2.018	0.767	316.3	-2.632	n.s
Voiceless Unaspirated							
Alveolar – Palatal		-1.677	0.546	532.2	-3.074	0.0333	
Intensity	Voiceless aspirated						
	Alveolar – Palatal	-2.589	0.910	84.3	-2.845	0.0837	
	Voicing	Voiced					
		Alveolar – Palatal	-1.763	1.83	73.2	-0.963	n.s
Voiceless Unaspirated							
Alveolar – Palatal		-1.872	1.45	109.6	-1.294	n.s	
Voicing	Voiceless aspirated						
	Alveolar – Palatal	-2.346	1.71	44.9	-1.375	n.s	

4.5 Discussion & Conclusion

In this chapter, the acoustic correlates of fricatives and affricates of the Angami were investigated. The fricatives occur in four places of articulation, namely, labio-dental, alveolar, palatal, and glottal. Out of these, only the first three were considered for the acoustic study. The fricatives contrast in two types of voicing: voiceless and voiced. The affricates in Angami occur in three places of articulation: labio-dental, alveolar, and palatal. The affricates have three voicing distinctions: voiceless unaspirated, voiceless aspirated, and voiced unaspirated. Amongst these, six affricates were considered for acoustic analyses, namely, alveolar and palatal, as they make three voicing distinctions.

The overall pattern noticed in the study of fricatives is that the spectrographs of the palatal fricatives have a darker frication area in the higher frequencies when compared to alveolar fricatives (as seen in Figure 4-2. Regarding the acoustic-phonetic feature of fricatives, voicing (laryngeal contrasts) and place of articulation (PoA) interactions significantly affect skewness, kurtosis, and CoG. The findings of this study regarding the significant effect of Center of Gravity (CoG) on fricatives are in agreement with existing studies on cross-linguistic study in languages such as Aleut, Apache, Chickasaw, Scottish Gaelic, Hupa, Montana Salish, and Toda (Gordon et al., 2002), as well as in the investigation of Chinese voiceless sibilants (Svantesson, 1986). This study shows that alveolar fricatives /s/ exhibit the highest CoG, while voiced labio-dental fricatives exhibit the lowest CoG value. This study further showed that the spectral peak of fricatives is higher in palatal than the alveolar, the opposite result of the reports in English (Jongman et al., 2000). It is also seen that the kurtosis values are also higher in the voiced fricatives than the voiceless fricatives. In English and Greek fricative studies, temporal measurements, and intensity distinguish between voiced and voiceless fricatives (Jongman et al., 2000; Nirgianaki, 2014). In this study, duration, intensity, and voicing percentage were deemed insignificant.

In investigating affricates, the acoustic-phonetic features of affricates, voicing (laryngeal contrasts), and place of articulation (PoA) interactions significantly affect skewness and kurtosis. Figure 4-13 shows that the skewness peaks highest in voiceless unaspirated and aspirated palatal affricates. The skewness and kurtosis values are higher in palatal than in alveolar affricates; this result is similar to the findings reported in Kalasha, where skewness and kurtosis have higher values for dental and alveolopalatal affricates (Kochetov and Arsenault, 2020). In Angami, CoG is highest in alveolar fricatives and palatal affricates.



Chapter 5

Acoustics of Angami Laterals and Approximants

Voiceless laterals and approximants are relatively rare cross-linguistically, with Angami being an exception where they exhibit aspirated characteristics. This investigation delves into the distinctive traits of voiced and voiceless laterals and approximants in the language produced in three contexts. The result of the study showed that there is a clear distinction between voicing laterals and approximants, and the acoustic-phonetic properties that are significant in voicing distinction are formants (F1- F2), Harmonics to Noise Ratio (HNR), and amount of voicing. F1 is statistically significant for voicing in laterals /l/ and /l^h/, as well as in palatal approximants /j/ and /j^h/. Whereas, F1-F2 is distinct for bilabial approximants /w/ and /w^h/, while F1-F4 shows significance for voicing in alveolar approximants /ɹ/ and /ɹ^h/. The study also shows that voiceless laterals and approximants differ in the contexts. Statistical analyses provide further support for the observed distinctions.¹

¹Part of this chapter is published as proceedings of a peer-reviewed conference- “Voiced and voiceless laterals in Angami”. In Proceedings of Interspeech 2024, 3704-3708.

5.1 Introduction and literature review

Cross-linguistically, laterals are primarily voiced. However, there is evidence that some languages have a phonemic distinction between voiced and voiceless laterals in languages such as Burmese, Tibetan, and Non-Tibeto-Burman languages (Estonian Swedish, Klamath) (Ladefoged and Maddieson, 1996; Maddieson and Emmorey, 1984; Asu et al., 2015; Blevins, 1993). Among the languages that have voicing contrasts in laterals, the Tibeto-Burman language forms a large cluster (Moran and McCloy, 2019). However, there are only a few studies that have described the acoustic properties associated with the voiceless laterals (Maddieson and Emmorey, 1984; Asu et al., 2015; Steed and Hardie, 2004). Considering that this work provides an acoustic analysis of voiced and voiceless laterals in Angami.

Lateral consonants are produced by positioning the tongue against the palate to permit continuous airflow around the sides of the tongue. This configuration produces a sound characterized by lateral airflow alongside the tongue, creating an approximant quality. Across languages, it is common for a language to have a lateral sound but not common to have more than one. For example, the Australian aboriginal language, Kaititj, has four places of articulation (PoA): laminal dental, apical alveolar, apical post-alveolar, and laminal post-alveolar laterals (Ladefoged and Maddieson, 1996; Tabain et al., 2016a). Albanian distinguishes between apical dental and alveolar, whereas Tamil and Telegu distinguish between apical and sub-apical lateral retroflex (Ladefoged and Maddieson, 1996). While laterals seem to contrast in terms of PoA, the contrast in voicing across languages is rare.

In terms of voicing contrasts in laterals, it was shown that languages like Burmese, Tibetan, Klamath, Laai, Kuy, and several Irish dialects have voiced and voiceless lateral approximants as contrastive segments (Ladefoged and Maddieson, 1996). It is also argued that Hindi distinguishes between breathy and plain-voiced lateral approximants (Ohala, 1983). In the case of the Tibeto-Burman languages spoken in

India, voicing contrast in laterals is reported in Angami, Mizo, Lotha, and Mongsen Ao (Chase, 1992; Fanai, 1992; Bruhn, 2014). In the existing literature, the voiceless laterals are often described as aspirated laterals or as voiceless lateral fricatives. However, Maddieson and Emmorey (1984), in their study of Zulu, Navajo, Taishan, Burmese, and Tibetan, state that voiceless lateral fricatives (VLF) and voiceless lateral approximants (VLA) exhibit distinct phonetic characteristics. According to the authors, voiceless lateral fricatives are characterized by longer duration, broad high-frequency peaks, and a larger noise amplitude than voiceless lateral approximants. Phonotactically, VLA is constrained to the initial position of the syllable and co-occurs with its voiced counterpart in the language's phonetic inventory (Maddieson and Emmorey, 1984). In light of the above discussion, we believe that the voiceless laterals in Angami are voiceless lateral approximants because they are restricted to syllable onsets and contrast from their voiced counterpart in the language.

Further, Tibeto-Burman languages have a higher prevalence of voiceless lateral approximants. In the characterization of lateral consonants in three central Australian languages (Arrernte, Pitjantjatjara, and Warlpiri), it was observed that the place of articulation in laterals are determined by spectral measures such as Center of Gravity (CoG), Standard Deviation (SD), and the first four formants (Tabain et al., 2014, 2016b). Similarly, in a study on Welsh, Icelandic, and Estonian Swedish, acoustic features such as duration and intensity were explored to distinguish VLA and VLF (Asu et al., 2015). Maddieson and Emmorey explored the spectral shape characteristics to characterize the VLA and VLF (Maddieson and Emmorey, 1984).

In the study of English approximants, Ladefoged and Disner (2012) reported that in the production of the word /wet/, the second formant of the approximant /w/ rises in the onset of the word, while in the production of /yet/, /y/ has a rise in the first formant and fall in the second formant. For the /l/ sound in the word /let/, there is a break in between before the onset of the vowel. The /ɹ/ (r) for the word

/retch/, the third formant has a low frequency. The authors note that whenever there is a /ɹ/, the third formant is always below 2000 Hz. Considering the discussion on the acoustic-phonetic features used to characterize laterals, we will employ both temporal and spectral features in the current study to characterize the voiced and voiceless laterals in Angami.

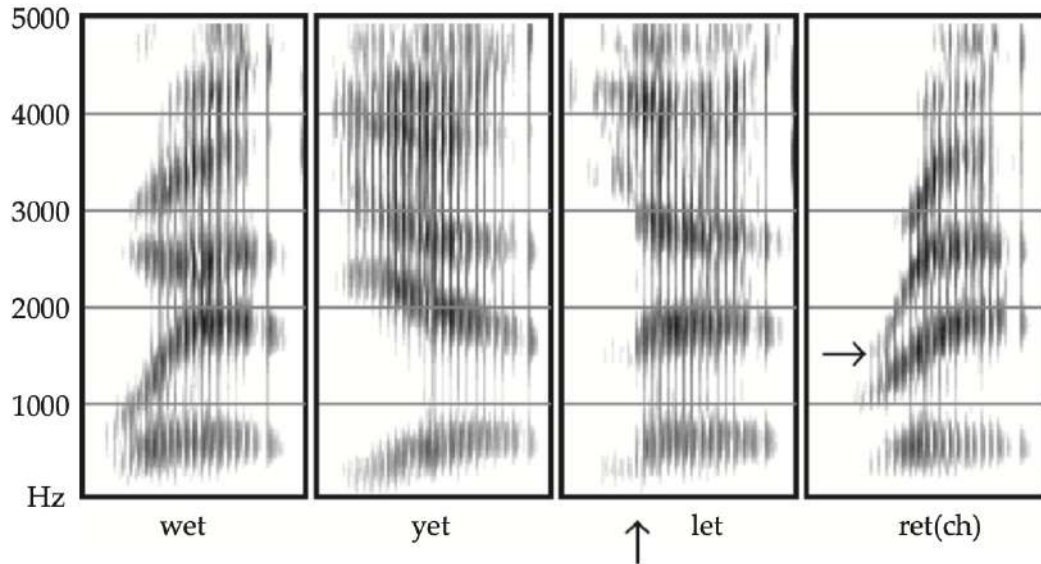


Figure 5-1: Spectrograms of English approximants for the word *wet*, *yet*, *let* and *ret(ch)* in Ladefoged and Disner (2012)

5.2 Literature review of Angami laterals & approximants

The literature on laterals in the Angami language is primarily descriptive and is divided into describing the voiced and voiceless contrast in laterals. Except for Blankenship et al. (1993) and Meyase (2016), who categorize the laterals as voiced and voiceless, other authors predominantly describe /l^h/ as an aspirated lateral sound (Burling, 1960; Ravindran, 1974; Giridhar, 1980; Kuolie, 2006). In case of the place of articulation, the laterals have further been described either as alveolar (Giridhar, 1980;

Kuolie, 2006; Blankenship et al., 1993; Ravindran, 1974; Meyase, 2016), apico-dental (Burling, 1960) and dental Suokhrie (2015). The lateral in Angami is produced by placing the tongue tip touching the alveolar ridge and, in some instances, the back of the teeth while air flows through the sides; it is close to denti-alveolar. This production of the lateral sound, in particular, the position of the tongue, is similar to the denti-alveolar (t, t^h & d) plosives stated in Chapter 2.

Various authors have provided different descriptions for the sounds represented by < r > and < rh > in Angami. While few have described the sound as a trill (Burling, 1960; Ravindran, 1974; Giridhar, 1980; Kuolie, 2006), others have considered it as a tap (Suokhrie, 2015; Meyase, 2016) and an approximant (Blankenship et al., 1993). In the case of place of articulation, < r > and < rh > are described to be produced in the post-alveolar region, retroflex, and as alveolar. Similar to the description of the laterals, the authors predominantly described the contrast as unaspirated voiced and aspirated voiced segments. Trill sounds are produced through a rapid series of closures where the articulators quickly come into contact. Spanish has two types of < r >, one of which is a trill (*perro-pero* ‘dog’) and the other a tap (*pero- pero* ‘but’) (Ladefoged et al., 2006). In contrast, the Angami < r > production does not involve such a series of closures, indicating that it does not qualify as a trill. Blankenship et al. (1993) has defined the r_l as a voiced and voiceless retroflex approximant in describing sounds in Khonoma Angami. To produce retroflex sounds, the tongue’s tip curls backward and is positioned near the back of the alveolar ridge (Stevens and Blumstein, 1975). Since the tongue does not curl in the production of r_l in Angami, the sound is not a retroflex. The Angami < r > sound is produced the same as the English, < r > which is an approximant /ɹ/, the production involves elevating the tongue’s tip toward a point on the palate well beyond the upper front teeth without making contact (Ladefoged and Disner, 2012).

In the production of approximants represented by < w > and < wh >, au-

thors agree that they are bilabial in their production, except for Blankenship et al. (1993), who describes /w/ as a voiced labio-velar approximant, and /ɱ/ as a voiceless labio-velar approximant. As sonorant in nature, these sound segments have been predominantly characterized either as semi-vowels or, more broadly, as approximants in the existing literature. Finally, in the case of < j > and < jh > sound, all authors agree that the sound is a palatal approximant (Burling, 1960; Ravindran, 1974; Giridhar, 1980; Blankenship et al., 1993; Kuolie, 2006; Suokhrie, 2015; Meyase, 2016). However, as noted earlier (Section 1.2.4), in the orthography of Angami, it is often written with the grapheme < y > and < yh >. Table 5.1 summarizes the different descriptions of the approximant sounds described by various authors. The minimal pair of voiced laterals, voiceless laterals, and voiced approximants, voiceless approximants in Angami is shown in Table 5.2.

Given the description of the lateral and approximant sounds in Angami, the goal of this paper is to provide an acoustic-phonetic characterization of the laterals in order to help describe the consonants in an unambiguous manner. The rest of the paper is organized as follows, Section 5.3 provides the methodology used in the work, while, Section 5.4, the results obtained by various acoustic measures are reported. Finally, in Section 5.5, we conclude the study.

Table 5.1: Approximants described in Angami by various authors. Here, Approx. stands for approximants and L-velar for labial velar. Original transcription are retained for the cited references.

Author	Type	Bilabial	Alveolar	Post-Alveolar	Retroflex	Palatal	L-velar
Burling (1960)	Trill Approx.	w wh	r rh			y yh	
Ravindran (1974)	Trill Semi-vowel	w wh		R Rh		y yh	
Giridhar (1980)	Trill Semi-vowel	w wh		r rh		y yh	
Blankenship (1992)	Approx.				ɻ ɻ̣	j j̣	w ɰ
Kuolie (2006)	Trill Approx.	w wh	r rh			y yh	
Suokhrie (2015)	Tap Approx.	w wh	r rh			y yh	
Meyase (2016)	Tap Approx.	w ɰ ^h	r r̥ ^h			j j̥ ^h	

Table 5.2: Laterals and Approximants minimal pairs in Angami.

Word	Meaning	Word	Meaning
le	peel	l ^h e	healthy
li	pot	l ^h i	business
wi	Bos frontalis	w ^h i	around
we	pronoun (we)	w ^h e	hover
.ra	place	r ^h a	redeem
.ru	bone	r ^h u	dirt
ja	cover	j ^h a	lift
ju	jokes	j ^h u	desicate

5.3 Data and Methodology

This Section describes the participants, the materials, and the recording procedure. Apart from that, the acoustic features used to characterize the laterals and approximants are discussed in this Section. Finally, the statistical methods used to model the data are discussed.

5.3.1 Speakers

Speech data were collected from 10 native Angami speakers (5M, 5F) in this study. The participants are all residents of Kohima Village. Speakers' ages range from 27 to 43 years, with an average age of 31.4 ($\sigma = 6.3$) at the time of recording. In addition to Angami, all participants were fluent in English (the state's official language) and Nagamese (a *lingua franca*).

5.3.2 Materials

In order to collect speech data, the speakers were provided a list of Angami words and phrases that contrast voiced and voiceless laterals. Meaningful words were used to capture the voicing contrast in laterals and approximants. Each sound was combined

with all the possible vowels /i, e, a, ə, u, o/. The words were produced in three contexts, as described in Section 1.3.3. Hence, 300 lateral tokens and 784 approximant tokens were recorded. The data recordings were conducted in Kohima village, and the speech data were recorded in October 2022. The recordings were conducted following the standard procedure as discussed in Section 1.3.3.

5.3.3 Acoustic features and statistical analyses

The speech recordings were transferred to a computer and were manually segmented and annotated by the first author, a native speaker of Angami, using Praat 6.0.43 (Boersma, Paul and Weenink, David, 2020). Five acoustic-phonetic features, namely, formant frequencies (F1-F4), intensity, amount of voicing, Harmonics to Noise Ratio (HNR), and segment duration, were extracted to characterize the laterals and approximants.

In order to visualize the lateral and approximant formant contours, formant values (F1-F4) are extracted at every 10% of the total duration. The previous study (Chapter 3) on Angami nasals showed that the voiceless and voiced nasal formants differ significantly. Hence, we attempt to capture the lateral and approximant voicing distinction using formant values in this study. HNR measures the ratio between periodic and non-periodic components of a speech sound (Fernandes et al., 2018). Hence, we also measured the HNR values for voiced and voiceless laterals & approximants, as we expected the voiced to have a higher HNR value than the voiceless. Similarly, fully sonorous segments, such as the voiced laterals & approximants, should have higher intensity values than the voiceless laterals & approximants. Considering that, we measure the intensity of the laterals and approximants at every 10% of the total duration. The literature on voiceless laterals & approximants has argued that they are not entirely voiceless but partially voiced (Asu et al., 2015; Fanai, 1992). In order to confirm that, we determined the percentage of voicing in each lateral & approxim-

ant consonant by calculating the percentage of voiced frames among the total frames in a sound segment. All measures were calculated using a Praat script written for the purpose.

Descriptive and exploratory statistics were conducted using statistical packages on R (R Core Team, 2022). Descriptive statistics and exploratory statistical modeling were employed as described in Section 1.3.6. All plots were generated using the *ggplot2* package on R (Wickham, 2016). To investigate the statistical significance, separate linear mixed-effect models were created for the acoustic-phonetic features across the laterals and approximants. The LME models consist of the mean values of the acoustic features such as duration, CoG, intensity, and voicing as the dependent variable. Laryngeal contrasts (LC), PoA, three contexts, and their interactions were the fixed effects, while the speaker and types of words were random effects. The model is shown in 5.1.

$$Features \sim LC + PoA + Context + LC * PoA * Context + (1 | Speaker) + (1 | word) \quad (5.1)$$

5.4 Results

The study's findings are divided into two broad sections: the first section examines the acoustic properties of laterals in Angami, while the second Section discusses the acoustic properties of approximants in Angami.

5.4.1 Acoustics of laterals in Angami

To visualize the differences between the voiced and voiceless laterals, the waveforms for the word /li/ and /l^hi/ are plotted in Figure 5-2. The figure illustrates that the voiced lateral /l/ has periodic waveforms, whereas the voiceless lateral, /l^h/ has aperiodic ones. Moreover, the voiceless laterals have a period of aspiration following

the voiceless region and before the initiation of the vowel. Hence, we consider the voiceless laterals as aspirated in Angami. Further, the spectrograms of the voiceless lateral produced in a sentence, carrier phrase, and isolation are shown in Figure 5-3. As seen in the figure, voiceless laterals embedded in sentences and carrier phrases have traces of voicing in the initial parts. However, the voicing diminishes as the consonant progresses. Compared to that, the voiceless laterals in isolation are entirely voiceless. All voiceless laterals, nevertheless, terminate with aspiration. Even though there is no voicing in the voiceless portions, formants are present as they characterize the vocal tract (filter) characteristics, where the source is the turbulence of air rather than the vibration of the vocal cords. The following sections provide insights into the acoustic characteristics of voiced and voiceless laterals using various acoustic-phonetic features.

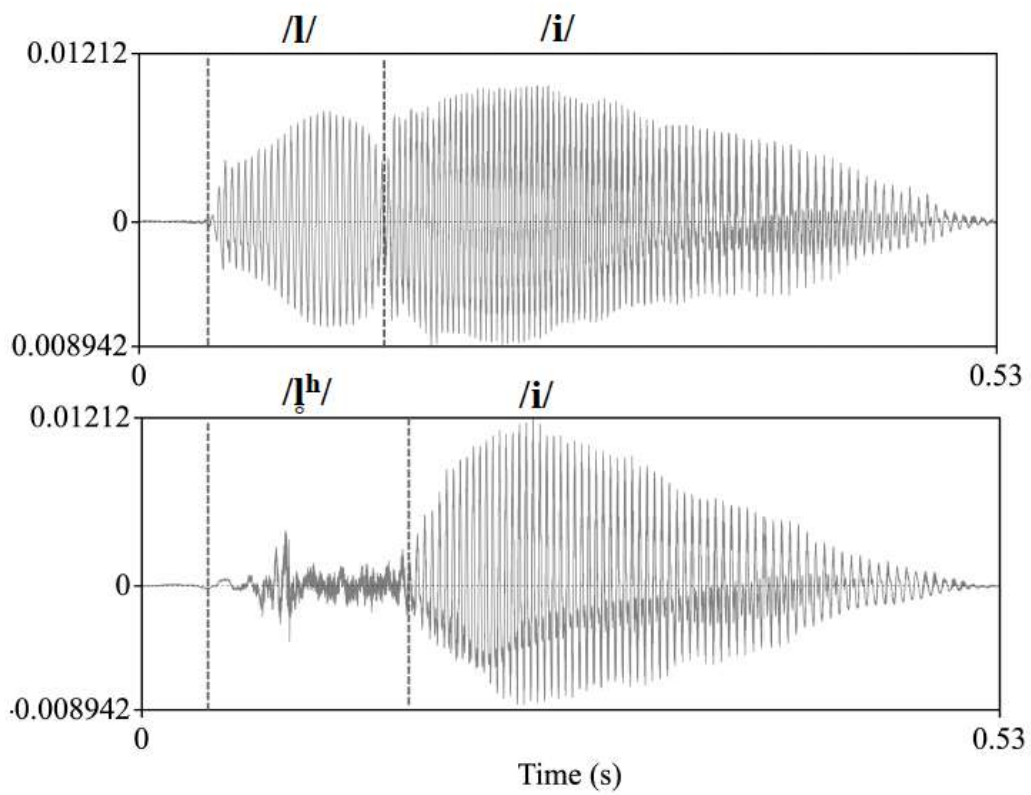


Figure 5-2: An example of the lateral and vowel boundary for the words /li/ ‘pot’ and /li^h/ ‘business’ as produced by a female Angami speaker in isolation.

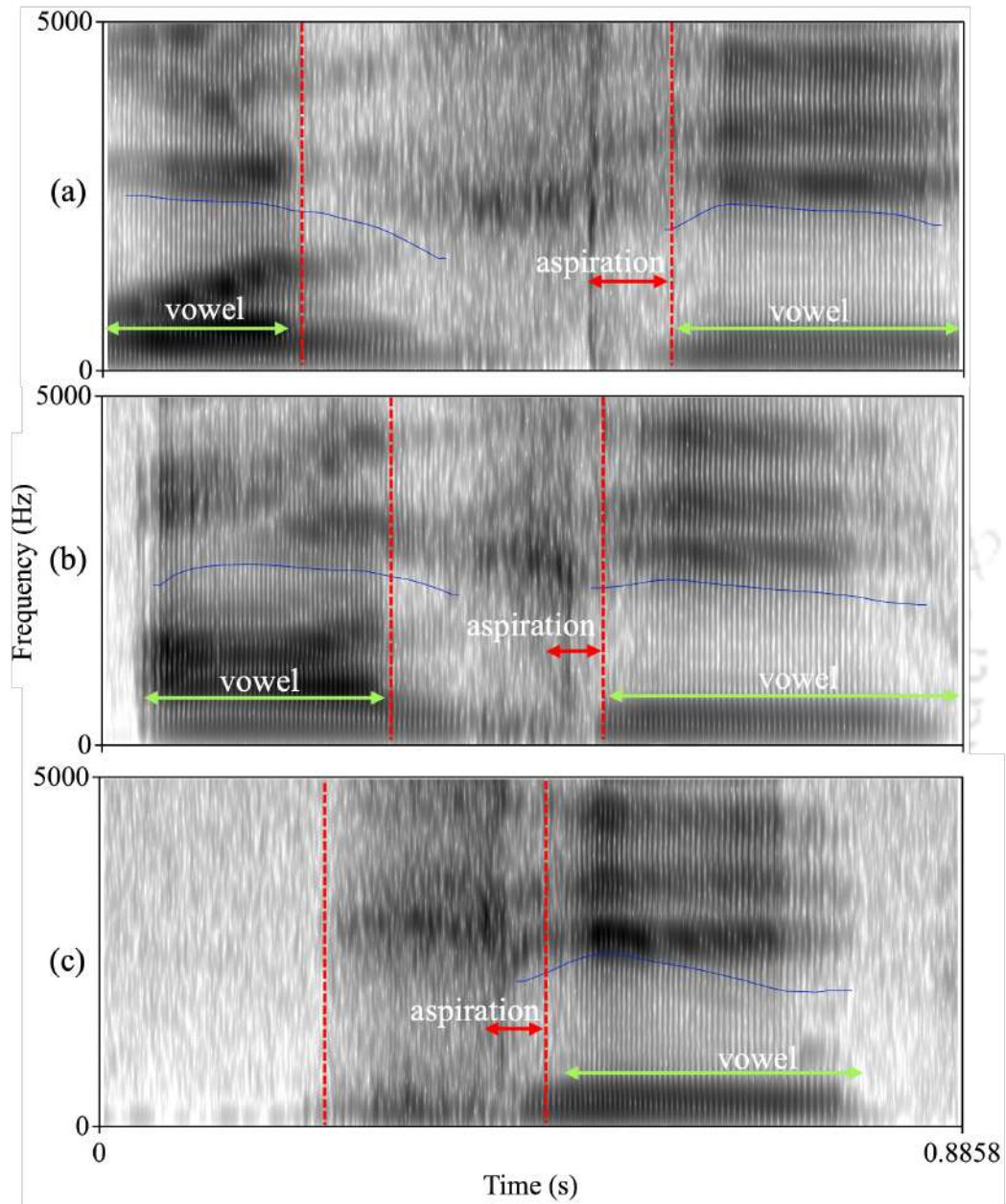


Figure 5-3: Voiceless lateral / $l̥i$ / 'business' as produced by a female speaker in three contexts.

5.4.1.1 Duration and voicing

The duration of the laterals across the three contexts is shown in Figure 5-4 and Table 5.3, where it is observed that voiceless laterals are marginally longer than their voiced counterparts. This seems to be true for all three contexts. However, these differences are not statistically significant in any of the contexts, as seen in Table 5.4. Hence, it can be concluded that lateral voicing does not have any significant implication on the duration of the laterals.

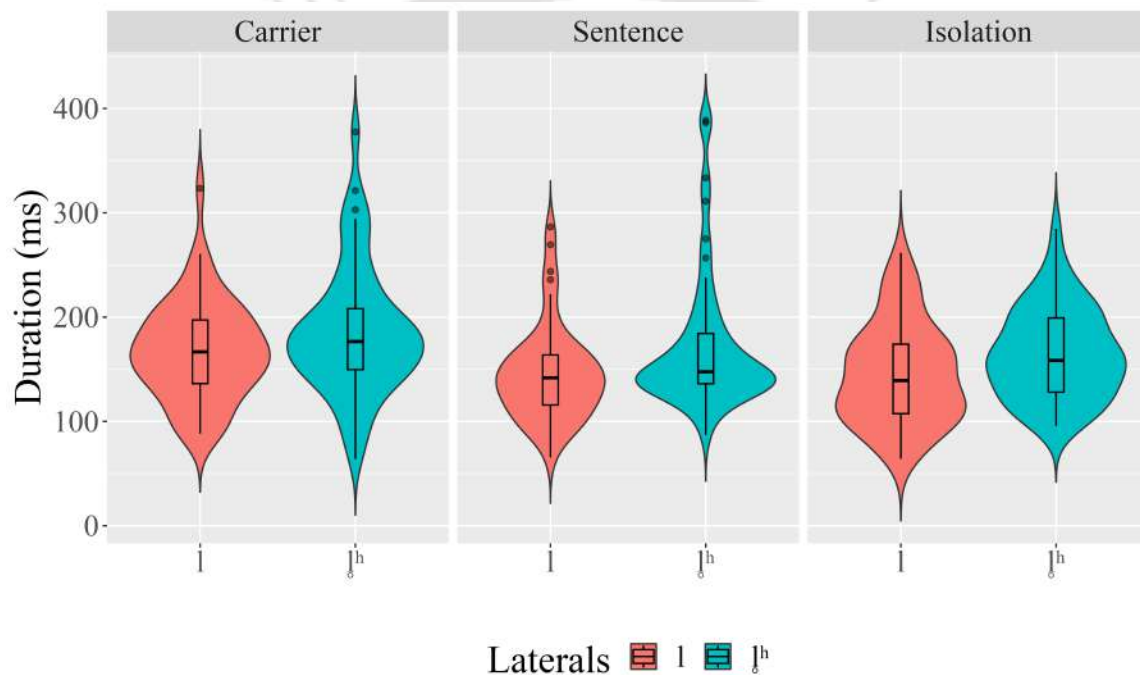


Figure 5-4: Violin plots showing the duration (ms) in voiced and voiceless laterals across the three contexts.

As seen in Table 5.3, the average voicing contrast values in the three contexts clearly distinguish between voiced and voiceless laterals. The voiced lateral /l/ exhibits a higher percentage of voicing than its voiceless counterpart. Notably, the percentage of voicing in the voiceless lateral varies across the three contexts. In sentence and carrier phrase contexts, /l^h/ shows voicing levels at 56% and 38% of the total voicing, respectively. However, when produced in isolation, the percentage of

voicing is minimal at 4%. The distinction in voicing is supported by statistical tests (see Table 5.4), confirming its significance across all three contexts.

5.4.1.2 Spectral measures

The formants of the lateral consonants are plotted for visualization in Figure 5-5. The figure distinguishes between voiced and voiceless laterals for the first and the second formants (F1 & F2) in all three contexts. However, the distinction decreases in the case of formants F3 and F4, with these formants overlapping in both carrier and isolation. Voiced laterals exhibit lower formant values, while voiceless laterals demonstrate higher ones. The LME models built for the first four formants were subjected to the post-hoc Bonferroni test for pair-wise comparison across the three contexts. Table 5.4 shows that the voiced and voiceless laterals maintain statistically significant differences in formant F1 and F2 across all contexts.

The HNR values across the duration of the laterals are depicted across the three contexts in Figure 5-6. The figure reveals that voiced lateral /l/ exhibits higher HNR values than its voiceless counterpart /l^h/. The lower HNR values observed in /l^h/ are attributed to its association with aperiodicity. Voiced and voiceless laterals maintain statistically significant differences in HNR across all contexts, as seen in Table 5.4.

The intensity values across the duration of the voiced and voiceless laterals across the three contexts exhibit subtle differences, as seen in Figure 5-7, and similar results in the average intensity values (Table 5.3). The lateral /l^h/ displays a lower intensity contour than its voiced counterpart. Intensity decreases because the voiceless lateral has lesser energy and is associated with aspiration. Statistical tests indicate that the voicing distinction in laterals is not statistically significant in intensity across the three contexts. The detailed results of the Type II Wald chi-square tests for analysis of deviance conducted on the LME models for all the acoustic features, including the results of the post-hoc Bonferroni tests, are provided in Table 5.5.

Table 5.3: Average (in bold) and standard errors of the acoustic features in Angami voiced and voiceless laterals.

Features	Carrier		Sentence		Isolation	
	/l/	/l ^h /	/l/	/l ^h /	/l/	/l ^h /
Duration	167	185	145	172	145	164
	6	8	6	9	6	6
Voicing	98	38	100	56	98	4
	1	5	0	5	1	1
F1	373	772	374	781	377	847
	12	30	13	33	16	32
F2	1645	1999	1645	1988	1714	1992
	35	41	39	41	36	39
F3	2798	2924	2704	2904	2723	2929
	40	27	47	33	40	26
F4	3719	3875	3631	3881	3701	3854
	61	31	58	42	54	38
HNR	19	8	19	10	16	2
	1	1	1	1	.5	.4
Intensity	40	36	43	37	39	32
	1	1	1	1	1	1
CoG	337	532	382	544	346	910
	9	59	17	69	11	92

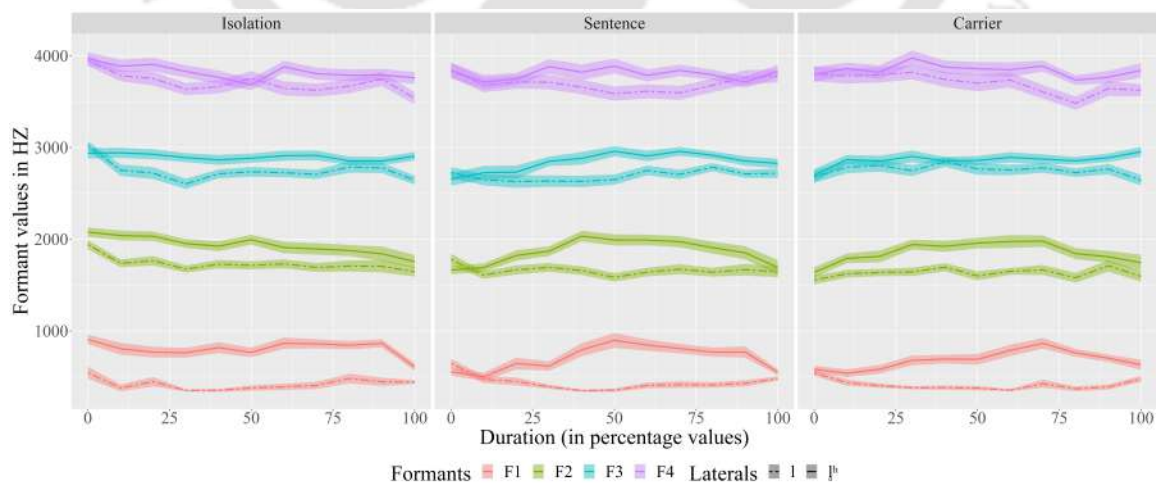


Figure 5-5: Formants (F1-F4) in voiced and voiceless laterals across the three contexts.

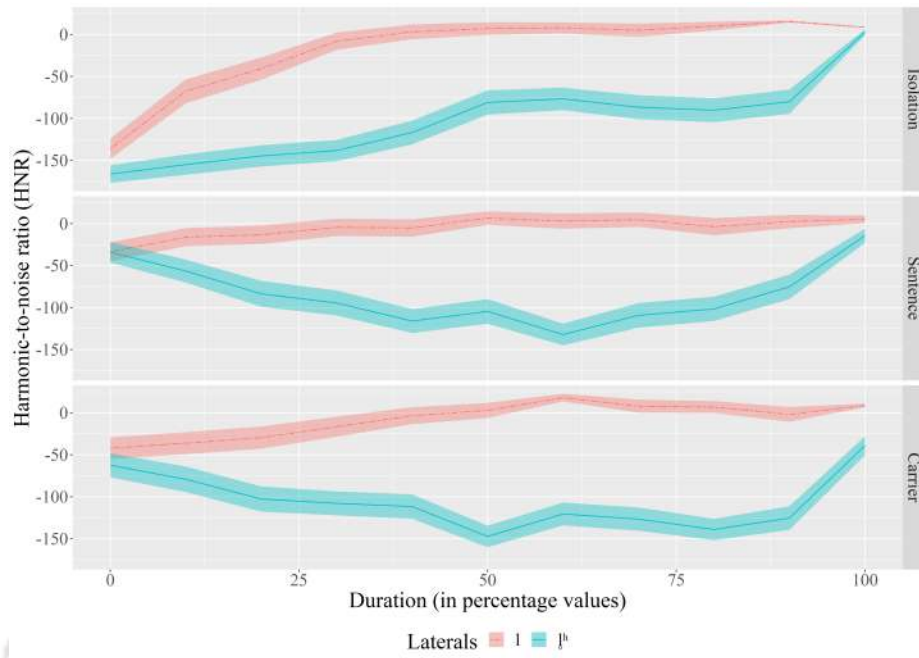


Figure 5-6: HNR in voiced and voiceless laterals across the three contexts.

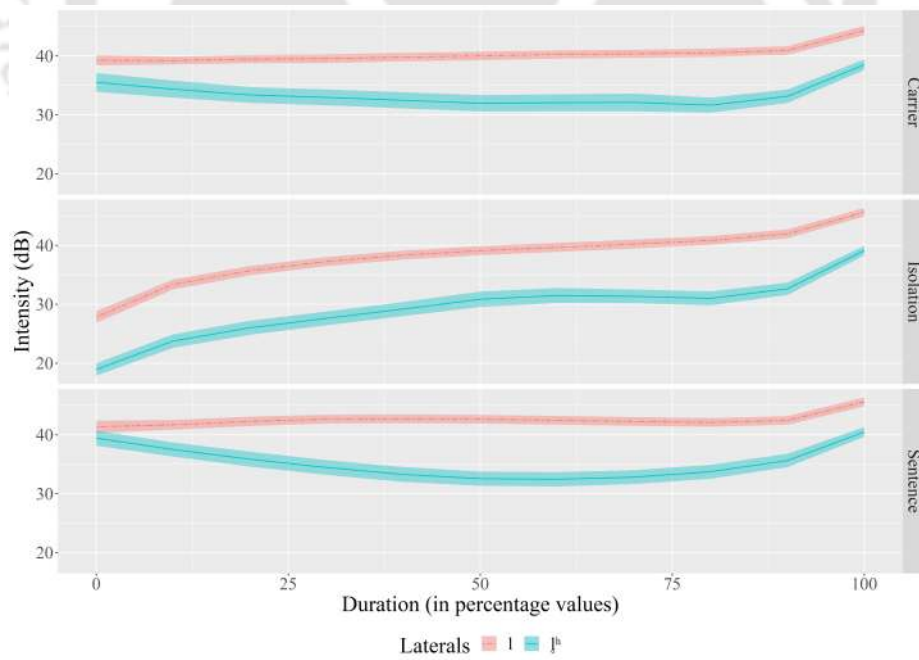


Figure 5-7: Intensity in voiced and voiceless laterals across the three contexts.

Table 5.4: Results of Bonferroni post-hoc pairwise comparison conducted on LME models with acoustic features as the dependent variable and voicing contrasts, contexts and their interactions as fixed effects.

Features	Carrier	Sentence	Isolation
Duration	n.s	n.s	n.s
Voicing	***	***	***
F1	***	***	***
F2	**	**	*
F3	n.s	*	*
F4	n.s	n.s	n.s
HNR	***	***	***
Intensity	n.s	n.s	n.s

n.s.: not significant, *: $p < 0.01$, **: $p < 0.001$ ***: $p < 0.0001$

Table 5.5: Analysis of Deviance (ANOVA) tests on the LME models with acoustic features as the dependent variable and voicing contrasts, contexts and their interactions as fixed effects, pair-wise contrasts estimated from post-hoc Bonferroni tests. The abbreviation ISO, ST and CR represent isolation, sentence and carrier phrase, respectively.

Feature	χ^2	Contrasts	Estimate	SE	df	t -ratio	p -value
F1	Laterals: 491.3***	/1/ - /1 ^h / ISO	-470.6	35.1	13.1	-13.4	< .0001
	Context: 3.3	/1/ - /1 ^h / ST	-406.6	35.1	13.1	-11.6	< .0001
	Laterals x Context: 2.8	/1/ - /1 ^h / CR	-399.3	35.1	13.1	-11.4	< .0001
F2	Laterals: 67.6***	/1/ - /1 ^h / ISO	-296.7	66.2	18.0	-4.5	0.0043
	Context: 2.4	/1/ - /1 ^h / ST	-357.8	66.3	18.1	-5.4	0.0006
	Laterals x Context: 2.7	/1/ - /1 ^h / CR	-372.5	66.2	18.0	-5.6	0.0004
F3	Laterals: 62.9***	/1/ - /1 ^h / ISO	-205.9	42.5	12.7	-4.848	0.0051
	Context: 4.5	/1/ - /1 ^h / ST	-199.4	42.7	12.4	-4.7	0.0074
	Laterals x Context: 2.7	/1/ - /1 ^h / CR	-125.5	42.5	12.7	-2.9	0.1719
F4	Laterals: 31.7***	/1/ - /1 ^h / ISO	-152.8	60.9	8.4	-2.5	0.5262
	Context: 1.0	/1/ - /1 ^h / ST	-199.4	42.7	12.4	-4.7	0.0074
	Laterals x Context: 1.8	/1/ - /1 ^h / CR	-125.5	42.5	12.7	-2.9	0.1719
HNR	Laterals: 770.7***	/1/ - /1 ^h / ISO	14.0	0.8	7.0	18.1	< .0001
	Context: 107.6***	/1/ - /1 ^h / ST	9.7	0.8	6.8	12.4	0.0001
	Laterals x Context: 19.1***	/1/ - /1 ^h / CR	11.1	0.7	6.9	14.4	< .0001
Voicing	Laterals: 736.5***	/1/ - /1 ^h / ISO	93.7	4.4	10.4	21.2	< .0001
	Context: 83.1***	/1/ - /1 ^h / ST	43.2	4.4	10.3	9.8	< .0001
	Laterals x Context: 76.5***	/1/ - /1 ^h / CR	59.2	4.4	10.4	13.4	< .0001
Duration	Laterals: 18.8***	/1/ - /1 ^h / ISO	-19.0	10.0	46.2	-1.9	0.9809
	Context: 15.2***	/1/ - /1 ^h / ST	-26.7	10.1	45.9	-2.6	0.1703
	Laterals x Context: 0.6	/1/ - /1 ^h / CR	-17.7	10.1	46.2	-1.8	1.0000
Intensity	Laterals: 30.5***	/1/ - /1 ^h / ISO	7.6	1.6	14.3	4.7	0.0042
	Context: 51.1***	/1/ - /1 ^h / ST	5.9	1.6	14.3	3.7	0.0312
	Laterals x Context: 7.5*	/1/ - /1 ^h / CR	4.5	1.6	14.3	2.8	0.1894
CoG	Laterals: 22.3***	/1/ - /1 ^h / ISO	-539.5	105.8	29.9	-5.1	0.001
	Context: 17.7***	/1/ - /1 ^h / ST	-142.5	106.1	29.9	-1.3	1.0000
	Laterals x Context: 19.7***	/1/ - /1 ^h / CR	-170.7	105.8	29.9	-1.614	1.0000

n.s.: not significant, *: $p < 0.01$, **: $p < 0.001$, ***: $p < 0.0001$

5.4.2 Acoustics of approximants in Angami

There are six approximants in Angami which are produced in three PoA, namely, bilabial, alveolar and palatal, and two laryngeal contrasts: voiceless and voiceless aspirates - /w, w^h, ɹ, ɹ^h, j, j^h/. To visualize the differences between voiceless laterals approximants across the three PoA, spectrograms of the word /w^he/ ‘cover’, /ɹ^h/ ‘redeem’ and /j^ha/ ‘lift’ has been plotted in Figure 5-8. As seen in the figure, the approximant segments are voiceless portions accompanied by a brief aspiration before the onset of the vowel. Voiceless bilabial /w^h/ appear to have a longer duration than /ɹ^h/ and /j^h/.

5.4.2.1 Duration and voicing in approximants

The duration of the approximants across the three contexts is shown in Figure 5-9, and the average value is detailed in Table 5.6. The result shows that voiceless approximants are higher than their voiced counterparts. The duration of the palatal is higher than the alveolar and bilabial. The duration of the approximants was further analyzed by the three contexts. Figure 5-10 and Table 5.7 show that the duration produced in a carrier is marginally higher than the duration produced in sentence and isolation. Hence, the result shows that the approximant voicing contrast and the context affect the duration of the laterals.

Figure 5-11 and Table 5.6 provide the average values of the amount of voicing contrast in approximants to clearly distinguish between voicing in approximants. The voiced approximants, as they are sonorous, have a < 95% of voicing, whereas the percentage of voicing in the voiceless is > 15%, which is likely voiced at the beginning or termination boundary, and it may be due to coarticulatory effects. This is confirmed when we look at the voicing values produced in three contexts: the voiceless approximants produced in isolation have the least % of voicing as shown in Figure 5-12 and the values in Table 5.7.

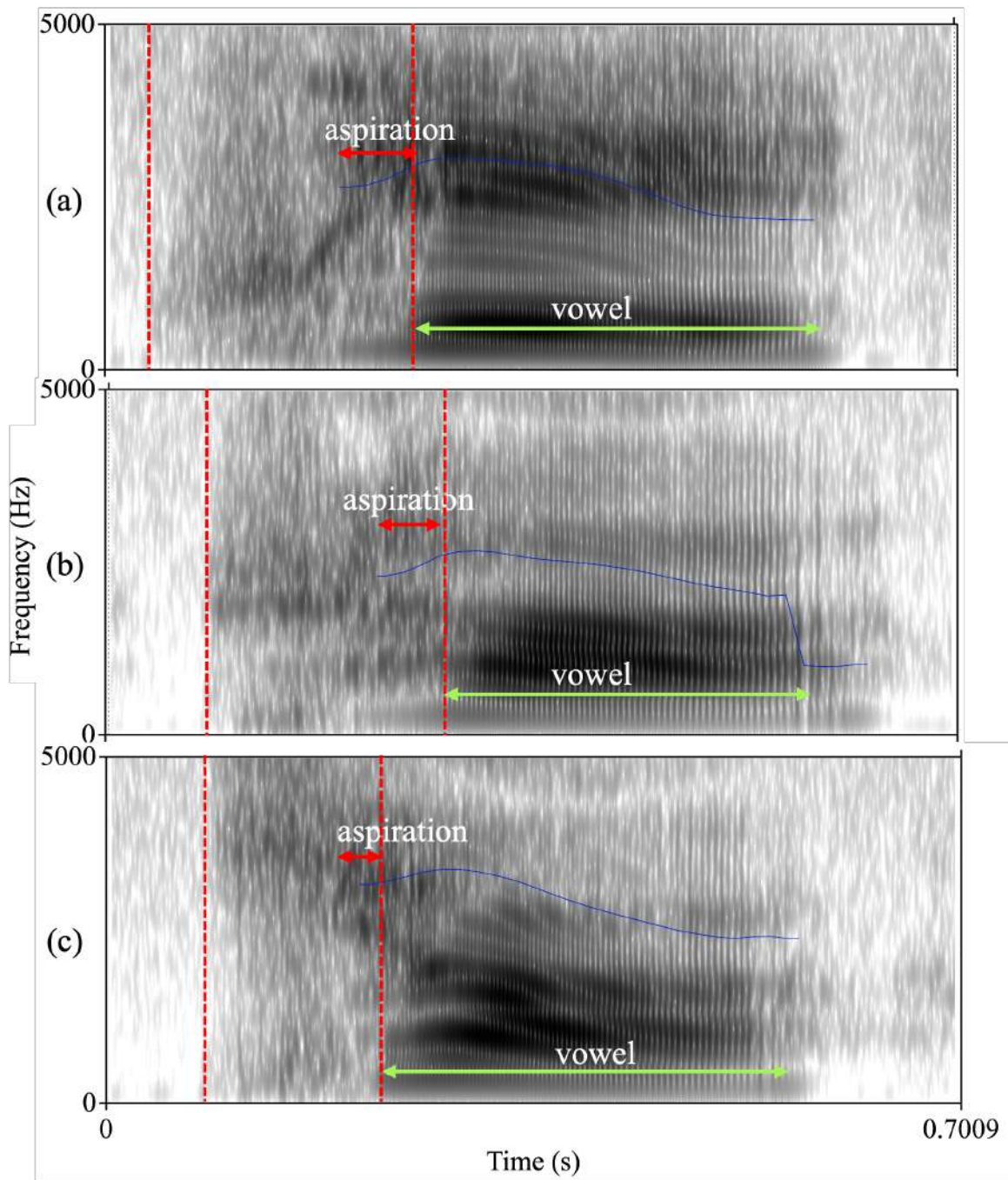


Figure 5-8: Voiceless aspirated approximants $/w^h e/$ 'cover', $/ɪ^h a/$ 'redeem', and $/j^h a/$ 'lift' as produced by a female speaker in isolation.

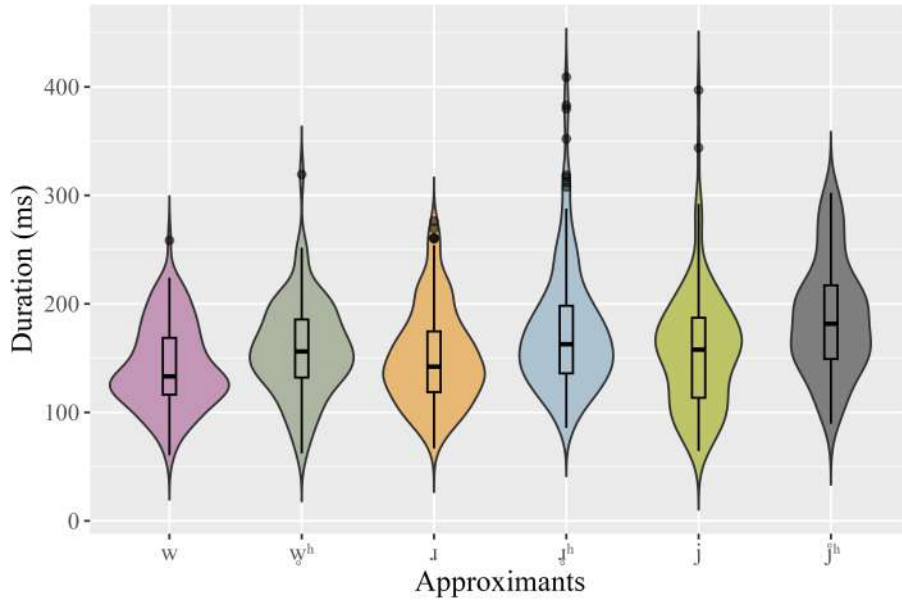


Figure 5-9: Violin plots showing the duration (ms) in voiced and voiceless approximants

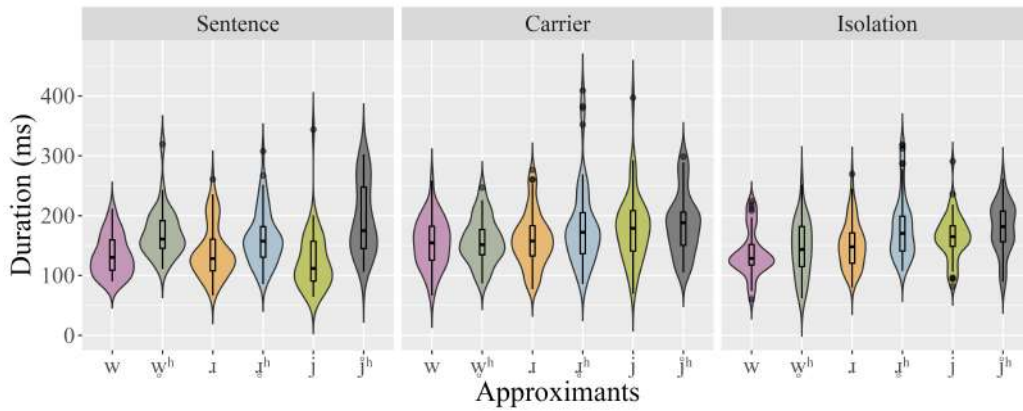


Figure 5-10: Violin plots showing the duration (ms) in voiced and voiceless approximants across the three contexts

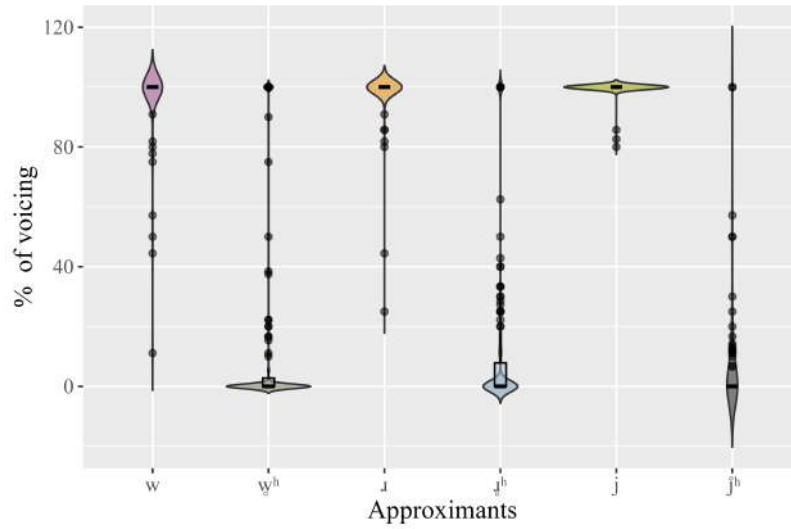


Figure 5-11: Violin plots showing the percentage of voicing in the voiced and voiceless approximants.

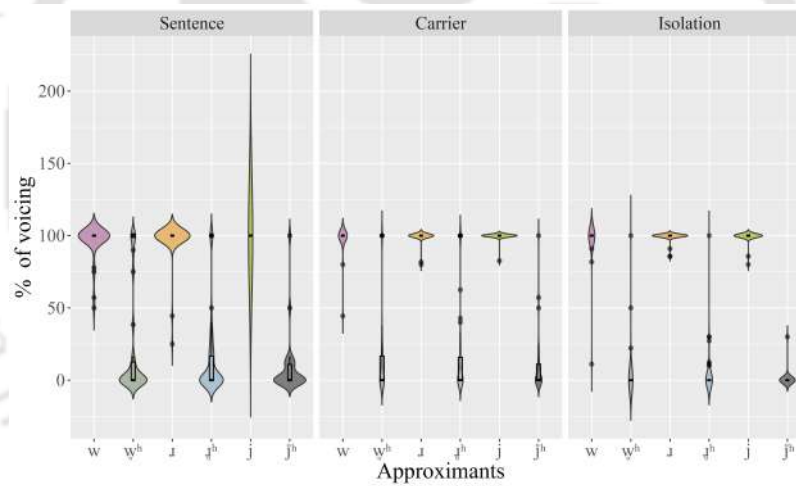


Figure 5-12: Violin plots showing the percentage of voicing in the voiced and voiceless approximants across the three contexts.

5.4.2.2 CoG measures in approximants

The average values of the center of gravity (CoG) in the six approximants are shown in Figure 5-13 and Table 5-14. It is evident that the voiceless have higher CoG values than their voiced counterpart. The CoG values for voiced bilabial, alveolar, and palatal approximants are at 716 Hz, 1032 Hz, and 1590 Hz, respectively. Similarly, the CoG values for voiceless bilabial, alveolar, and palatal sounds are 716 Hz, 1032 Hz, and 1590 Hz. Interestingly, the voiceless aspirated palatal has the highest CoG, while the voiced bilabial has the lowest. Further analysis based on context shows that for all the approximants, those produced in isolation have higher CoG values than those produced in carrier or sentence frames. This is seen in Figure 5-14 and the values in Table 5.7.

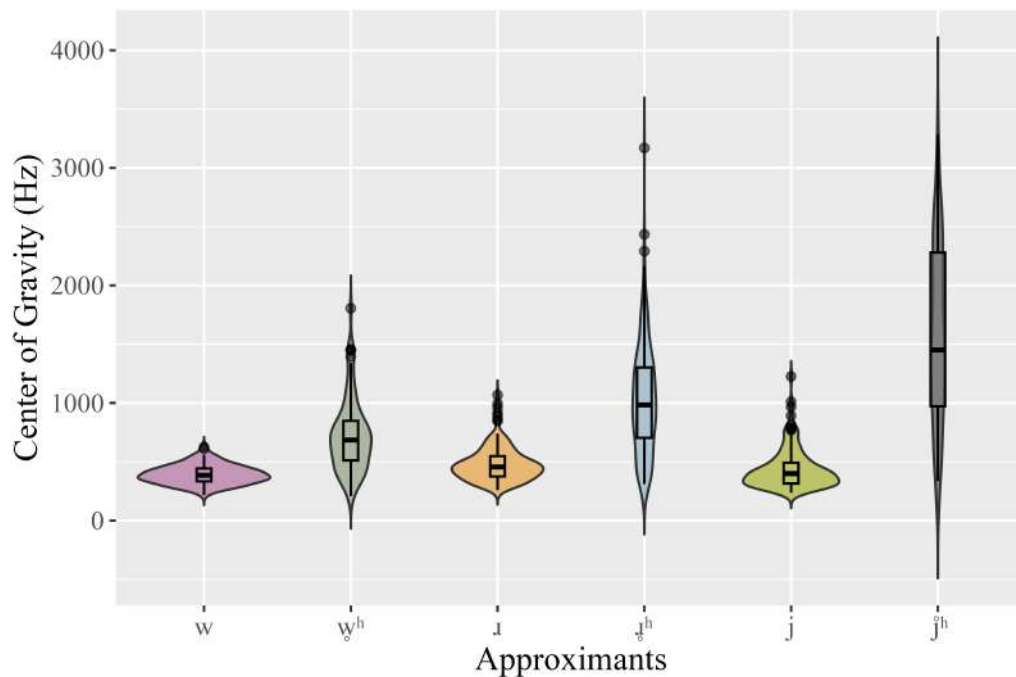


Figure 5-13: Violin plots showing the center of gravity in voiced and voiceless approximants

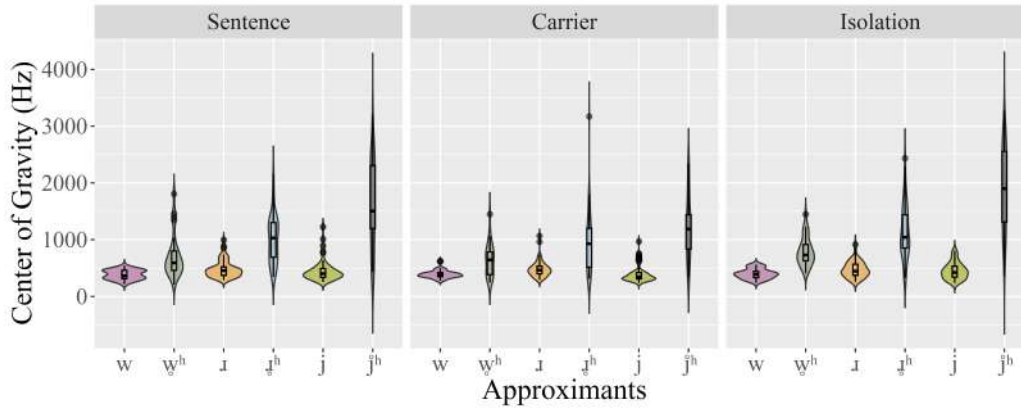


Figure 5-14: Violin plots showing the center of gravity in voiced and voiceless approximants across the three contexts

5.4.2.3 Spectral measures in approximants

The formants of the voiced and voiceless bilabial approximant across the three contexts are plotted for visualization in Figure 5-15. The figure distinguishes voiced and voiceless for the first and the second formants (F1 & F2) in all three contexts. However, the distinction decreases in the case of formants F3 and F4, with these formants overlapping in all the contexts. Voiced bilabial approximant exhibits lower formant values, while voiceless counterparts demonstrate higher ones. The average values of the formants (F1-F4) are provided in Table 5.7. The LME models built for the first four formants were subjected to the post-hoc Bonferroni test for pair-wise comparison across the three contexts. Table 5.4 shows that the voiced and voiceless bilabial approximant maintain statistically significant differences in formants F1 and F2 across all contexts.

Similarly, we look into the four formants of voiced and voiceless alveolar approximants across the three contexts. The results are plotted in Figure 5-15. The figure shows a clear distinction between voiced and voiceless in all four formants (F1 -F4) in all three contexts. There is a close overlap of the F4 in isolation frames. Voiced alveolar approximants exhibit lower formant values, while voiceless approximants demonstrate higher ones. The average values of the formants (F1-F4) are provided in Table

5.7. Pair-wise comparison across the three contexts shows that voiced and voiceless are significantly different in all the four formants, which have p-value $< .0001$, except for F4 produced in isolation, which is not significant. The results are displayed in Table 5.4.

Finally, for the spectral measurements, we look into the formants of the palatal approximants. Figure 5-15 is plotted for visualization, while the values are shown in Table 5.7. The result shows that only formants F1 have a distinction in voiced and voiceless palatal approximants. This is further shown in the statistical test as well (Table 5.4). The spectral measurements clearly distinguish the voicing and the place of articulation in approximants.

The HNR in the voiced and voiceless approximants across the three contexts show significant differences in voicing. The voiceless approximants have lower HNR values, while the voiced approximants have higher HNR values. This is seen in Figure 5-16 and the values in Table 5.7. Statistical tests show that the voicing distinction is statistically significant in harmonic-to-noise ratio values across the contexts. The findings in the intensity of the voiced and voiceless approximants correspond to the result of HNR, as shown in Figure 5-17.

Similar to the voiceless laterals discussed in the previous section (Section 5.4.1), the voiceless approximants exhibit lower energy due to the aspiration associated with the sound segment. Additionally, it is observed that alveolar approximants have a higher intensity than palatal and bilabial approximants. The intensity pattern for the approximants can be summarized as alveolar $>$ palatal $>$ bilabial. The intensity shows statistically significant effects on voicing, as illustrated in Table 5.4.

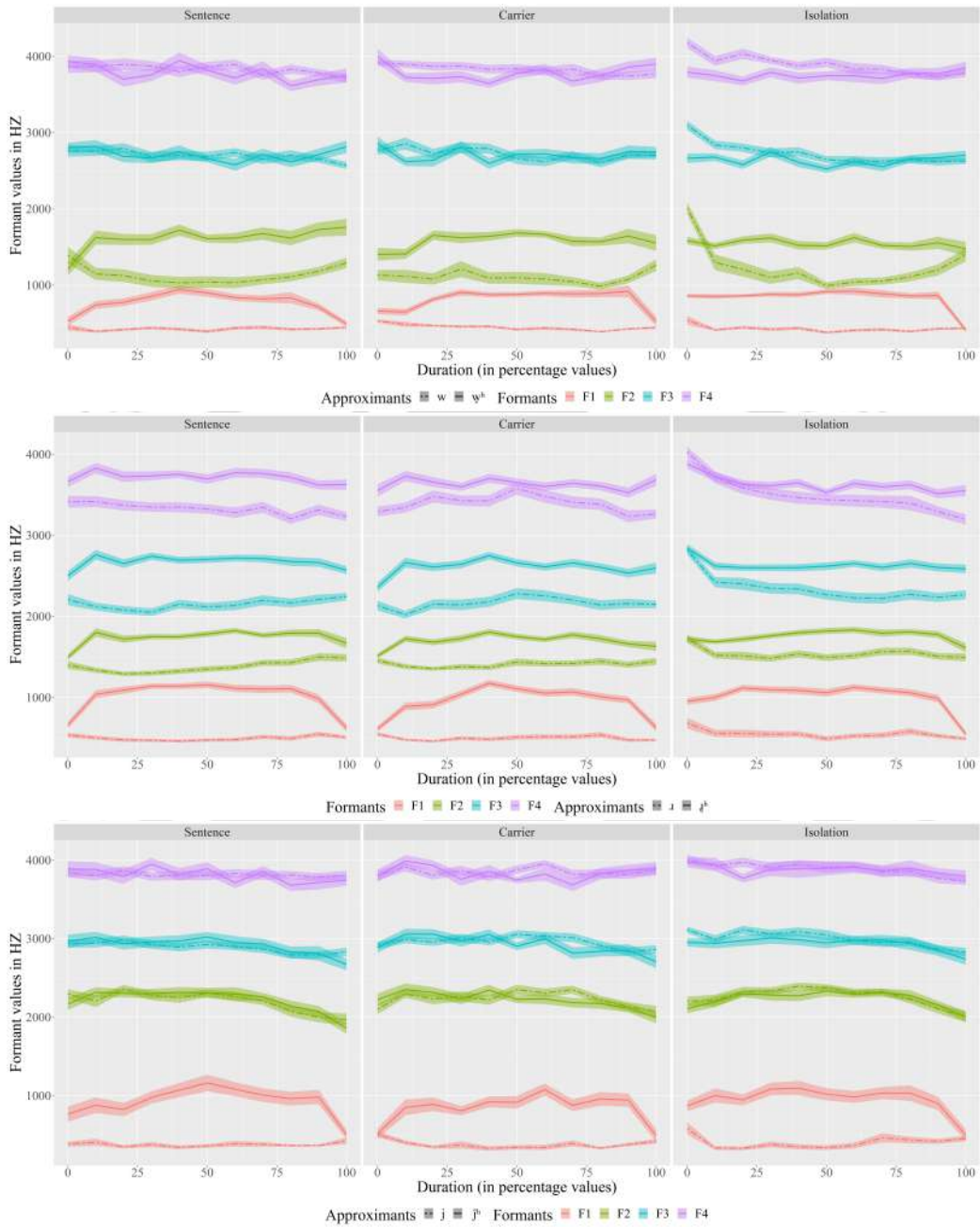


Figure 5-15: Formants (F1-F4) in voiced and voiceless bilabial (top), alveolar (mid) and palatal (bottom) approximants across the three contexts in Angami.

Furthermore, the approximants' Center of Gravity (CoG) average values were analyzed. The results indicate that CoG values are higher in voiceless approximants than their voiced counterparts. Among them, bilabial approximants exhibit the lowest CoG values, while palatal approximants show the highest. Statistical tests reveal that the distinction in voicing for bilabial approximants produced in sentence and carrier frames is not significant, whereas it is significant in other contexts for bilabial, alveolar, and palatal approximants.

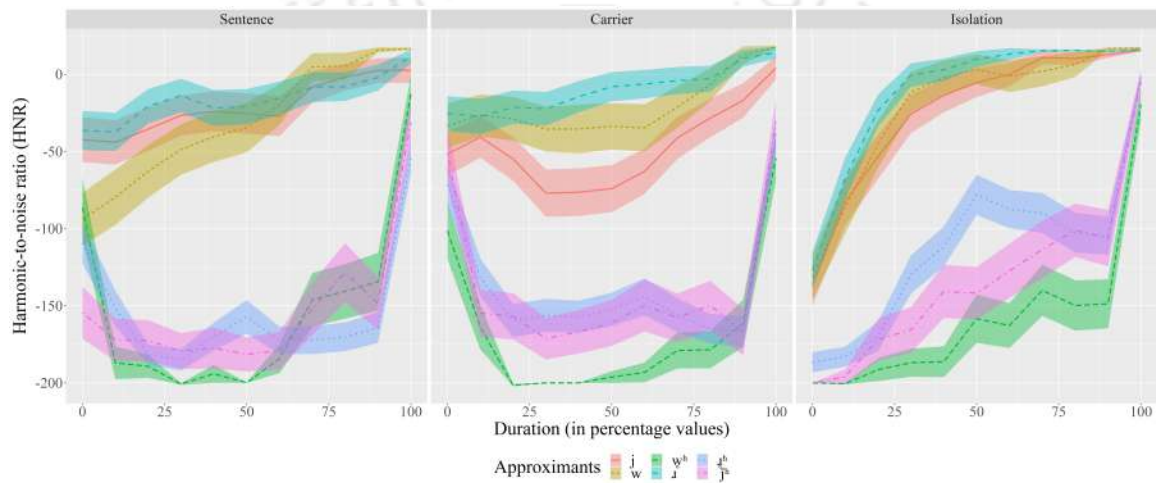


Figure 5-16: Harmonic to noise ratio in voiced and voiceless approximants across the three contexts.

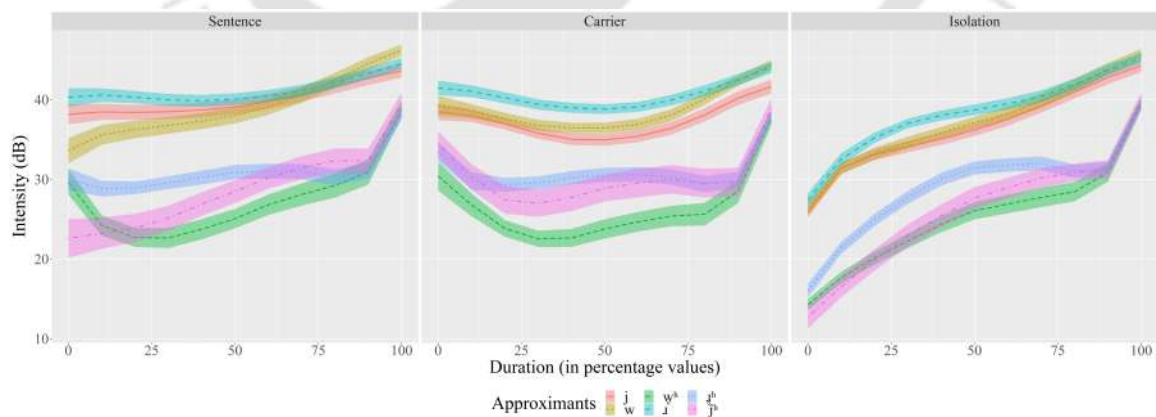


Figure 5-17: Intensity in voiced and voiceless approximants across the three contexts.

Table 5.6: Average (in bold) and standard errors of the acoustic features in Angami voiced and voiceless approximants

Features	/w/	/w ^h /	/ɹ/	/ɹ ^h /	/j/	/j ^h /
Duration	142	157	150	175	156	186
	3	4	3	4	4	5
Voicing	97	13	98	11	99	6
	1	3	0	1	0	1
CoG	391	716	480	1032	437	1590
	7	33	11	35	13	80
F1	423	890	507	1101	364	1002
	6	17	11	16	9	23
F2	1057	1589	1419	1774	2310	2293
	24	18	17	11	20	20
F3	2711	2646	2205	2666	2995	2979
	23	20	25	12	25	28
F4	3879	3738	3459	3661	3882	3846
	22	23	26	16	24	26
HNR	18	5	17	4	16	5
	0	0	0	0	0	0
Intensity	40	29	41	32	39	31
	0	0	0	0	0	0

5.5 Conclusion

This study examined the voicing distinction in Angami laterals and approximants produced in three different contexts. We first look into the production of voiceless laterals and approximants produced in isolation and observe that the sound segment is completely voiceless in the entire production with an aspiration towards the termination of the sound. Similar findings were also reported in Angami and Xumi voiceless nasals, which are characterized by entirely voiceless nasal portions followed by aspiration (Bhaskararao and Ladefoged, 1991; Chirkova et al., 2019; Terhiiija and Sarmah, 2020). Hence, phonemically, in Angami, the voiceless laterals and approximants are devoid of any voicing and are characterized by aspiration towards the termination of the consonants. However, when produced embedded in carrier frames, the voiceless

laterals and approximants may be initially voiced if preceded by vowels. As reported in languages such as Mien (Yao) and Sedang, voiceless laterals are claimed to be composed of either /l + h/ or /h+l/ (Purnell, 1965; Smith, 1968). However, evidence from Angami in the current work suggests that the voiceless laterals and approximants can be phonemically entirely voiceless, and any voicing in the voiceless segments is phonetically induced.

The acoustic evidence obtained in the current study supports the distinctness of voiceless laterals. The results show that the amount of voicing ($\chi^2(1) = 736$, $p < 0.001$), F1 ($\chi^2(1) = 491$, $p < 0.001$), F2 ($\chi^2(1) = 67$, $p < 0.001$), and HNR ($\chi^2(1) = 770$, $p < 0.001$) significantly differs between the voiced and the voiceless laterals in all phonetic contexts. Duration, intensity, and the fourth formant are statistically insignificant across the contexts for the voicing distinction in laterals.

In approximants, spectral measurements (F1-F4) effectively differentiate between place of articulation and voicing. Voiceless approximants consistently exhibit higher formant values than their voiced counterparts. Specifically, in bilabial approximants, the statistical analysis shows significant voicing effects on F1 and F2, whereas in alveolar approximants, F1-F4 showed statistically significant differences. In palatal approximants, the findings indicate that only F1 exhibits a statistically significant effect in voicing. The result further shows that HNR, voicing, intensity, and CoG have statistically significant effects on the voiced and voiceless approximants. Furthermore, context also played an important role in terms of voicing contrasts.

The results of the current study can be compared with the results of the studies on Angami nasals, where it was shown that the voiceless nasals in Angami are phonemically voiceless, followed by a brief aspiration (Chapter 3). Similar to the current results, it was shown that the amount of voicing differs depending on the phonetic context where the voiceless nasal is produced. To summarize, we can conclude that the voiceless sonorants in Angami have a similar production mechanism. In general,

this may be true for other Tibeto-Burman languages that have voiceless sonorants with aspiration. Overall, this study contributes to understanding the phonetic properties of voiceless laterals and approximants in Angami. This study's findings may apply to other languages of the Tibeto-Burman family with voiceless approximants in its inventory.



Table 5.7: Average (in bold) and standard errors of the acoustic features in Angami voiced and voiceless approximants across the three contexts.

Contexts	Features	/w/	/w ^h /	/ɹ/	/ɹ ^h /	/j/	/j ^h /
Carrier	Duration	155	157	159	182	179	186
		6	7	6	9	8	9
	Voicing	98	14	99	17	99	10
		1	5	0	4	0	4
	CoG	393	641	483	945	404	1218
		11	52	19	66	22	103
	F1	440	904	495	1102	347	992
		8	19	16	25	13	38
	F2	1081	1648	1384	1764	2294	2295
		37	22	24	20	30	34
	F3	2743	2679	2201	2674	3013	2998
		41	35	43	21	42	46
	F4	3879	3738	3459	3661	3882	3846
		22	23	26	16	24	26
HNR	18	6	17	5	16	7	
	0	1	0	0	0	0	
Intensity	40	30	41	32	38	32	
	0	1	0	0	0	0	
Sentence	Duration	133	171	139	162	124	195
		5	8	5	6	7	11
	Voicing	96	19	97	13	100	8
		1	6	1	3	0	4
	CoG	383	710	482	1002	458	1646
		13	71	20	53	29	148
	F1	417	863	485	1111	390	1029
		10	42	14	27	22	43
	F2	1060	1584	1358	1752	2307	2298
		52	42	26	18	39	37
	F3	2704	2662	2150	2674	2943	2969
		43	36	43	21	43	50
	F4	3841	3773	3393	3666	3869	3828
		43	42	43	28	42	49
HNR	17	6	18	5	16	5	
	0	0	0	0	0	0	
Intensity	41	30	41	32	40	31	
	0	1	0	0	0	1	

Table 5.7: Contd. . . Average (in bold) and standard errors (in reduced font) of the acoustic features in Angami voiced and voiceless approximants across the three contexts.

Contexts	Features	/w/	/w ^h /	/ɹ/	/ɹ ^h /	/j/	/j ^h /
	Duration	137	145	151	179	164	179
		6	8	5	6	5	7
	Voicing	97	5	99	3	99	1
		2	3	0	1	0	1
	CoG	397	794	475	1146	451	1911
		14	46	18	61	20	138
	F1	410	901	539	1089	357	986
		12	24	25	33	14	41
Isolation	F2	1031	1536	1512	1804	2330	2287
		39	25	34	22	30	36
	F3	2684	2599	2264	2659	3025	2969
		35	32	46	21	45	52
	F4	3894	3693	3501	3646	3896	3837
		34	41	47	28	43	48
	HNR	17	2	15	2	14	3
		0	1	0	0	0	0
	Intensity	39	29	40	31	39	30
		0	0	0	0	0	1

Table 5.8: Analysis of Deviance (ANOVA) tests on the LME models with acoustic features of the approximants as the dependent variable and voicing contrasts, contexts and their interactions as fixed effects, pair-wise contrasts estimated from post-hoc Bonferroni tests. The abbreviation ISO, ST and CR represent isolation, sentence and carrier phrase, respectively.

Feature	χ^2	Contrasts	Estimate	SE	df	<i>t</i> -ratio	<i>p</i> -value
F1	Approx.: 2870 *** Context: 0 Approx. x Context: 10	/w/ - /w ^h / ISO	90.3	4.2	680	21.3	< .0001
		/w/ - /w ^h / ST	76.3	4.3	664	17.4	< .0001
		/w/ - /w ^h / CR	82.2	4.2	665	19.3	< .0001
		/ɹ/ - /ɹ ^h / ISO	-294.8	34.7	779	-8.5	< .0001
		/ɹ/ - /ɹ ^h / ST	-402.3	35.4	777	-11.3	< .0001
		/ɹ/ - /ɹ ^h / CR	-387.1	35.0	779	-11.0	< .0001
		/j/ - /j ^h / ISO	-98.5	4.0	784	24.3	< .0001
		/j/ - /j ^h / ST	91.7	4.2	790	21.8	< .0001
		/j/ - /j ^h / CR	-683.1	36.5	788	-18.7	< .0001
F2	Approx.: 4168 *** Context: 6 * Approx. x Context: 29 **	/w/ - /w ^h / ISO	-470.7	47.6	778	-9.8	< .0001
		/w/ - /w ^h / ST	-487.3	49.3	775	-9.	< .0001
		/w/ - /w ^h / CR	-532.9	48.0	775	-11.1	< .0001
		/ɹ/ - /ɹ ^h / ISO	-470.7	47.6	778	-9.8	< .0001
		/ɹ/ - /ɹ ^h / ST	-402.3	35.4	777	-11.3	< .0001
		/ɹ/ - /ɹ ^h / CR	-387.1	35.0	779	-11.0	< .0001
		/j/ - /j ^h / ISO	-26.8	44.3	789	-0.6	1.0000
		/j/ - /j ^h / ST	-58.0	45.8	789	-1.2	1.0000
		/j/ - /j ^h / CR	-69.9	44.3	789	-1.580	1.0000
F3	Approx.: 1018 *** Context: 3 Approx. x Context: 13	/w/ - /w ^h / ISO	128.0	57.9	765	2.212	1.0000
		/w/ - /w ^h / ST	87.6	60.0	760	1.4	1.0000
		/w/ - /w ^h / CR	103.0	58.3	760	1.767	1.0000
		/ɹ/ - /ɹ ^h / ISO	-375.7	42.2	774	-8.8	< .0001
		/ɹ/ - /ɹ ^h / ST	-517.4	43.2	773	-11.9	< .0001
		/ɹ/ - /ɹ ^h / CR	-452.3	42.6	774	-10.6	< .0001
		/j/ - /j ^h / ISO	14.0	54.0	789	0.2	1.0000
		/j/ - /j ^h / ST	-69.7	55.8	789	-1.2	1.0000
		/j/ - /j ^h / CR	-27.0	54.0	789	-0.5	1.0000
F4	Approx.: 455 *** Context: 3 Approx. x Context: 14	/w/ - /w ^h / ISO	213.4	48.8	706	4.3	0.0022
		/w/ - /w ^h / ST	77.8	50.5	693	1.5	1.0000
		/w/ - /w ^h / CR	20.7	46.1	789	0.4	1.0000
		/ɹ/ - /ɹ ^h / ISO	-137.0	36.0	768	-3.8	0.0236
		/ɹ/ - /ɹ ^h / ST	-282.1	36.8	766	-7.6	< .0001
		/ɹ/ - /ɹ ^h / CR	-189.1	36.4	767	-5.2	< .0001
		/j/ - /j ^h / ISO	40.9	46.1	789	0.8	1.0000
		/j/ - /j ^h / ST	29.4	47.7	790	0.6	1.0000
		/j/ - /j ^h / CR	20.7	46.1	789	0.4	1.0000

Table 5.8: Contd...

Feature	χ^2	Contrasts	Estimate	SE	df	<i>t</i> -ratio	<i>p</i> -value	
HNR	Approx.: 2275 *** Context: 115 *** Approx. x Context: 36 ***	/w/ - /w ^h / ISO	14.3	0.8	662	16.2	< .0001	
		/w/ - /w ^h / ST	10.2	0.8	676	12.0	< .0001	
		/w/ - /w ^h / CR	11.4	0.8	651	12.7	< .0001	
			/ɹ/ - /ɹ ^h / ISO	12.2	0.5	683	20.9	< .0001
			/ɹ/ - /ɹ ^h / ST	12.8	0.6	706	19.8	< .0001
			/ɹ/ - /ɹ ^h / CR	11.2	0.6	695	18.6	< .0001
			/j/ - /j ^h / ISO	12.0	0.7	711	15.7	< .0001
			/j/ - /j ^h / ST	11.6	0.8	710	13.7	< .0001
			/j/ - /j ^h / CR	9.1	0.7	710	12.0	< .0001
Duration	Approx.: 87.9 *** Context: 35.9 *** Approx. x Context: 48.0 ***	/w/ - /w ^h / ISO	-4.2	10.0	708	-0.4	1.0000	
		/w/ - /w ^h / ST	-36.5	10.3	695	-3.5	0.0671	
		/w/ - /w ^h / CR	-0.1	10.0	698	-0.01	1.0000	
		/ɹ/ - /ɹ ^h / ISO	-27.3	7.4	610	-3.6	0.0406	
		/ɹ/ - /ɹ ^h / ST	-23.1	7.6	626	-3.0	0.3764	
		/ɹ/ - /ɹ ^h / CR	-23.3	7.5	589	-3.1	0.3059	
		/j/ - /j ^h / ISO	-14.1	9.4	716	-1.4	1.0000	
		/j/ - /j ^h / ST	-70.4	9.7	738	-7.1	< .0001	
		/j/ - /j ^h / CR	-5.2	9.4	716	-0.5	1.0000	
Voicing	Approx.: 4993 *** Context: 15 *** Approx. x Context: 21 *	/w/ - /w ^h / ISO	90.3	4.2	680	21.3	< .0001	
		/w/ - /w ^h / ST	76.3	4.3	664	17.4	< .0001	
		/w/ - /w ^h / CR	82.2	4.2	665	19.3	< .0001	
		/ɹ/ - /ɹ ^h / ISO	95.5	3.1	765	30.1	< .0001	
		/ɹ/ - /ɹ ^h / ST	83.8	3.2	765	25.8	< .0001	
		/ɹ/ - /ɹ ^h / CR	82.1	3.2	763	25.6	< .0001	
		/j/ - /j ^h / ISO	98.5	4.0	784	24.3	< .0001	
		/j/ - /j ^h / ST	91.7	4.2	790	21.8	< .0001	
		/j/ - /j ^h / CR	-89.6	4.0	784	22.1	< .0001	
Intensity	Approx.: 1014 *** Context: 13 *** Approx. x Context: 12	/w/ - /w ^h / ISO	9.5	0.9	772	10.4	< .0001	
		/w/ - /w ^h / ST	9.9	0.9	768	10.4	< .0001	
		/w/ - /w ^h / CR	6.2	0.8	785	7.3	< .0001	
		/ɹ/ - /ɹ ^h / ISO	8.5	0.6	767	12.8	< .0001	
		/ɹ/ - /ɹ ^h / ST	9.4	0.6	765	13.7	< .0001	
		/ɹ/ - /ɹ ^h / CR	8.6	0.6	767	12.7	< .0001	
		/j/ - /j ^h / ISO	8.8	0.8	785	10.3	< .0001	
		/j/ - /j ^h / ST	9.3	0.8	786	10.5	< .0001	
		/j/ - /j ^h / CR	6.2	0.8	785	7.3	< .0001	

Table 5.8: Contd...

Feature	χ^2	Contrasts	Estimate	SE	df	<i>t</i> -ratio	<i>p</i> -value
CoG	Approx.: 1050 ***	/w/ - /w ^h / ISO	-378.9	83.2	722	-4.5	< .001
	Context: 26 ***	/w/ - /w ^h / ST	-313.8	86.1	711	-3.6	0.0441
	Approx. x Context: 57 ***	/w/ - /w ^h / CR	-233.1	83.8	713	-2.7	0.8477
		/ɹ/ - /ɹ ^h / ISO	-668.8	61.7	606	-10.8	< .0001
		/ɹ/ - /ɹ ^h / ST	-511.6	63.0	622	-8.1	< .0001
		/ɹ/ - /ɹ ^h / CR	-456.2	62.5	585	-7.3	< .0001
		/j/ - /j ^h / ISO	-1481.2	78.5	709	-18.8	< .0001
		/j/ - /j ^h / ST	-1199.5	81.2	733	-14.7	< .0001
		/j/ - /j ^h / CR	-834.6	78.5	709	-10.6	< .0001

n.s.: not significant, *: $p < 0.01$, **: $p < 0.001$, ***: $p < 0.0001$

Chapter 6

Acoustics of Angami Vowels

This work investigates the acoustic properties of Angami vowels with the aim of definitively establishing the phonologically contrasting vowels in the language. Contrary to some previous studies that report seven monophthongs and multiple diphthongs, this study concludes that there are six monophthongs and two diphthongs in the language. The acoustic characteristics associated with the monophthongs and the diphthongs are explored and reported in this work. For both monophthongs acoustic characteristics, such as, the first three formants ($F1$, $F2$, $F3$), and duration were explored. For diphthongs, acoustic features, such as, the first two formants ($F1$, $F2$), their discrete cosine transforms (DCT) and duration were explored. The salient of the vowels in terms of their acoustic properties was substantiated by statistical analyses.

6.1 Introduction and literature review

This chapter aims to investigate the acoustic properties of monophthongs and diphthongs in Angami. Besides a preliminary study reported in a dialect of Angami (Khonoma) by Blankenship et al. (1993), no other acoustic studies have been conducted on the Angami vowels. This paper also aims to bridge the research gap between the scientific and descriptive analyses of Angami vowels. The descriptions of

Angami vowel inventories vary in the total number of vowels, from six to seven. The disagreements in vowel description of Angami arise due to the status of the schwa vowel, sometimes described as a close mid-central vowel [ə] and as an open mid-central vowel [ü]¹ at other times. Some studies have also assigned phonemic status to both [ə] and [ü]. Apart from the monophthongs, diphthongs in the language have not received any attention in the linguistic literature. Hence, this chapter also aims to provide an acoustic-phonetic description of the diphthongs in Angami.

The mid-central vowel, or schwa, is denoted by the <ə> symbol. The schwa vowel occurs in languages with at least 1 to 7 vowels, with the most common inventory having 5 (Schwartz et al., 1997a). Recasens (2022a) observed in recent cross-linguistic research of 110 languages that the majority of schwa occurs in languages with an inventory of seven vowels (20.9%), followed by languages with inventories of six, eight, and nine (18.7%), and lastly languages with inventories of four and ten vowels (12.1%). The same study showed that when mid-vowels in the language system increase, the first formant (F1) is more variable among languages than the second formant (F2). Male speakers' formant frequency values for schwas are estimated to range between 500 Hz (F1) and 1500 Hz (F2). This range reflects the resonances of a vocal tract 17.5 cm long (Fant, 1970).

According to cross-linguistic reports, the vowel schwa's properties vary depending on the context. In languages such as English and Dutch, schwa is restricted to unstressed syllables due to vowel reduction. According to Flemming and Johnson (2007); Flemming (2009), there are two kinds of schwa in English: schwa that occurs word finally and schwa that occurs in word-internal. The authors observed that the F2 of word-internal schwa is highly variable due to the assimilation of the contextual sounds. A similar case of F2 variability has also been reported in Dutch (Koopmans-

¹The correct IPA for open mid-central vowel correct is /ɜ/, however, Kuolie (2006) described the vowel as open mid-central vowel and represented it by /ü/. For the uniformity in the study /ü/ will be used to differentiate it from mid-central vowel /ə/.

van Beinum, 1994). The duration of schwa in English when produced word-finally is relatively longer than the schwa produced word-internally. Flemming (2009) reports that non-final schwa vowels averaged 64 ms in duration, while the word-final schwa vowels had a mean duration of 153 ms. In Dutch, the duration of the schwa vowel is reported to be shorter than the other vowels (Koopmans-van Beinum, 1994). This study focuses on the schwa vowel in Angami as it is described to occur in two different heights and with distinct durations (Kuolie, 2006).

As mentioned earlier, there are no dedicated studies that reported the acoustic characteristics of the vowels in Angami. Hence, in the current study, we provide a description of the Angami vowels as spoken in the Kohima village variety. We also address the inconsistencies in terms of the number and type of vowels in the standard Kohima village variety. Considering that, it is pertinent that we conduct a survey of the literature on the monophthongs and diphthongs in Angami. Hence, in the following section 6.1.1, we provide a literature review on the monophthongs in Angami, followed by a literature review on the diphthongs in Section 6.1.2. In Section 6.1.3, we provide a description of the representation of Angami vowels in the standard orthography of the language. Following this, the methodology of the current work is discussed in Section 6.2, followed by the results and discussion in Section 6.3 and Section 6.4, respectively.

6.1.1 Literature Review of Angami monophthong Vowels

Except for the study on the Khonoma dialect, a western variety of Angami, studies on Angami vowels have been impressionistic and without any acoustic description (Blankenship et al., 1993; Giridhar, 1980; Kuolie, 2006; Ravindran, 1974). These studies have also resulted in disagreements regarding the phonemic inventory of the Angami language. For example, in previous studies, the vowel inventory size proposed for Angami ranges from six to seven phonemic vowels (Burling, 1960; Ravindran,

1974; Giridhar, 1980; Kuolie, 2006). All studies agree on the five peripheral vowels in Angami /i, e, u, o, a/ but they have described the central vowel either as a close mid-central vowel or as an open mid-central vowel and in some cases as both.

The seventh vowel arises when authors differentiate between the mid-central vowel /ə/ and with open-mid central vowel [ü]. In one of the earliest works on the language by McCabe (1887), the language is described to have six vowels / i, e, u, ü, o, a /. He described the central vowel as a high vowel. Studies on the Kohima variety by Burling (1960), Marrison (1967), Ravindran (1974), and Giridhar (1980) agree that there are six vowels in Angami / i, e, u, ə, o, a/, and the central vowel is mid-central. In the study of the Khonoma Angami variety, Marrison (1967) and Blankenship et al. (1993) also maintain that there are six vowels in the inventory. As opposed to these studies, Kuolie (2006) has described that there are seven vowels in the inventory. The author states that there are two central vowels, one is the open mid-central vowel /ü/, and the other is the mid-central vowel /ə/. Meyase (2016), suggested that Kuolie's open mid-central vowel is an allophone of the close mid-central vowel. The descriptions of Angami vowels in the literature are summarized in Table 6.1.

While Ravindran (1974) and Giridhar (1980), have described the central vowel for the word 'poetry' as [ə], Kuolie (2006) have made the distinction between mid and open mid-central vowels. He assigned [ü] for poetry and [ə] for 'small'. Note that in the standard Angami orthography, both [ü] and [ə] are ascribed a single grapheme < ü > or < ii >.

6.1.2 Literature Review of Angami Diphthong Vowels

Previous literature on the explanation of the diphthong vowels is inconsistent, as the inventory size ranges from 1 to 6 diphthongs in Angami. McCabe (1887) reports in Angami is one of the earliest works on the language. Although the author did not list out the inventory size of the diphthongs, from analyzing his list of words in Angami,

Table 6.1: Description of monophthong vowels in Angami by various authors

Author (year)	Variety	No. Vowel	description
McCabe (1887)	NA ²	7	/ i, e, u, ü, ö ³ , a/
Burling (1960)	Kohima	6	/ i, e, u, ə, o, a/
Marrison (1967)	Khonoma	6	/ i, e, u, ə, o, a/
Ravindran (1974)	Kohima	6	/ i, e, u, ə, o, a/
Giridhar (1980)	Kohima	6	/ i, e, u, ə, o, a/
Blankenship et al. (1993)	Khonoma	6	/ i, e, u, ə, o, a/
Kuolie (2006)	Standard	7	/ i, e, u, ə, ü, o, a/
Meyase (2016)	Standard	6	/ i, e, u, ə, o, a/
Suokhrie (2015)	Kohima	7	/ i, e, u, ə, ü, o, a/

we find 1 diphthong [ie]. Marrison (1967) has listed 1 diphthong [ie] in the Khonoma variety of Angami. While Kuolie (2006) and Suokhrie (2015) have listed 2 diphthongs [ie, uo] for the Kohima variety, Burling (1960), Marrison (1967), and Ravindran (1974) have reported 4 diphthongs, although the description differs. Giridhar (1980) has listed 6 diphthongs. Table 6.2 summarizes the number and description of vowels as described in previous works.

Table 6.2: Description of Diphthong vowels in Angami by various authors

Author (year)	Variety	No. Vowel	description
Burling (1960)	Kohima	4	/ ie, uo, əi, uu /
Marrison (1967)	Khonoma	1	/ ie /
Marrison (1967)	Kohima	4	/ ie, uo, ei, ou /
Ravindran (1974)	Kohima	4	/ ie, uo, ou, əi /
Giridhar (1980)	Kohima	6	/ ie, uo, əi, io, eo, ou /
Chase (1992) ⁴	Khonoma	7	/ ie, ei, ai, ia, io, oi, ui/
Kuolie (2006)	Standard	2	/ ie, uo/
Suokhrie (2015)	Kohima	2	/ ie, uo/

As seen in Table 6.2, [ei, ou] are two of the most commonly occurring diphthongs in the previous studies. Besides the two common diphthongs, there are other diphthongs which are described. The inconsistencies in the description of Angami diphthongs arise partially due to the orthographic convention used for the language⁵. The Angami

⁵We examine words from one of the earliest written documents in Angami, such as the Hymns by Rev. S.W. Rivenburg (1903), where vowels are represented in two ways to signify a single sound.

orthography uses [ei] for the /i/ sound, while [ou] is used for /u/ sound. This has prompted Marrison (1967), Ravindran (1974), and Giridhar (1980); and in Marrison (1967) to describe the orthographic < ei > and < ou > as diphthongs. This issue of orthography in Angami is further addressed in Section 6.1.3.

Secondly, the other issue in the available descriptions of the diphthong is the [əi] sound which is described by Burling (1960); Ravindran (1974); Giridhar (1980). The author's sound [əi] is described in words such as [zəi] 'dark'; [təi] 'sky' and [pʰhenəi] 'clothes, however the phonemic status of the [əi] sound in Angami is a close high vowel /i/. The IPA for the words is /zi/, /ti/, and /pʰeni/, respectively. Burling (1960) also describes [uu] as a diphthong in words such as [suu] /su/ 'deep' and [rəluu] /rəlu/ 'war dance'. Angami does not differentiate between vowel duration, hence [uu] described as a diphthong is a monophthong /u/. Finally, in Angami, a vowel sound can act as a suffix to the free morpheme and these suffixes have specific grammatical functions. Kuolie (2006) addressed them as vowel sequences. Examples are [ha+u] /hau/ 'this-DEF' and [a+e] /ae/ '1st person singular + NOM'. Hence, in Angami, vowel sequences do not qualify as diphthongs. Taking into consideration the existing literature on diphthongs, it is evident that there are only two diphthongs in Angami, namely, [ie] and [uo]. In the current study, as we also agree with the number of diphthongs in Angami as two, we consider only [ie] and [uo] for acoustic analysis.

6.1.3 Representation of Angami vowels in orthography

When describing the phonological of the languages, it is common to be influenced by the writing system. Angami, lacking its own script, is instead transcribed using the Roman orthography. To understand the writing system in Angami versus the sound produced, examples of the word with the IPA symbols are listed in Table 6.3. As seen in the table, the vowel /i/ is represented both with < ei > and < i >; and for /u/ with < ou > and < u >. It should be noted that the authors did not find any vowel

duration differences or tone differences among the various ways of writing. This can be examined in future studies. More examples of the words with the same sounding vowel with different orthography are attested in Table 6 of Appendix A.

Table 6.3: Orthography vs. IPA of the vowels in Angami

Orthography	IPA	Meaning	Tone
mei	/mi/	tail	T1
mi	/mi/	fire	T1
nuo	/no/	child	T4
nou	/nu/	heart	T4
nu	/nu/	inside	T4
pfü	/pfə/	carry	T5

This study looks into the six monophthong and two diphthong vowels, the minimal pairs are shown in Table 6.4 and Table 6.5 in a CV structure. As for the monophthong vowels, the mid central vowel /ü/ was examined as well in comparison with vowel /ə/ and other peripheral vowels. For this, the example stated in the dictionary Liezietsu et al. (2019) as well as Kuolie (2006) were considered. It is to be noted that [ü] occurs in a restricted manner. It is either a stand-alone vowel, an onset, or a coda in the VCV structure. The vocabularies are also few and limited. The findings regarding the two central vowels are discussed in detail in the results (Section 6.3.1).

Table 6.4: Example of monophthong vowel minimal pair in Angami

IPA	Word	Meaning	Tone
a	ṭsa	less	T1
e	ṭse	break	T1
ə	ṭsə	sprout	T3
i	ṭsi	over	T2
o	ṭso	reach	T3
u	ṭsu	went	T3

Considering the discussion above, we will look into acoustic-phonetic characterization of the monophthong and diphthongs in Angami. The monophthongs in the language are analyzed using the formants features and the duration, whereas, the

Table 6.5: Example of diphthong vowels minimal pairs in Angami

IPA	Word	Meaning	Tone
ie	tie	lips	T3
uo	tuo	Future marker	T4

diphthongs are captured using the first two formants (F1, F2), their discrete cosine transforms (DCT) and duration.

6.2 Methodology

This section describes the methodology of the current study which includes the participants, the materials, the recording procedure, and the annotation of the data. Apart from that, the extraction of formant values (F1, F2 & F3) and the statistical analyses are also discussed.

6.2.0.1 Speakers

33 Angami native speakers (16 Female and 17 Male) contributed speech data for the monophthongs study. Out of the 33 speakers, 28 were born and raised in the Kohima village, Nagaland. The speakers were chosen from Kohima village, as the variety spoken in this area is regarded as the standard variety of the Angami language. While the additional 5 speakers were language teachers who have a minimum of master's degree in the language. From the subset of the data, 21 speakers' speech data were used for the diphthong study. The study area is in the Kohima district of Nagaland, North-East India. The average speaker's age is 34.9 years (Standard Deviation = 8.6) at the recording time. The mother tongue (L1) of all the speakers is Angami (Tenyidie). All participants can also speak English (state official language) and Nagamese (lingua franca) besides Angami. All participants had a minimum education of a bachelor's degree, and some were working professionals. They were all proficient readers of Angami. The speech data recording was conducted following the

standard procedure discussed in Chapter 1 (Section 1.3.3).

6.2.0.2 Materials

The data for the vowel study comes from the collection of databases that were used for the study of nasals, fricatives, and affricates approximants and tones (see Section 1.3.2). The data combined CV and CVCV, where V is either a monophthong or a diphthong. The databases that consist of meaningful words in Angami were considered. Hence, database codes Anl, Anaso, Aaf, Apx, Amt, and Act were analyzed for vowel study. The data incorporate all six monophthongs and two diphthongs in Angami.

In a subset of the data, the vowels /i/ and /ə/ were produced only in isolation frame for the vowels preceding by nasal series in CV syllables; hence, the number is relatively higher than other vowels. Each vowel was produced in three contexts: sentence, carrier, and isolation. This study investigates 12,066 monophthong tokens in total, and the distribution is shown in Table 6.6. Additionally, we examined 2936 diphthong tokens. The distribution is shown in Table 6.7. Angami has 4-5 level tones. Each of the 5 tones in Angami, namely, T1-T5 as reported in the Angami dictionary, MKS Dieda (Liezietsu et al., 2019).

Table 6.6: Distribution of six monophthong vowels in Angami across the three contexts in the data.

Vowels	Carrier	Phrasal	Isolation	Total
a	545	552	680	1,777
e	545	569	723	1,837
ə	640	655	1290	2,585
i	725	750	1032	2,507
o	531	537	655	1,723
u	484	491	662	1,637
Grand Total				12,066

Table 6.7: Distribution of the two diphthong vowels in Angami

Vowels	Carrier	Phrasal	Isolation	Total
ie	131	132	183	446
uo	85	89	114	288
Grand Total				734

6.2.0.3 Annotation

For analysis, the speech recordings were uploaded to a computer. The sound files were manually segmented for the vowel boundaries using Praat 6.0.43 to listen and visually inspect the waveforms and spectrograms (Boersma, Paul and Weenink, David, 2020). The segmented files were cross-checked three times to avoid any errors in annotation and segmentation. As seen in Figure 6-1, the vowel boundaries, along with the consonant boundaries, are annotated. A Praat script was used to extract the F1, F2, F3, and duration values which were transferred to an Excel for further analysis.

6.2.0.4 Acoustic measurements

Vowel qualities of languages are best represented by the first three formants, where the first formant (F1) is inversely correlated with the height of the vowel. The second formant (F2) is directly proportionate to the frontness of the vowel. Finally, the third formant (F3) is inversely proportionate to the roundness of the vowel. To see the acoustics of Angami vowels, we have extracted the F1, F2, and F3 values of the vowels from the midpoint of the vowels. Apart from the vowel quality acoustics (F1, F2, and F3), we also extracted vowel quantity information as expressed by the duration of the vowels. Formant ceilings for females were at 5500, while for the male were 5000.

For diphthongs, F1 and F2 values were extracted at every 2% of the total duration of the vowels. In order to estimate the slope of the F1 and F2 trajectories, the values were converted to discrete cosine transform (DCT) coefficients. The first three coef-

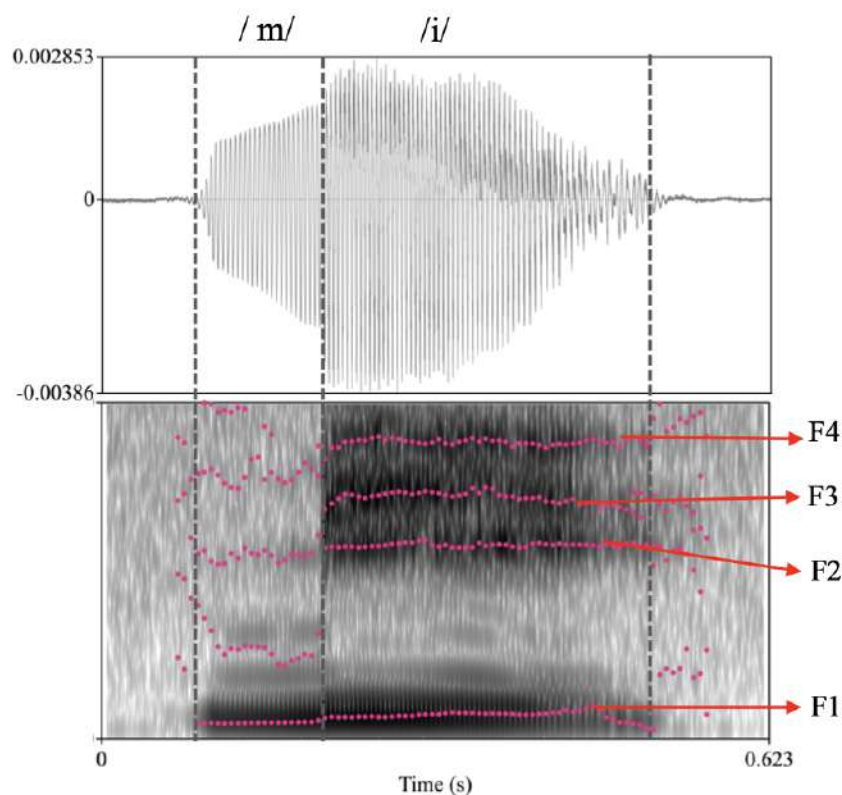


Figure 6-1: Annotation example showing the vowel boundary of the word /mi/ 'fire' as produced by a female speaker

ficients, namely, C0, C1 and C2 were analyzed for diphthong characteristics. While, C0 provides the average formant values calculated from the entire formant trajectory, C1 provides information about the slope of the formant trajectory by comparing it to a falling half cycle of a cosine curve. C2 provides values corresponding to the formant trajectory when it is compared to a full cycle of a cosine curve. Hence, a falling trajectory has high positive C1 values and a raising trajectory has high negative values. Similarly, a convex contour results in high positive C2 values whereas a concave contour results in a high negative C2 values.

6.2.0.5 Statistical analyses

To eliminate speaker effects, the raw formants were normalized using the z-score normalization method as describes in Chapter 1 (Section 1.3.6). Hence, in this study,

before proceeding with the analysis of the formant frequencies, the data were normalized with the Lobanov method for speaker effects by using the `normLobanov` function from the `phonR` package on the open-source R platform (R Core Team, 2022). Utilizing R's statistical packages, qualitative and exploratory statistics were calculated as described in Chapter 1 (Section 1.3.6).

6.3 Results

This section discusses on result of the monophthongs and diphthong vowels using spectral and temporal features.

6.3.1 Acoustic Features of Monophthong Vowels in Angami

To visualize the three formants' differences in the six Angami monophthong vowels, density plots for normalized values of the F1 and F2 formants are presented in Figure 6-2. The figure also accompanies the kurtosis and skewness values for each vowel, representing the normalcy of the distribution of the normalized F1 and F2 values. Skewness indicates the symmetry of the data, whereas kurtosis indicates the magnitude of the tails. A symmetrically distributed data has skewness between -0.5 and 0.5 . While kurtosis values around 3 indicate a normal distribution, values more than 3 and less than 3 are short-tailed and heavy-tailed distribution indicators, respectively.

As seen in the figure, vowels /a/, /e/, and /ə/ have distributions peaking at 2, 0, and 0.5 hertz in normalized F1, respectively, which are symmetrical considering their skewness values. As indicated by the kurtosis values, /e/ and /ə/ have normal normalized F1 distribution, while /a/ is slightly short-tailed. Similarly, the vowel /i/, /o/, and /u/, have F1 distributions peaking at -1 , 0, and -1 hertz, respectively. The F1 values for all three vowels are highly skewed. The high kurtosis values indicate that the F1 of these three vowels has a short-tailed distribution. In Figure 6-2, we

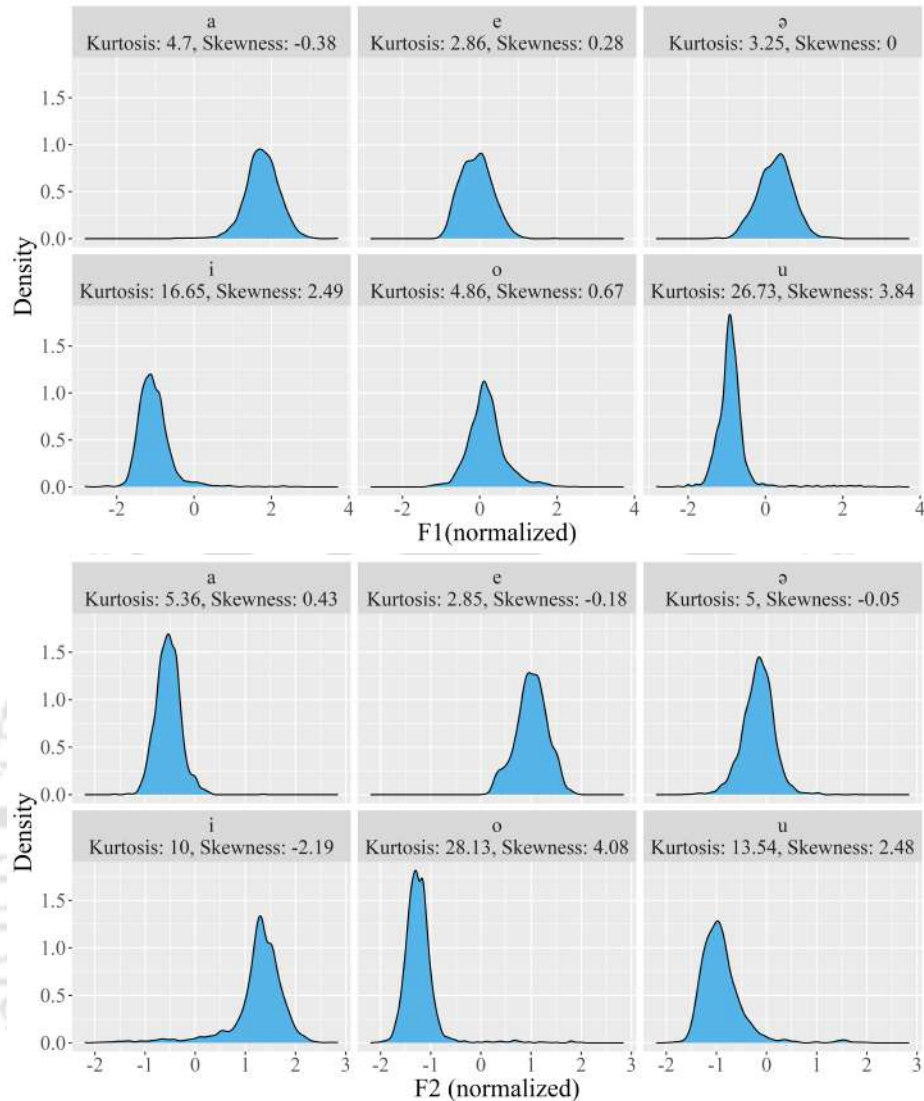


Figure 6-2: Skewness and Kurtosis graph of normalized F1 (top) and F2 (bottom) in the six vowels of Angami

also looked into the distribution of the normalized F2. Vowels /a/, /e/, and /ə/ have distribution peaking at 0.5, 1, and 0 hertz, respectively, and they have symmetrical distribution based on their skewness values (between -0.5 and 0.5). Except for vowel /e/, the other vowels in F2 values have short-tailed distribution.

In order to examine the vowel characteristics, the first two formant frequencies for each vowel were measured from the beginning of the voicing to the end. While the first formant (F1) frequency gives information about vowel height, the second formant

(F2) frequency provides information on vowel backness (Cox, 1996). Both F1 and F2 frequencies of the vowels were calculated at the vowel midpoint in hertz (Hz) values. To capture the vowel-specific formant feature without the co-articulatory features arising due to the preceding and the following consonant contexts, vowel features are obtained only in the midpoint of the vowel. After extracting the vowel formants, vowels were normalized and formants were plotted (one standard deviation ellipses).

6.3.1.1 Vowel density

Vowel Density helps us to see how uniformly the vowels are distributed in terms of F1 and F2, and also in identifying whether the vowels are pronounced in one distinct way or two. When the vowels are distributed uniformly, we get one center point. Vowel density further helps in understanding the pattern of vowel dispersion. In our present study, we used vowel density not only to determine the pattern, but also to help us confirm the status of individual vowels. The vowel density plot of the six vowels has been plotted for visualization in Figure 6-3. As seen in Figure 6-3 (top), the values without normalization have been where two center points are observed in vowels such as /i/, /e/, and /ə/. This different center point arises due to gender effects. Hence, the formant values were normalized using the Labonov normalization and were plotted as shown in Figure 6-3 (bottom). As seen in the figure, after normalization of the values, the vowels are distributed uniformly.

6.3.1.2 Vowel quality

To see if there are seven vowels in Angami with the distinction of the two central vowels, we differentiated the schwa vowel that occurs alone as V, symbolized by <ü>, and schwa that occurs in a CV or CVCV syllable, symbolized by /ə/. They were plotted along with the other peripheral vowels such as /a, e, i, o, u/. For visualization, the normalized values of F1 and F2 of the seven (left) and six (right)

vowels are plotted in Figure 6-4. As noticed in Figure 6-4 (top), the schwa that is produced in isolation occupies the same space F1 and F2 of /ə/. Hence, this confirms that there are six vowels in Angami as seen in the Figure 6-4 (bottom). The front close vowel /i/ and front close-mid vowel /e/ show proximity of the ellipses, whereas the back close vowel /u/, central vowel /ə/, back close-mid vowel /o/ and front open vowel /a/ are distinct from the rest of the vowels and are independent vowels without any overlapping ellipses. The mean values and standard deviation of the F1, F2, and F3 of the six vowels in Angami are displayed in Table 6.8. As seen in the table, the F1 value is inversely related to the height of the vowel, where the low vowel /a/ has a high F1 and the two high vowels /i/ and /u/ have a low F1. As for F2 values, the front vowels /i, e/ have higher F2 values whereas the back vowels have lower F2 values.

We look further into the variation of the vowel spectral dispersion based on gender, as plotted in Figure 6-5. It is clear from the figure that the vowels produced by male speakers do not have any overlapping ellipses and are distinct from each other. However, in the female speaker production of the vowel, the front close vowel /i/ and front close-mid vowel /e/ have minimal overlapping ellipses. Table 6.9 provides the mean values and standard deviation of the three formants across the six vowels by gender. We see that female speakers have higher formant Frequencies (F1, F2 & F3).

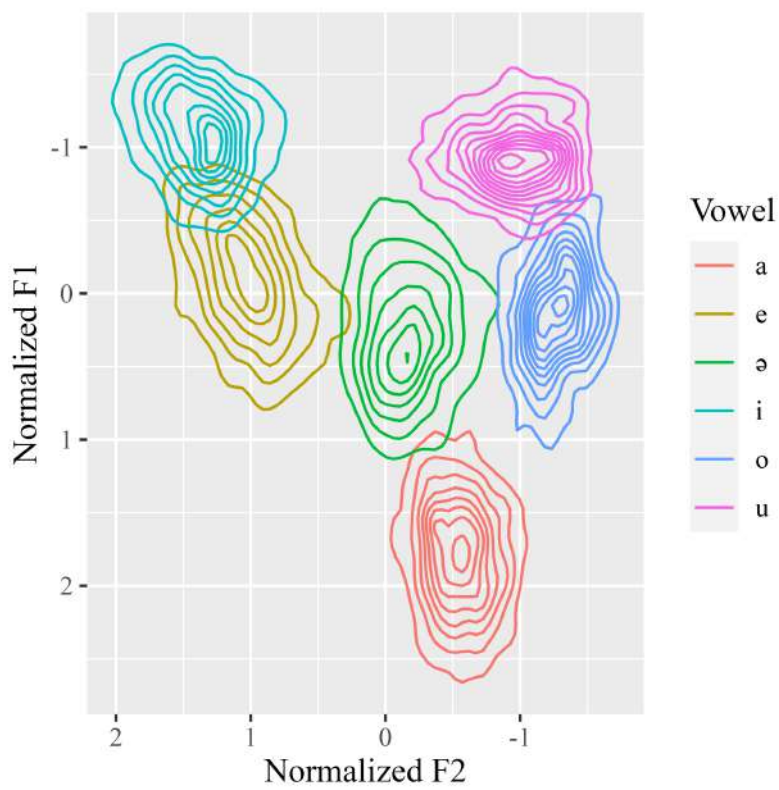
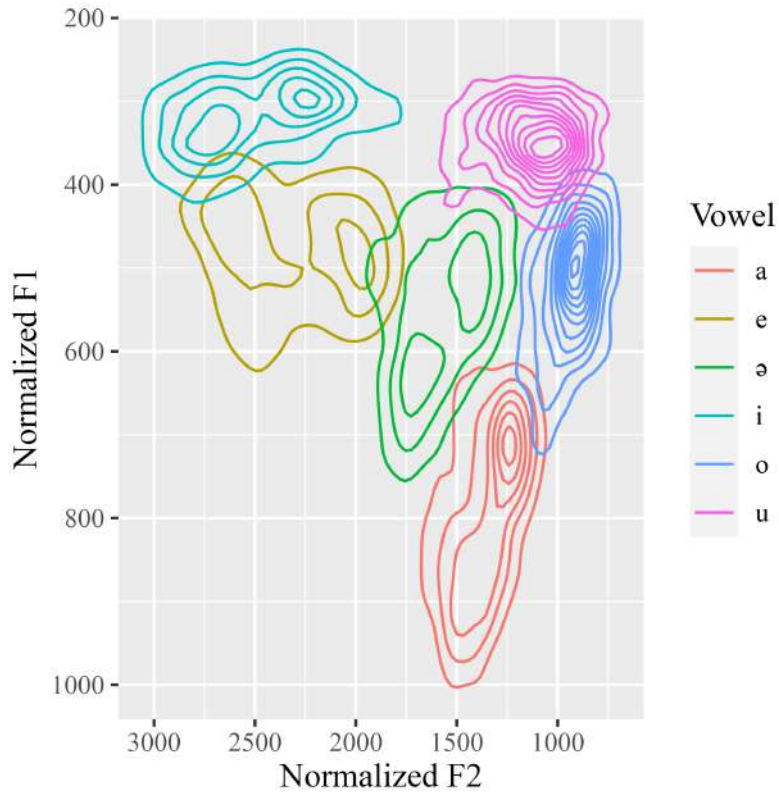


Figure 6-3: Mean F1 and F2 plots for the six Angami vowels (top); Normalized average F1 and F2 plots of the six Angami vowels (bottom)

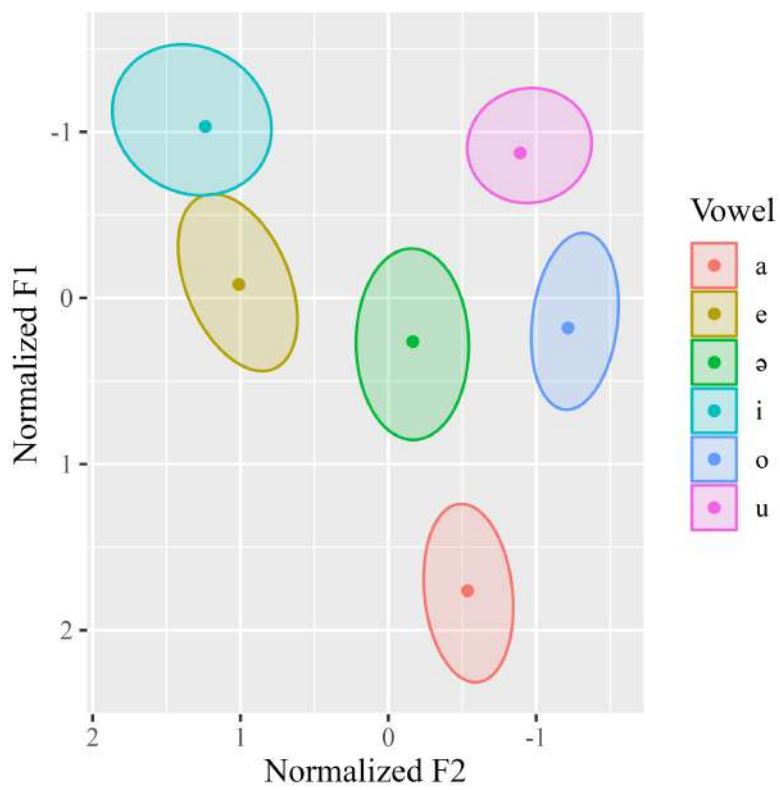
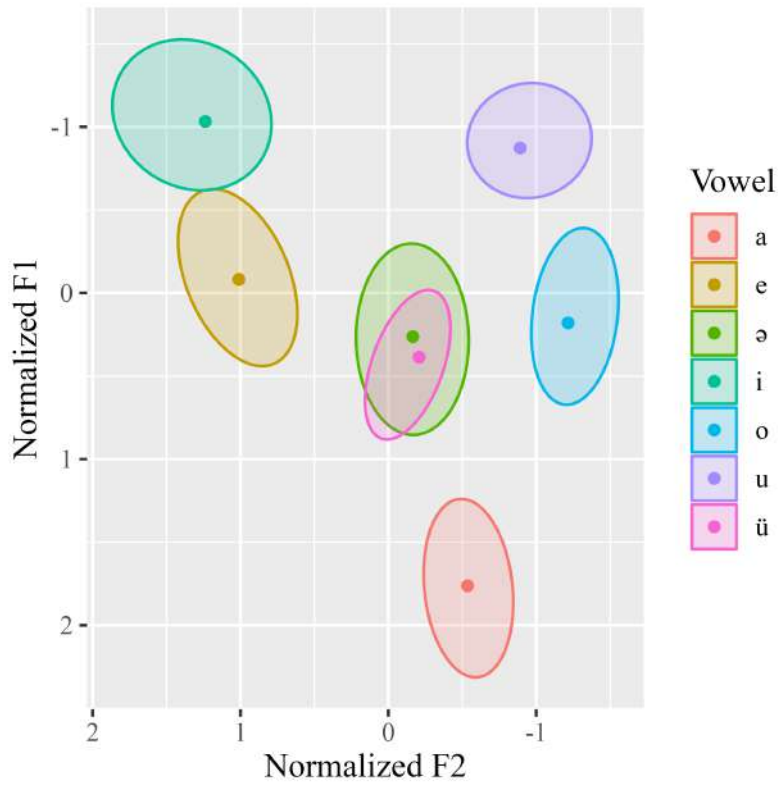


Figure 6-4: Normalized F1 and F2 with one standard deviation ellipses for 7 vowels (top) and 6 vowels (bottom)

Table 6.8: The mean values and standard deviation (SD) of F1, F2, and F3 of the six vowels in Angami.

Vowels	F1	F2	F3
i	334 (67)	2441 (417)	3071 (303)
e	488 (75)	2304 (315)	2908 (325)
ə	555 (97)	1605 (222)	2748 (394)
a	793 (104)	1371 (164)	2667 (395)
o	535 (87)	966 (220)	2794 (358)
u	362 (81)	1167 (261)	2670 (316)

To confirm the statistical significance of the formant frequencies (normalized F1 and F2) among the six vowels, we constructed LME models. The dependent variables were the formant frequencies. Hence, vowels (V), context (C), gender (G), tones (T), and their interactions were fixed effects, while speaker (Sp), syllable type (St), structure type (StrT), and individual word codes were the random effect, as in 6.1.

$$\text{Formants} \sim V + C + G + T + V * C * G * T + (1|Sp) + (1|St) + (1|StrT) + (code) \quad (6.1)$$

A backward reduction model (step model) was conducted where the random effects were removed since the vowel values were already normalized. The summary of the LME model for normalized F1 is presented in Table 6.10. The constructed models were subjected to Type II Wald Chi-square tests for analysis of deviance (ANOVA) on LME models separately. The analysis indicates that the effect of gender, tone, and vowel significantly impacts the first formant (F1). The interaction demonstrates statistical significance ($p < 0.001$). However, upon reviewing the LME model for normalized F2, as illustrated in Table 6.11, it becomes evident that the fixed factors, including gender, vowel, and tone, do not significantly influence F2. Furthermore, the interaction between these factors significantly impacts F3, as seen in Table 6.12. Figure 6-6 displays the interaction plot of the LME models, providing a visual representation of how the variables of interest, such as gender, tone, and vowel, interact and impact the formant frequencies.

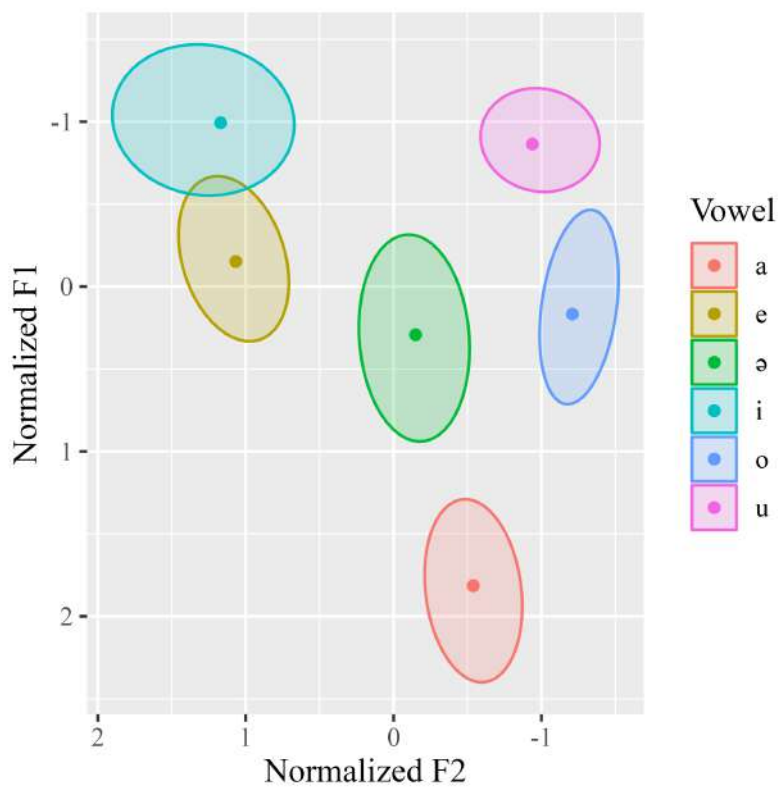
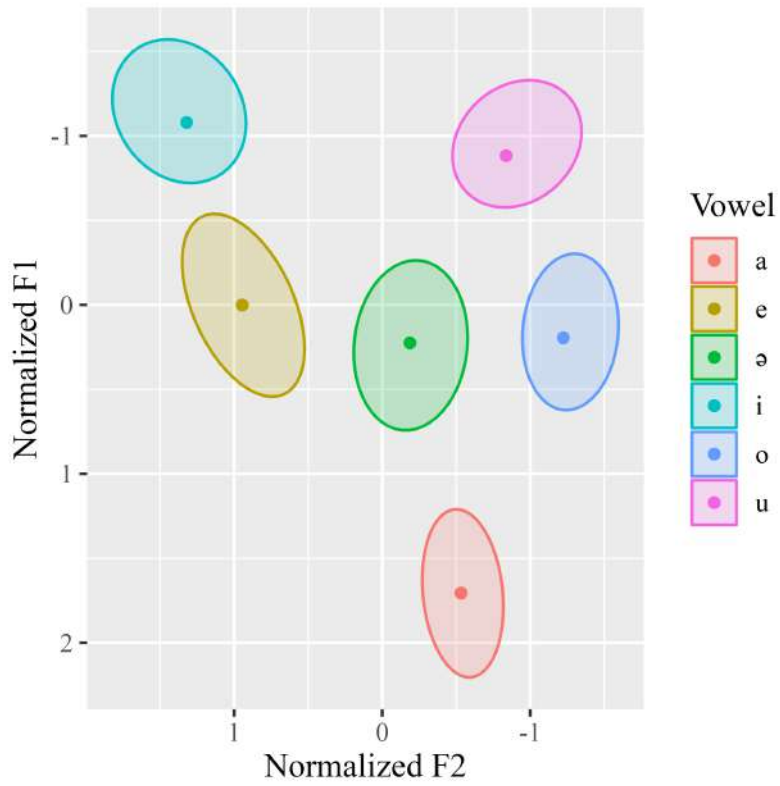


Figure 6-5: Normalized F1 and F2 with one standard deviation ellipses for Male (top) and Female (bottom) speakers)

Table 6.9: The mean values and standard deviation (SD) of F1, F2, and F3 of the six vowels in Angami as produced by Male and Female speakers.

Vowels	Male			Female		
	F1	F2	F3	F1	F2	F3
i	320 (67)	2250 (285)	2885 (222)	344 (63)	2609 (435)	3237 (271)
e	477 (62)	2043 (172)	2669 (199)	497 (81)	2523 (217)	3105 (267)
ə	513 (73)	1469 (163)	2574 (219)	588 (95)	1701 (199)	2891(418)
a	720 (68)	1284 (126)	2502 (217)	864 (85)	1452 (154)	2819 (444)
o	507 (64)	918 (218)	2562 (246)	564 (98)	1003 (208)	2999 (307)
u	352 (85)	1131 (262)	2469 (247)	370 (76)	1195 (250)	2884 (228)

Table 6.10: Results of an ANOVA conducted on an LME model with normalized F1 as the dependent variable and vowel, context, gender, tone, and their interactions as fixed effects. Only statistically significant interactions are shown in this table.

	χ^2	(df)	p-value
vowel	41674.0	5	< 0.001
context	574.2	2	< 0.001
gender	0.17	1	0.67
tone	70.4	4	< 0.001
vowel*tone	245.13	20	< 0.001
vowel*context	315.1	10	< 0.001
vowel*gender	102.3	5	< 0.001
vowel*context*tone	92.4	40	< 0.001
vowel*gender*tone	68.9	20	< 0.001
Conditional R ² = 0.85, Marginal R ² =0.79			

Table 6.11: Results of an ANOVA conducted on an LME model with normalized F2 as the dependent variable and vowel, context, gender, tone, and their interactions as fixed effects. Only statistically significant interactions are shown in this table.

	χ^2	(<i>df</i>)	<i>p</i> -value
vowel	48316.5	5	< 0.001
context	517.1	2	< 0.001
gender	0.2	1	0.60
tone	8.7	4	0.06
vowel*tone	159.2	20	< 0.001
vowel*context	82.2	10	< 0.001
vowel*gender	165.3	5	< 0.001
vowel*context*tone	78.8	40	< 0.001
vowel*context*gender	21.2	10	< 0.05
Conditional $R^2 = 0.86$, Marginal $R^2 = 0.83$			

Table 6.12: Results of an ANOVA conducted on an LME model with normalized F3 as the dependent variable and vowel, context, gender, tone, and their interactions as fixed effects. Only statistically significant interactions are shown in this table.

	χ^2	(<i>df</i>)	<i>p</i> -value
vowel	2257.4	5	< 0.001
context	13.5	2	< 0.001
gender	0.19	1	0.65
tone	131.4	4	< 0.001
vowel*tone	231.2	20	< 0.001
vowel*context	59.0	10	< 0.001
vowel*gender	229.0	5	< 0.001
vowel*gender*tone	80.8	20	< 0.001
Conditional $R^2 = 0.86$, Marginal $R^2 = 0.83$			

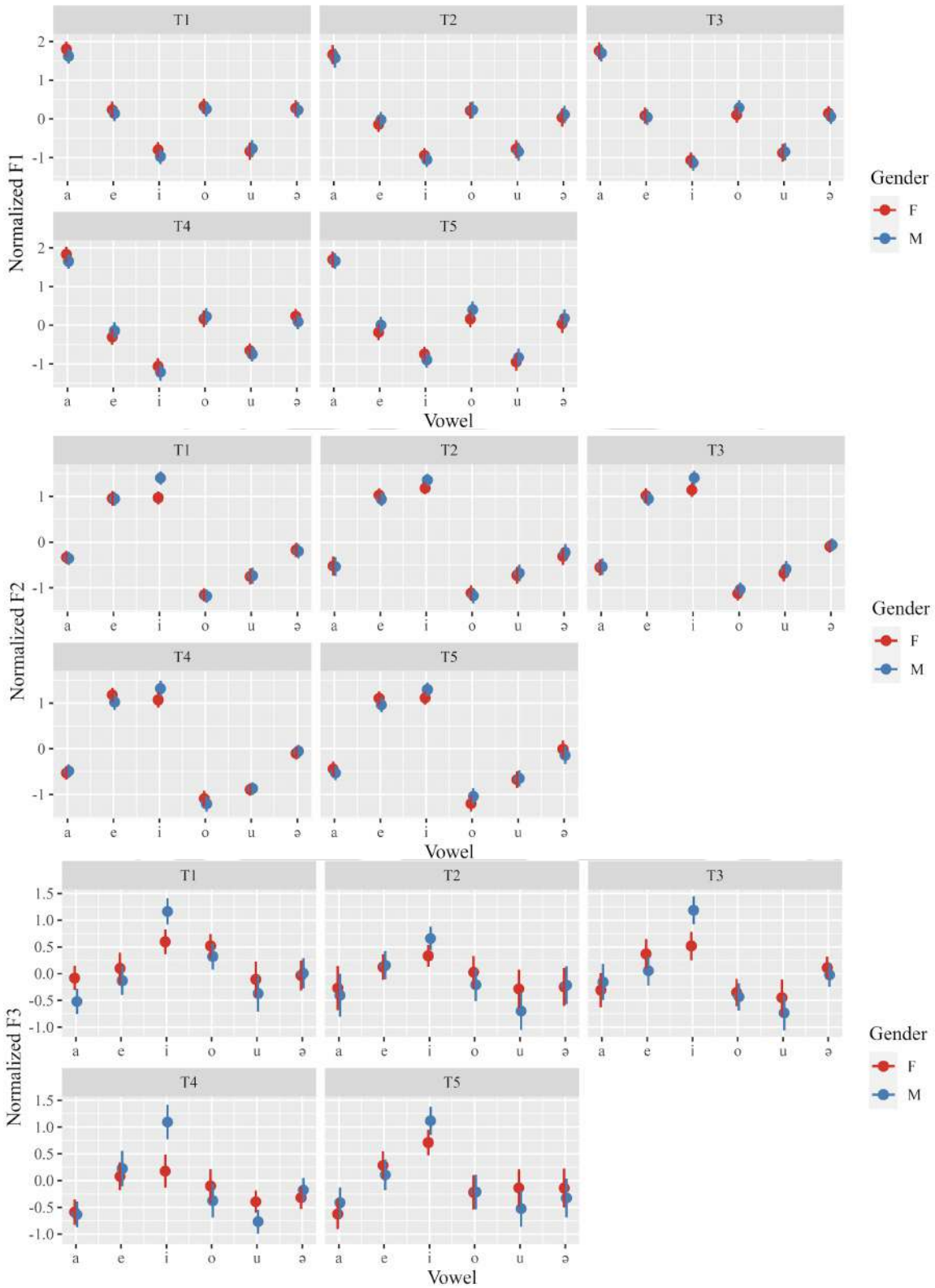


Figure 6-6: Interaction plot for the LME models where normalized F1 (top), F2 (mid), and F3 (bottom) are dependable variables, whereas the interactions of gender, tones, and vowels are the fixed effects.

6.3.1.3 Monophthong Vowel Duration

Apart from the vowel quality, we also extracted vowel quantity information, as expressed by the duration of the vowels. There are two parts to the vowel duration study. One was a preliminary test, where three female speakers were asked to produce the four meaningful vowels /a, ə, o, u/ that can appear in lexical words without an onset or a coda. This preliminary vowel duration study is to confirm if the four vowels that occur in isolation (V) have similar or temporal differences. The example used in the study is shown in Table 6.13.

Table 6.13: Example of a minimal pair of vowels where they can appear in a lexical word without an onset or a coda.

Vowel	Meaning	Tone
/a/	1st person pronoun	T3
/ə/	song, poetry	T1
/o/	okay, agreement	T3
/u/	3rd person pronoun	T1

Table 6.14: Pairwise comparison of the vowels that appear as a lexical word without an onset or a coda.

Contrast	Estimate	SE	df	t-ratio	p-value
a - ə	2.31	21.9	57.1	0.106	1.0000
a - o	7.65	21.5	57.1	0.356	1.0000
a - u	79.44	21.5	57.1	3.701	0.0029
ə - o	5.34	14.2	55.0	0.377	1.0000
ə - u	77.13	14.2	55.0	5.441	< 0.0001
o - u	71.79	13.7	54.8	5.237	< 0.0001

Figure 6-7 presents the results of the preliminary test on vowel duration, illustrating the average mean and standard error of the four vowels when produced in isolation. The findings indicate that three out of four vowels exhibit similar durations, falling within the 260-280 milliseconds range, while the duration of vowel /u/ is at 190 milliseconds. The result showed that vowels, when produced in isolation, have higher duration, except for /u/, a back close vowel. The results indicate that vowels

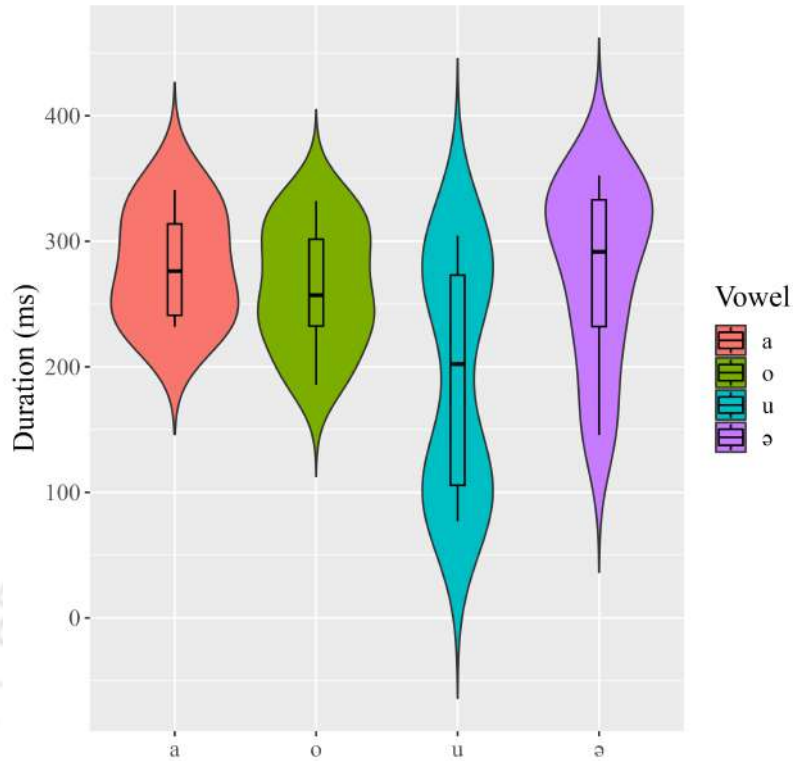


Figure 6-7: Violin plots of the average duration of vowels where they can appear in a lexical word without an onset or a coda.

produced in isolation exhibit similar durations, particularly for the central vowel. To investigate the significance of vowel duration across the vowels, an LME model was constructed, with vowel duration as the dependent variable and vowel as the fixed effect, while considering speaker, context, and repetition as random effects. The Type II Wald Chi-square tests for analysis of deviance revealed that vowel duration is indeed significant across the four vowels ($\chi^2 = 43.59$, $p < 0.001$). Subsequently, the LME model underwent post-hoc Bonferroni analysis to identify significant differences between vowels. The results, as presented in Table 6.14, indicate that the vowel duration of /a, ə, o/ is not statistically significant, suggesting that these vowels share similar durations when produced without an onset or coda.

The second part of the vowel duration study involved analyzing 12,066 tokens. Figure 6-8 illustrates the average duration of the six vowels and their standard errors,

while the specific values are given in Table 6.15. As seen in the figure and table, vowels /a, e, i, & o/ are in the mid-range with a duration ranging from 185 to 195 ms, while vowels /u/ (175 ms) and /ə/ (202 ms) are the shortest and longest, respectively. We conducted further analysis to explore vowel duration differences concerning gender and context. The result, as displayed for visualization in Figure 6-9 and the duration values in Table 6.16, indicates that vowels produced in isolation exhibit longer durations compared to when they are produced within sentences or carrier frames. The duration is short in sentence and carrier frames because as the vowels are embedded in sentences, the contextual effect (carryover or anticipatory) or the rate of the speech will reduce the vowel duration. On the other hand, vowels produced in isolation are longer as there are no anticipatory effects. The duration of the vowels is gender-specific too, as we see in Figure 6-9, the female speakers have higher duration in both sentence and carrier frames. In the isolation frame, the /ə/ is relatively higher in females than males.

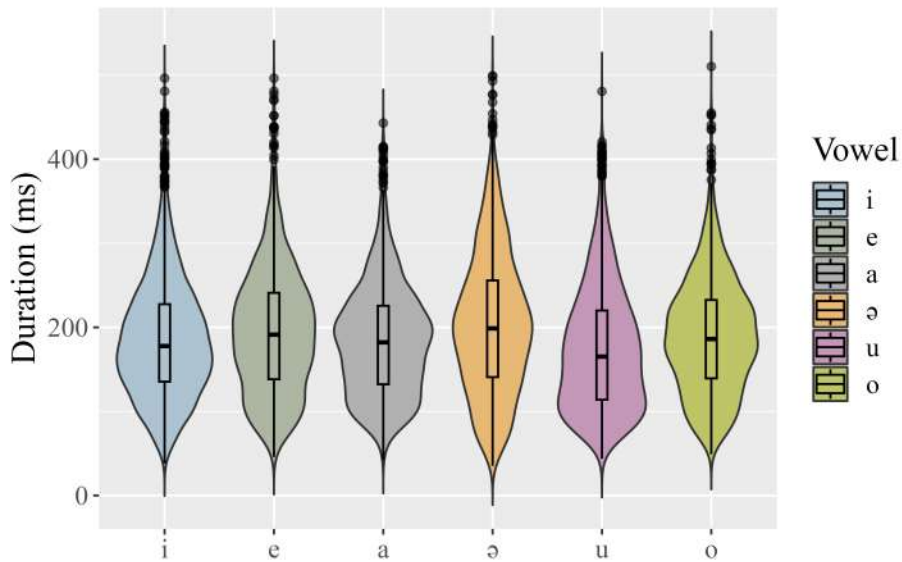


Figure 6-8: Violin plot of the average duration of the six Angami vowels.

Table 6.15: Average vowel duration and standard deviation in six Angami vowels.

Vowels	Mean	SD
a	184.2	66.0
e	194.0	74.1
ə	202.3	83.2
i	186.0	71.0
o	190.0	69.0
u	175.0	75.0

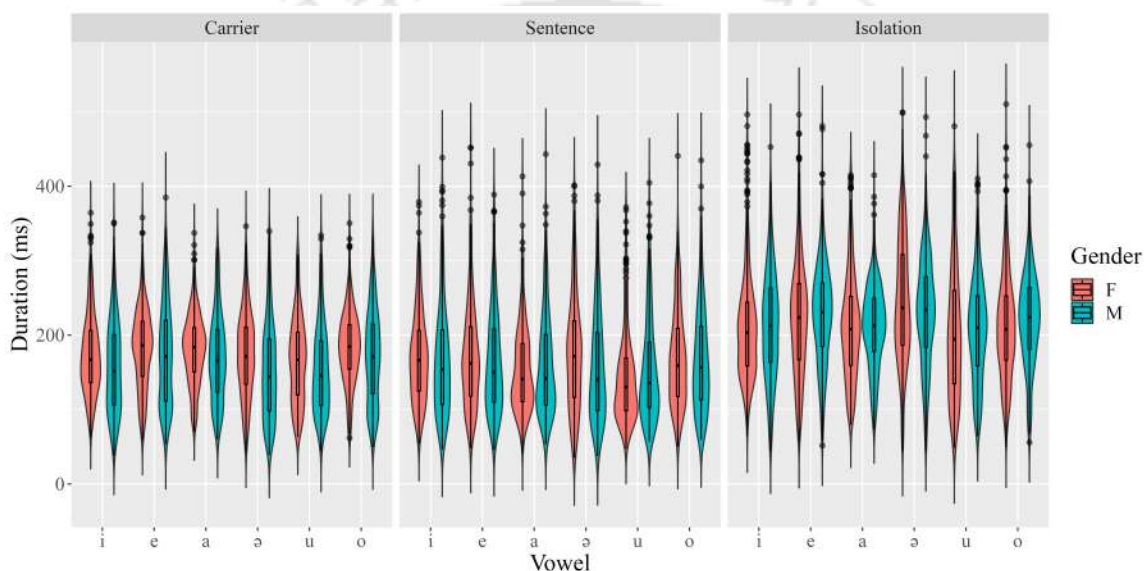


Figure 6-9: Violin plots of the average duration of the six Angami vowels by contexts and gender

Table 6.16: Average vowel duration and standard deviation in six Angami vowels by three context and gender.

Vowels	Male			Female		
	Carrier	Phrasal	Isolation	Carrier	Phrasal	Isolation
a	Mean (SD)	Mean (SD)	Mean (SD)	Mean (SD)	Mean (SD)	Mean (SD)
e	154.5 (59)	181.8 (47)	212.5 (71)	172.5 (71)	169.7 (59)	212.3 (57)
ə	172.5 (71)	182.2 (54)	223.3 (79)	163.2 (69)	174.0 (68)	228.4 (69)
i	175.0 (76)	174.0 (57)	246.0 (84)	158.0 (76)	153.0 (66)	230.2 (73)
o	171.1 (61)	174.5 (55)	211.9 (72)	166.1 (75)	157.6 (62)	212.4 (73)
u	169.1 (65)	184.0 (49)	212.3 (73)	167.2 (71)	172.0 (63)	221.0 (64)
u	144.2 (63)	165.7 (56)	203.2 (89)	154.0 (71)	155.0 (59)	207.0 (70)

Table 6.17: Results of an ANOVA conducted on an LME model with vowel duration as the dependent variable and vowel, context, gender, tone, and their interactions as fixed effects. Only statistically significant interactions are shown in this table.

	χ^2	(df)	p-value
vowel	99.1	5	< 0.001
context	1595.1	2	< 0.001
gender	1.3	1	0.24
tone	53.5	4	< 0.001
vowel*tone	2845.3	20	< 0.001
vowel*context	67.1	10	< 0.001
vowel*gender	36.5	5	< 0.001
vowel*context*tone	200.5	40	< 0.001
vowel*gender*tone	45.4	20	< 0.001
vowel*context*gender*tone	62.3	40	0.05
Conditional $R^2 = 0.47$, Marginal $R^2 = 0.17$			

An LME (Linear Mixed Effects) model was constructed to assess the significance of vowel duration within and across various contexts. This model utilized duration as the dependent variable, with vowels (V), gender (G), tone (T), context (C) as the fixed effects, speaker (Sp), syllable (Sy), structure (Strs), and code (Co) of the words as the random effects, as indicated in 6.2. The results revealed a significant effect on vowel duration resulting from the interaction of vowels, context, gender, and tone ($\chi^2(40) = 62.3$, $p > 0.05$), as shown in Table 6.17. A post-hoc Bonferroni test was conducted on the model where duration was the dependent variable and vowels served as the fixed effects for further examination of pairwise differences in duration across vowels. The results are shown in Table 6.18. Notably, the vowel duration of /a/ significantly differs from that of /u/; /e/ significantly differs from /ə, o, u/; /ə/ significantly differs from /u/; /i/ significantly differs from /o, u/; and /o/ significantly differs from /u/. These findings provide insights into the subtle differences in vowel duration across different vowel categories.

$$Duration \sim V + C + G + T + V * C * G * T + (1|Sp) + (1|Sy) + (1|Str) + (code) \quad (6.2)$$

Table 6.18: Results of a Bonferroni post hoc test conducted on an LME model with vowel duration as the dependent variable and vowel as fixed factors (df=Inf).

Contrast	Estimate	SE	z-ratio	p-value
a-e	-6.4577	2.35	-2.75	0.0894
a-ə	-0.0771	2.34	-0.03	1.0000
a-i	-5.4517	2.36	-2.31	0.3135
a-o	3.8710	2.46	1.57	1.0000
a-u	13.6765	2.51	5.44	< 0.0001
e-ə	6.3806	2.20	2.90	0.0556
e-i	1.0060	2.14	0.47	1.0000
e-o	10.3287	2.39	4.31	0.001
e-u	20.1342	2.51	8.03	< 0.0001
ə-i	-5.3746	2.08	-2.58	0.1473
ə-o	3.9481	2.35	1.67	1.0000
ə-u	13.7536	2.40	5.72	< 0.0001
i-o	9.3227	2.31	4.03	0.001
i-u	19.1282	2.52	7.57	< .0001
o-u	9.8055	2.54	3.86	0.001

6.3.2 Acoustic Features of Diphthong Vowel in Angami

As discussed in Section 6.1.2, there are two diphthongs in Angami, namely, /ie/ and /uo/. The discrete cosine transformed (DCT) of F1 and F2 formants was conducted to examine the diphthong characteristics. In contrast to the approach used for monophthongs, where vowel data is plotted from the midpoint of the vowel, the analysis of diphthongs involved extracting F1 and F2 values at every 5% interval of the total duration of the diphthong contour. This method was employed to capture the entirety of the diphthong glides. Additionally, to minimize contextual effects, both the onset and offset (10% each) were excluded from the plotting process. To visually represent the location of the diphthongs within the vowel space, the F1 and F2 values of the vowels were plotted, overlaid by 80% of the F1 and F2 values of the diphthong vowels, as depicted in Figure 6-10. The figure illustrates that the onset of the diphthong /ie/ occupies the vowel space typically associated with the vowel /i/, while its offset aligns with the vowel /e/. Similarly, the onset of the diphthong /uo/ corres-

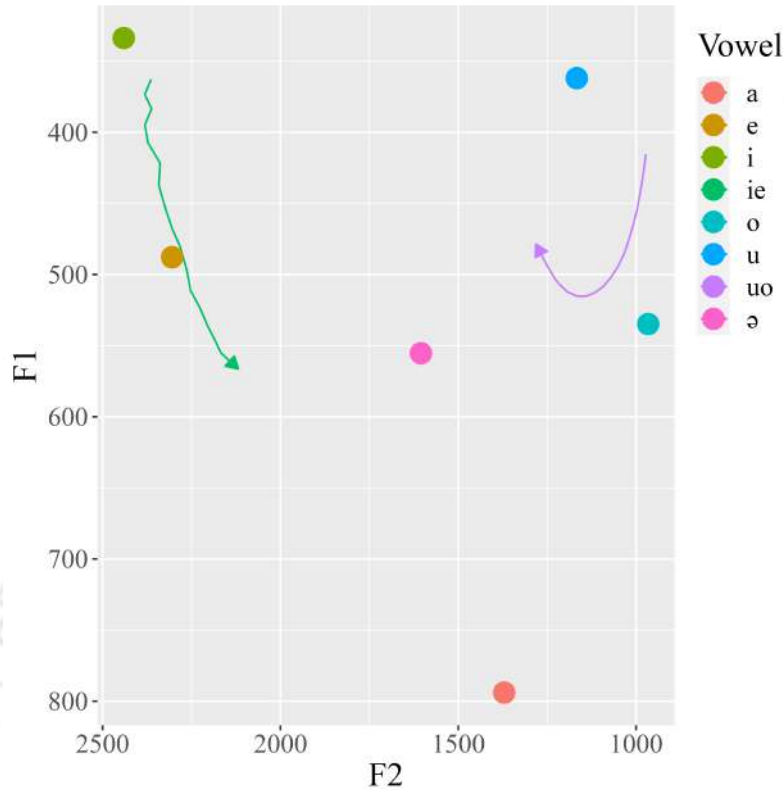


Figure 6-10: The formant values of F1 and F2 for monophthong and diphthong showing the space diphthong vowel occupies.

ponds to the vowel space of /u/, with its offset aligning with the vowel /o/. The figure effectively highlights the trajectory of the diphthongs within the vowel space. Table 6.19 displays the mean values and standard deviations (SD) of F1 and F2 at 30% and 70% of the two diphthong vowels in Angami. Diphthong calculated from the mid-80% is plotted for visualization in Figure 6-11. The variation in diphthong production by male and female speakers was further examined by plotting the DCT values of F1 and F2 in Figure 6-12. The mean values are summarized in Table 6.20. As observed, male speakers tend to have lower F1 values compared to female speakers at 30% and 70% of the total duration for both diphthongs. Conversely, in both diphthongs, the F2 values of female speakers are higher than those of male speakers. These findings underscore the gender-related differences in diphthong production within the Angami language, particularly in terms of F1 and F2 characteristics.

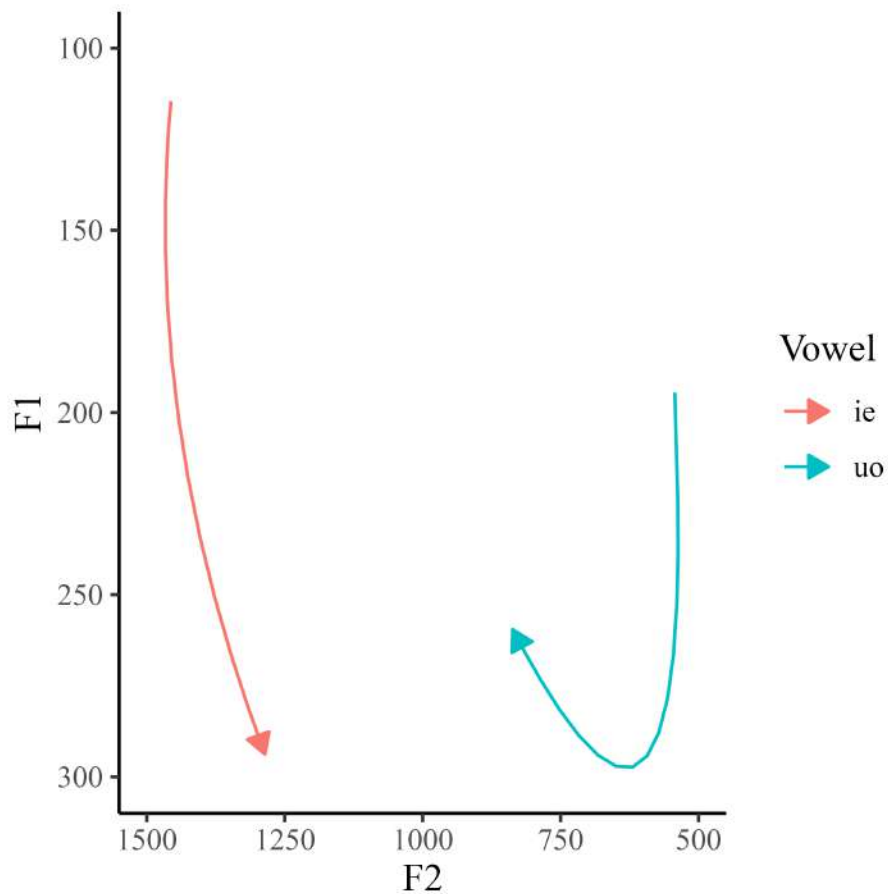


Figure 6-11: Inverse DCT contours of F1 and F2 calculated from the mid-80% of the total duration of the two diphthongs in Angami.

To analyze the differences in formant values across the diphthongs, we constructed six Linear Mixed Effects (LME) models. These models were built with C0, C1, and C2 as dependent variables for both F1 and F2. Vowel, context, gender, and their interactions were the fixed effects, while speaker and unique words were the random effects. The analysis results as seen in Table 6.21, Table 6.22 and Table 6.23, indicates that both vowel type and gender significantly affect C0 ($\chi^2 = 31.2$, $p < 0.001$), C1 ($\chi^2 = 33.0$, $p < 0.001$), and C2 ($\chi^2 = 33.2$, $p < 0.001$) for F1 as well as in C0 ($\chi^2 = 71.0$, $p < 0.001$) and C1 ($\chi^2 = 15.1$, $p < 0.001$) for F2 formants. Furthermore, the distinctiveness of the diphthongs is evident from the analysis, indicating that they exhibit unique formant patterns that differentiate them from one another.

Table 6.19: The mean values and standard deviation (SD) of F1 and F2 of the two diphthong vowels in Angami.

Vowels	30%		70%	
	F1	F2	F1	F2
ie	407 (51)	2371 (342)	523 (100)	2224 (312)
uo	464.5 (59)	1016 (142)	518 (93)	1169 (251)

Table 6.20: The mean values and standard deviation (SD) of F1 and F2 of the two diphthong vowels in Angami as produced by Male and Female speakers.

	Vowels	30%		70%	
		F1	F2	F1	F2
Male	ie	381 (36)	2155 (138)	454 (64)	2044 (165)
	uo	431 (51)	942 (101)	458 (54)	1060 (185)
Female	ie	425 (53)	2516 (362)	570 (93)	2346 (330)
	uo	483 (55)	1057 (145)	552 (94)	1230 (262)

Table 6.21: Results of an ANOVA conducted on an LME model with C0 (F1 & F2) as the dependent variable and vowel, context, gender, tone, and their interactions as fixed effects. Only statistically significant interactions are shown in this table.

	(df)	F1		F2	
		χ^2	<i>p</i> -value	χ^2	<i>p</i> -value
vowel	1	3.1	0.1	1250.3	< 0.001
context	2	181.6	< 0.001	2.3	0.31
gender	1	26.8	< 0.001	37.3	< 0.001
vowel*context	2	56.5	< 0.001	0.0	0.99
vowel*gender	1	31.2	< 0.001	71.9	< 0.001

Table 6.22: Results of an ANOVA conducted on an LME model with C1 (F1 & F2) as the dependent variable and vowel, context, gender, tone, and their interactions as fixed effects. Only statistically significant interactions are shown in this table.

	(df)	F1		F2	
		χ^2	<i>p</i> -value	χ^2	<i>p</i> -value
vowel	1	81.6	< 0.001	48.0	< 0.001
context	2	116.9	< 0.001	20.4	< 0.001
gender	1	30.0	0.53	0.3	0.53
vowel*context	2	27.3	< 0.001	25.2	< 0.001
vowel*gender	1	33.0	< 0.001	15.1	< 0.001

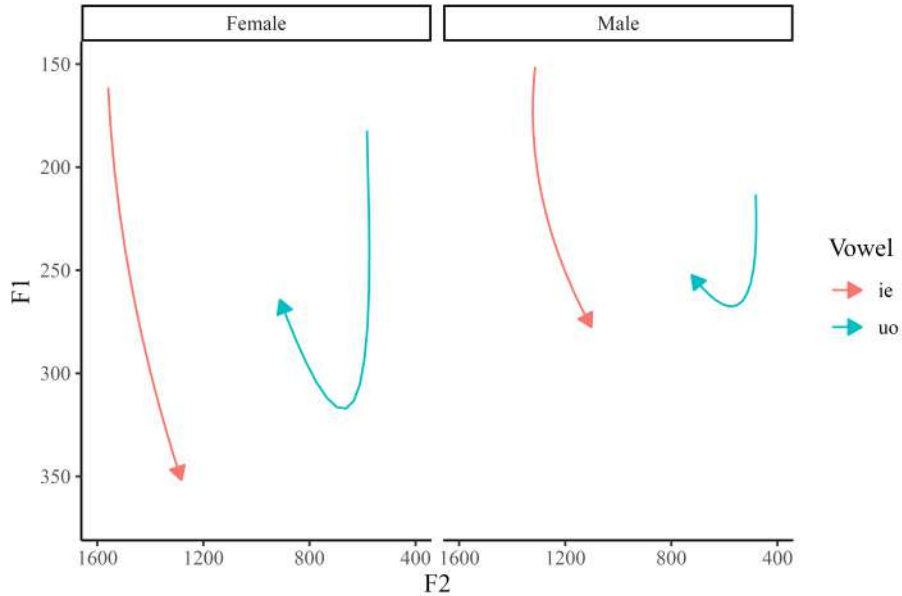


Figure 6-12: Inverse DCT contours of F1 and F2 calculated from the mid 80% of the total duration of the two diphthongs for male and female Angami speakers.

Table 6.23: Results of an ANOVA conducted on an LME model with C2 (F1 & F2) as the dependent variable and vowel, context, gender, tone, and their interactions as fixed effects. Only statistically significant interactions are shown in this table.

	(df)	F1		F2	
		χ^2	<i>p</i> -value	χ^2	<i>p</i> -value
vowel	1	93.3	< 0.001	24.5	< 0.001
context	2	15.6	< 0.001	98.4	0.52
vowel*context	2	5.4	0.06	46.2	< 0.001
vowel*gender	1	33.2	< 0.001	2.9	0.09

6.3.2.1 Diphthong Vowel duration

The average duration of the diphthongs for both male and female speakers is visually presented in Figure 6-13, with corresponding mean values detailed in Table 6.24. As depicted in the figure, the average duration of the /ie/ diphthong is longer than that of the /uo/ diphthong. Additionally, for both diphthongs, the duration tends to be relatively longer in female speakers than in male speakers. Further diphthong durations were further analyzed based on three contexts and two genders. As illustrated in Figure 6-14 and detailed in Table 6.25, diphthongs produced in isolation exhibit

Table 6.24: Average duration of the diphthongs in Angami.

Vowel	Mean (SD)
ie	209.4 (59)
uo	163.4 (82)

longer durations compared to those produced in carrier and sentence frames. Notably, except for the /ie/ diphthongs produced in an isolation context, diphthongs articulated by female speakers tend to have longer durations across all contexts. These findings about the temporal characteristics of diphthong production in the Angami show that context and speaker gender play an important role.

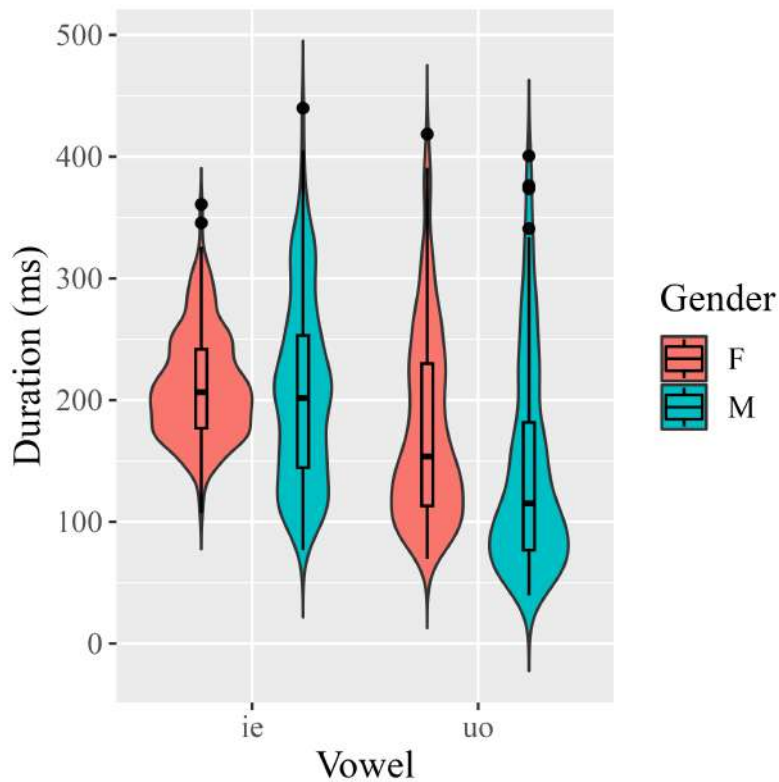


Figure 6-13: Violin plots of the average diphthongs duration across male and female speakers.

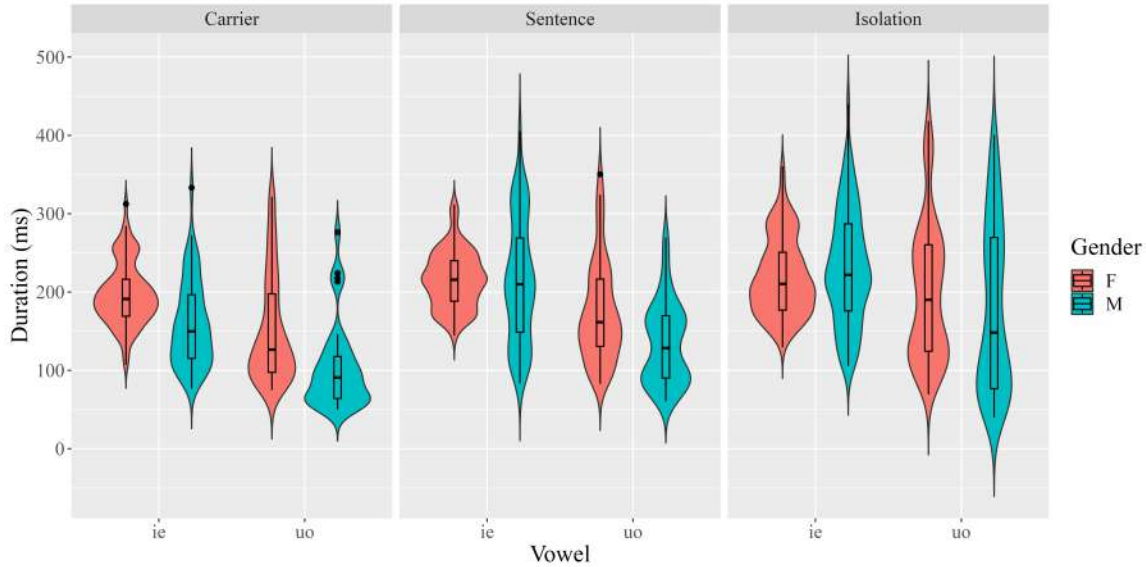


Figure 6-14: Violin plots of the average diphthongs duration by contexts and gender.

Table 6.25: Average duration of the two diphthongs with SD in three contextual frames divided by speakers in Angami.

Vowels	Female			Male		
	Carrier	Phrasal	Isolation	Carrier	Phrasal	Isolation
	Mean (SD)	Mean (SD)	Mean (SD)	Mean (SD)	Mean (SD)	Mean (SD)
ie	216.0 (36)	197.5 (40)	219.9 (49)	214.7 (79)	159.2 (55)	231.5 (72)
uo	174.2 (64)	151.3 (68)	199.2 (87)	134.2 (52)	103.5 (53)	178.5 (106)

6.4 Discussion

In this chapter, we examined the vowel system of Angami, encompassing both monophthongs and diphthongs. Our analysis confirms the presence of six phonemic monophthongs and two phonemic diphthongs in the language. Cross-linguistically, the most common vowel systems languages exhibit are inventories containing 5 to 6 distinct vowels. This observation is further supported by the data from the UPSID database (317 entries) and subsequent studies on vowels in the World Atlas of Language Structures (WALs) (564 entries). In both these databases, languages most commonly reported 5 to 6 vowels (Maddieson, 1984b, 2013c).

This holds true for several Naga languages in particular and for Tibeto-Burman

languages in general. For example, In case of other Naga languages, such as, Chokri (Neinu, 1990), Khezha (Yabu, 1994), Mao (Marrison, 1967), Lotha (Acharya, 1975), Sema (Teo et al., 2014), and Ao (Coupe et al., 2003), it becomes apparent that Naga languages commonly have vowel systems consisting of 6 to 7 vowels on average. Notably, Yabu (1994) documented that Khezha specifically features a vowel inventory encompassing seven vowels. In case of Angami, contrary to Kuolie (2006) and Suokhrie (2015), this study found only a six vowel system. The vowels /ü/ and /ə/ as reported in Kuolie (2006) and Suokhrie (2015), did not exhibit any distinctive spectral or temporal characteristics. Hence, in the current study, these two vowels were considered as belonging to the same category. Nevertheless, it was also noticed that the vowel /ə/ can be used as affix⁶ where they acquire substantially longer duration.

Furthermore, in addition to the number of vowels, cross-linguistically, a preference is also noticed for specific vowel qualities. Among the languages, that have five vowel systems, there is a preference for the inventory of /i, e, a, o, u/. Similarly for six vowel systems, there is an overwhelming preference for /i, e, a, o, u, ə/ (Schwartz et al., 1997b; Recasens, 2022b). In the UPSID database, among 54 languages that have six vowel systems, 26 follow the inventory with schwa in it. Only 12 out of 54 languages prefer having a /ü/ instead of a /ə/ (Maddieson, 2013c). In case of NE Indian TB languages, it is noticed that the sixth vowel is described with various symbols and descriptions in the literature. Positional restrictions and poor orthographic representation has always marred the true identity of the sixth vowel in the TB languages (Sarmah et al., 2015a; Burling, 2013). Even in the current study it was shown that some of the discrepancies in the description of segments in Angami (also in several other TB languages of NE India) arise due to the inadequate orthography adopted for the language.

⁶For example, as a prefix, /ə-tsali/ for ‘folk-song’; as a suffix /dzə-ə/ ‘water-draw’.

Chapter 7

Acoustics of suprasegmentals in Angami

This chapter examines the contrastive tones in Angami. It addresses issues related to tone description and inventory size. The chapter is divided into two parts. First, the production of lexical tones produced in three contexts is investigated. The findings of this study reveal the existence of four contrastive tones, which depend on relative F_0 height, and the tones in Angami are level. Secondly, the interaction of tones and segments was explored. The CF_0 effect is observed where the extent of consonantal perturbation depended on the tonal categories. Voiceless consonantal perturbations were less as tone categories got higher. Conversely, voiced stop consonantal perturbations were less as tone categories got lower. This study also found a systematic relationship between tone height and VOT duration. High tones induced shorter VOT, while low tones induced longer VOT. The study further explores the intrinsic F_0 , investigating the positive interaction between vowels and tone. Finally, the contextual effects of tones in disyllables are investigated by determining how the tones in the second syllable affect the tones in the first syllable and vice versa. The findings indicate that

*both anticipatory and carryover effects depend on the height of the tone.*¹

7.1 Introduction and literature review

Tone is a linguistic feature of the language that can be perceived through pitch patterns. A language is considered tonal if pitch differences exist in the same word with different lexical meanings. The vocal fold is the primary source of pitch differences, and the pitch differences occur due to adjusting the mass and stiffness of the vocal cord (Yip, 2002; Hirose, 1997). Hyman (2006) states, ‘A language with tone is one in which an indication of pitch enters into the lexical realization of at least some morphemes.’ Thus, Tone refers to a phonological category that distinguishes two words or utterances and is relevant for languages in which pitch plays some linguistic roles. According to Yip (2002), the fundamental frequency (F_0) of a sound, which we perceive as pitch, is primarily determined by the frequency of vibration of the vocal folds inside the larynx. F_0 is an acoustic term that refers to the amount of pulse within a single vibration of the vocal fold. The frequency of these pulses is measured in Hertz (Hz), where one Hertz is one cycle per second.

The two major types of tones are register tones (also known as level tones) and contour tones. According to Maddieson (1978a), a register tone is ‘one for which a level pitch is an acceptable variant.’ In register tone systems, a syllable has high, mid, or low pitch contour tones; on the other hand, they are the combination of levels because, in contour tones, the syllable tone glides from one pitch to another, high to low, mid to high, etc. Yip (2002) claims that contour tones are added to tonal inventories if and only when there are a large number of tonal contrast. If the

¹Part of this chapter has been published - “Vowel-Tone Interaction in Two Tibeto-Burman Languages”. In INTERSPEECH (pp. 3970-3974), 2019. doi: 10.21437/Interspeech.2019-2808; “Tonal coarticulation in Angami level tones”. Proceedings of the Second International Conference on Tone and Intonation, 2023; “Voice Onset Time in Angami”, in Linguistics of the Tibeto-Burman Area, volume 47, issue 1 (2024). In press for publication in June 2024. John Benjamins Publishing Company.

language has two contrastive tones, it will likely be register in nature and not contour tones. Counter tones are often found only on the word-final syllable because the word-syllable is often lengthened cross-linguistically (Leben, 1973; Goldsmith, 1976). Tones that fall and rise (concave) or tones that rise and then fall (convex) are common in Asian languages. Convex tones are less common as compared to concave tones. The accentual tonal system is one where tonal contrast is reduced to a minimum and the presence and absence of high (or low) on a particular syllable. There is a one-way contrast between accentual tones in Japanese and Bantu languages. Accentual languages are those that have a small number of contrastive tones. Languages like Japanese, Serbo-Croatian, and some type of Dutch are considered as accentual tones. In Europe, most of the languages considered tonal are accentual. For example, Serbo-Croatian and Lithuanian are tonal (Zec, 1999; Kenstowicz, 1971).

In Southeast Asia, where the dominant language groupings are Sino-Tibetan and Mon-Khmer families, tonal languages are abundant. Japanese languages are accentual, as mentioned earlier. In contrast, Koreans do not have lexical pitch differences in the form of tones, but pitch differences are determined by laryngeal properties (voicing, aspiration, and glottalization). In India, a few Indo-Aryan languages like Punjabi (Bahl, 1957; Bhatia, 1957, 1993) and Kalam Kohistani (Baart, 1997) are tonal, else majority are non-tonal. However, the Indian Northeastern regions where the occupants belong to the Tibeto-Burman group have rich inventories of Tonal. There are some tonal languages where it is more systematic in assigning tones. In a language like Changki, a bisyllabic word can be first low, then high, or the reverse and the meaning changes completely. However, for some languages, as in Chiziguála (Bantu language), the change of tones depends on the morphology of the complex word where the tone is marked on the penultimate syllable only (Yip, 2002; Coupe et al., 2003).

In World Atlas of Language Structures Online (WALS), a resource curated by

linguists at the Max Planck Institute for Evolutionary Anthropology, Maddieson (2013a), reported that there are three hundred and seven which are non-tonal languages ('no tone'), one hundred and thirty-two 'simple tone' languages, and eighty-eight languages with 'complex tone' systems. Interestingly, complex tone systems are predominantly found in Southeast Asia, Sub-Saharan Africa, and a few North and South American languages. Conversely, continents such as Europe and Australia tend to need tonal distinctions. Sub-Saharan Africa is particularly notable for hosting the largest concentration of tonal languages globally (Yip, 2002).

7.1.1 Tone notation

There are various methods by which tones are marked. Traditionally, there are three tonal notation styles, namely, Africanists who use a set of accent marks to convey tone, Asians especially the Chinese, who use a numerical system known as the 'Chao tone letters' (Chao, 1930), and Americans use numerical for tone marking as well. The difference between the Chinese and American way of tone notation is that for Chinese, in level high tones, the number is superscript fifty-five, whereas for the Americans level high tones, the number is superscript one; the digits are reversed here.

7.1.2 The physiology of tone production

In order to understand tone, we look at the physiological factors contributing to the Production of tone. In the Production of tone, Hombert et al. (1979a) gives a detailed description of the anatomy involved in producing tones. The larynx in the neck is a valve for controlling airflow into and out of the lungs and a source of sound Production. It comprises four principal cartilages: the thyroid, the cricoid, and a pair of arytenoids. Ligaments and muscles interconnect these cartilages, including the adductor, abductor, tensor, strap, constrictor, and other pharyngeal muscles. At the core of the larynx lies the vocal fold, which consists of ligaments that stretch

from the inner lower front surface of the thyroid cartilage to the front faces of the separate arytenoid cartilages. Rotation of the arytenoid cartilages allows for adjusting vocal fold positioning, bringing them closer together for voicing or breath-holding (adduction) or separating them (abduction), influencing the quality and pitch of the produced tones.

In tone production, the adductor muscles bring the vocal folds together, creating a narrow opening through which air from the lungs is forced. Combined with the Bernoulli effect, this action generates a suction force that draws the vocal folds together. The vibration of the vocal folds occurs when pressure from the lungs builds up behind the closure and eventually bursts through, releasing a puff of air and reducing the sub-glottal pressure. This cycle repeats continuously, with each burst of air representing one cycle of vocal fold vibration. In typical speech, the cycle of vocal fold vibration can range from as low as eighty times per second to as high as four hundred per second for a female voice (Yip, 2002).

7.1.3 Theories on the vocal-fold tension (pitch perturbation)

7.1.3.1 Aerodynamic hypothesis

During voiced stops, oral pressure gradually builds up, thus decreasing the pressure drop across the vocal cords, which in turn decreases the F_0 upon release of the stop, producing an initial low and rising F_0 contour after voiced stops (Ladefoged, 1967). Contrary to the study of Aerodynamic study, Ohala (1974) presents subglottal pressure data and shows that glottal airflow and subglottal pressure may be less at vowel onset after voiceless aspirated stop than voiced stops.

7.1.3.2 Horizontal and vertical vocal fold tension (theories)

Halle and Stevens (1971) propounded the theory of Horizontal tension of the vocal folds. According to the authors, consonants produced with stiff vocal folds are voice-

less in stop consonants. Since the vocal folds are stiff, the following vowel is produced with a raised pitch. The authors further state that if vocal folds are slackened, they vibrate more easily, making it possible for voicing to continue. Because of the slackness of the vocal fold, the following vowel has a lower pitch. Voiceless obstruents and high pitch, as well as voiced obstruents and low pitch, are widely attested in natural languages. Dixit (1975) supports this theory as he found a consistently higher activity of the cricothyroid muscle during the production of Hindi voiceless stops.

Ewan and Krones (1974) proposes another hypothesis known as vertical vocal fold tension, where the author states that larynx height is slightly higher for voiceless than voiced stops. Vocal fold tension varies with larynx height; the different pitch on vowels following the different obstruents is accounted for. The lower larynx during voiced stops is a consequence of the vocal tract's need to expand to accommodate more air in the oral cavity to maintain the positive transglottal pressure drop necessary for voicing. Ewan and Krones (1974) found the differences in larynx height between voiced and voiceless stops to be greater at stop release than at stop onset. These differences in larynx height persist for some time ($> 100\text{ms}$) into the following vowel.

The study of phonological processes in which the language does not have tone features but acquires them in the later stage is known as Tonogenesis, a term coined by Matisoff (1970, 1973). French botanist and orientalist turn linguist Haudricourt (1954) is considered the first to study the process and development of tones in Vietnam. According to Haudricourt, three-syllable types gave rise to tones; all three had distinctive voicing in initial stop consonants.

- Open syllables (those ending in a vowel or nasal)
- Syllables formerly checked with a final voiceless spirant that had become /h/
- Syllables formerly checked with some stop, symbolized by /x/ that had become a glottal stop.

One of the best-documented works on the emergence of tone in language has been reported by Hombert et al. (1979b) in the Phonetic explanation for the development of tones. The study throws light on the development of contrastive tones on vowels because of the loss of a voicing distinction on obstruents in the prevocalic position and the relative lowering of tone or relatively high tone after a previously voiced series or voiceless series, respectively. In short, voiceless stops activate high tone, and voiced stops activate low tone. This effect of F_0 variation (pitch) after voiced and voiceless stops in tonal language is attested in Yoruba, a tonal language (Hombert, 1975, 1976) and non-tonal language English. Findings in this study suggest that the duration of perturbation caused by prevocalic consonants on the F_0 of vowels is shorter in Yoruba (40ms to 60ms) than in English (>100ms). To supplement the effect, Gandour (1974) from his investigation of Thai tones also reported that a shorter part of the vowel was affected by the preceding consonant (30ms for voiceless consonants and 50ms for voiced). The development of tone on vowels after voiced and voiceless obstruents can be explained by reference to articulatory and auditory facts, especially voicing distinction in prevocalic position cause small perturbations that are perceptible and whose physiological origin can be accounted for, as the position of the larynx position is higher for voiceless stops (Dixit, 1975).

Besides the effect of loss of voicing distinction for the development of contrastive tones, there are other consonant types in tone development as well. In the Punjabi language, breathy-voiced consonants become voiceless unaspirated, leaving a low tone on the following vowel (Hombert et al., 1979b; Gill and Gleason, 1969; Haudricourt, 1972). In Tibeto-Burman languages, Glover (1970) reports that breathy voiced consonants are stronger F_0 depressors than (simple) voiced obstruents. One explanation for such effect is given by Ohala and Ohala (1972) where it states that a breathy voice involves less forceful contraction of the laryngeal adductor muscles, thus bringing the vocal fold together and regulating the F_0 . In a speech, some lexical tones may change

their tonal category in tonal context than in isolation. In Mandarin, the low tone changes into a rising tone when followed by another low tone. When such cases occur, the derived rising tone is perceptually indistinguishable from the lexical rising tone. This categorical shift due to tonal context is usually called ‘tone sandhi’ (Xu, 1994b). In the Mizo language, too, tone sandhi is reported, where the rising tone changes to a low tone when followed by a high or falling tone (Sarma et al., 2015).

Studies conducted under the Max Planck Institute suggest a correlation between complex tone systems and regions with high humidity, contrasting with simpler tone systems prevalent in drier areas (Moran and McCloy, 2019). As discussed earlier regarding tone production, the larynx plays a pivotal role, and the researchers propose that increased humidity affects the vocal folds by maintaining moisture in the mucous membranes, rendering them more elastic. This elasticity facilitates adequate oscillation of the vocal folds, resulting in the production of distinct tones.

However, one limitation highlighted in the article and website is the need for more representation of tonal languages from India. The map of tone languages includes only Boro, Nyishi, Meithei, and Ao, which fail to capture the region’s linguistic diversity adequately. India boasts a rich tapestry of languages, with more than a hundred spoken across various communities. Consequently, drawing generalized conclusions based on a limited sample size could lead to inaccuracies and overlook the nuances present in the linguistic landscape of the region. Hence, the study of Angami tones aims to bridge the gap in the understanding of tones in Northeast of India and in general, the Tibeto-language family.

7.2 Literature Review of Angami tones

Angami is a tone language, however, not much is known about the acoustic characteristics of the tones in the language. Moreover, previous studies have differing

views on the types and number of tones in the language. Hence, in this chapter, we intend to provide a basic acoustic characterization of the Angami tones. While doing that, we also provide a definitive answer regarding the types and number of tones in the language. In the existing literature on the Angami language, not many studies have focused on the phonetics and phonology of tones in the language. Whatever works exist in the domain of Angami tones, their descriptions are neither exhaustive nor uniform. For example, the number of tones in Angami has been reported to be either four or five in the literature (Burling, 1960; Giridhar, 1980; Ravindran, 1974; Kuolie, 2006; Chase, 1992; Suokhrie, 2017; Meyase, 2021). Apart from the number of tones, studies have also not agreed on the types of tones in the language. For example, while a group of authors has reported Angami to have both contour and level tones (Burling, 1960; Giridhar, 1980; Ravindran, 1974; Kuolie, 2006; Liezietsu et al., 2019), others have settled with only level tones in the language (Chase, 1992; Suokhrie, 2017; Lalhminghlui et al., 2019; Meyase, 2021). In a standard dictionary of Angami named the *MKS Dieda*, the entries are associated with five contrastive level and contour tones (Liezietsu et al., 2019). The example of tone minimal set as provided in the dictionary is shown in Table 7.1. The number and tones reported in the previous studies are summarized in Table 7.2.

Table 7.1: Example of tonal minimal set in Angami (Liezietsu et al., 2019)

Word	Tone	Meaning
/pe/	T1	incline
	T2	bridge
	T3	march
	T4	shiver
	T5	shoot

Angami's number and types of tones are of interest, as it is comparatively rare to have languages that contain four phonological tones. Angami becomes typologically interesting, as it may be one of the few languages with four contrastive level tones in its inventory. In a cross-linguistic study of over 200 tone languages, it is reported

Table 7.2: Tones described in Angami by various authors

Authors	No.	Types	Descriptions
Burling (1960)	5	Contour & Level	low falling, low, mid resonant, mid normal & high
Ravindran (1974)	5	Contour & Level	low, low falling, mid, mid rising & high
Giridhar (1980)	5	Contour & Level	low, low-fall, mid, mid-fall & high
Chase (1992)	4	Level	low, mid low, mid high & high
Kuolie (2006)	5	Contour & Level	low, low-high, mid, high-low & high
Liezietsu et al. (2019)	5	Contour & Level	high, unmarked, mid, low-high & low
Suokhrie (2017)	4	Level	t1, t2, t3 & t4 (t1 : low, t4 : high)
Meyase (2021)	4	Level	low, mid, high & extra high

that the pitch contrast or multi-level tone contrast varies from two to five-level tones, although four to five-level tones are rare in the world’s languages (Maddieson, 1978b).

To determine the number and types of tones in Angami, several acoustic-phonetic studies have been carried out. Section 7.3 examines the analysis of lexical tones in Angami, meanwhile Section 7.4 investigates the interaction of tones and segments.

7.3 Tones in Angami

One of the primary objective of this study is to investigate the number and types of tones in Angami. As seen in the previous section, there are discrepancies in the description of tones. This study aims at understanding the Angami lexical tones.

7.3.1 Methodology

This section discusses the methodology for the acoustic analysis of tones in Angami. The following subsections briefly describe the speech corpus, the F_0 extraction method, and statistical measurements.

7.3.1.1 Materials

Lexical tones in Angami were recorded for this study, comprising a list of 45 unique words (see Table 5). Additionally, a selection of speech data contrasting in tones

Table 7.3: Distribution of tone tokens in Angami.

Tones	tokens
T1	1333
T2	993
T3	1263
T4	1232
T5	1056
Total	5877

with CV structure where C includes nasals (Anl), approximants (Apx), fricatives and affricates (Aaf), and V includes / i, e, a, ə, u, o/ was collected. As discussed previously, for the collection of data, all possible five tones in Angami, as per MKS Dieda, the dictionary (Liezietsu et al., 2019). The tones are marked by T1-T5; here, T1 represents the high tone, T5 represents the low tone, whereas T2, T3 & T4 are the intermediate tones. The corresponding representation of tones in Chao’s tone numerals is 55 (T1), 44 (T2), 33 (T3), 22 (T4), and 11 (T5). This study included data from 24 native speakers (9 male, 15 female) of Angami as their first language; they also have English, Nagamese, and Hindi proficiency. The age range of the speakers during data collection was between 25 and 35 years. The dataset consists of 5877 tokens, with the distribution presented in Table 7.3. Data was recorded following the standard procedure described in Chapter 1 (Section 1.3.3).

7.3.1.2 Acoustic measurements

Tones were manually annotated and segmented at the syllable level with the help of visual pitch representation in Praat and by listening. An illustration of the tone boundary for the word /le/ across five tones, as produced by a female speaker, is shown in Figure 7-1. In this context, the word /le/ carries different meanings depending on the tone: T1 signifies ‘to peel,’ T2 signifies ‘really,’ T3 signifies ‘think,’ T4 signifies ‘hot,’ and T5 signifies ‘go in.’ The fundamental frequency (F_0) was extracted at the tone-bearing unit, which in this case was the syllable. The (F_0) values are extracted

from the beginning to the end of the TBU at every 2% of the point (0%, 2%,4%...). A Praat script was written to extract the values, and an interpolate function in Praat was used for any values that were “undefined”. Due to contextual effect, the nucleus (V) position could be affected by the onset (C), so only mid-80% of the total tone duration was considered for analyses.

7.3.1.3 Statistical measurements

The fundamental frequency underwent normalization to eliminate speaker effects using the z-score normalization method outlined in Chapter 1 (Section 1.3.6). Therefore, before conducting the fundamental frequency analysis, the data were subjected to normalization for speaker effects using the Lobanov method, employing the `normLobanov` function from the *phonR* package within the open-source R platform (R Core Team, 2022). Subsequently, qualitative and exploratory statistics were computed using R’s statistical packages, as detailed in Chapter 1 (Section 1.3.6).

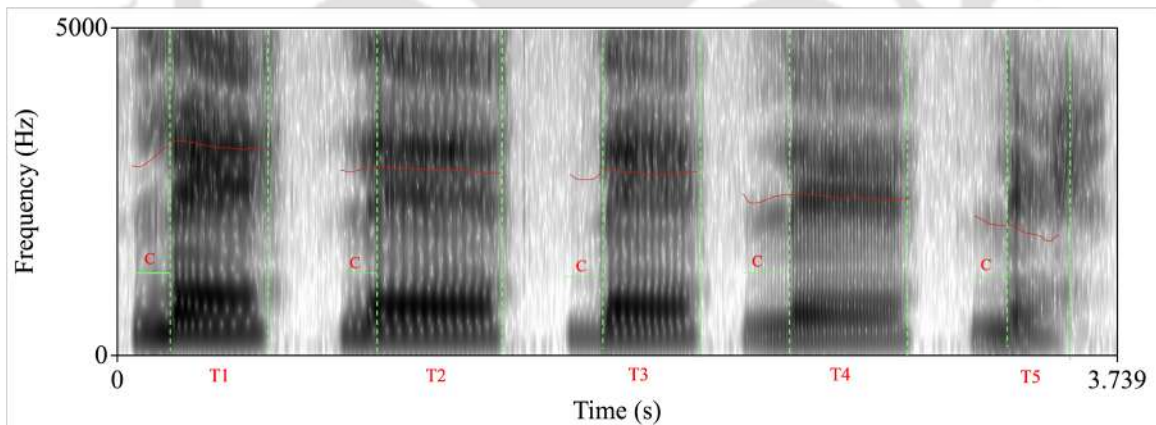


Figure 7-1: Spectrographs of the word ‘le’ across the five tones as produced by a female speaker.

7.3.2 Results

The following subsection discusses the result of the study, which includes the acoustic measurements as well as the temporal measurements. The results are further analyzed

using statistical tests.

7.3.2.1 Acoustic of Angami tones

To visualize the mean F_0 differences among the five Angami tones, density plots have been generated and plotted in Figure 7-2. The figure illustrates the F_0 range and the distribution of tones. Specifically, Tone T5, characterized as a low tone, occupies the lower F_0 range, while Tones T2, T3, and T4, intermediary tones, occupy the mid-range of F_0 . Tone T1, representing a high tone, occupies the higher F_0 range. For each tone, there are two peaks, which is likely because of the gender differences.

The result of the tones produced by male and female speaker is shown in Figure 7-3. The mid 80% of the total F_0 contour has been plotted. The result shows a huge difference between the male and female production of the tones. While the female speakers have a high F_0 range for all five tones, the male has a lower F_0 range. The F_0 range of the female speaker is around 150 Hz to 240 Hz, while the male F_0 range is about 120 Hz to 175 Hz. This figure further shows that the two intermediary tones, T2 and T3, overlap for both male and female speakers.

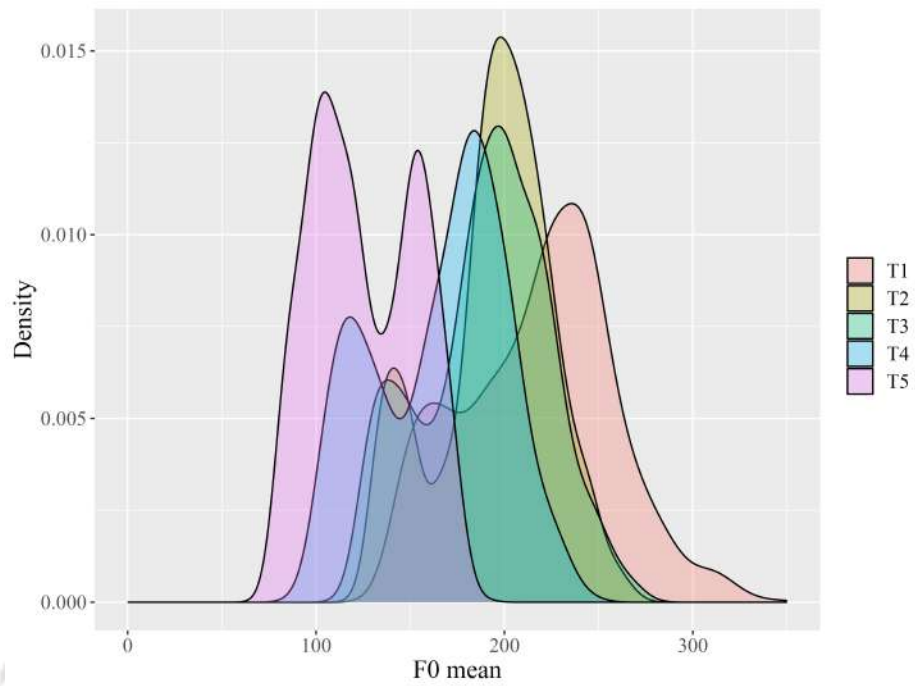


Figure 7-2: Kernel density estimation plots for the F_0 extracted from five Angami tones.

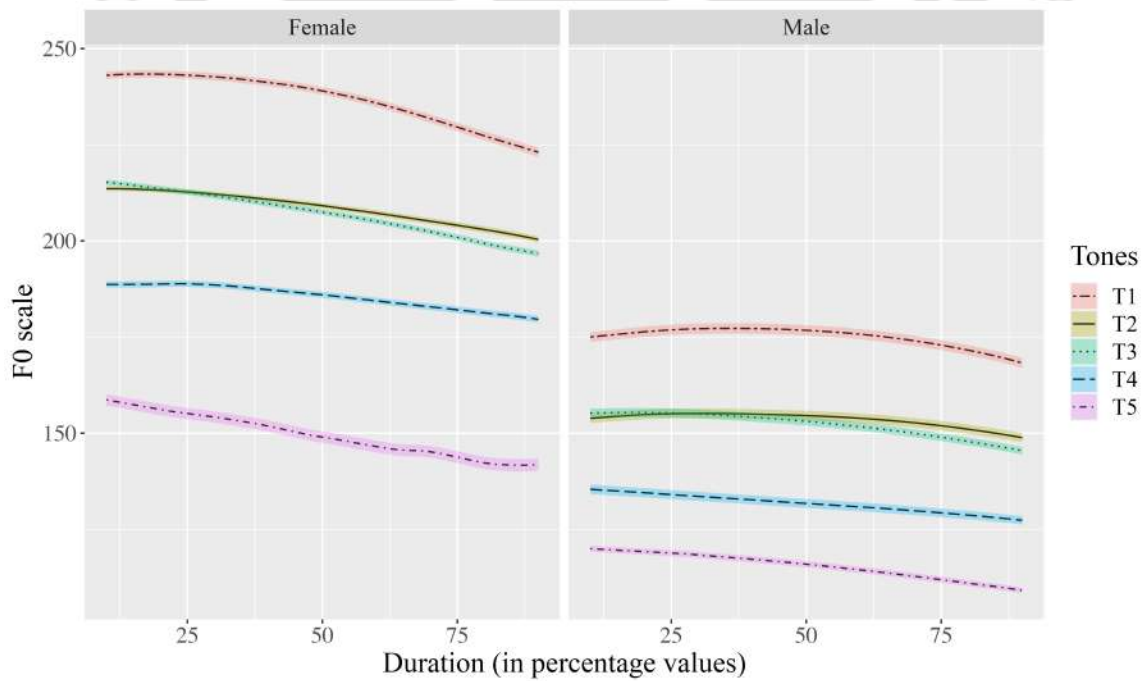


Figure 7-3: F_0 contour plot tones in Angami by male and female speakers.

Individual differences or gender differences could have an effect on the tones, so the F_0 values were normalized with the Lobanov method as discussed in Section 1.3.6. The average mid 80% of the total F_0 contour for the five tones is depicted in Figure 7-4. The top figure represents the average raw F_0 values, while the bottom figure illustrates the average normalized F_0 values. The F_0 scale indicates that T1 can reach as high as 220 Hz, T2 and T3 range between 200-180 Hz, T4 centers around 170 Hz, and T5 can descend as low as 100 Hz. Both figures display four distinct tones, with T2 and T3 being marginally separated. Moreover, the figure corroborates the observation that tones in Angami exhibit a level F_0 pattern.

Furthermore, the normalized F_0 of Angami tones were analyzed based on the three environments, i.e. carrier, isolation & sentence. As Figure 7-5 shows, tones differ based on the environment, where the tones produced in sentences are relatively higher than the tones produced in isolation or carrier frames. In the sentence frame, the target tones occur in a sentence, where the preceding or the following sound could be any of the consonants or vowels in Angami. Additionally, the tones are not controlled. Hence, the sentence frames have a greater chance of having a tonal co-articulatory effect. The onset of the syllable boundary or at the off could be affected. This effect is studied and discussed in detail in the following Section. Whereas, in carrier frames, there is less variability of the contextual effect as the same frame is used for all tones, i.e., / a x puba / 'I x said,' where the tone pattern is T3 x T2T3. The high tone T1 produced in isolation shows declination as it is produced after sentence and carrier and the speakers do not sustain the tones. Overall, the contextual effect on tones can be summarized as sentence > isolation > carrier.

To confirm the statistical significance of the fundamental frequencies (F_0) among the five tones, we constructed LME models. The dependent variables are the normalized values of either of the three mid- F_0 points (40%, 50% & 60%). Tones (T), three contexts (C), and their interactions are the fixed effects while, vowel (V), the unique

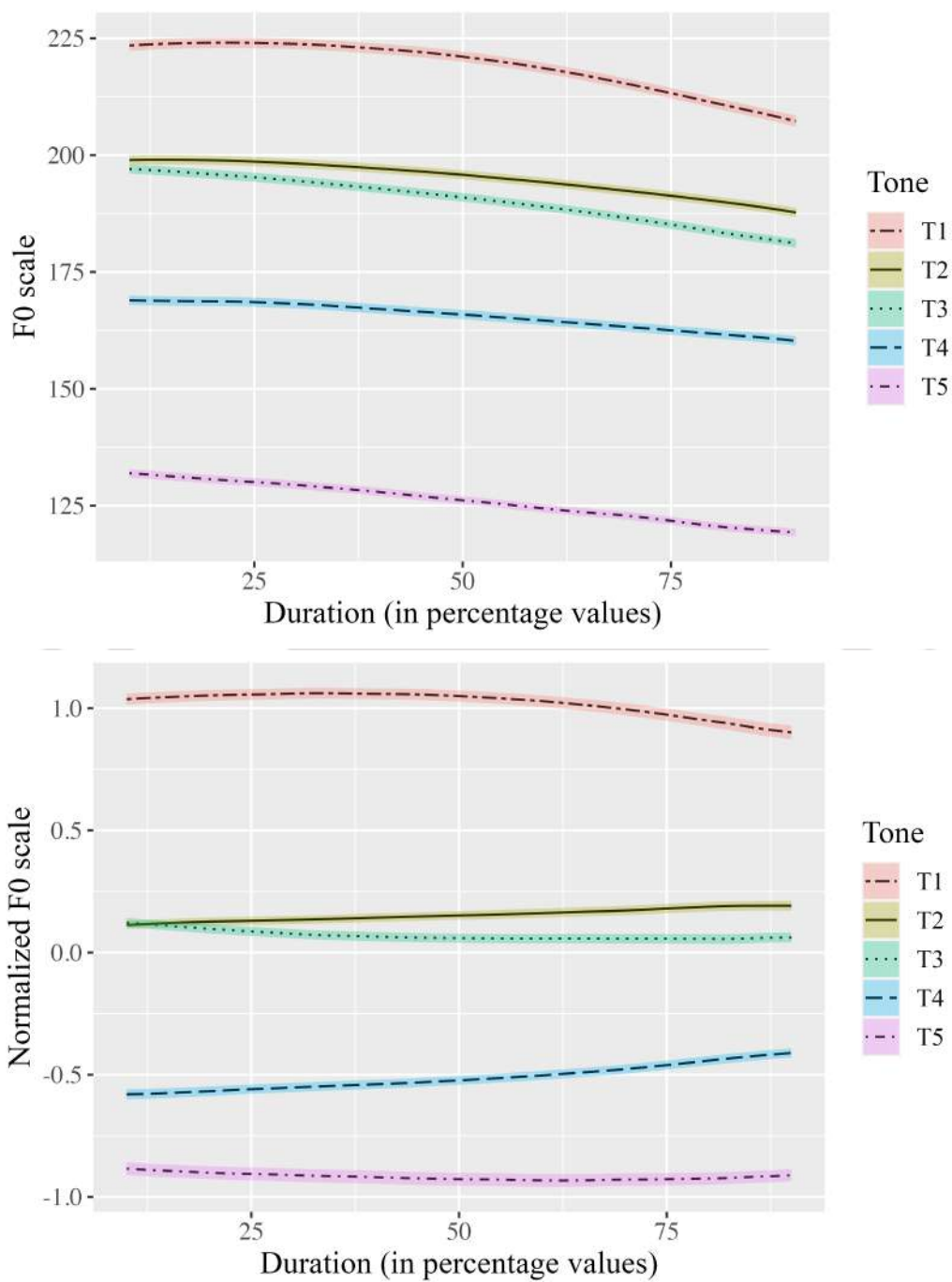


Figure 7-4: F_0 contours across the five tones in Angami, (top) is the average raw F_0 values while, (bottom) is the average normalized F_0 values.

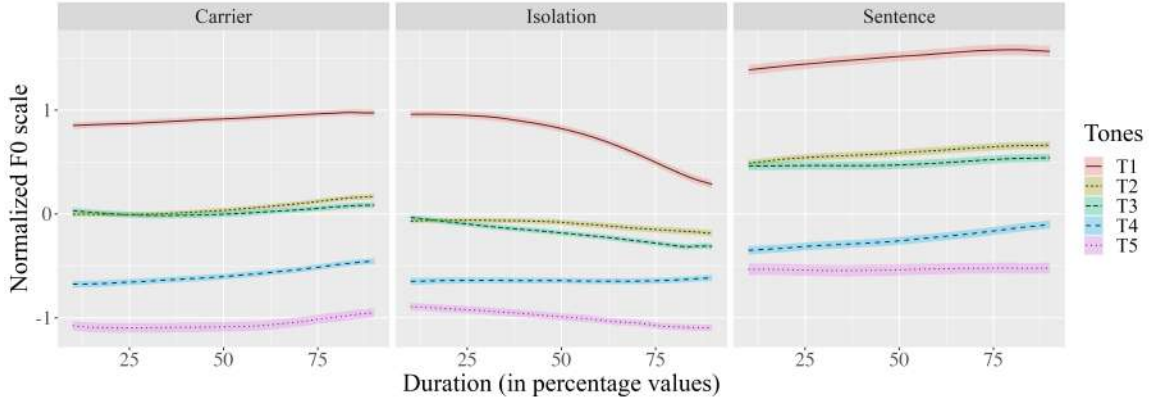


Figure 7-5: Normalized F_0 contour plot of the tones in Angami across the three context.

words (W), the types of data (D) are the random effects as shown in 7.1. The summary of the results of an Analysis of Deviance (ANOVA) test on the LME models and estimates of post-hoc Bonferroni tests for pair-wise contrasts is shown in Table 7.4. In all cases, the standard error is 0.0 indicating little variation in the F_0 values. The result shows that tones have a significant effect on the mid of the F_0 contour. It is also observed that the interaction of tones and context also have significant effect. The χ^2 value, degrees of freedom (df), and significance for each feature are shown in Table 7.4 (2nd column). Hence, looking into the pairwise comparisons of F_0 by tone contrasts is pertinent. The LME models were further subjected to post-hoc Bonferroni analysis to see the significant differences between the different tones. The result shows that at all the three mid-points, T1-T2, T3-4, T4-T5 are significant which means they are distinct. However, tones T2-T3 are not significant. This result confirms that the two overlapping tones as seen in Figure 7-4 are not distinct.

$$mid - F_0 \sim Tones + Context + Tones * Context + (1|V) + (1|W) + (1|D) \quad (7.1)$$

Table 7.4: Results of an Analysis of Deviance (ANOVA) test on the LME model for estimating the effect of tones, contexts and their interactions on the mid-points (40%, 50% & 60%) of the normalized F_0 contour and estimates of post-hoc Bonferroni tests for pair-wise contrasts.

Feature	Fixed effects: χ^2	Contrasts	Estimate	df	<i>t</i> -ratio	<i>p</i> -value
Norm. F_0 at 40%	Tones: 2523 ***	T1-T2	0.9	2397	17.8	< .0001
	Contexts: 736 ***	T2-T3	0.1	2582	1.9	0.4584
	Tones x Contexts: 57 ***	T3-T4	0.5	2826	10.4	< .0001
		T4-T5	0.5	2754	11.2	< .0001
Norm. F_0 at 50%	Tones: 2574 ***	T1-T2	0.9	2502	17.7	< .0001
	Contexts: 859 ***	T2-T3	0.1	2688	2.1	0.3350
	Tones x Contexts: 68 ***	T3-T4	0.5	2937	10.4	< .0001
		T4-T5	0.6	2865	11.7	< .0001
Norm. F_0 at 60%	Tones: 2553 ***	T1-T2	0.9	2559	17.1	< .0001
	Contexts: 1029 ***	T2-T3	0.1	2742	2.1	0.3142
	Tones x Contexts: 86 ***	T3-T4	0.5	2984	10.4	< .0001
		T4-T5	0.6	2926	12.0	< .0001

7.3.2.2 Temporal measurements

The graph in Figure 7-6 visualizes the duration of tones in Angami; the values are detailed in Table 7.5. The result shows that tones T1, T2, and T3, has a duration of 178 ms, 180 ms, and 180 ms, respectively. Notably, T3 exhibits considerable variability, as evidenced by the high standard deviation. Conversely, tones T4 and T5 are relatively shorter, with a duration of 158 ms and 166 ms. The tone durations are further examined across the three contexts. The result is displayed in Figure 7-7. The result shows that tones produced in isolation frames have the longer duration than the tones produced in carrier and sentence. Tones produced in sentence frames are the shortest. Hence, the result of the duration across the three contexts can be summarized as isolation > carrier > sentence. The intermediary tones T2 and T3 produced has the marginally higher duration. The temporal measurements are further subjected to statistical tests to determine the statistically significant effects.

LME model were constructed where the duration is the dependent feature, tone, context and their interactions are fixed effects, vowel (V), the unique words (W), the types of data (D) and speakers (Sp) are the random effects as shown in 7.2. The

Table 7.5: Mean and standard deviation of the duration (ms) of the five Angami tones

Tones	Mean	SD
T1	178	72
T2	180	70
T3	180	92
T4	158	62
T5	166	66

Table 7.6: Mean and standard deviation of the duration (ms) of the five Angami tones across the three contexts

Tones	Carrier	Sentence	Isolation
T1	167 (55)	155 (60)	209 (87)
T2	172 (59)	165 (71)	203 (75)
T3	162 (69)	154 (77)	219 (110)
T4	159 (57)	148 (61)	166 (68)
T5	161 (51)	158 (77)	177 (71)

result is summarized in Table 7.7 where it is seen that the types of tones, contexts and their interaction have an effect on the duration of the tones ($p < 0.001$). The pairwise comparison shows that only T1-T2 and T4-T5 are significant. Further, the pair-wise comparison of the tones were analyzed based on the contexts, displayed in Table 7.8. The effect of tones and context are significant in T4-T5 pair in isolation and in carrier frames. Whereas, T1-T2, T2-T3 and T3-T4 pair are not significant in all the three contexts.

$$Duration \sim Tones + Context + Tones * Context + (1|V) + (1|W) + (1|D) + (1|Sp) \quad (7.2)$$

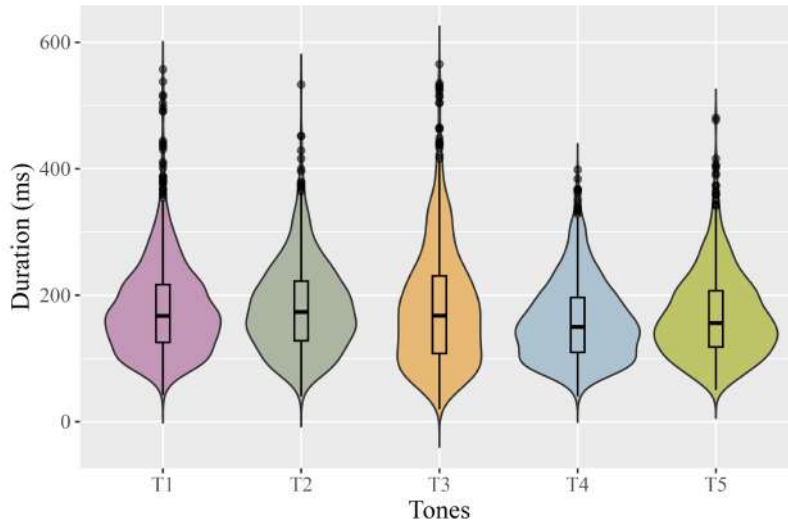


Figure 7-6: Violin plots of the average duration of the five Angami tones

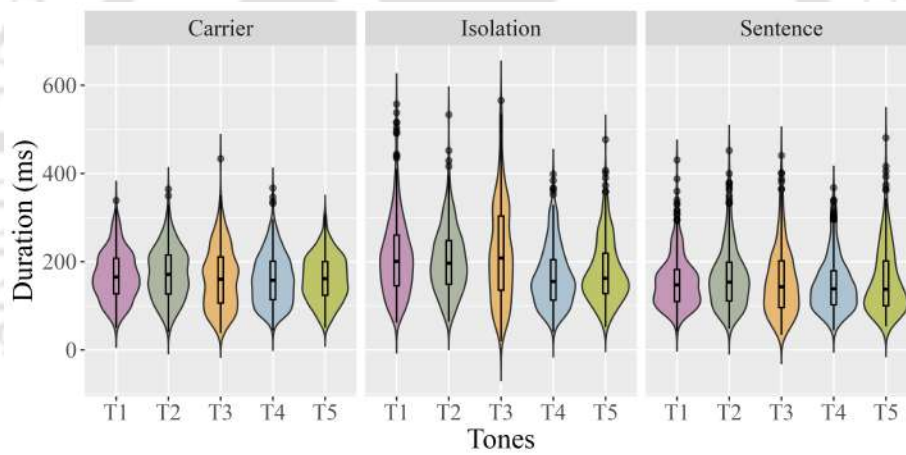


Figure 7-7: Violin plots of the average duration of the five Angami tones across the three contexts.

Table 7.7: Results of an Analysis of Deviance (ANOVA) test on the LME model for estimating the effect of duration of tones, contexts and their interactions on the mid-points (40%, 50% & 60%) of the normalized F_0 contour and estimates of post-hoc Bonferroni tests for pair-wise contrasts.

Fixed effects: χ^2	Contrasts	Estimate	SE	df	<i>t</i> -ratio	<i>p</i> -value
Tones: 76 ***	T1-T2	-14.59	5.0	1791	-2.8	0.0409
Contexts: 331 ***	T2-T3	-4.7	4.9	1965	-0.9	1.0000
Tones x Contexts: 45 ***	T3-T4	11.6 4.5	2162	2.5	0.1006	
	T4-T5	24.2	4.9	2104	4.9	< .0001

Table 7.8: Results of post-hoc Bonferroni tests for pair-wise contrasts across the three context.

	Contrasts	Estimate	SE	df	<i>t</i> -ratio	<i>p</i> -value
Isolation	T1-T2	-16.8	6.4	2377	-2.6	0.9426
	T2-T3	-3.6	6.2	2487	-0.5	1.0000
	T3-T4	18.3	5.8	3022	3.1	0.1734
	T4-T5	34.4	6.2	2725	5.5	< .0001
Sentence	T1-T2	-18.9	7 6.5	2918	-2.8	0.4249
	T2-T3	-4.3	6.4	3318	-0.6	1.0000
	T3-T4	7.7	6.0	3468	1.2	1.0000
	T4-T5	18.1	6.7	3434	2.697	0.7376
Carrier	T1-T2	-8.0	5.1	3351	-1.5	1.0000
	T2-T3	-6.1	5.1	3692	-1.2	1.0000
	T3-T4	8.7	4.7	3888	1.8	1.0000
	T4-T5	20.1	4.9	3826	4.0	0.0049

7.3.3 Conclusions

In this chapter, we examined the tones in Angami. While the existing literature on Angami suggests the presence of four to five tones, our study reveals that Angami exhibits four distinct tones. The intermediary tones T2 and T3 exhibit overlap and lack statistical significance. The canonical tones in Angami is proposed as shown in Figure 7-8 by merging the two overlapping tones. During our fieldwork data collection, language instructors mentioned differentiating between tone T2 and T3 based on duration, with T2 being longer than T3. Therefore, we conducted temporal measurements and found that the duration difference between T2 and T3 is not significant. However, tone durations show a significant contrast between T4 and T5. This study further shows that the different contexts in which the tones are produced significantly affect the tones. In order to determine the canonical tones in Angami, tones produced in carrier frames are the most reliable as the context of the target tones is controlled. Whereas in sentence frames, there are higher chances of contextual effects, and declination of the speech in the case of isolation frames. Finally, this study confirms that tones in Angami are register tones and not contour tones. The F_0 contour of the tones clearly showed a rather straight line than a dynamic line.

The existence of either four or five-level tones in a language poses unique challenges, as native speakers cannot either produce contrastive level tones or have difficulties in perceiving level tones close. This difficulty arises when the F_0 of the two-level tones have less difference in the tonal space. For instance, Black Miao (Qingjiang Miao), spoken in China, features five level tones (as shown in Figure 7-9 and three contour tones (Kuang, 2013). Using multidimensional scaling, the author reported that the mid-level tones (T22, T33) in Black Miao lacked distinctiveness in the tonal space and were rarely contrastive with T44 (level tone). Similar results are reported in the production of Cantonese, where young speakers tend to merge mid-level (T33) and low-level (T6) (Mok and Wong, 2010).

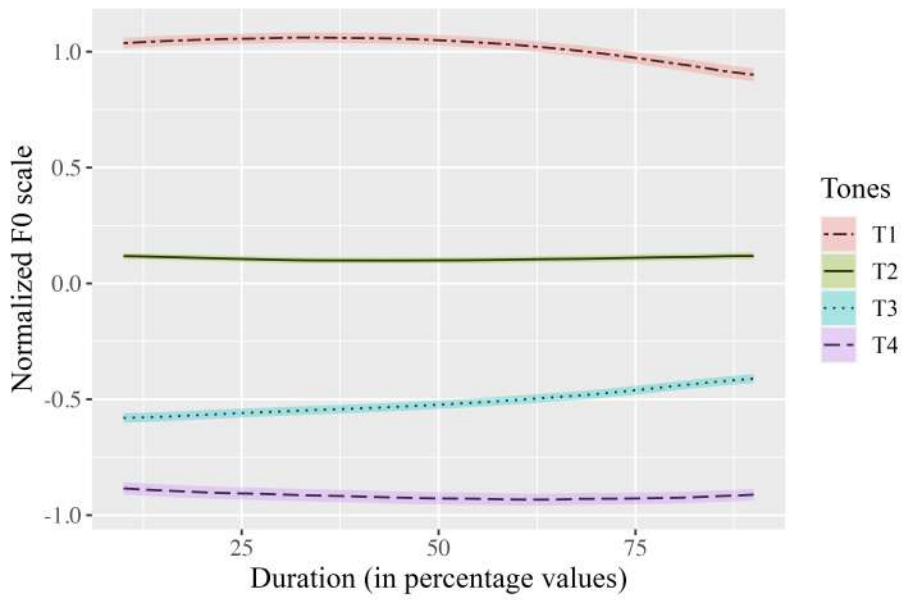


Figure 7-8: Normalized F_0 contour plot of the four tones in Angami

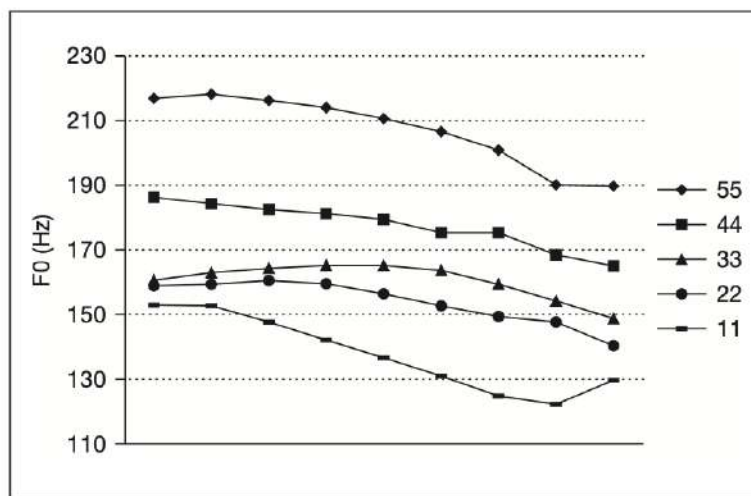


Figure 7-9: F_0 contour tones of the five level tones in Black Miao as produced by male speaker (Kuang, 2013)

7.4 Interaction of tone and segments

The subsequent sections examine how tones interact with both segmental and supra-segmental features. The first study investigates the interactions of tones and consonants, the second study examines the interaction of tones and vowels. Lastly, the third study delves into contextual tones.

7.4.1 Tone and Angami consonants

Angami is a tone language and has at least four tones in its inventory. Thus, in this study, we also provide an analysis of the interaction of tones and VOT in the language. The interaction of VOT with segmental and prosodic features is widely attested in the literature (Gandour, 1974; Ewan and Krones, 1974; Maddieson, 1978a; Hombert, 1978; Xu and Xu, 2003; Lai and Jongman, 2004; Lalhminghlu and Sarmah, 2020). The effect of VOT types of onset consonants on the following F_0 is widely attested. Generally known as co-intrinsic pitch or CF_0 , the fundamental frequency is measured at the vowel onset following onsets with various VOT types. CF_0 perturbations on the F_0 are reported equally in non-tonal and tonal languages. One early study of CF_0 perturbations on vowel onset shows that the duration of effect on non-tonal language such as English is longer (100 ms) than in tonal Yoruba language (Hombert, 1978). As reported in the study, for English, F_0 is raised after voiceless consonants and lowered after voiced consonants.

In a study of CF_0 , Lai and Jongman (2004) examined CF_0 of Taiwanese voiceless unaspirated and aspirated stops in initial stop consonants in three places of articulation (POA), namely, bilabial, alveolar and velar. The tones selected for the study are high level (T1), high falling (T2), low level (T3), low-rising (T5), and low-falling (T7). The study reports that F_0 of tones following aspirated stops is higher than tones following unaspirated stops. The raising effect is seen at about 30% of the total tonal contour/duration. In terms of POA: F_0 after velar stops were found to be significantly

higher than after bilabial or alveolar stops. They also reported on the interaction of POA x tone, where it is found that the velar stops were found in correlation with higher F_0 in tone 1 (high level) and tone 2 (high falling) than the other tones. On the other hand, Chen (2011) examines the effect of consonant-tone interactions in Shanghai Chinese, which reports that with regard to the effect of lexical tone on F_0 perturbation, preceding low-rising (24) tones were found to have a more significant F_0 perturbation (CF_0) than high-rising (35) and high-falling (52) tones both in terms of magnitude and temporal extent.

Of late, in Kirby (2018), CF_0 is noticed to be higher in voiceless stops across the languages. In his study, CF_0 effects in tone languages were observed when produced in isolation (as opposed to sentential form), especially the high falling in Thai and the high-level and mid-falling tones in Vietnamese, which were more prominently affected than the low tones. The observation in Hombert is true in many studies discussed above, where for tonal language, the perturbation caused by voiced consonants on a following high tone or by a voiceless consonant on a following low tone is greater than the effect of these two series of consonants on a mid-tone. The effect of a voiced consonant on a following high tone is greater than that of a voiceless consonant on a following low tone (Hombert, 1978).

The effect of tones on VOT is also investigated across languages. Maddieson (1978a) showed that tones could affect the adjacent consonants in three ways, namely, diachronic change in which tone affects consonants, synchronic rules where consonant differences depend on tones, and finally, the phonetic effects of tones on consonants. High tone condition longer consonants in Thai (Maddieson, 1978a). In a study of tone effects of VOT in Hong Kong Cantonese, Tse (2005) looked into the effects of tone on the VOT of Cantonese stops. It was shown that the VOT of the aspirated stops was affected by tonal differences. Tones in the lower pitch ranges induced higher VOT in Cantonese (Tse, 2005).

7.4.1.1 Methodology

For the study of the interactions of VOT and tones, the speech data used for analyses is from speech database Avot which is discussed in Chapter 2 (Section 2.2). For the extraction of F_0 values, ZFF technique was employed as discussed in Chapter 1 (Section 1.3.5).

7.4.1.2 Tone effects on VOT in Angami

To see the effects of tone, VOT values were analyzed in the data reported above, categorized by the five tones categories as reported in (Liezietsu et al., 2019) for Angami. The VOT values of stops produced by Angami speakers are presented in Table 7.9. The normalized F_0 plot of each of the five Angami tones in the present data is shown in Figure 7-10. As noticed in the Figure, while the T1, T4, and T5 are distinctly located in the tonal space, T2 and T3 tend to merge into a single category. Hence, we believe that Angami speakers have not maintained any systematic difference between these two tonal categories. This was also noticed in Lalhminghlui et al. (2019) where it was argued that Angami has four tone categories. However, a confirmation that Angami has only four-way tonal contrast would require further phonetic and phonological investigation. Hence, at this time, we will follow the tone categories as defined in Liezietsu et al. (2019), taking the fact into consideration that speakers may not make any pitch distinction between T2 and T3.

In the previous study of VOT in Chapter 2 (Section 2.3.1), it was seen that laryngeal contrasts and tone interactions have statistically significant effect on VOT. This implies that within each laryngeal category, VOT is distinct for each tone. Hence, to explore these interactions in detail, the LME model 2.1 (as discussed in Chapter 2) was subjected to a pairwise comparison using a Bonferroni post-hoc test. The results of the post-hoc test showed that the significant effect of tones on VOT is only restricted to the voiceless aspirated stops, as shown in Table 7.11. Nevertheless, the

Table 7.9: Mean VOT and standard deviation for Angami stops across all tones.

Stops	T1	T2	T3	T4	T5
/p/	15(7)	15(7)	15(7)	16(7)	15(6)
/t/	17(5)	17(6)	17(6)	17(6)	18(5)
/k/	38(14)	39(13)	39(13)	40(12)	41(14)
/p ^h /	75(29)	82(29)	84(28)	83 (29)	89(31)
/t ^h /	67(22)	75(28)	76(27)	81(29)	88(28)
/k ^h /	96(25)	105(29)	106(27)	107(27)	110(30)
/b/	-109(30)	-111(31)	-109(31)	-107(28)	-115(30)
/d/	-117(30)	-115(30)	-114(32)	-117(34)	-120(32)
/g/	-119(42)	-126(45)	-131(42)	-129(40)	-133(40)

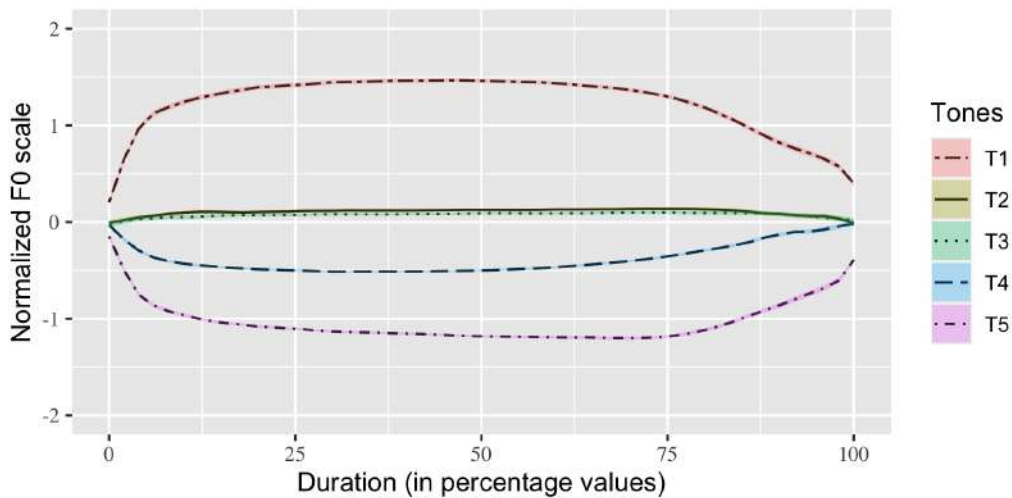


Figure 7-10: Normalized mean F₀ tone contours of five Angami tones with standard errors

estimates reported in the table clearly show that as tone goes lower, VOT increases for all laryngeal contrasts. This inverse relationship between tone height and VOT values is captured in Figure 7-11, where for all three laryngeal categories, T1 is associated with lower VOT, and T5 is associated with higher VOT.

$$Toneduration \sim Tones + (1|Speaker) + (1|Contrasts) + (1|Vowels) + (1|POA) + (1|Repetition) \quad (7.3)$$

In order to make sure that the tone-induced VOT differences are not due to

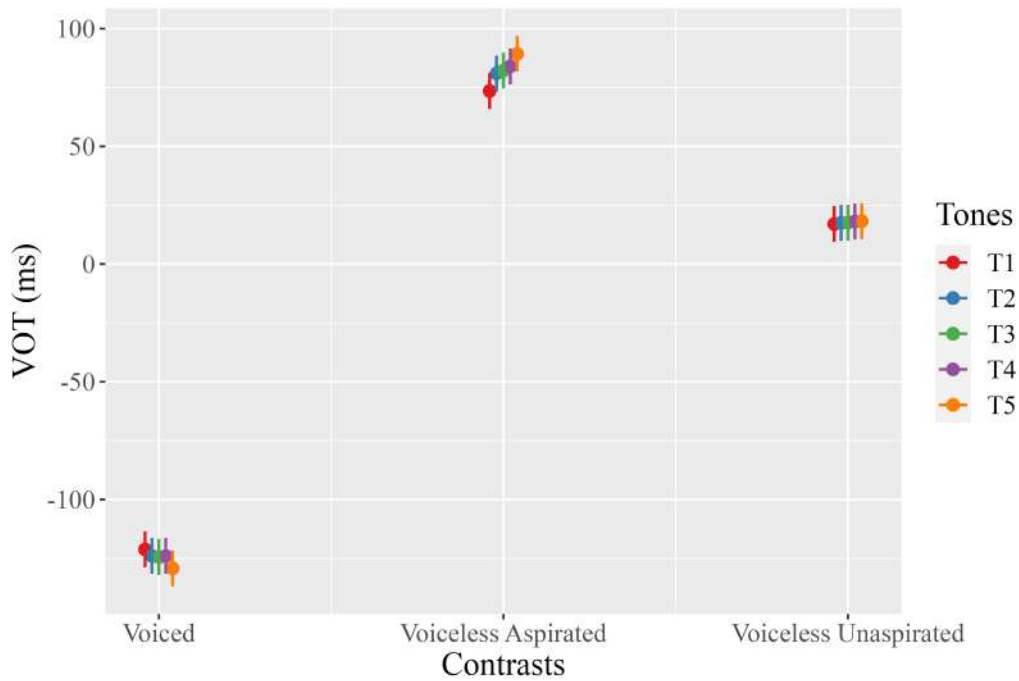


Figure 7-11: Variable interaction plot generated from an LME model with VOT as the dependent variable and laryngeal contrast and tones as fixed effects.

Table 7.10: Mean and standard deviation of the duration of the five Angami tones

Tones	Mean	SD
T1	286	90
T2	294	86
T3	290	93
T4	314	92
T5	275	79

the inherent tone durations in Angami, we analyzed the duration of the tones by tone categories. As seen in Table 7.10, there is no systematic relationship between tone height and tone duration. We conducted an explorative LME modelling where tone duration was the dependent variable, and tone categories were the fixed effects. Laryngeal contrasts, vowels, speaker, POA, and repetition were kept as random effects as seen in 7.3.

While the LME model showed a significant effect of tones on duration ($\chi^2(4)=349.5, < 0.001$), the effect was not consistent with the height of the tones. The model was fur-

ther subjected to a pairwise comparison using a Bonferroni post- hoc test which showed the following pattern of tone duration derived from the estimates of the contrasts as shown in Table 7.12: $T4 > T2 > T3 > T1 > T5$. Considering that tone duration does not correspond with tone height, we can conclude that the systematic VOT differences noticed for distinct tone categories are not due to differences in tone durations, rather due to differences in tone height in Angami.

7.4.1.3 VOT effects on tones in Angami

The effect of onset consonants on the following tone is discussed widely in the literature. Generally, it has been shown that laryngeal distinctions of onset consonants have an effect on the following F_0 (Hombert, 1978; Kirby and Hyslop, 2019; Lee et al., 2019). It is generally agreed that while voiced stops lower the following F_0 , voiceless stops raise it (Hombert, 1978; Tse, 2005). Apart from that, voiceless aspirated stops are known to induce even higher F_0 on the following pitch. While two-way laryngeal contrast affecting the following F_0 is widely reported, three-way laryngeal contrasts are relatively less reported. In case of tone languages, consonantal effect on the F_0 of the following tone is reported in Yoruba, Shanghai Chinese, Taiwanese, Thai, Vietnamese etc.(Chen, 2011; Hombert, 1978; Kirby, 2018; Lai and Jongman, 2004; Maddieson, 1978a). Most of these tone languages have contour and register tones in their tonal inventories. Yoruba is the only language that has three register tones, where consonantal effect on the tonal F_0 is shown. In the current study, we have investigated the consonantal effect on F_0 of tones in Angami that have at least four distinct register tones.

The tone's effect at the onset of each of the stop consonants are shown in Figure 7-12. In order to see the laryngeal effects on the Angami tones, pitch tracks of the five tones in the language, separated by laryngeal contrasts, are plotted in Figure 7-13. As reported in the previous studies, Angami tonal contours make four distinct tonal

categories, with T2 and T3 merging into a single category. This pattern is uniform regardless of whether the host syllable is meaningful in the language or not. Figure 7-13 also shows that the initial rise of the tone contour is distinct, depending on the laryngeal feature of the preceding consonant. As extensively reported in previous studies, the voiced onset induces lowering of the F_0 contour in the initial parts of the tone. On the other hand, voiceless onsets are known to raise the initial F_0 of the following pitch (Hombert, 1978; Kirby, 2018)

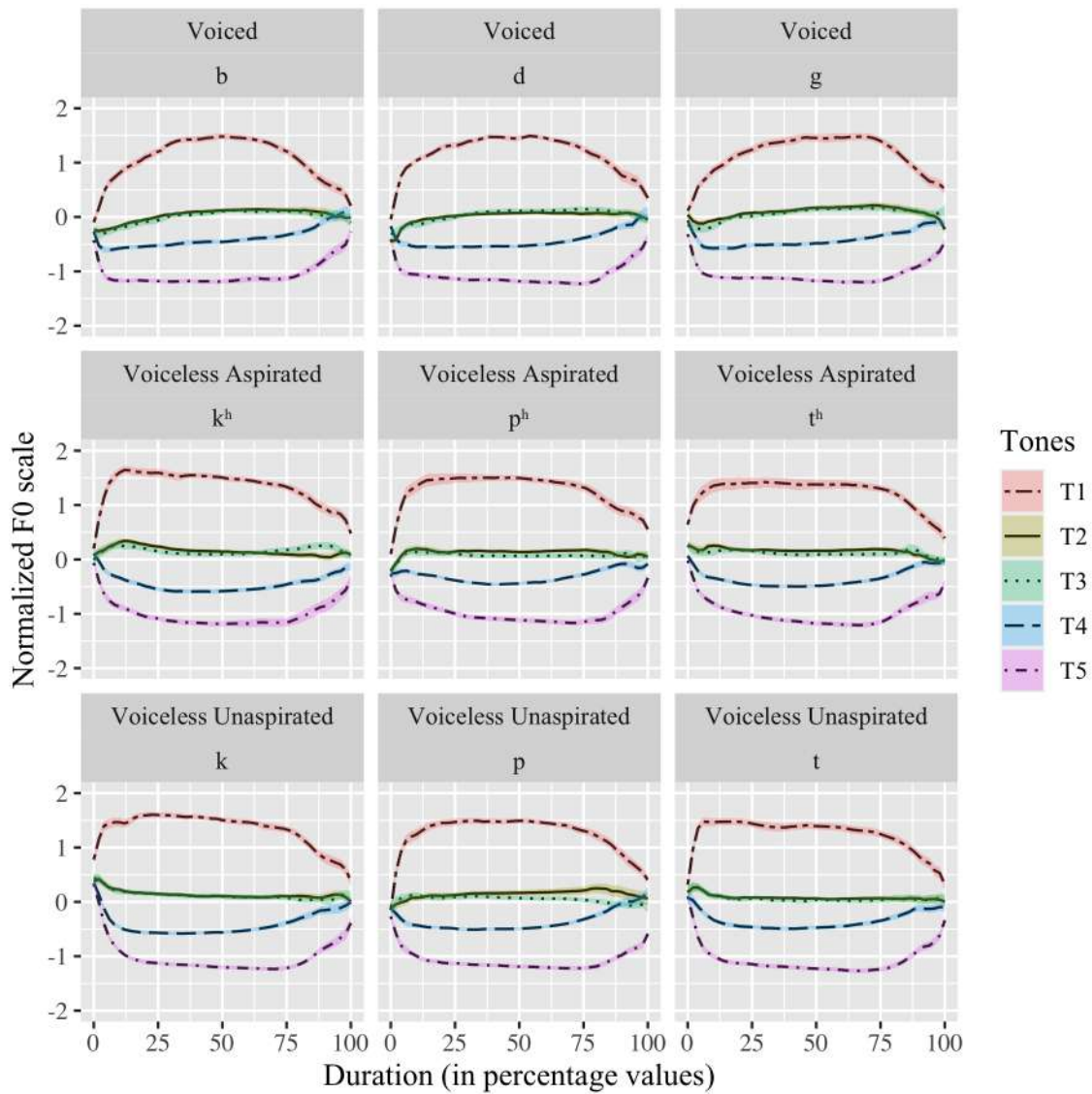


Figure 7-12: Normalized F_0 contours of the five Angami tones in the nine Angami stops.

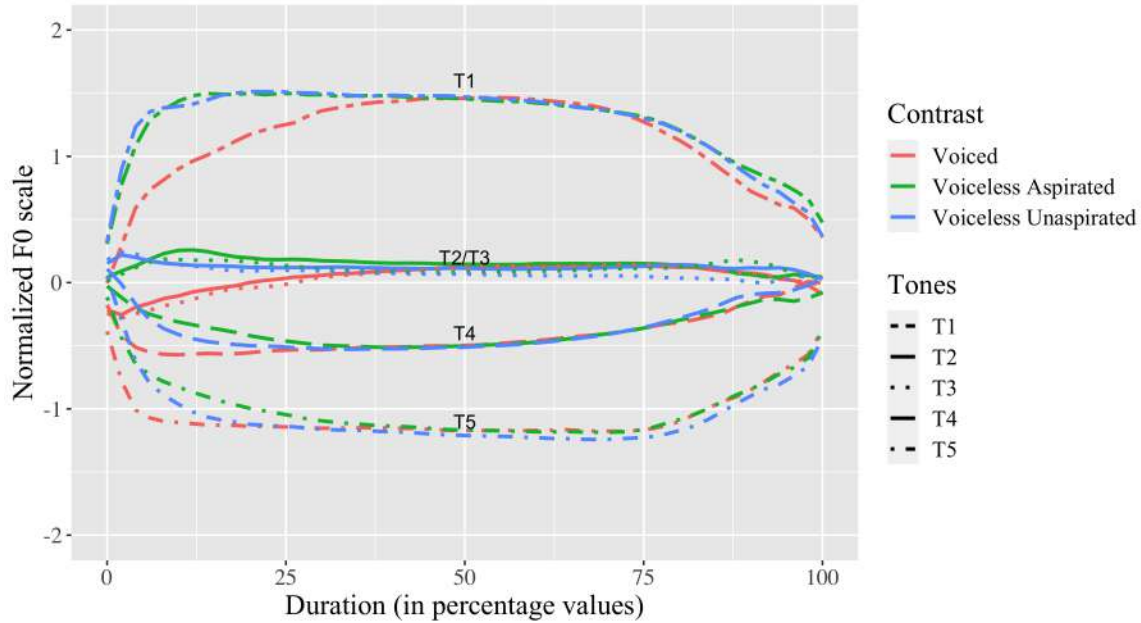


Figure 7-13: Normalized F_0 contours for Angami tones grouped by three laryngeal features.

Figure 7-13 shows that the effect of preceding consonants differs according to the tonal categories. As seen in the figure, the low tone followed by a voiced stop has the least effect on the onset. On the other hand, the voiceless onsets raise the F_0 of the low tone, T5, until about 30% duration of the tone overall. As voiced onsets induce low F_0 onto the following tone, the tone contour for T5 followed by voiced stops reaches the low tone target earlier than the voiceless unaspirated and aspirated stops. Similarly, in the case of T4, the tone contour preceded by a voiced stop reaches the low target at about 5% of the tonal duration. However, when preceded by voiceless stops, the tone contours retain consonantal effects until about 25% of the total duration of the tone. In the case of tones T2 and T3, the F_0 contour does not exhibit a significant effect of the preceding voiceless onsets. However, the target for T2 and T3 is much higher than the natural F_0 induced by the voiced onsets. Hence, the F_0 followed by the voiced onset has consonantal perturbation until about 30% of the total duration of the tone. Finally, in the case of the high tone, T1, the voiceless onsets reach

the high F_0 target much earlier as these onsets naturally induce high F_0 onto the following tone. However, when the preceding onset is voiced, it takes longer to reach the high F_0 target for T1. As observed in Figure 7-13, the F_0 contour following the voiced onset reach the F_0 target for T1 only at about 40% of the total tone duration. Hence, as tone categories get higher in Angami, consonantal perturbations appear to be lesser for the voiceless stops. Conversely, as tone categories get lower, consonantal perturbations from the voiced stops appear lesser. Figure 7-13 also shows declining F_0 for T1 and rising F_0 for T4 and T5 towards the termination of the tones. This change correlates with the end of energy in the production of the syllables. As such, a change in the F_0 is purely motivated by the termination of the syllable; the rising or the falling F_0 contours do not show any consonant specific differences.

7.4.1.4 Discussion

This study looked into the interaction of tones and onset consonants. As reported in previous studies, the voiced onsets induce F_0 lowering of the following tone. On the other hand, voiceless onsets induce raising in the following F_0 . Angami has four phonetic tones differing in register level heights, namely, high (T1), mid (T2 and T3), mid-low (T4), and low (T5). Hence, as tone categories get higher in Angami, voiceless consonantal perturbations are less noticeable as the tone specific F_0 height is achieved faster due to the voiceless stop specific F_0 perturbations. At the same time, when produced with T4 or T5, the effect of the voiceless consonants appear more pronounced as voiceless consonant induced high F_0 requires longer to lower to the tone specific F_0 . Conversely, for T4 and T5, consonantal perturbations of the voiced stops appear shorter as they induce low F_0 , which reaches the target F_0 of the low tones faster.

One novel finding in the current analyses is the confirmation of the effect of tones on VOT in Angami. In previous studies, it has been shown that as speakers' pitch is

inversely related to the duration of the VOT in aspirated stops (Narayan and Bowden, 2013). Tse (2005) has shown that in the case of Cantonese, lower pitch ranges induce higher VOT. In case of Mizo, aspirated stops, tones that begin with higher F_0 , namely, the falling and high tone, induce shorter VOT in the language (Lalhminghlui and Sarmah, 2020). In the current study, we have observed a systematic relationship between tone height and VOT values. High tones induce shorter VOT, while low tones induce longer VOT. While the correlation is noticed for all three laryngeal types, a significant effect of tones on VOT is restricted to the voiceless aspirated stops.

It was found that there is a statistically significant durational difference among the four phonetic tone categories. However, the durational characteristics do not follow the universally attested patterns of short high tones and long low tones. In the case of Angami, no systematic relationship between tone height and tone duration was found. As shown in Section 7.4.1.2, the descending order of tones by their duration is $T4 > T2 > T3 > T1 > T5$. This study also confirmed that contrary to some of the previous reports, Angami has only four phonetically distinct tones.

Table 7.11: Result of Bonferroni post-hoc pairwise comparison conducted on an LME model with normalized VOT as the dependent variable and tone and laryngeal contrast as fixed effects (VD= voiced, VLA= voiceless aspirated, VLU = voiceless unaspirated)

	Contrast	Estimate	SE	z-ratio	p-value
VD	T1 - T2	2.76	2.0	1.38	n.s
	T2 - T3	0.49	2.0	0.25	n.s
	T3 - T4	-0.46	2.0	-0.23	n.s
	T4 - T5	5.23	2.0	2.59	n.s
	T2 - T3	3.25	2.0	1.62	n.s
	T2 - T4	0.03	2.0	0.02	n.s
	T3 - T5	4.77	2.0	2.37	n.s
	T1 - T4	2.79	2.0	1.39	n.s
	T2 - T5	5.26	2.0	2.63	n.s
	T1 - T5	8.03	2.0	3.98	**
VLA	T1 - T2	-7.46	1.98	-3.76	< 0.05
	T2 - T3	-1.19	1.98	-0.60	n.s
	T3 - T4	-1.82	1.98	-0.92	n.s
	T4 - T5	-5.32	1.98	-2.67	n.s
	T2 - T3	-8.66	1.98	-4.36	< 0.01
	T2 - T4	-3.02	1.98	-1.52	n.s
	T3 - T5	-7.14	1.99	-3.59	< 0.05
	T1 - T4	-10.47	1.99	-5.27	< 0.001
	T2 - T5	-8.34	1.99	-4.19	< 0.01
T1 - T5	-15.79	1.99	-7.92	< 0.001	
VLU	T1 - T2	-0.53	1.99	-0.26	n.s
	T2 - T3	-0.04	1.99	-0.02	n.s
	T3 - T4	-0.49	1.99	-0.25	n.s
	T4 - T5	-0.14	2.0	-0.07	n.s
	T2 - T3	-0.57	1.99	-0.26	n.s
	T2 - T4	-0.54	1.99	-0.27	n.s
	T3 - T5	-0.63	2.0	-0.32	n.s
	T1 - T4	-1.06	1.99	-0.53	n.s
	T2 - T5	-0.67	2.0	-0.34	n.s
	T1 - T5	-1.19	2.0	-0.59	n.s

Table 7.12: Result of Bonferroni post-hoc pairwise comparison conducted on an LME model with tone duration as dependent variable and tone as fixed effects.

Contrast	Estimate	SE	df	<i>t</i>-ratio	<i>p</i>-value
T1-T2	-8.94	2.13	5293	-4.19	< 0.001
T1-T3	-5.57	2.14	5293	-2.605	n.s.
T1-T4	-28.96	2.14	5293	-13.55	< .0001
T1-T5	9.91	2.14	5293	4.629	< .0001
T2-T3	3.38	2.13	5293	1.585	n.s.
T2-T4	-20.03	2.13	5293	-9.392	< .0001
T2-T5	18.85	2.14	5293	8.816	< .0001
T3-T4	-23.41	2.13	5293	-10.964	< .0001
T3-T5	15.47	2.14	5293	7.230	< .0001
T4-T5	38.88	2.14	5293	18.147	< .0001

7.4.2 Tone and Angami vowels

In this work, we look into the interaction of tones and vowels in Angami. This work aims to identify and account for the changes in tone characteristics due to the influence of the vowels. The positive correlation of vowel height and F_0 has been referred to as intrinsic F_0 (IF_0) in the literature, and we investigate how IF_0 affects the pitch of different tones in Angami.

According to Lieberman, IF_0 occurs due to a coupling effect, where the shape of the vocal tract during a vowel production determines the amount of vibration of the vocal folds (Lieberman, 1970). On the other hand, it is also argued that there is a direct relationship between the raising of the tongue in high vowel production and the tension on the vocal folds. According to this argument, in the production of high vowels, the tongue pulls on the larynx, giving rise to the high tension of the vocal folds, resulting in higher F_0 values (Lehiste, 1970; Ladefoged, 1964). Ohala and Eukel provided substantial evidence for the tongue pull hypothesis through their bite block experiment, making it a dominant explanation for IF_0 in the years to follow (Ohala and Eukel, 1987). However, the amount of control a speaker exerts on IF_0 is unclear.

There is also an enhancement hypothesis that posits that IF_0 variation is deliberate and is aimed at improving auditory salience (Diehl and Kluender, 1989). It argues that IF_0 is used as an additional feature to enhance the perception of vowel contrasts. In other words, languages with large vowel inventory may need IF_0 to enhance the perception of different vowel categories. While earlier studies did not find any correlation between vowel inventory size and IF_0 differences (Whalen and Levitt, 1995), a later study confirms that there is a direct relationship between the two (Van Hoof and Verhoeven, 2011). Contrary to that, a later work investigating IF_0 differences in American English varieties found that the sources of the control of F_0 use may not be the size of vowel inventory but may be socio-cultural, reflecting regional variation in speech (Jacewicz and Fox, 2015). Moreover, the same study found clear evidence of

IF₀ only in the vowels of prominent syllables and not so much in the non-prominent syllables.

Intrinsic vowels of tone languages further complicate the situation, as speakers also need to control the F₀ for the best realization of the lexical tones. In the case of tone languages, a few interesting observations are made in the literature (Whalen and Levitt, 1995). While IF₀ is noticed for the tone languages, the IF₀ differences observed are negligible in low tones and in the lower F₀ ranges. In languages with more than two-tone levels, the IF₀ should decrease as tones become lower. For Standard Chinese, it is reported that a high tongue position, while articulating vowels, induces higher tones (Hoole and Hu, 2004) as in non-tone languages. Clear evidence for IF₀ in three African register tone languages, namely, Ibibio, Kunama, and Dschang, are found in a previous work (Connell, 2002). The same study failed to find significant evidence of IF₀ in Mambila. However, IF₀ was found to be significant in Shona, a language related to Mambila (Gonzales, 2009).

While some studies have investigated the evidence of IF₀ in Chinese and African languages, not many studies have reported the interaction between vowel and IF₀ in tone languages of the Tibeto-Burman family, specifically the ones spoken in India. Hence, this study addresses this gap in the literature by investigating the interaction of vowels and F₀ in Angami. This study tries to answer the following questions:

- (i) Is there any evidence of vowel effect on the F₀ of tones in Angami?
- (ii) Do vowel effects manifest equally along the duration of the tone?
- (iii) Considering the differences in the tone inventory of Angami, is there a difference in the vowel effect on F₀ of tones?

The rest of the study is arranged in the following manner: Section 7.4.2.1 describes the methodology adopted in this study. Section 7.4.2.4 presents the results, and finally, Section 7.4.3 discusses the findings and concludes the study.

7.4.2.1 Methodology

6 native speakers (3 male, 3 female) participated in the study. Speakers are from the Kohima village, and Angami is their first language; they also speak and understand English, Nagamese, and Hindi. At the time of data collection, their age ranged between 20 and 30 years.

7.4.2.2 Materials

Table 7.13: Distribution of vowel tokens in Angami.

	T1	T2 / T3	T4	T5	Total
/a/	20	26	16	7	71
/e/	08	29	07	18	70
/i/	24	46	22	23	115
/o/	18	47	20	19	104
/u/	23	44	23	24	114
/ə/	26	47	23	25	121
Total	119	239	111	116	585

The data for Angami consists of 6 Angami vowels /a, i, e, u, o, ə/, produced in isolation with each of the 5 tones in Angami, namely, T1, T2, T3, T4, T5. This resulted in 30 unique tokens, of which only 2 could be potentially associated with a lexical word in Angami. The token /a/ with T2/T3 for ‘me/ I’ and the token /u/ with T2/T3 for ‘yes.’ However, all participants produced the vowels with sustained, long pronunciation, reducing any association with a lexical word. As with Mizo, the participants were asked to mimic the tones of Angami in the vowels using the tone of a real lexical item as a cue. The word /pe/, with five different meanings for each tone, i.e., T1 - ‘to incline,’ T2 - ‘bridge,’ T3 - ‘fat,’ T4 - ‘shiver’ and T5 - ‘shoot’ were displayed on a computer screen along with the vowel. Each token was displayed randomly to avoid the listing effect. The participants were requested to mimic the tone of the /pe/ syllable and produce the vowels. Each participant produced each token five times, resulting in 900 tokens. However, 585 tokens are

considered a large number of tokens produced with creaky phonation, making pitch estimation impossible and many others mispronounced. The tokens considered for analysis in this study are presented in Table 7.13.

7.4.2.3 Acoustic and statistical analysis

The speech recordings were transferred to a computer for analysis. The sound files were segmented and annotated as described in Chapter 1 (Section 1.3.3). The tone boundaries are marked at the initiation of the F_0 till the point of termination. A script was used to automatically extract F_0 values at every 25% of the total duration of the pitch contour and export them to a spreadsheet for further analysis. To eliminate speaker effects, the raw F_0 values were normalized using the z-score normalization method Rose (1987). The averaged normalized F_0 values for each vowels were plotted for visual examination. Statistical analysis were explored as described in Chapter (Section 1.3.6). For the LME test, separate analysis were conducted with F_0 values at 0%, 25%, 50%, 75% and 100% as dependent variables. A full model was constructed considering vowel categories, gender and tone as fixed effects and speaker as random effect and null models were built with vowel categories removed. The null and the full models were compared using ANOVA.

7.4.2.4 Results

In the present study, Angami's register tones tend to fall when produced in isolation. The average F_0 contours of the five Angami tones are provided in Figure 7-14. As seen from Figure 7-15 till Figure 7-18, the vowel effect on the tones is more pronounced at the point of initiation of the tones. However, as the pitch contour falls to lower F_0 levels, the vowel-specific differences diminish. T1, the highest tone in Angami, shows the largest vowel-specific F_0 differences until 50% of the pitch contour.

The height of the vowel /u/ corresponds with higher F_0 in T1, T2, and T5,

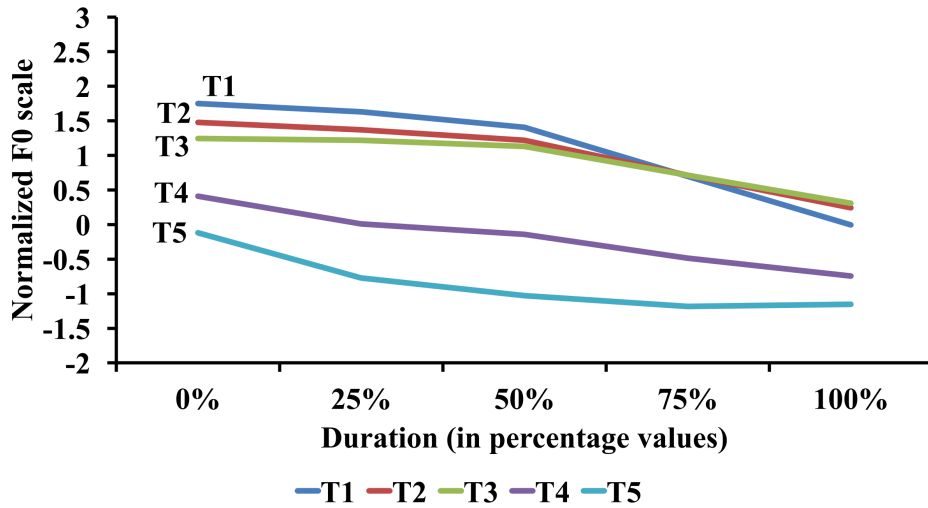


Figure 7-14: Average F_0 contours of five Angami tones.

indicating robust vowel and tone interaction. At around 25%, high vowel /i/ and mid-vowel /e/ overlap in all the tones except in T4, which is a low tone. Back vowels /a/ and /o/ have similar pitch contours in all the tones except in T2/T3. Across all tones, two clear effects of vowels emerge: vowels /u/, /i/, and /e/ induce higher F_0 , and vowels /a/ and /o/ induce lower F_0 . The vowel /ə/ stands out in most of the tones; it reaches its tone target only at 50% of the duration of the tone.

The results of the LME test, reported in Table 7.14, showed that vowels interact significantly with the F_0 at all points of the F_0 contour, except the final one. The comparison of the full LME model (with vowel categories, gender, and tone as fixed effects and speaker as random effect) with the null model (excluding vowel categories) showed significant differences in ANOVA.

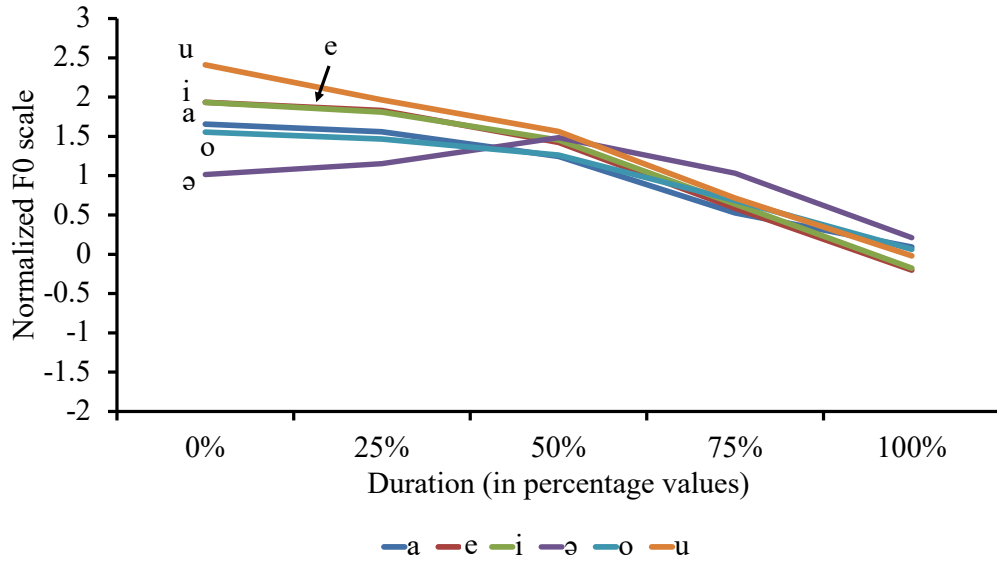


Figure 7-15: F₀ contour plot of T1 with all the six Angami vowels.

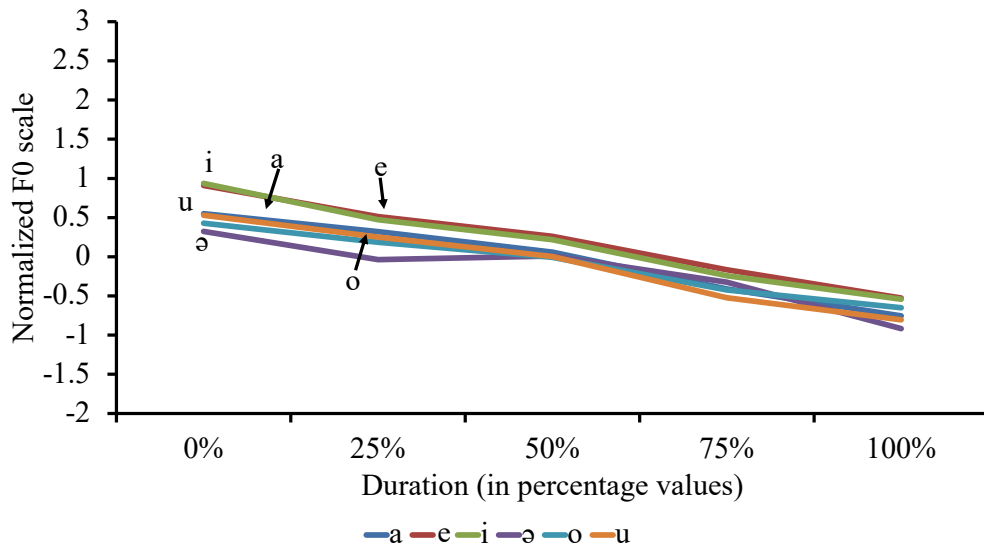


Figure 7-16: F₀ contours of T2/T3 with all Angami vowels.

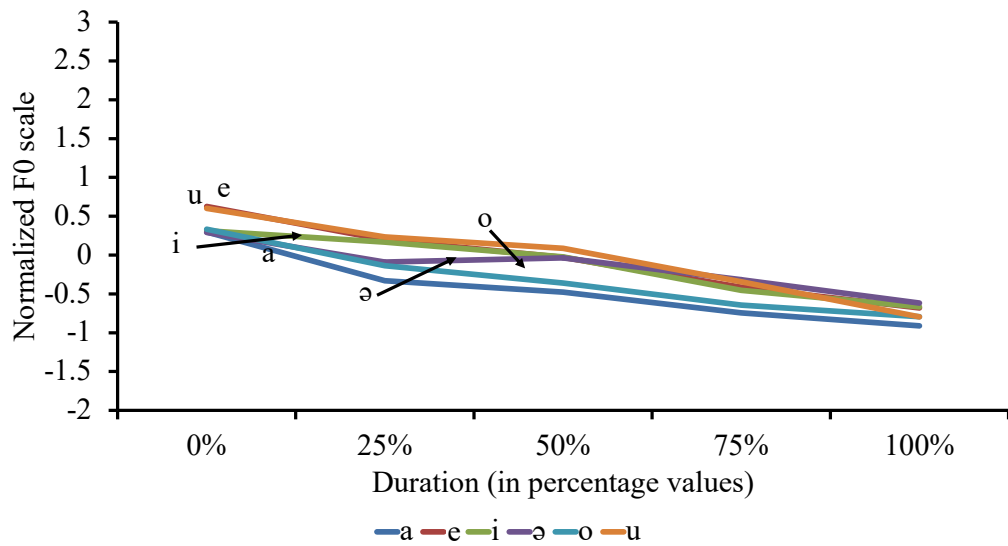


Figure 7-17: F₀ contour plot of T4 with all the six Angami vowels.

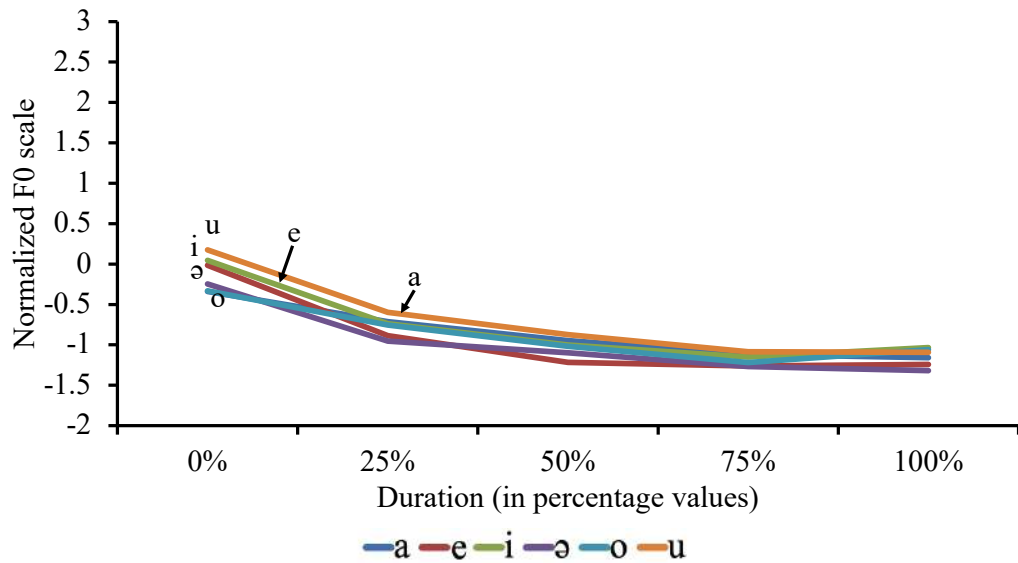


Figure 7-18: F₀ contour plot of T5 with all the six Angami vowels.

Table 7.14: Result of LME test in Angami tones

	0%	25%	50%	75%	100%
$\chi^2(5)$	46.93	59.67	26.90	19.78	7.50
p	<0.001	<0.001	<0.001	<0.01	0.186

7.4.3 Discussion and conclusions

This study suggests a clear correlation between F_0 and vowel quality. High vowels induce higher F_0 , and low vowels induce F_0 lowering in Angami. The effect of vowels on tones is prominent in both languages at the initial part of the F_0 contour. At the initial points (0%), the speakers do not control the tone-specific pitch in the utterance; hence, IF_0 is prominent in that part. However, by the next quarter of the pitch contour, speakers begin to control the pitch for tone-specific values, resulting in the reduction of vowel-specific F_0 differences. The results of this study also show that the vowel-specific differences are more pronounced in the high tones or in the higher pitch registers. However, vowel-specific differences are negligible as the tone or the pitch contour lowers. The results also corroborate the results observed in earlier studies on tone-vowel interaction. It supports the earlier views that tone languages do show evidence for IF_0 . However, as speakers try to reach the tone targets, the IF_0 differences are subdued by tone-specific values. Hence, the effect of vowels may not be manifested equally along the duration of the tone. No significant vowel effect difference was observed on F_0 based on the salient pitch contour dynamics of Angami.

7.4.4 Tonal coarticulatory effects in Angami

This study attempts to understand the effects of contextual tones in Angami. This work investigates if carryover and anticipatory effects on tones exist in Angami. Secondly, it also investigates the nature of the contextual effects, viz. they are assimilatory or dissimilatory. Considering Angami has at least four level tones and

no contour tones, this study provides a unique insight into the tonal coarticulatory phenomena in a level tone-only language. Apart from that, Angami belongs to the Tibeto-Burman language family which is less studied, and even less is known about the acoustic features of the tones of the family. Hence, the current work contributes towards the understanding of tones in Tibeto-Burman languages in particular and of tonal coarticulation in the world's languages in general.

While F_0 is the primary cue in distinguishing tones in lexical tone languages (Yip, 2002), the F_0 itself may be affected by the F_0 of the tones produced in the preceding or the following syllable (Han and Kim, 1974; Abramson, 1979). When the following tones (syllable 2) affect the preceding tones, the effect is known as anticipatory, while, the phenomenon of the preceding tones (syllable 1) affecting the following tones is referred to as the carryover effect. The nature and the direction of such effects are language-specific (Gandour et al., 1994; Xu, 1994a). As seen in Mandarin and Cantonese, the coarticulatory effects are bidirectional, with dissimilatory anticipatory effects and assimilatory carryover effects (Xu, 1994a; Wong, 2006). Cantonese tones get higher when followed by a low onset tone, showing the dissimilatory nature of anticipatory effects. In contrast, onset is higher when preceded by a high offset showing assimilatory carryover effects. Similar results are also attested in Mizo, a Tibeto-Burman language, and in Thai, where the effects were bidirectional, however unlike Mandarin and Cantonese, the anticipatory effects were assimilatory or dissimilatory depending on the following tones (Sarmah et al., 2015b; Gandour et al., 1994). In Mizo, anticipatory effects are more significant on contour tones where falling tones are lowered when followed by high, rising, or low tones, which assimilate or dissimilate depending on the following tones. Mizo carryover effects are assimilatory where high or low onsets (falling or high) are lowered by preceding low offsets (low or falling). The bidirectional effects in Mizo have a symmetric pattern (Sarmah et al., 2015b). In Thai, a high tone gets higher when followed by a rising or a low tone, and rising

tones get higher when followed by a falling tone or a high tone (Gandour et al., 1994). Tonal coarticulation effects are also reported to vary depending on tone height and tone slope. In Southern and Northern Vietnamese, not only the tone height, but also the slope is changed. The entire tone contour is shifted upwards or downwards. On the other hand, in the case of Mandarin, there is no significant change in the slope of the tones (Brunelle, 2009a). In all the studies reported, it is shown that carryover effects are more common and robust than anticipatory effects.

While there is a considerable number of studies on the tonal coarticulation or contextual tonal characteristics in contour tone languages, the contextual effects of tones in level tone languages are relatively unexplored. We could locate only the abstract of the study that reported contextual tone variations in Hausa, Bole, and Yoruba (Yu, 2009). Hence, this study attempts to fill the research gap by reporting the results of a study on tonal coarticulation in the level tones of Angami.

The sections of this paper are organized as follows: Section 7.4.4.1 provides a description of data collection, the language interest of study, the participants; and the acoustic and the tools used for statistical analysis. In Section 7.4.4.2, the results of the study are discussed, finally, the conclusion in Section 7.4.4.3.

7.4.4.1 Methodology

This section describes the speech corpus along with the recording procedure, the F_0 extraction and visualization methods.

7.4.4.1.1 Speakers

Native speakers of Angami residing in the Kohima district of Nagaland contributed speech data with the aim of capturing tonal coarticulation. All the speakers are multilingual and can speak English and Nagamese (*Lingua Franca*), apart from their mother tongue, Angami (Tenyidie). A total of 24 speakers consisting of 15 females

and 9 males were recorded for this study. The mean speakers' age was 33.1 years (SD = 3.3) at the time of recording. Data was recorded in October 2019 and June 2022. 10 speakers produced only the noun combination data (25 combinations x 3 contexts x 10 speakers = 750 tokens) whereas, 14 speakers contributed to both noun and verb combination data (50 combinations x 3 contexts x 14 speakers = 2100 tokens).

7.4.4.1.2 Speech corpus

The material was designed to capture the production of tonal coarticulation in Angami. For this study, we labeled the material with five tones as described in *MKS Dieda Liezietsu et al. (2019)*. We labeled the high tone as T1, and the low tone as T5, whereas the intermediate tones were labeled as T2, T3 & T4. The material contained 50 distinct meaningful disyllabic words in Angami in CVCV syllable structure, only in one case, the syllable structure was VCCV. The disyllabic words include combinations of two nouns (NN, n = 25), and also combinations of verbs and verb morphemes (VV, n = 25). The target syllables are embedded in three syllables, as described in Chapter 1 (Section 1.3.2).

Table 7.15: Example of the Noun-Noun combination and the Verb - Grammatical category (GC) combination used in the study.

Tones	Noun - Noun		Verb - GC	
T1	/tso/ elephant	/mi/ tail	/ta/ run	/zie/ PrPROG
T2	/wi/ mithun	/ne/ ear	/w ^h e/ cover	/i/ CC
T3	/k ^h uo/ fish	/tsə/ head	/t ^h ə/ work	/tjie/ IMP-ASP
T4	/ut/ camel	/zə/ skin	/p.u/ jump	/ja/ HB
T5	/keɹ/ horse	/m ^h i/ eye	/pe/ shoot	/lie/ IMP-ASP

7.4.4.1.3 Recording of the data

The recordings were made with a head-mounted Shure SM10A microphone and a TASCAM DR-100 MKII portable digital solid-state recorder at a sampling rate of 44.1 KHz. Using Praat software, speech files were subjected to analysis and saved in the .wav format (Boersma, Paul and Weenink, David, 2020). The sound files were manually segmented and annotated on the tone bearing unit of the tones.

7.4.4.1.4 F_0 extraction

To obtain tone information, we extracted F_0 values at the onset of the F_0 contour of the first syllable (V_1) and then continued onto the onset consonant of the second syllable (C) and finally ending at the termination of the F_0 contour in the second syllable (V_2). In order to interpolate F_0 values in the voiceless portions of C, we used the *interpolate* function in Praat, which provides a linear interpolation of the missing F_0 values. F_0 values were extracted at every 2% of the total duration of the V_1CV_2 , resulting in 51 points across the total duration where F_0 were extracted. In order to remove speaker effects, z-score normalization of the F_0 values was conducted (Rose, 1987). The normalized values, along with speaker, context, and data type (NN or VV) information, were saved as comma-separated values for further visualization and statistical analyses. Further qualitative and exploratory statistics were conducted as described in Chapter 1 (Section 1.3.6).

7.4.4.2 Results

To visualize the tonal coarticulation effects in Angami, F_0 contours of the disyllables were plotted. The boundaries of the first syllable, V_1 , consonants, and the second syllable, V_2 , are marked by dashed lines. The following section discusses the anticipatory and carryover effects in Angami tones.

7.4.4.2.1 Anticipatory effects

The normalized mean F_0 of the disyllables show anticipatory effects of the following tones, as seen in Figure 7-19. A visual examination shows that T1 (high tone) has a greater anticipatory effect. In figure, the F_0 of T1 (high tone) in the first syllable is raised when followed by a T5 (low tone). On the other hand, the F_0 of T1 (high tone) in the first syllable is the lowest when the tone in the following syllable is T1.

Visually, the intermediary tones (T2, T3, T4) induced similar effects on T1, making the F_0 contour of T1 somewhere in between, compared to the previous two contexts. This confirms the dissimilatory anticipatory effect on T1. Compared to T1, the intermediate tone, T2 shows less pronounced anticipatory effects. T2 is raised when followed by T5 (low tone). A marginal lowering of T2 is noted when T4 follows. Figure 7-19 (mid) shows that the anticipatory effects on T3 are less pronounced. However, a clear dissimilatory pattern is observed: when followed by T1 (high tone), T3 is lowered, and it is raised when followed by T5 (low tone). In the subsequent figure, T4 demonstrates an assimilatory anticipatory effect, being raised when followed by T3 in the second syllable, while other tones (T1, T2, T4, T5) have marginal effects. Nevertheless, on close observation, it is noticed that T4 and T5 induce the lowering of T4 in the first syllable. The tone contour of T5 is higher when followed by T2 and T3 than when followed by T4 and T5, as seen in Figure 7-19 (low). In that case, T5 has more assimilatory anticipatory effects. However, when followed by T1, the highest tone, T5 is lower than when followed by T2 and T3. We presume that the extra height required to achieve the target for T1 makes it lower its F_0 to the minimum possible level in the preceding syllable.

7.4.4.2.2 Carryover effects

Figure 7-20 illustrates the normalized mean F_0 of the disyllable tone contours, highlighting carryover effects from the preceding tone. In Angami, carryover effects appear more prominently than the anticipatory effects. Specifically, Figure 7-20 (top) demonstrates the carryover effects of the preceding tones on T1 (high tone) in the

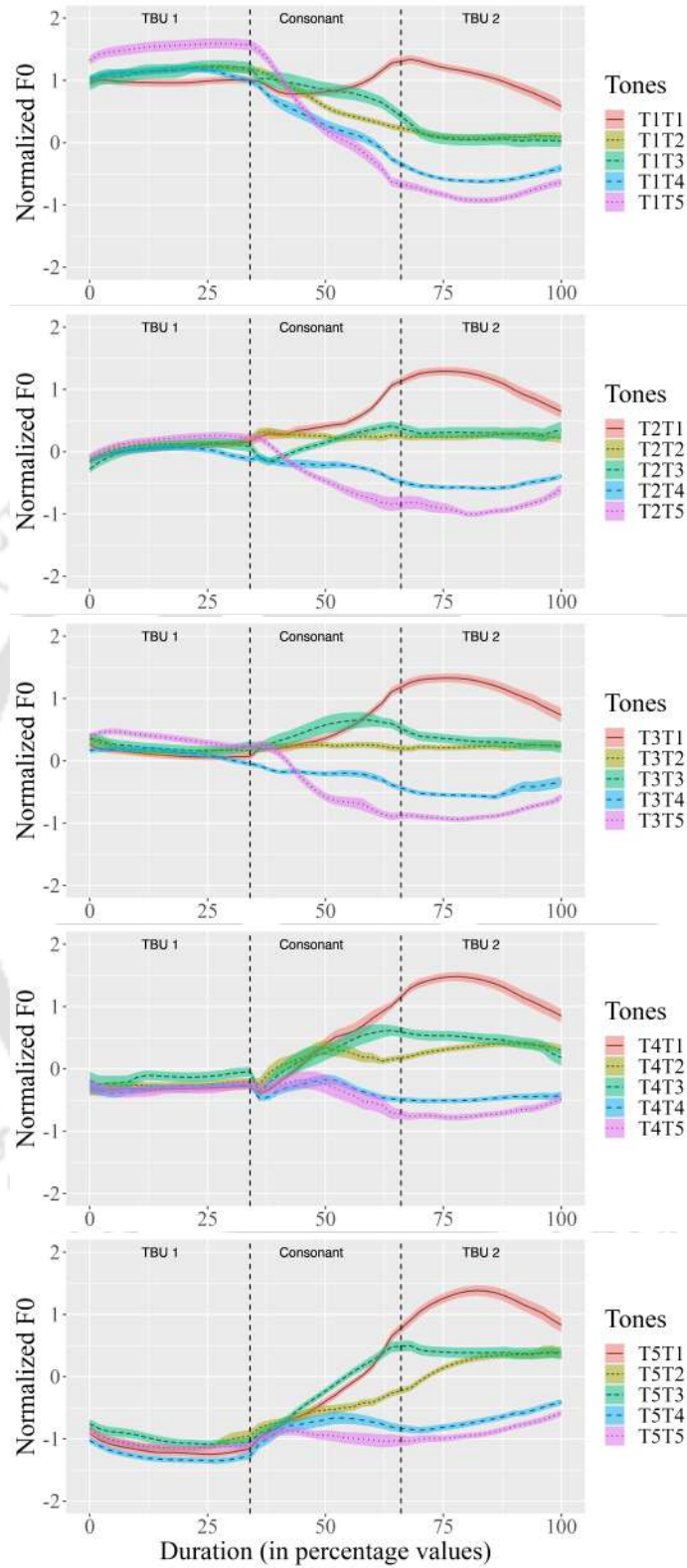


Figure 7-19: Normalized mean F_0 of the disyllable tone contours showing anticipatory coarticulatory effects.

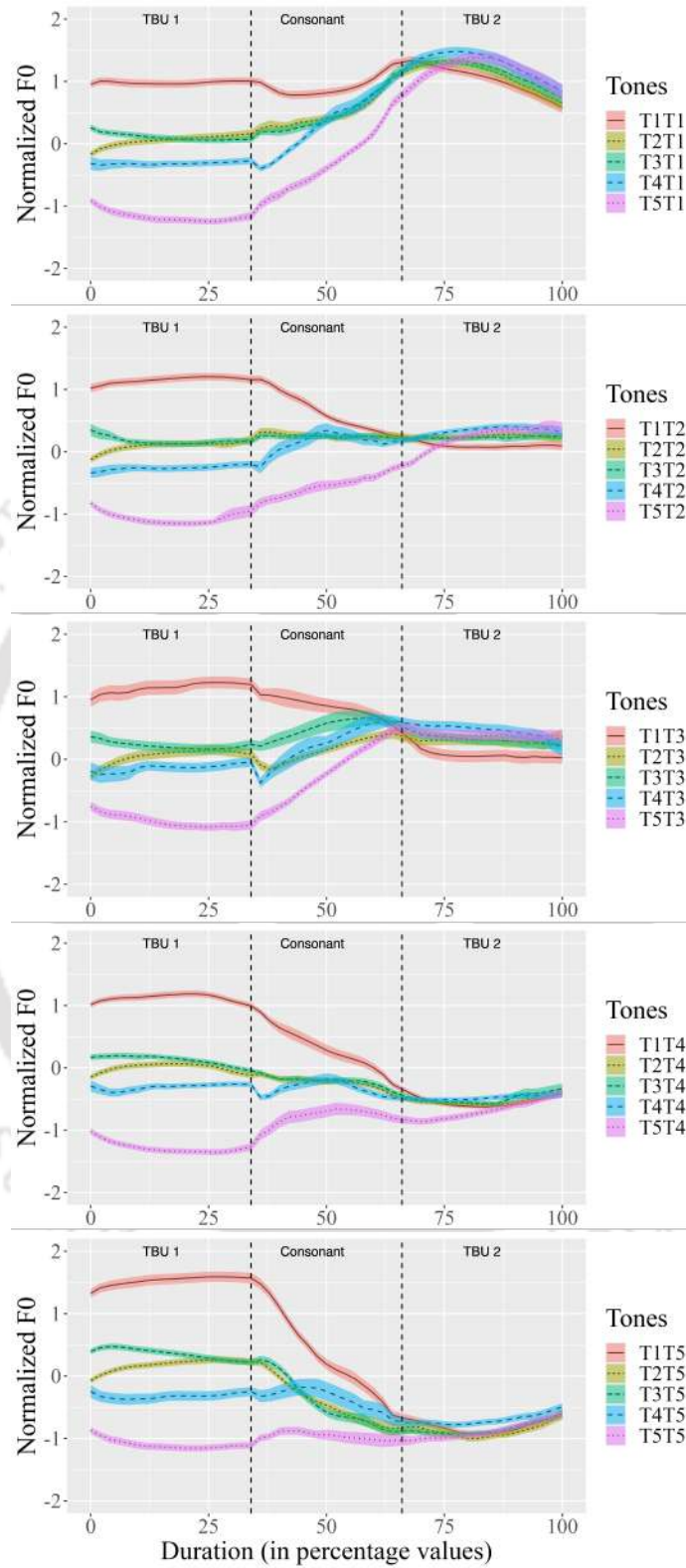


Figure 7-20: Normalized mean F_0 of the disyllable tone contours, showing carryover coarticulatory effects.

second syllable. T1 exhibits a dissimilatory pattern, being lowered when preceded by another T1 in the first syllable, and raised when preceded by T4 and T5 at around 80% of the F_0 contours. In the subsequent figure, the carryover effect of T2 is dissimilatory, with T2 lowering when preceded by tone T1 and raising when preceded by tones T4 and T5 in the first syllable. T3 also shows a pattern similar to T2 in terms of carryover effects, as seen in Figure 7-20 (mid).

Contrary to the previous three cases, T4 and T5 do not show dissimilatory patterns. The figure confirms assimilation, as T4 is lowered when preceded by T5. A marginal carryover effect is observed when T4 is preceded by high tone T1, up to about 50% of the total duration of the second syllable. In Figure 7-20 (bottom), a potential assimilatory carryover effect is noted, with T5 being raised when preceded by T4 in the first syllable. The following section reports the statistical analyses conducted to confirm the contextual effects on tones in Angami.

7.4.4.2.3 Statistical analyses result

To confirm contextual effects on tones, we constructed LME models with the dependent variable being the three mid- F_0 points (44%, 50% & 56%) of either the first syllable or of the second syllable. The tones in the first or in the second syllables were considered fixed effects for anticipatory or carryover effects, respectively. Speaker, context, and type of data (NN or VV) were considered random effects. The results of the Analysis of Deviance Tests on the LME model for estimating anticipatory effects are shown in Table 7.16. Statistical significance of anticipatory effect is seen on T1 (high tone) and on T3 (intermediate tone). This confirms the visual representation as seen in subsection 7.4.4.2.1 where anticipatory effects were observed on T1 and T3. Although minimal effects were observed in the anticipatory effect of tones T2, T4 & T5, statistical tests could not confirm those effects. Statistical analysis in Table 7.17 shows the result of Analysis of Deviance Tests on LME models for estimating

carryover effects at three middle points. T2, T3, and T5 had statistically significant carryover effects.

Table 7.16: Results of Analysis of Deviance Tests on LME models for estimating anticipatory effects at three middle points, showing χ^2 values ($df=4$) with statistical significance indicated by an asterisk *.

Following Tones	44%	50%	56%
T1	30.9*	34.0*	36.2*
T2	27.0 ^{n.s.}	2.6 ^{n.s.}	2.9 ^{n.s.}
T3	31.1*	28.9*	27.8*
T4	2.1 ^{n.s.}	0.8 ^{n.s.}	0.8 ^{n.s.}
T5	6.8 ^{n.s.}	7.1 ^{n.s.}	7.0 ^{n.s.}

Table 7.17: Results of Analysis of Deviance Tests on LME models for estimating carryover effects at three middle points, showing χ^2 values ($df=4$) with statistical significance indicated by an asterisk *.

Preceding Tones	44%	50%	56%
T1	8.7 ^{n.s.}	8.1 ^{n.s.}	7.3 ^{n.s.}
T2	10.8*	11.2*	10.8*
T3	10.6*	9.9*	9.7*
T4	9.7 ^{n.s.}	6.4 ^{n.s.}	5.4 ^{n.s.}
T5	15.2*	15.6*	14.3*

7.4.4.3 Conclusions

The current study shed light on tonal coarticulation in a level tone-only language, Angami. The results confirm bi-directional tonal coarticulation in Angami. Both assimilatory and dissimilatory effects were observed in anticipatory and carryover contextual effects. In Angami, the three highest tones (T1, T2, T3) showed distinct coarticulatory patterns compared to the two lowest tones (T4 and T5). Among the highest tones, T1 stood apart as it is an exceptionally high (extra high) tone.

In terms of anticipatory tonal effects in Angami, we see that the high tone (T1) is affected more by the following tones as raising and lowering of F_0 are more pronounced. While the prominent effect of the following tones is also observed in the

case of the other tones, the effects are not as pronounced. Regarding both anticipatory and carryover effects, the contextual high tone (T1) and low tone (T5) exert the highest effect on the preceding or following tones. The effect on higher tones (T1, T2, and T3) is dissimilatory in nature, whereas, on lower tones (T4 and T5) it is assimilatory. This is true for both anticipatory and carryover effects. As far as we know, no previous study has attested to such a mixed effect on contextual tones. For example, studies on contextual tones in Mandarin (Wong, 2006), Cantonese (Wong, 2006), Thai (Gandour et al., 1994), and Mizo (Sarmah et al., 2015b), do not show any evidence of dissimilatory tone carryover effect. Hence, we conclude that the contrasting assimilatory and dissimilatory patterns in Angami are due to the height of tones in the language. As Angami is a register tone language, such contrastive patterns within carryover or anticipatory effects are observed distinctly.

Another observation we report in this paper is the coarticulation of the high-tone T1. It is noticed that T1 requires to be produced at a much higher F_0 , recognized as an extra high tone. In order to prepare for the production of the F_0 corresponding to the extra high tone, the larynx needs to be lowered to the extent that a lower than normal F_0 is produced. This results in the low F_0 preceding the production of T1 in Angami, as observed in the current study.

The tendency of low tones to be more assimilatory than high tones may have a physiological explanation. The effects observed in Angami contextual tones may be due to physiological constraints where high tones are more prone to variations than low tones. This asymmetry in F_0 variation in high and low tones arises due to the involvement of greater muscle activity in the production of low tones than in high tone (Ohala, 1972). Production of low pitch is more complex than the high tone production as a cricothyroid muscle must be relaxed before the strap muscles activity (Erickson et al., 1982). The complexity in the production of low tones may induce more assimilatory effects than dissimilatory ones.

In this chapter, conclusive evidence emerges regarding the production of Angami tones. The results reveal that Angami possesses four distinct tones in its inventory. Furthermore, the studies indicate that all four tones in Angami are level in nature. Co-articulatory effects on the tones and the interaction of vowel height effect on tones were observed. The following chapter delves into the F_0 -based features of Angami tones and employs machine learning tools for tone classification.



Chapter 8

Acoustic Correlates and Classification of Tones in Angami

This study reports the acoustic characteristics of tones in Angami. Previous works did not agree on the types and number of tones in Angami. With speech data collected from ten Angami speakers, it is established in this work that Angami has four distinct tones that are level in nature. A classification of the five tone categories conducted using the Random Forest (RF) classifier yielded an accuracy of 66.4 %. Considering the merger of the two tones (T2 and T3), the accuracy increases to 81.6 %. It was noticed that the highest and the lowest tones were recognized more accurately than the intermediate tones. In terms of acoustic features for tone identification, it was concluded that F_0 mean, F_0 height, F_0 max, and F_0 min contributed most towards the correct classification of tones.

8.1 Introduction

As seen in the previous chapter, it has been observed that Angami has four level tones, which is relatively rare as only few languages has reported to have four contrastive level tones. The existence of four or five level tones in a language has its unique

challenges, both in production and perception. To distinguish the five level tones in Black Miao it was found that only average F_0 is required to distinguish the level tones from one another. Other measures associated with the tones, such as ΔF_0 , F_0 Onset, F_0 offset & Duration were deemed unimportant for tone classification (Kuang, 2013). In the case of perception by native speakers, it was noticed that young speakers consistently merged two categories of level tones (Mok and Wong, 2010). In the paper on the perception of merging tones in Hong Kong Cantonese, it was shown that young speakers merge level tones T22 and T33, although there is 30 Hertz difference between the two tones (Mok and Wong, 2010).

Various studies have explored the automatic classification of tones using different fundamental frequency (F_0) features, aiming to distinguish between different types and numbers of tones in tone languages. For instance, research on the Mizo language, which features four lexical tones (two level and two contour tones), utilized six F_0 features derived from F_0 contours. These features include F_0 height, slope, initial and final F_0 , F_0 variance, and F_0 difference. Two classification models, Support Vector Machine (SVM) and Deep Neural Network (DNN), were employed for tone recognition tasks, resulting in accuracy rates of 74.11% for DNN and 73.39% for SVM (Gogoi et al., 2020). Similarly, in the classification of tones in the Ao dialects (Changki, Mongsen, and Chungli), which exhibit three level tones (high, mid, and low), F_0 features such as F_0 height, initial and final F_0 , and mean F_0 were used. Automatic classification using SVM achieved accuracy rates of 53%, 62%, and 52% for Changki, Mongsen, and Chungli dialects, respectively (Gogoi et al., 2020). These studies highlight the importance of F_0 features in distinguishing between tones in different languages and dialects, with varying success rates depending on the specific characteristics of each language's tonal system. In this study, we opt for Random Forest (RF) for tone classification due to the relatively smaller size of our dataset. Support Vector Machines (SVM) and Deep Neural Networks (DNN) typically demand

more training data and computational resources, which may not be feasible given our limited dataset size and computational constraints. Therefore, Random Forest offers a suitable alternative for achieving accurate tone classification with our available resources.

The objectives of the current study are as follows:

- Contributes toward the understanding of the production and distinctiveness of level tones when they are crowded in tonal space.
- Secondly, to definitively establish the acoustic phonetic properties of tones in Angami.
- Finally, an automatic tone classification experiment to determine the crucial phonetic correlates for the identification of Angami tones.

The following sections of this paper are organized as follows. Section 8.2 describes the methodology of data collection, materials selection and the acoustic features used in the study. Section 8.3 presents the results of the study and finally, Section 8.4 discusses and concludes the study.

8.2 Methodology

This section discusses the methodology for the acoustic analysis and detection of tones in Angami. For this study, the speech data recorded for the plosives study has been utilized. Detailed information regarding the speakers, speech corpus, and data recording can be found in Section 2.2. A total of ten speakers, comprising 5 males and 5 females, contributed speech data for the study. In total, 5,377 speech tokens were analyzed. The following subsections describe the F_0 extraction method and the classifier. Figure 8-1 illustrates the overall framework of the Angami tone detection system. The input speech signal is initially pre-processed by removing the trend and

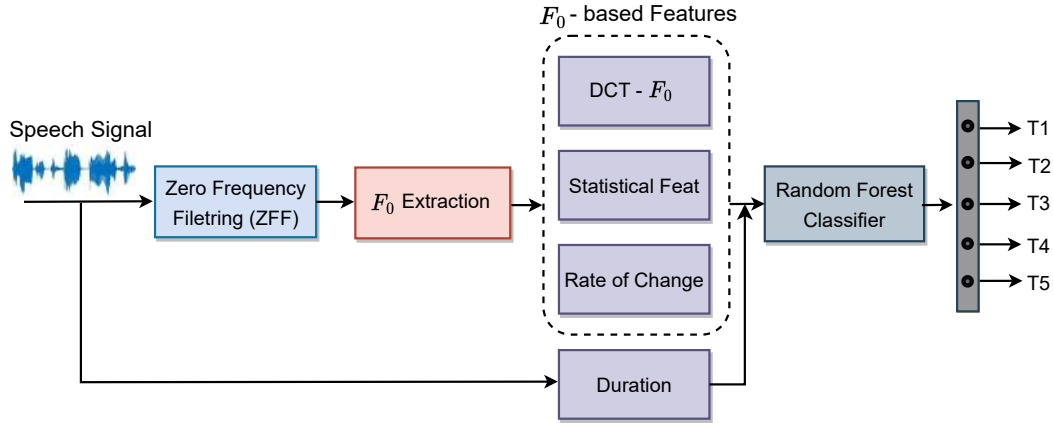


Figure 8-1: Overall framework of the Angami tone detection system.

resampling to 8 kHz. Next, the zero frequency filtering (ZFF) method is used to extract the F_0 -based features. The extracted features are then fed to the random forest (RF) classifier for the Angami tone classification task.

8.2.1 F0 Extraction Using Zero-frequency filtering (ZFF)

In this study, by employing Zero-frequency filtering technique (as discussed in Section 1.3.5), the F_0 values are extracted using the ground truth from the TBU regions. Therefore, the ZFF method is used in this work to extract the F_0 values of each monosyllabic tone utterance. The speech signal is initially pre-processed to sample at 8 kHz. Next, the difference of the speech signal is computed to remove any time-varying low frequency in the speech signal. The extracted ZFF F_0 contour is linearly time-aligned to 51 data points (P_1, P_2, \dots, P_{51}). To remove the speaker effects, z-score normalization is computed to the F_0 contour (Rose, 1987). Furthermore, to remove consonantal effects and enhance the stability of the mean estimates, the middle 39 data points (P_7, P_8, \dots, P_{45}) are considered for the analysis.

This work explores a comprehensive set of F_0 based features for the Angami tone classification. The F_0 based features are extracted from the five level tones of the Angami language such as F_0 Mean, F_0 Max, F_0 Min, F_0 Range, F_0 Height, F_0 Slope,

ΔF_0 , $\Delta\Delta F_0$, F_0 Initial, F_0 Final, F_0 Variance and 7 discrete cosine transform (DCT) coefficients (C_0, C_1, \dots, C_6) extracted from the F_0 contour. F_0 Range is computed by taking the difference between F_0 Max and F_0 Min. In order to calculate the F_0 Slope and F_0 Height, the F_0 contour of every monosyllable is fitted with a line to see if tones differ in their characteristics (Gogoi et al., 2020b). Fitting is done using linear regression, which estimates the coefficients (a, b) of a polynomial of degree one that fits the values best in a least-squares sense. The estimated line is

$$y = ax + b \quad (8.1)$$

where x is a vector ($x = 1, 2, \dots, N$), N is the length of the segment, a is the F_0 Slope and b is the F_0 Height (Sarma et al., 2015). The Δ features are considered to see the rate of change of the F_0 contour. ΔF_0 is determined by taking the successive difference between F_0 values and $\Delta\Delta F_0$ is determined by taking the successive difference between ΔF_0 . To determine the F_0 Initial and F_0 Final, the middle 39 data points (P_7, P_8, \dots, P_{45}) of the F_0 contour are considered, and the mean F_0 points at the starting and the ending of that syllable are computed as

$$F_0 \text{ Initial} = \frac{P_7 + P_8}{2} \quad (8.2)$$

$$F_0 \text{ Final} = \frac{P_{44} + P_{45}}{2} \quad (8.3)$$

8.2.2 Random Forest

Random forest (RF) is a discriminative and supervised machine learning algorithm used for classification and regression tasks. It is a popular algorithm as it has many commanding characteristics such as better generalization capability, feature pruning ability and simplicity to perform non-linear classification. The idea of RF is to construct a number of decision trees by bootstrapping training data with a replacement

for each tree. In RF, rather than choosing the entire features from the training set, only a subset of randomly selected features is considered (Kumar et al., 2016). Based on the majority vote from each decision tree, the test set predicts the final class. The RF classification experiment was set up using the R software with the randomForest, caret and e1071 packages (Liaw and Wiener, 2002; Kuhn, 2021; Meyer et al., 2021). A grid search was conducted and using the best nodes, the RF parameters were set. The RF parameters set the number of trees at 300, maxnodes at 372.

8.3 Results

This section discusses the results obtained for acoustic analysis, feature importance, classification and statistical analysis of tones in Angami.

8.3.1 Acoustic analysis of tones

The normalized F_0 of the ZFF derived values for the five Angami tones (T1-T5) is shown in Figure 8-2. In this figure, the tone contours are plotted by extracting F_0 values at regular intervals, at every 2% of the total duration of the tones, resulting in a total of 51 data points. As outlined in Section 8.2.1, to mitigate the influence of consonantal effects (also known as co-intrinsic pitch (CF0) at the onset), a method was employed wherein six initial and six final F_0 values were omitted. Consequently, only the mid 80% of the total duration of the tones was taken into account for the subsequent analysis, ensuring a focus on segments less impacted by consonantal effects. It is clear that the tonal space of T2, T3 & T4 are condensed as compared to T1 and T5 in Angami. Moreover, the tonal space between tone T2 & T3 is minuscule and overlapping at times. After conducting statistical analyses, no significant difference between T2 and T3 was found; hence they were considered as belonging to a single tonal category. Although we determined that Angami has only four dis-

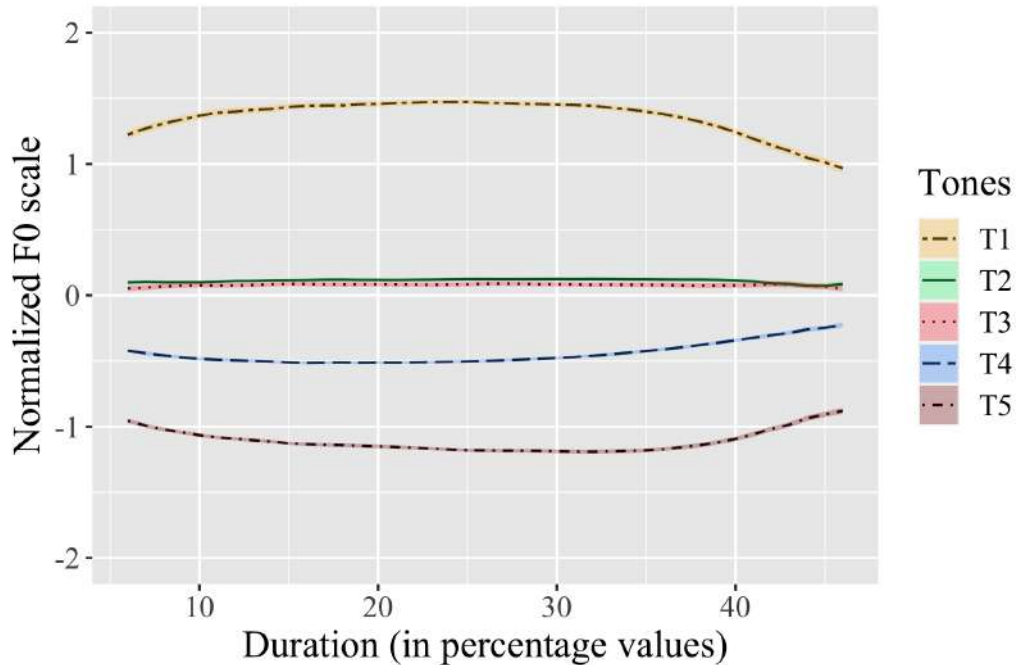


Figure 8-2: Mean (mid 80%) of normalized F_0 contours of Angami tones.

tinct level tones, for the following classification of tones using Random Forest, all five tone categories will be analyzed. This is partly due to the fact that Random Forest classifications work the best when the class sizes are comparable.

8.3.2 Feature Importance using Random Forest

Based on the variable importance plot derived from the Random Forest (RF) analysis, it elucidates the significance of F_0 -based features in discriminating among the five tones within Angami. Among the 18 proposed features, this plot emphasizes the top 10 features in terms of their importance within the model. Notably, the plot reveals that the features related to F_0 Slope and rate of change demonstrate comparatively lower importance in distinguishing between the five tones. Contrarily, features such as F_0 mean, DCT coefficient, and F_0 Height emerge as the most pivotal attributes in this discrimination process. This robustly supports the conclusion that the tones within Angami primarily adhere to a level tone structure, dismissing the significance

of contour tones within this tonal system.

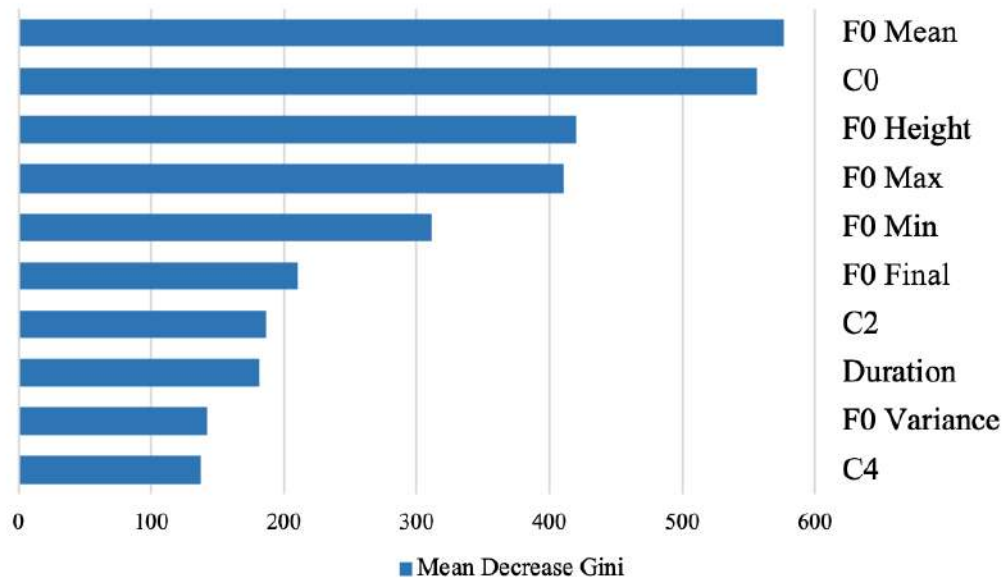


Figure 8-3: Variable importance of top 10 features for tones in Angami.

Table 8.1: Tonal space of the five tones in Angami between male and female speakers in Hertz.

	T1-T2	T2-T3	T3-T4	T4-T5
Female	30.8	1.8	11.9	20.6
Male	22.3	0.3	10.1	9.2

8.3.3 Classification of tones using Random Forest

The experimental setup and the results computed to classify the five level tones in Angami are discussed in this subsection. Figure 8-1 shows the overall framework of the Angami tone classification system. The F_0 -based features such as DCT coefficients from F_0 contours, statistical features and rate of change along with duration are considered as the whole feature set. This feature set is then passed to the RF classifier to distinguish the five level tones of Angami. Table 8.2 shows the tone distribution tokens across the five tones. As the RF classification was conducted with the set

Table 8.2: Distribution of the number tones in the tone classification study.

	T1	T2	T3	T4	T5	Total
no. tokens	1075	1078	1077	1075	1072	5377

parameters as discussed in section 8.2.2, 5-fold cross-validation was performed where the tokens were split into a 80 : 20 for training and testing in each of the folds. The overall accuracy across the five tones gives a decent performance of 66.4%, which is around three times above chance probability. Further analysis of the results shows that the performance of T1 and T5 are significantly high in comparison to the three intermediate tones. The distinctiveness of T4 within the tonal system of Angami contributes significantly to higher accuracy in classification compared to T2 and T3. Figure 8-4 shows the confusion matrix across the five level tones using RF classifier, where the accuracy rates are reported in percentage (%). The confusion matrix confirms our earlier conclusion that T2 and T3 are of the same category, as we see that there are a great deal of in between misclassification. T2 and T3 exhibit similarities in their acoustic features, the classifier might struggle to differentiate between them, leading to lower accuracy rates. This could be due to the proximity of these tones in the tonal space, or the subtle nuances that the classifier finds challenging to discern.

By treating T2 and T3 as a single category, the model doesn't struggle with differentiating between these closely related tones, resulting in a more accurate classification across the dataset (see Figure 8-4). The rise in average accuracy to 80.8% following the merging of T2 and T3, demonstrates the effectiveness of simplifying the tonal categories for the classifier to achieve a better overall performance in tone recognition.

8.3.4 Statistical Analysis

Qualitative and exploratory statistics were calculated using R's statistical packages. Statistical modeling, like the Linear Mixed Effects (LME) models, was carried out

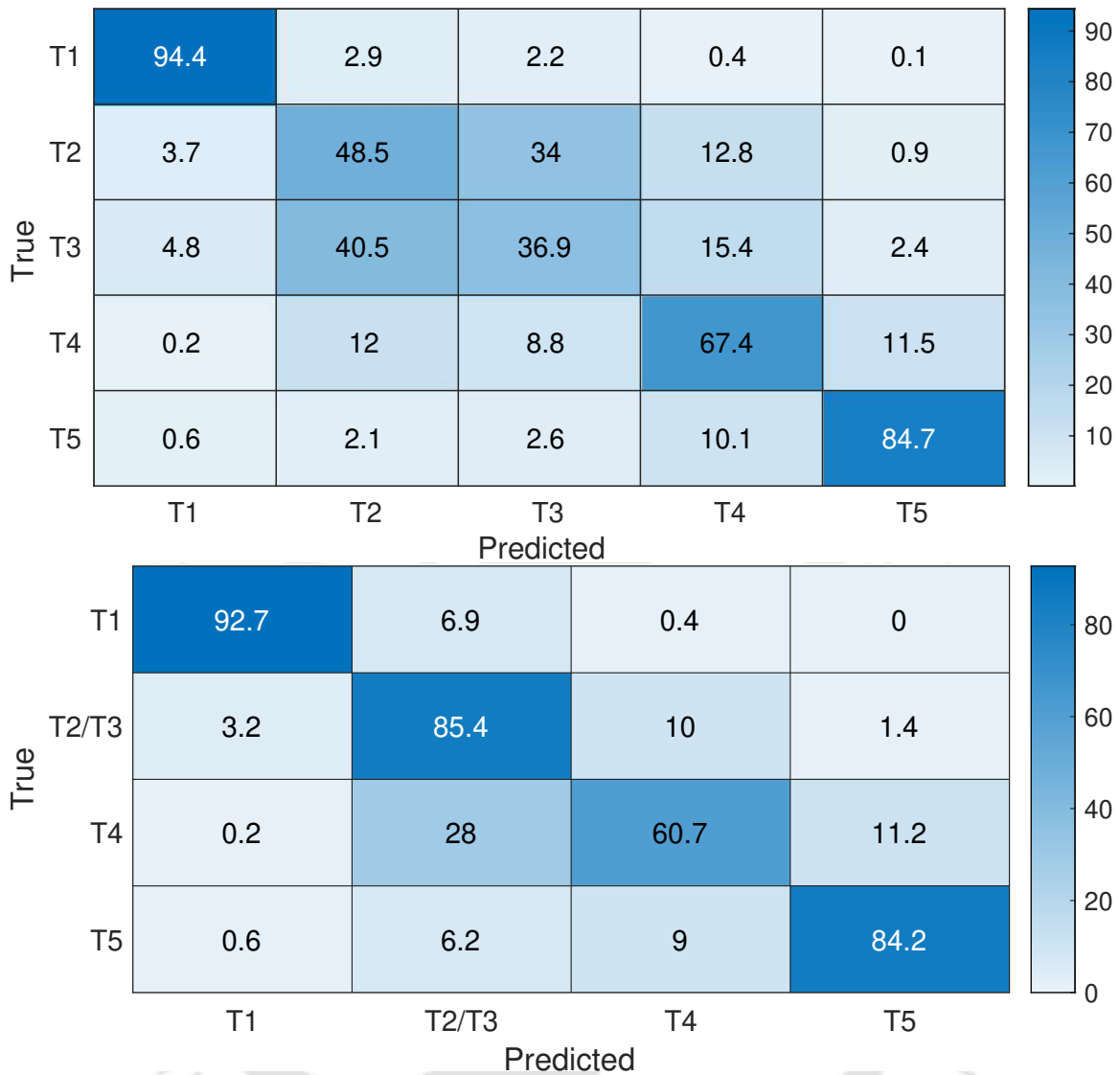


Figure 8-4: Confusion matrix in %. (top) Five level tones, (bottom) Tones T2 and T3 merged as one unit.

using the lme4 package on R (Bates et al., 2015; R Core Team, 2022). The F_0 contour of the tone plots were generated on R with the ggplot2 package (Wickham, 2016). Using the ANOVA function in the *car* package, Type II Wald chi-square tests were performed to determine the significance of the fixed effects (`car`). To estimate the contrasts between the levels in the fixed factors, Bonferroni post-hoc tests were performed using the emmeans package in R (Lenth, 2019).

In order to confirm the statistical significance of the F_0 features on the Angami tones, we conducted separate exploratory modeling where the top 10 F_0 features

($\Delta features$) were the dependent variables, tones were the fixed effects, and the speaker was the random effects as shown in 8.4.

$$\Delta features \sim Tones + (1|Speaker) \quad (8.4)$$

Type II Wald Chi-square tests for analysis of deviance test conducted on the ten LME models separately showed that tones have a significant effect on all F_0 features, as seen in Table 8.3. The LME models were further subjected to Bonferroni post-hoc test to see the differences between pair-wise contrasts of the tones. Table 8.3 shows the estimates, standard errors (SE), degrees of freedom (df), t-ratios, and p-values of tone contrasts. Statistically significance effect is seen in the F_0 features such as F_0 mean, C_0 , F_0 height, F_0 max, F_0 min and F_0 final for the tonal pairs, T1-T2, T3-T4 and T4-T5. Whereas, in all cases, tones T2-T3 do not have any significant effect.

8.4 Discussion and Conclusion

The analysis of pitch contours in Angami reveals the presence of four distinct level tones. However, to confirm that, a perception experiment with the native speakers maybe be necessary. Nevertheless, acoustic evidence indicates an absence of differentiation between tones T2 and T3 within the Angami tonal system. This finding underscores the potential convergence or similarity between these two tonal categories, emphasizing the need for further investigation through auditory perception studies within the native Angami speaker community to solidify these observations. The pitch trajectories of these two tones exhibit significant overlap, with no discernible differences in duration or voice quality to distinguish between these categories. This observation in the acoustic domain was reinforced by conducting automatic classification experiments. Employing a random forest classifier specifically designed for categorizing the Angami tones, the results indicated a high overall accuracy in tone re-

cognition. However, notably, the classifier faced a challenge in distinguishing between T2 and T3, underscoring the difficulty in differentiating these specific tonal categories based on the parameters utilized within the classification model. It appears that within the tonal space of Angami, the tones situated farthest apart exhibit higher classification accuracy. However, the intermediate tones pose a challenge for classification accuracy. Notably, the tonal spaces among the intermediate tones, specifically T2/T3 and T4, demonstrate a smaller magnitude of difference compared to the gap between the highest tone, T1, and the subsequent highest, T2. The divergence in hertz between tone categories categorized by gender is outlined in Table 8.1. This discrepancy suggests that while the extremes in tonal categories present clearer distinctions, the intermediate tones pose a greater difficulty in classification due to their closer proximity in the tonal space.

It has been established that Angami primarily consists of level tones, despite the subtle contours observed at the onset and conclusion of the pitch tracks, which have been attributed to consonantal perturbation. This conclusion aligns with the findings from the random forest classifier, highlighting crucial insights: the key features contributing significantly to tone classification are F_0 mean, F_0 height, and F_0 Max. This reaffirms that factors such as F_0 slope or variation do not significantly contribute to aiding in the classification process. Hence, the prominence of specific fundamental frequency attributes underscores the dominance of level tones within the Angami tonal system.

In one of the Tibeto-Burman language, Mizo, the F_0 features which were found to be of importance were F_0 slope, F_0 height and F_0 variance. Since Mizo have both level (high, low) and contour tones (rising, falling). For the level and contour tones, F_0 variability were less and high in Mizo, respectively. F_0 height were distinct for level tones, whereas the F_0 slope differs for the falling and rising tones (Gogoi et al., 2020a).

Table 8.3: Analysis of Deviance (ANOVA) tests on the LME models, with the top six F_0 features as dependent variables and pair-wise contrasts estimated from post-hoc Bonferroni tests.

Model*	Contrasts	Estimate	SE	df	t -ratio	p -value
F_0 mean $\chi^2(4)=16726^*$	T1-T2	1.2	0.02	5373	63.0	< 0.0001
	T2-T3	0.03	0.02	5373	1.6	1.0000
	T3-T4	0.5	0.02	5373	26.4	< 0.0001
	T4-T5	0.8	0.3	1511	3.0	< 0.05
C0 $\chi^2(4)=16726^*$	T1-T2	7.6	0.1	5373	63.0	< 0.0001
	T2-T3	0.2	0.1	5373	1.6	1.0000
	T3-T4	3.3	0.1	5373	26.4	< 0.0001
	T4-T5	4.2	0.1	5373	33.5	< 0.0001
F_0 height $\chi^2(4)=11090^*$	T1-T2	1.4	0.0	5373	52.3	< 0.0001
	T2-T3	0.0	0.0	5373	1.4	1.0000
	T3-T4	0.6	0.0	5373	24.2	< 0.0001
	T4-T5	0.6	0.0	5373	22.0	< 0.0001
F_0 max $\chi^2(4)=11112^*$	T1-T2	1.4	0.2	5371	55.32	< 0.0001
	T2-T3	0.0	0.0	5371	0.4	1.0000
	T3-T4	0.5	0.0	5371	21.0	< 0.0001
	T4-T5	0.6	0.0	5371	24.0	< 0.0001
F_0 Min $\chi^2(4)=3852^*$	T1-T2	0.8	0.0	5371	26.3	< 0.0001
	T2-T3	0.0	0.0	5371	1.1	1.0000
	T3-T4	0.3	0.3	5371	11.1	< 0.0001
	T4-T5	0.7	0.3	5371	22.0	< 0.0001
F_0 Final $\chi^2(4)=3630^*$	T1-T2	0.9	0.0	5373	29.0	< 0.0001
	T2-T3	0.0	0.0	5373	0.1	1.0000
	T3-T4	0.3	0.0	5373	10.0	< 0.0001
	T4-T5	0.6	0.3	5373	20.2	< 0.0001
C2 $\chi^2(4)=676^*$	T1-T2	-0.5	0.0	5373	-12.6	< 0.0001
	T2-T3	-0.0	0.0	5373	0.9	1.0000
	T3-T4	-0.3	0.0	5373	-7.5	< 0.0001
	T4-T5	-0.1	0.0	5373	-2.5	< 0.1419
Duration $\chi^2(4)=387^*$	T1-T2	-4.7	1.9	Inf	-2.3	< 0.1904
	T2-T3	5.3	2.0	Inf	2.7	0.0781
	T3-T4	-19.9	2.0	Inf	-10.0	< 0.0001
	T4-T5	38.7	2.0	Inf	19.4	< 0.0001
F_0 variance $\chi^2(4)=296^*$	T1-T2	0.2	0.0	5371	13.0	< 0.0001
	T2-T3	-0.0	0.0	5371	-0.9	1.0000
	T3-T4	0.3	0.0	5371	2.9	< 0.0344
	T4-T5	-0.0	0.0	5371	-1.6	1.0000
C4 $\chi^2(4)=476^*$	T1-T2	-0.2	0.0	5373	-9.6	< 0.0001
	T2-T3	-0.0	0.0	5373	-1.3	1.0000
	T3-T4	-0.0	0.0	5373	-4.7	< 0.0001
	T4-T5	-0.0	0.3	5373	-5.0	< 0.0001

p < 0.001 is indicated by an asterisk

Chapter 9

Perceptual correlates of tones in Angami

This chapter investigates the perception of tones among native Angami listeners. The study aims to explore the primary cues for categorical perception of Angami tones, particularly in cases where the tone system is purely level. Two separate experiments were conducted: the first involved natural speech stimuli produced by male and female Angami speakers, while the second involved a synthetic tone continuum with synthesized fundamental frequency (F_0) and duration. Results were based on listeners' judgments, and error patterns were also analyzed to determine which tones were confused. The result also includes the reaction time of the responses. Both experiments yielded similar results, demonstrating that native Angami listeners could perceive four tone categories in the language. The perceptual correlates of tones in Angami as seen with the synthesized stimuli includes both relative F_0 and duration of the tones. This study further corroborates earlier findings regarding tone production in Angami, as well as the classification of tones within the language.

9.1 Introduction

Categorical perception (CP) of speech occurs when listeners categorize speech sounds into discrete categories. This phenomenon was demonstrated by Liberman et al. (1957), where features of stop consonants, such as voicing and place of articulation (PoA), were perceived categorically. In short, Categorical perception refers to the phenomenon where small acoustic differences lead to perceptual distinctions between phonetic categories but not within a phonetic category, resulting in sharp boundaries between categories in perception (Liberman et al., 1957). Existing study on the categorical perception of tones shows that pitch height (F_0) and pitch movement (F_0 contour) are an important cue to distinguish level and contour tones (Gandour, 1978). In a cross-linguistic investigation of tone perception encompassing tonal languages such as Thai and Yoruba, the fundamental frequency is found to be the primary cue for perceiving level tones (Gandour, 1978). Specifically, in Thai, mid and low tones are categorized as level tones, while in Yoruba, all three tones (high, mid, and low) are considered level tones. The duration or increase in amplitude did not result to help in the identification of the tones.

In regard to pitch movement of the level tone (one level to another), Abramson (1977) noted that the level tones continuum were not perceived categorically. Abramson's investigation into the categorical nature of tone perception focused on three Thai level tones: low, mid, and high tones (Abramson, 1977). Synthetic syllables were created by imposing sixteen fundamental frequency levels ranging from 92 Hz to 152 Hz in steps of 4 Hz onto the syllable type [kha:]. These synthetic syllables were presented to 33 native speakers of Thai for identification, who were asked to choose one of five possible words, with the intended words being the three-level tones. The discrimination study was conducted using the 4-interval forced-choice (4IAX) test of pair similarity. Abramson's study revealed that the perception of Thai tonal categories is not categorical, which suggests that listeners do not rely solely on discrete

categories to perceive and distinguish between tones (Abramson, 1977). On the contrary, Wang (1976), found that Mandarin ‘high level’ and ‘high rising contour’ tones were perceived categorically. Sun and Huang (2012) argues that the result of the two studies i.e Abramson (1977) and Wang (1976) are complimentary and not contradictory, where the authors argue that the categorical perception of contour tones is explained by the rapid changes in pitch height, such as rising or falling contours. Conversely, the non-categorical perception observed in Thai for level tones can be attributed to their continuous patterns, resembling vowels, where the pitch height remains constant throughout their duration. Khouw and Ciocca (2007) investigates the perception of Hong Kong Cantonese contrastive lexical tones, encompassing three contours and three level tones. The level tones in Cantonese include high level (Tone 55), mid-level (Tone 33), and low level (Tone 22), while the contour tones consist of high rising (Tone 25), low falling (Tone 21), and low rising (Tone 23). The study measured F_0 at eight consecutive sections of the vocalic segment, determined by nine positions with 12.5 Hz differences in each position. Results indicated that contour tones were identified more accurately, with a percentage ranging from 75% to 84%, compared to level tones, which ranged from 58% to 70%. Additionally, the study revealed that the direction of F_0 change and the magnitude of F_0 change serve as perceptual cues for distinguishing between contour and level tones and among contour tones. Furthermore, it demonstrated that the sixth and seventh sections at the end of the vocalic segment play a significant role in tone perception.

Tone perception is not only influence by the height and the movement of pitch but also by the phonation types. In the study of the Mon-Khmer language Kammu, including its Northern, Western, and Eastern dialects, dialectal variation in tone perception is evident. While Northern and Western Kammu speakers rely on fundamental frequency (F_0) differences to distinguish words (e.g., High tone for ‘eagle’ and Low tone for ‘stone’), even with minimal F_0 differences, Eastern Kammu speakers

rely on contrasting voicing (voiced and voiceless) of the initial consonants (Svantesson and House, 2006). A similar study of dialectal perception differences has been reported in Northern and Southern Vietnamese as well. In addition to pitch height (F_0), phonation types were found to be important perception cues for Northern Vietnamese speakers (Brunelle, 2009b).

Influence of the language (tone and non-tonal language) also impact the perception of tones. Research on tone perception among speakers of tone and non-tone languages consistently demonstrates that individuals who speak tone languages exhibit a significant ability to perceive lexical tones compared to those who do not (Li, 2016; Gandour and Harshman, 1978; Sun and Huang, 2012; Xu et al., 2006). Li (2016) conducted a perception study involving English and Thai speakers' perception of four Mandarin tones: high level (T1), mid-rising (T2), falling rising (T3), and high falling (T4). In this tonal language context, Thai speakers outperformed non-tonal English speakers in the perception of Mandarin tones. Thai speakers exhibited more confusion between Mandarin tones T2 and T3, which have shorter durations than their counterparts in Thai tones. Conversely, English speakers displayed confusion between tones T1-T2 and T2-T3, as these tones share phonetic similarities. This study revealed that non-tonal language speakers, such as English speakers, rely primarily on pitch height, while tonal language speakers depend on pitch height and direction. Similar findings have also been shown where Taiwanese speakers perceived the high-level to high falling tones better than the English speakers (Sun and Huang, 2012). Gandour and Harshman (1978) using a multidimensional scaling method (MDS), noted that tonal language (Thai & Yoruba) speakers paid more attention to the pitch contour direction than the non-tonal language (American English) in the tone perception.

Given the existing literature on tone perception and the diverse findings, it is essential to investigate the perception of tones in languages that have purely four or five level tones. It is important to determine whether pitch height (fundamental

frequency) alone serves as a cue, or if other factors such as duration also influence the identification of the tones. As seen in the previous chapter (Chapter 7), the native speakers of Angami makes four distinct tone production as opposed to the five stated in the existing literature. In the production of Angami tones, it is noted that there is overlapping of F_0 for the intermediary T2 and T3 and the tones are not distinct. Hence, two distinct perception tests were conducted to ascertain whether native speakers perceive four or five tones. The primary aim was to investigate if the native Angami listeners are able to perceive four or five tones. The overarching objective of the experiment was to explore whether there is a concurrence between the production and perception of tones by Kohima village natives. The perception test employed natural stimuli in the first experiment, designated as E1 hereafter. Conversely, synthetic stimuli were utilized in the second experiment, denoted as E2. Both experiments utilized a forced-choice identification task as the method for the perception study. Through these experiments, the research sought to uncover any discrepancies or similarities between how tones are produced and perceived by native speakers in Kohima village.

The study accounted for both correct and incorrect identifications of tones by the participants. Incorrect identifications provided valuable information regarding which tones (options) the stimuli were confused with. Additionally, the reaction times of the participants were analyzed for further insights.

- Shorter reaction times might be expected for stimuli where the identification is more consistent or where participants easily differentiate between tones, perhaps due to distinctiveness in frequency ranges or duration.
- Conversely, longer reaction times could occur when stimuli are perceived ambiguously or when participants find it challenging to categorize tones due to overlapping frequency ranges or duration.

9.2 Perception of Angami tones using natural stimuli (E1)

In this first perception study, natural stimuli of the tones in Angami were presented to the listeners. This section is organized as follows: Section 9.2.1 accounts for the procedure and methodology used, whereas Section 9.2.2 accounts for the result. Finally, Section 9.2.3 discusses the findings.

9.2.1 Methodology

The stimuli for the perceptual experiment were contributed by two native speakers from the Kohima village. The male speech stimuli were provided by a 43-year-old individual, while the female speech stimuli were contributed by a 53-year-old woman. Both individuals hail from Tsituonuomia Khel (T Khel) and hold professions as language educators. The male speaker has over 15 years of experience teaching at the undergraduate level, while the female speaker has been a high school teacher for the past 31 years. The selection of the speakers was carefully carried out, as their extensive teaching backgrounds underscore their familiarity and expertise in the language, contributing valuable authenticity to the stimuli used in the experiment.

For eliciting the tone stimuli, a set of five lexical tones, /pe/ encompassing all five tones was embedded into sentences, and phrases, and presented. The target tones were reiterated three times. The lexical tones produced in isolation were segmented and extracted for stimuli in this experiment. The frame for the stimuli remained the same as in 'I said X' ('*a X puba*'). A non speech sound along with the frame of the stimuli and the selected tone stimuli were concatenated using PRAAT. Male stimuli and female stimuli experiments were conducted separately. However, the structure of the experiment was the same. The mean F_0 contours' of the tones are shown in Figure 9-1. Since one of the goals of this study was to determine whether tone

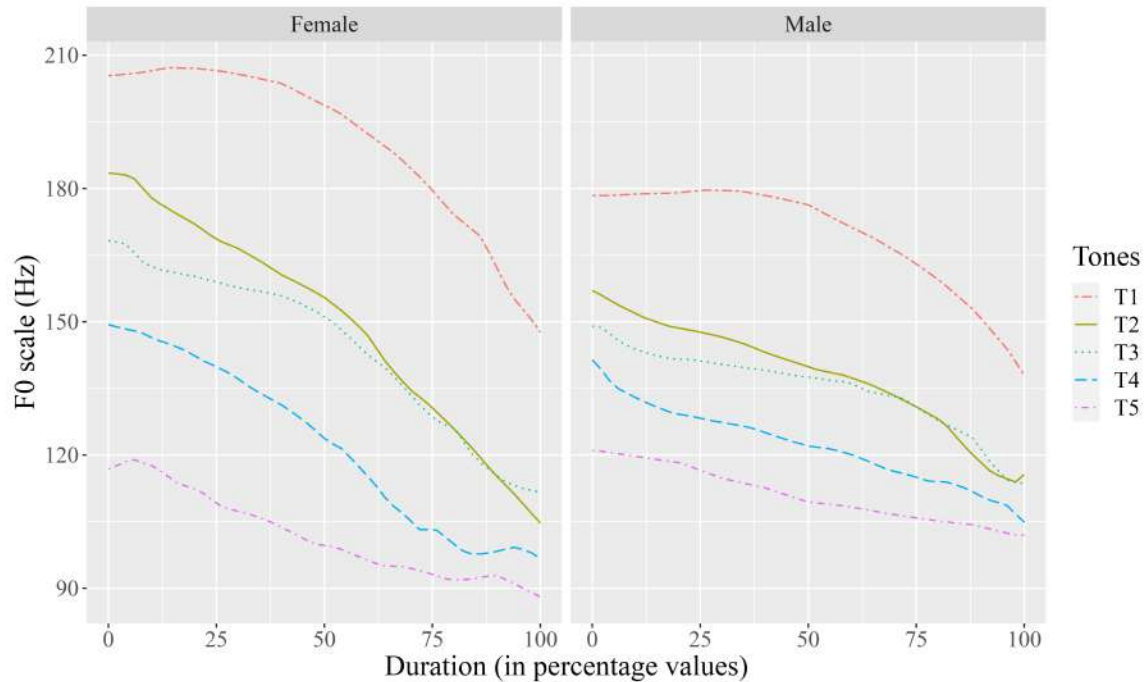


Figure 9-1: F₀ contour of the five tones stimuli produced by a female and male Angami speaker

duration acts as a cue for differentiating across tones, the stimulus duration was left uncontrolled. The duration of the tones is shown in Figure 9-2. There are five tones (T1, T2, T3, T4 & T5) and each tone were repeated 10 in random order. A total of 100 speech stimuli were used for the perception test. 100 speech stimuli were used in the perception test; the stimuli were presented in a random order. This large stimulus played a crucial role in providing a thorough and careful investigation of listeners' perceptions and categorization of the tones in Angami.

9.2.1.1 Participants

In the perception test (E1), 27 listeners who are native Angami from Kohima village participated, comprising 16 females and 12 males. The age range from 16 to 66 years, with an average of 28 years (SD = 13). The participants have a minimum degree from high school. All subjects were multilingual; they all know their mother tongue (L1) which is Angami and also English and Nagamese. The experiment garnered a total

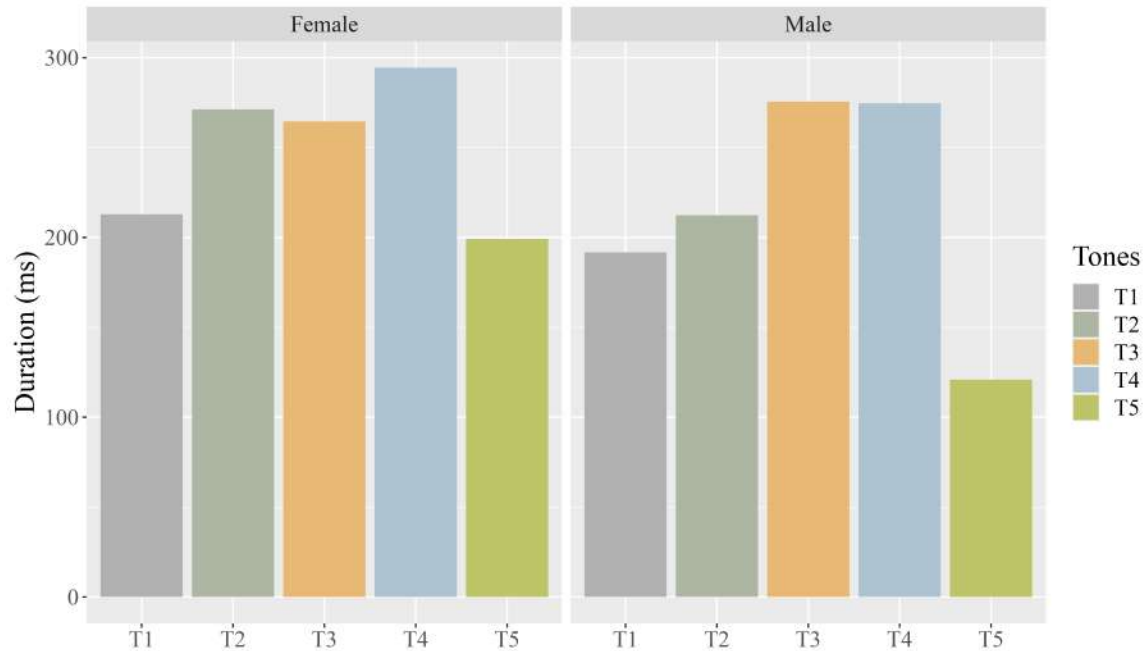


Figure 9-2: Duration of the five tones stimuli as produced by a female and male Angami speaker

of 2,800 responses from the participants, forming the basis for subsequent analyses.

9.2.1.2 Procedure

Each listener was trained to the stimuli using PRAAT. The listeners were asked to familiarize themselves with the tones and differentiate them. The training period ranges from 1 minute to 3 minutes. Using PRAAT, each listener was trained to respond to the stimuli. The participants were asked to become familiar with the tones and differentiate them. The duration of the training session ranges from one to three minutes. No specific threshold was established for familiarity with the stimuli, as all the listeners were already familiar with the provided stimuli. The perception test was conducted in the homes of the participants, church buildings, and community halls.

The experiment was conducted using a PRAAT script which was run on a PANASONIC laptop and the sound was played through SHURE professional headphones (model-

SRH1440) at a comfortable hearing range. The script design allowed subjects to initially listen to the stimuli with a blank display on the laptop. Following a two-second pause, the screen displayed the alternatives that corresponded with the stimuli. By purposefully arranging the stimuli in such a way, participants were able to concentrate only on auditory perception, since they initially heard them without any visual signals. After a short delay, options were displayed to them, making it easier for them to choose their response without any visual disruption and preserving the controlled experimental setting. The display consists of the meanings in English of the five lexical words for the sound /pe/ in Angami (incline, bridge, fat, tremble, shoot) as shown in Figure 9-3. The participants were instructed to select, by clicking on the screen, one of the five meanings that corresponded to the sound they heard. One-time replay functionality was included in the experiment so that participants could hear the stimuli again if necessary. Participant responses and reaction times (RT) in selecting the displayed options were recorded and analyzed to derive the results. These responses and duration of response taken contributed to understanding both the accuracy of responses and participants' decision-making processes while distinguishing the five tones. The perception study (E1) took an average 20 minutes.

9.2.2 Results of the perceptual study

The overall responses of the experiment (E1) is summarized in Table 9.1. It's evident from the table that the two extreme tones, T1 and T5, were accurately identified by participants, with correct perception percentages of 95% and 97%, respectively. Despite being an intermediary tone, T4, showed a higher response accuracy as compared to the other two intermediary tones (T2 and T3). Stimuli associated with T4 were correctly identified as T4 with a 64% accuracy rate, and they were occasionally mistaken as T5, accounting for 19% of the responses. This observation might be attributed to the similarity in the fundamental frequency (F_0) between T4 and T5,

particularly their lower F_0 ranges. We have seen in previous chapter 7, that in the production T2 and T3, the F_0 contour was overlapping in the tonal space and they were not distinct. Similarly, in this perception test, we see that participants were not able to correctly identify the two tones. Notably, stimuli associated with T2 were correctly identified as T2, with an accuracy rate of 54%. However, a significant portion, 32% of the time, of these stimuli were identified as T3. On the other hand, stimuli associated with T3 are corrected identified as T3 with an accuracy of 34% and 57% of the time, the stimuli were identified as T2. This classification might suggest a certain similarity or overlap in the acoustic characteristics between T2 and T3, leading to confusion or ambiguity in perception for some participants.

Table 9.1: Response of the perception test in percentage % values and average reaction time

	T1	T2	T3	T4	T5	<i>RT (ms)</i>
T1	93.21	0.54	0.54	5.18	0.54	<i>1.9</i>
T2	0.71	55.00	32.32	11.61	0.36	<i>2.0</i>
T3	0.36	57.68	34.11	6.96	0.89	<i>2.1</i>
T4	3.04	2.86	13.39	61.96	18.75	<i>2.4</i>
T5	0.00	0.36	0.71	1.07	97.86	<i>1.7</i>

Table 9.2: Reaction Time for every stimulus and responses in the experiment (E1)

	T1	T2	T3	T4	T5
T1	1.8	3.1	0.9	2.9	0.9
T2	3.9	1.9	2.0	2.3	0.8
T3	1.5	2.1	2.2	2.4	4.2
T4	3.0	4.0	2.6	2.2	2.2
T5	0.0	4.9	2.3	4.7	1.6

We also look into the reaction time while selecting the responses, which is shown in the right column of the Table 9.2. Interestingly, compared to the intermediate tones, participants were able to identify T1 and T5 in less time. Given the reduced reaction time, it is possible that the extreme tones, T1 and T5, are more distinct, which enable participants to identify them more quickly. Conversely, the intermediary tones (T2,

T3 and T4) might share certain acoustic similarities, which leads to a slightly longer processing time for participants discerned between these tones.

To see if the reaction time is statistically significant to the responses, an LME model was constructed where RT is a dependable variable, responses were fixed effects and stimuli types and listener were the random effects (see equation 9.1). One way ANOVA results shows that the RT is significant ($\chi^2(4) 28.8, < 0.001$). The model was further subjected to see the pairwise comparison, where we see that RT is significant only for the pair T4-T5, as seen in Table 9.3.

$$RT \sim Responses + (1|StimuliTypes) + (1|listener) + (1|CorrectResponses) \quad (9.1)$$

Table 9.3: Result of post-hoc Bonferroni pairwise comparison conducted on an LME model with reaction time (RT) as the dependent variable and the responses as fixed effects.

contrast	estimate	SE	df	t-ratio	p-value
T1-T2	-0.16	0.12	1895	-1.4	1.0000
T2-T3	-0.02	0.12	2780	-0.2	1.0000
T3-T4	-0.28	0.13	2681	-2.1	0.3406
T4-T5	0.61	0.11	2781	5.2	< 0.0001

The results were further analyzed to see if there is any tone perception differences based on male stimuli and female stimuli. Table 9.4 shows the result of the responses based on the stimuli types, indeed demonstrates intriguing findings. It shows that regardless of male or female stimuli, tones T1, T4, and T5 were recognized with comparable percentages. However, an interesting disparity arises concerning the two other intermediary tones, T2 and T3. Among the female speech stimuli, tone T2 is correctly identified as T2 in 58.6% of cases, while being identified as T3 and T4 in 21.1% and 18.6% of cases, respectively. On the other hand, tone T3 is accurately identified as T3 in 42.1% of instances and recognized as T2 in 51.8% of cases. In

the case of male speech stimuli, tone T2 is correctly recognized as T2 and T3 with accuracy rates of 21.4% and 43.6%, respectively. On the other hand, tone T3 is identified as T2 and T3 with accuracy rates of 63.6% and 26.1%, respectively. Overall, we see that there is a higher tendency for confusion between the male stimuli of tones T2 and T3. This suggests that distinguishing between T2 and T3 among male speech stimuli poses a challenge compared to other tones.

Table 9.4: Responses of the perception test based on stimuli produced by female speaker and male speaker in percentage % values

	Stimuli Female					Stimuli Male				
	T1	T2	T3	T4	T5	T1	T2	T3	T4	T5
T1	93.6	1.1	0.4	4.3	0.7	92.9	0	0.7	6.1	0.4
T2	1.1	58.6	21.1	18.6	0.7	0.4	51.4	43.6	4.6	0
T3	0.7	51.8	42.1	4.0	1.4	0	63.6	26.1	10.0	0.4
T4	1.8	1.1	14.3	62.1	21.0	4.3	4.6	12.5	61.8	16.8
T5	0	0	0.4	1.2	98.6	0	0.7	1.1	1.1	97.1

To see if the reaction time of the two types of stimuli differs, the RT was plotted against each stimulus and response as seen in Table 9.5. There are no significant RT differences that can be observed. Statistical test also showed that there is no significant effect of RT based on the interaction of stimuli types and the responses ($\chi^2 (4) 3.1, 0.54830$).

Table 9.5: Reaction time based on stimuli produced by female speaker and male speaker

	Stimuli Female					Stimuli Male				
	T1	T2	T3	T4	T5	T1	T2	T3	T4	T5
T1	1.6	3.1	0.0	2.1	0.5	1.9	0	1.5	3.5	1.7
T2	1.2	1.8	1.9	1.8	0.8	11.8	2.1	2.0	4.1	0
T3	1.5	2.2	2.1	2.9	5.2	0	2.1	2.5	2.2	0.3
T4	2.4	6.3	2.1	2.2	1.9	3.3	3.5	3.1	2.3	2.5
T5	0	0	0.0	6.4	1.6	0	4.9	3.0	2.9	1.6

9.2.3 Discussion

This perception experiment's (E1) results point to a noteworthy trend: participants demonstrated a higher accuracy in distinguishing between the extreme tones, T1 (high tone) and T5 (low tone), compared to their identification of the intermediary tones (T2, T3, and T4). This heightened accuracy in discerning between T1 and T5 suggests a distinctiveness in these extreme tones, indicating that the associated fundamental frequency (F_0) features are distinguishable and perceptually salient.

On the other hand, there may be less accuracy in differentiating between the intermediate tones (T2, T3, and T4) due to fewer differences in the tonal space and or more subtle F_0 properties. This finding confirms that the extreme tones in the tonal system are perceptually unique, supporting the theory that their F_0 characteristics are more easily perceived than those of the intermediate tones. As observed in Table 9.4, we see that stimuli T2 is identified as either T2 and T3; similarly, stimuli T3 is identified either as T3 or T2. Rather than framing it as a 'misclassification', the observed pattern might indeed support the hypothesis that the language comprises only four distinct tones rather than five. It's possible that T2 and T3 operate as a single category inside the language's tonal system and the native speakers and listeners of Angami are unable to distinguish them, both in production and perception. This finding supports the theory that these two tones may not have obvious perceptual borders and may instead be interpreted as a single category, strengthening the case for a four-tone system as opposed to a five-tone one.

In this study on tone perception in Angami has some interesting results, but it also has several limitations that call for more research. The tones that were employed as stimuli were produced by two people, each of whom probably has their own intrinsic fundamental frequency (F_0) ranges. As teachers, both the male and female Angami speaker have carefully tried to produce the five tones and tried to differentiate the tones T2 and T3. There could be subjective biases. Furthermore, there seemed to

be noticeable differences in tone duration within the stimuli set. Particularly, the durations of the low tone (T5) and high tone (T1) were shorter than those of the intermediate tones. Whether tone 1 and tone 5 have a high accuracy rate due to either the F_0 ranges or the duration of the tone, which act as a cue in identifying the tones. Hence, the follow-up experiment looks into synthesized Angami tones, which takes into consideration both F_0 ranges and duration.

9.3 Perception of Angami tones using natural and synthetic stimuli (E2)

To ascertain whether Angami listeners relied solely on F_0 (fundamental frequency) for tone identification or if duration also played a role, we conducted an identification study by manipulating the stimuli. This perception test involved creating various stimuli that altered specific acoustic properties like pitch (F_0) and duration.

The stimuli included:

- **Variations in Fundamental Frequency (F_0):** Synthesis of the F_0 stimuli. 120 Hz to 330 Hz with an increase of 5 Hz at every interval (120 Hz, 125 Hz, 130 Hz....)
- **Duration Manipulation:** While keeping the F_0 range constant, the durations were altered into three categories-input short (180 ms), input mid (270 ms) and input long (500 ms). This helps in assessing whether Angami listeners rely on duration cues in addition to pitch for tone identification.
- **Natural stimuli:** The original five tones produced by a female speaker were also randomly added with the synthesized input stimuli. A threshold was set, ensuring that only participants who accurately identified the original stimuli

with a success rate of 60% or higher would have their responses considered for analysis.

9.3.1 Methodology

The stimuli (original) for this perception test is produced by a 34-year-old female speaker from Kohima village. The speaker is a language teacher by profession, have also freelance in voice over educational videos. It is to be noted that the speaker produced the speech data with utmost care and may have exaggerated the duration. The word /pe/ in five tones as produced by the speaker is plotted in Figure 9-4 and the duration's in Figure 9-5. The sound was synthesized over the lexical word /le/ in Angami. The minimal pair of the word /le/ with the tones is shown in Table 9.6. As stated in the previous section 9.3, the stimuli were synthesized not only in the F_0 ranges but also the duration. Figure 9-4 represent the synthesized duration's of the stimuli. Unlike the experiment 1 where the stimuli were embedded in a phrase, in this study, the stimuli were provided in isolation. The synthesized stimuli were presented to the participants similar to experiment 1 (see Section 9.2.1).

Table 9.6: Tonal minimal pair for the word /le/ in Standard Angami as per MKS Dieda (dictionary)

Word	Tone	Gloss
/le/	T1	peel
	T2	question
	T3	really
	T4	hot
	T5	go in

9.3.1.1 Participants

A careful selection process was conducted for participants in the perception study, focusing on Angami listeners whose primary language is Angami (L1), as well as has exposure to or have background in music. Initially, 16 native Angami participated in

the study, however only 8 were considered for analyses. It is because as the perception test is tedious, the three older participants (< 60 years) could not complete the study, 5 participants could not reach the threshold of identifying the original stimuli.

9.3.1.2 Procedure

For the second perception study, 43 F_0 stimuli were generated, each with 3 different durations, and these were repeated 4 times. Additionally, there were 5 original tone stimuli, and these were repeated 4 times. In total 536 stimuli were presented to each participant. As the study took much time and effort, after every 100 stimuli, the participants were encouraged to take a few minutes break. As we took eight participants responses into final consideration, a total of 4,288 responses were subjected to analyses. It is also to be noted that the perception test was carried out at the 3rd wave of COVID, it was not feasible to have more people for participation. Each person took an average of 1.5 hours for the experiment. The experiment was conducted using PRAAT script which was run on an Apple laptop and the sound was played through a professional headphones Audio - Technica ATH-M50x, frequency range 15 – 28,000 Hz¹.

¹Courtesy of the proprietor of Beat Device studio, Kohima.

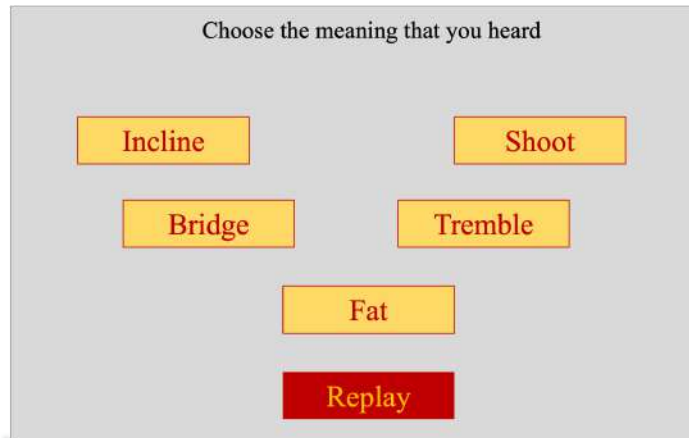


Figure 9-3: An example of the options displayed on the screen for perception study (E1)

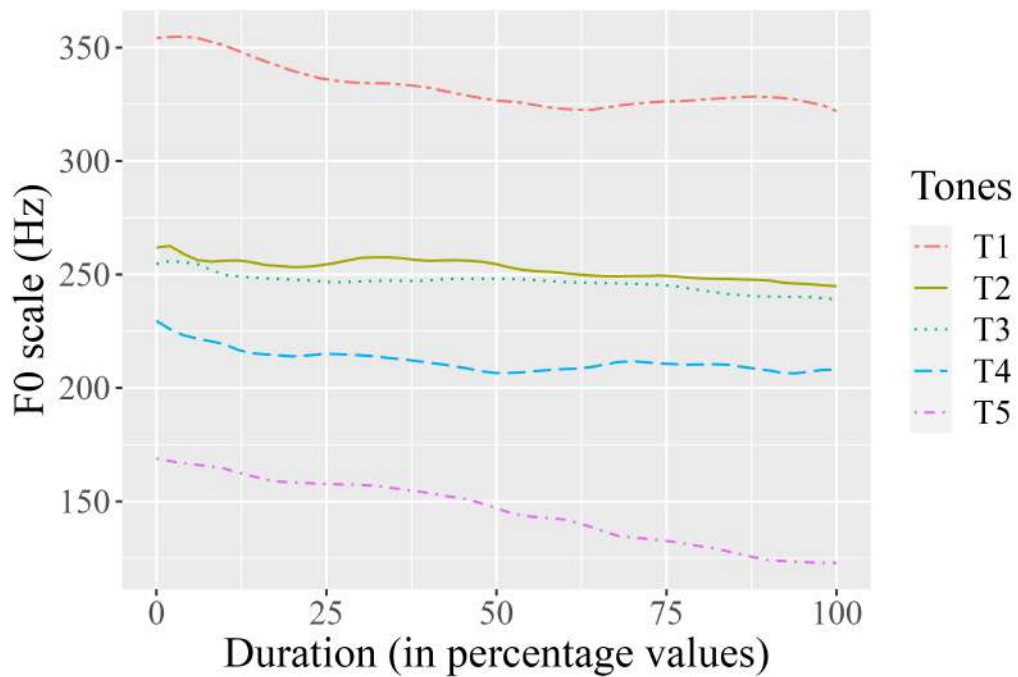


Figure 9-4: F₀ contour of the five tones stimuli for the word 'le' as produced by a female Angami speaker

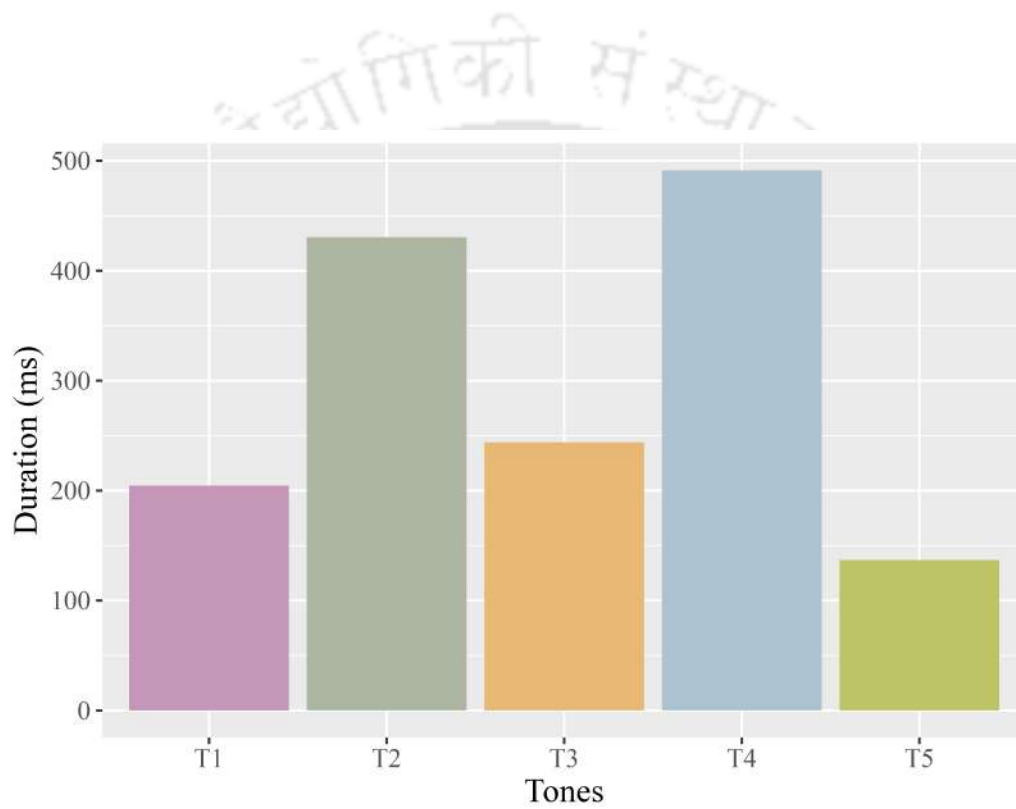


Figure 9-5: Duration of the five tones stimuli for the word 'le' as produced by a female Angami speaker

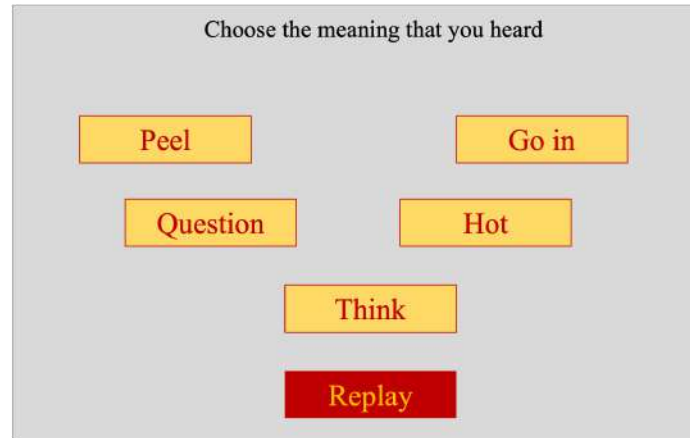


Figure 9-6: An example of the options displayed on the screen for perception study (E2)

9.3.2 Results of the perceptual study

The results will be segmented into two main sections. The initial section will focus on analyzing the responses elicited by the original stimuli produced by the female speaker. This analysis aims to understand the participants' reactions and perceptions specifically related to the stimuli produced by the female speaker. In the subsequent section, the focus will shift to examining the responses elicited by the manipulated or synthesized stimuli, particularly those involving variations in F_0 (fundamental frequency) and duration. This analysis aims to explore how participants perceived and responded to the modified stimuli compared to the original ones, shedding light on the impact of alterations in F_0 and duration on perception within the study. It will also help us to understand at what F_0 range, the participant start hearing and differentiating one tone from the other.

9.3.2.1 Natural stimuli

Table 9.7 shows the responses acquired from the five original tone stimuli. This table displays the recognition patterns and the accuracy rate represented in a percentage (%) values. The analysis of this data could provide insights into how participants

perceived and identified the original tones presented in the study. As seen in the table, high and low tones (T1 and T5) are identified relatively well at an accuracy rate of 84.4 and 82.8 %, respectively. Whereas, the intermediary tone, T4, has been identified with an accuracy of 65.6 %. However, tones T2 and T3 have low accuracy rates. The participants show confusion of the two tones as seen in the table. This result is similar to the previous perception test conducted and reported in section 9.2.2. When we look at the reaction time, it is seen that the stimuli T1 and T5 when identified correctly takes lesser time than the other three intermediary tones. It is also noted that the participants took longer time in identifying tone T2.

Table 9.7: Response of the perception test (natural stimuli) in percentage (%)

	T1	T2	T3	T4	T5
T1	84.4	12.5	14.1	0.0	0.0
T2	9.4	37.5	35.9	10.9	6.3
T3	10.9	21.9	34.4	10.9	3.1
T4	0.0	28.1	15.6	65.6	7.8
T5	0.0	0.0	0.0	12.5	82.8

Table 9.8: Reaction Time of the perception test (natural stimuli) in seconds (sec)

	T1	T2	T3	T4	T5
T1	2.7	3.9	3.7	0.0	0.0
T2	5.9	3.8	3.2	3.9	2.6
T3	2.8	3.0	3.0	3.5	4.3
T4	0.0	3.6	4.4	3.2	3.4
T5	0.0	0.0	0.0	3.0	2.4

9.3.2.2 Synthesized stimuli

As discussed earlier, this synthesized stimuli perception test looks into how different tones are perceived at varying durations and frequencies. The overall average identification of the tones based on the F_0 ranges and their three different duration is shown in Figure 9-7. When the stimuli duration is short, i.e, 180 ms, low tone T5 is identified from 120-160 Hz. Conversely, T4, usually linked with longer durations,

is identified only at higher frequencies (170, 180, and 200 Hz). Tone T2 occupies a larger F_0 range from 190, 210-250. As we have seen in the original stimuli that high tone T1 duration is also relatively short. Hence, in this table, we expect to see the identification range of T1 to be more than T2 and T4. T1 is identified from F_0 ranges 260-330 Hz.

As seen in the mid-duration stimuli, set at 270 milliseconds, tone T4 appears to be associated with frequencies between 150-200 Hz, this pattern shows that for tone T4, duration plays an important role in identifying the tones. Whereas, for high tone T1, it is identified within the 250-330 Hz range and identification of T1 may depend not only on duration but on the high fundamental frequency ranges. Additionally, the low tone T5, despite its low fundamental frequency range, is identified within the narrower frequency range of 120-140 Hz.

A distinct pattern emerges in the identification of different tones within specific frequency ranges when we look at the longer duration stimuli (set at 500 milliseconds). Interestingly, the low tone (T5), which typically isn't associated with long duration's, is consistently identified within a narrow F_0 window of 120-130 Hz. This suggests a clear distinction in low tone (T5) is perceived in longer-duration stimuli compared to its typical associations with shorter duration's. On the other hand, T4, known for its association with longer durations in the language, is identified within the frequency range of 140-200 Hz in these longer stimuli. This consistent identification aligns with its expected characteristics in longer-duration contexts. Additionally, T2 seems to be perceived in the range of 210-240 Hz, while T1 is identified within a broader frequency span of 250-330 Hz during the presentation of these longer-duration stimuli.

A detailed breakdown of how participants identified specific F_0 (fundamental frequency) ranges as one of the five tones and the percentages at which these ranges were confused with other tones is shown in Figures 9-8. This detailed analysis can provide insights into how participants perceived and categorized the tones within dif-

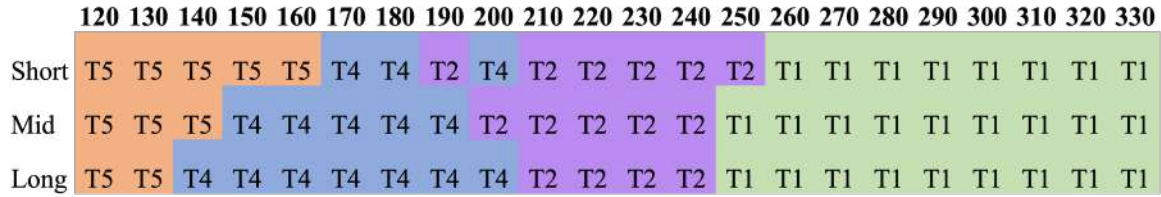


Figure 9-7: Average responses of the synthesized stimuli.

ferent frequency boundaries. For instance, if a particular F_0 range was predominantly identified as a specific tone (e.g., T1) and had a low confusion percentage with other tones, it suggests that this frequency range had distinct characteristics associated with that specific tone, making it easier for participants to correctly recognize it. The tone which is identified more amongst the other four tones is highlighted. Tone 5 is highlighted in orange, T4 in navy blue, T2 in purple and high tone T1 in light green. This figure shows that, participants did not fully eliminate tone T3, but the tone was confused especially at the range 190-250 Hz with neighboring tones such as T2 and T4 in short and mid-duration stimuli, and with T2 and T1 for long duration stimuli.

We also further look into the average reaction time for the different stimuli as responded by the participants as shown in Figure 9-9. The relatively shorter reaction times observed for lower (120-150 Hz) and higher (280-330 Hz) fundamental frequency (F_0) ranges, regardless of duration, indicate that participants might have more easily and quickly identified extreme tones (such as T5 and T1) within these frequency ranges. This suggests that these tones might have distinct and easily recognizable characteristics within these frequency boundaries, facilitating quicker perception and categorization by the participants. However, we see a longer reaction times within the F_0 ranges from 160-270 Hz, which suggests a more complex or ambiguous perceptual processing. It implies that tones within these frequency ranges, regardless of duration, might have characteristics that are harder to differentiate or categorize, leading to longer reaction times.

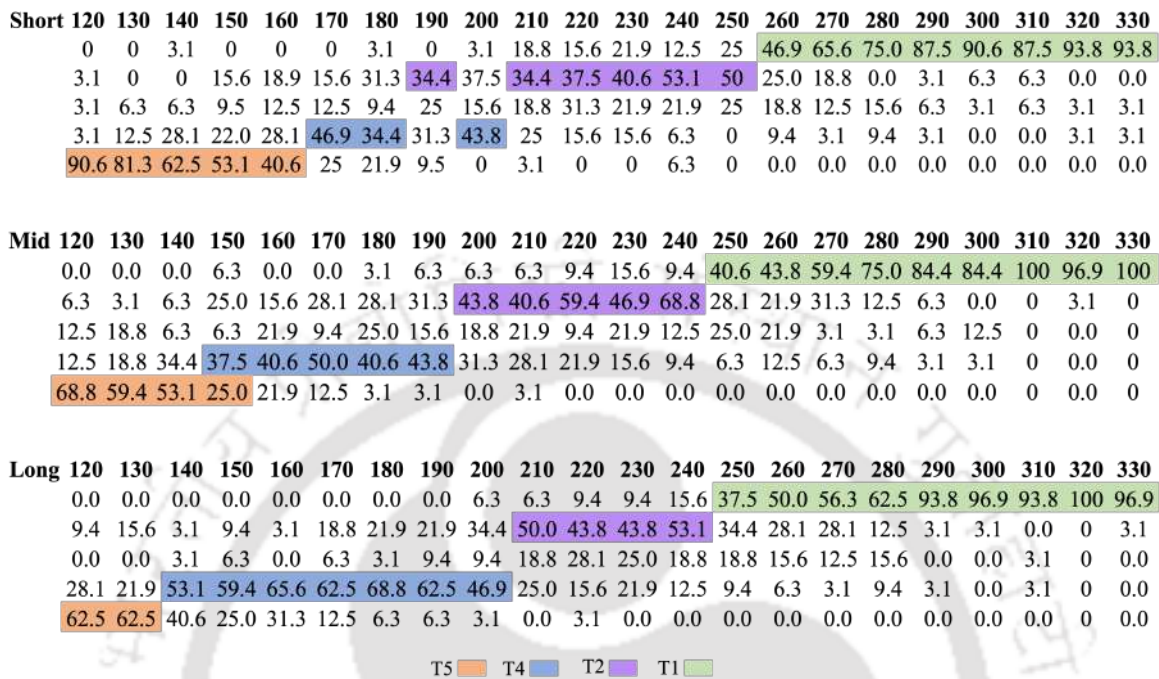


Figure 9-8: Average responses of the synthesized stimuli across short (top), mid (middle) and long (bottom) durations.

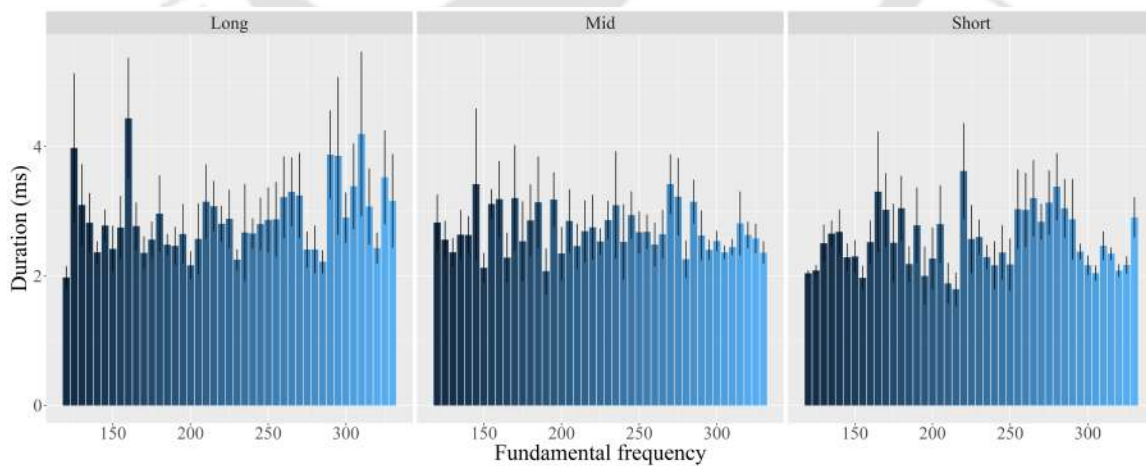


Figure 9-9: Average reaction time to the synthesized stimuli.

9.3.3 Discussion

The two perception study in this chapter shows similar result, although the methodology was different. In the first perception study (E1) using raw stimuli of a male and female speech, we see that tone T1 and T5 have higher correct responses (93.1% and 97%, respectively). This show that the relative F_0 of the two tones are distinct, and they act as a cue to the perception of tones. The study also showed that the intermediary tones occupy a lesser tonal space and hence have greater confusion rate. The intermediary tone, T4, has greater perception accuracy than tone T2 & T3, since the F_0 is distinct and as noticed, the duration of T4 is longest among the five tones. As for T2 and T3, in production of the stimuli by the native speakers, there is an overlap in the F_0 contours where the two tones share the same tonal space. In perception study, tone 2 and tone 3 were most confused, the listeners choose T2 over T3. While the female and male stimuli were analyzed, the result show that more of male T3 (63%) stimuli is identified as T2, which led to the hypothesis that duration of tone may act as a cue in identifying tones in Angami.

In the second perception study (E2), the result of the natural stimuli produced by the female speaker is similar to the first study, where the listener's had great confusion between T2 and T3. In synthesized short stimuli, low tone which is associated with short duration is identified more (120-160 Hz) than the next tone T4 (170,180 & 200 Hz) which is often associated with long duration. Whereas, in mid and long synthesized duration's, tone T4 occupies a wider low F_0 ranges. This show that the perceptual correlates of Angami tones are the relative F_0 and the duration of the tones. Meanwhile, we see that regardless of the duration types, high tone T1 and T2 are constantly identified in higher F_0 ranges. In the case of tones T5 and T4, it seems that duration serves as a crucial cue for identification. This indicates that the duration of the stimuli significantly influences how these tones are perceived or categorized. This might mean that within certain duration thresholds, these tones have distinct

characteristics that facilitate their differentiation, even if the F_0 ranges might overlap. Conversely, for tones T1 and T2, the emphasis is placed more on the fundamental frequency (F_0) ranges. This suggests that within specific F_0 boundaries, these tones are perceived, and variations in F_0 play a more pivotal role in differentiating between these tones, potentially regardless of the duration of the stimuli. Regarding reaction time, as expected the tones which were distinct, the listeners took relatively lesser time to respond while the intermediary tones took longer time.

In the perception of tone languages which has both contour and level tones, the distinction between the two tones is seen in the relative F_0 level, direction of F_0 change and magnitude of F_0 change (Wang, 1976; Sun and Huang, 2012; Khouw and Ciocca, 2007). However, this study shows that the perceptual correlates of an only level tone is the relative F_0 level and duration. The findings of this study that relative F_0 level as a cue for perception in level tones is similar to the findings reported in Thai & Yoruba (Gandour and Harshman, 1978) and Cantonese (Khouw and Ciocca, 2007). However, this study also shows that, in the perception of level tones, the tones are divided into two groups: mid and high tones (T1 and T2) which depend on the F_0 height, whereas mid-low and low tones (T4 and T5) where the perceptual correlates are F_0 height and also duration of the tones.

Chapter 10

Summary and Conclusion

10.1 General Conclusions

The primary goal of this thesis was to offer a comprehensive overview of the phonetic characteristics of Angami speech sounds. While doing that, the thesis examined the consonantal inventory of Angami in detail. This was followed by a comprehensive analysis of the vowels in Angami. Finally, the dissertation delved deep into the tones and their interactions with consonants, vowels, and other tones. While this work is expansive, it provides a holistic view of the phonetics of the Angami language at different analytical levels. While phonetic studies are aimed at specific phenomena in the language, this work goes beyond such specific goals. It provides more comprehensive phonetic analyses of various features and the interplay among them. In order to understand the workings of a language, it is important that all such nuances are brought to light through substantial data and rigorous experiments on the data. The current work was an attempt to do that.

The result of the current study showed that the segmental and suprasegmental features of the Angami language cannot be analyzed as isolated phenomena. The production of segments, such as the plosives, also depends on the lexical tones as-

sociated with the tone-bearing unit. At the same time, it was also seen that the primary cue of tones, namely, the F_0 , is also affected by the segments in its context as documented cross-linguistically (Gandour, 1974; Ewan and Kronen, 1974; Maddieson, 1978a; Hombert, 1978; Xu and Xu, 2003; Lai and Jongman, 2004; Tse, 2005; Lalhminghlui and Sarmah, 2020).

At the level of individual analysis of speech sounds in Angami, this study has several novel contributions. Firstly, the analysis of voiced and voiceless nasals is one of the few comprehensive studies on the acoustic properties of the voiceless nasals in the Tibeto-Burman languages. It was clear at the very outset that acoustic analysis of voiceless nasals is exceptionally challenging. Hence, this study used a novel method to analyze the voiceless nasals, where the acoustic data from the nasal and oral channels of the nanometer were separated and analyzed. The method provided a unique opportunity to study the acoustic signal from the nasal channel.

Distinct patterns are observed in the nasal and oral channels in the production of voiceless nasals. In the nasal channel, continuous energy is present from the onset of the nasal segment to its termination, with a peak at the beginning that gradually decreases during onset of aspiration. Conversely, the oral channel initially lacks energy at the beginning of the nasal segment but begins to rise concurrently with the onset of nasal aspiration. On the other hand, when analyzing the acoustic speech data where the oral and nasal data are not separated, it is not clear why the formant values see a sharp drop in the production of voiceless nasals. When the formant data is compared to the nasometer derived, nasal and oral channel separated energy profile, it is clear that the sharp drop in the speech acoustic signal is caused by the presence of the initiation of aspiration.

Any fine phonetic description of the speech sounds of the language should include fundamental analyses of the speech sounds. Considering this, the VOT and the vowel analyses were conducted in the current study. While the Angami vowel and

the plosives are primarily uncomplicated, we decided to conduct an in-depth analysis of these segments so that they serve as a strong foundation for further research in the language. Accordingly, further levels of experiments and analysis were conducted to uncover hitherto unreported phenomena of tone and VOT interactions in the language.

When looked into specific contributions, this work has provided an in-depth analysis of the segmental properties of the Angami language. Previously, no other study had taken into account such a large amount of data to phonetically analyze and characterize the speech segments of Angami. Specifically, the current work yielded the outcomes detailed in the following subsections.

10.1.1 VOT in Angami laryngeal contrasts

The stops have three laryngeal types in Angami, voiceless unaspirated, voiceless aspirated and voiced plosives. In the acoustic study of Angami stops, the laryngeal contrasts are determined through variations in VOT values. Specifically, voiced stops exhibit large negative VOT values, whereas voiceless unaspirated and voiceless aspirated stops display positive VOT values. The place of articulation of the stops also influences the VOT value. The study indicates that velar stops consistently exhibit longer VOT values than bilabial and denti-alveolar stops.

10.1.2 Acoustic correlates of Angami nasals

The nasals in Angami have two voicing contrasts, voiced and voiceless nasals. The voiced and voiceless nasals occur in the three places of articulation, that is, bilabial, alveolar, and palatal. The formants of the nasal distinguish the voiced and voiceless nasals. The result show that the voiced nasals have low nasal formant (N1-N3) values, while the voiceless have nasal formants. The voicing contrasts in the nasal formants vary depending on place of articulation. In the case of N1, the value of voiceless nasal

is highest for bilabial, followed by alveolar and palatal. While, in the case of N2, the values for voiceless nasals is highest for the palatal than alveolar and bilabial. Hence, the acoustic-phonetic features of nasals that distinguish voicing and PoA are the first three nasal formant N1, N2, N3, and the first formant Bandwidth BW1. As for N4, significant effect is observed in bilabial and alveolar voicing contrasts.

This study also examined the effect of voiced and voiceless nasals on following vowels in terms of vowel nasality. The result show that for vowel nasality, average A1-P0, compensated A1-P0, and A1-P1 showed effects of nasal voicing. It is also seen that nasality is higher after the voiceless nasals than after the voiced nasals.

10.1.3 Acoustic correlates of Angami fricatives and affricates

The fricatives in Angami have two voicing contrasts: voiced and voiceless fricatives. The fricatives occur in four places of articulation, namely, labio-dental, alveolar, palatal, and glottal. The study found that the acoustic properties of Angami fricatives are the spectral moments, (i.e skewness, kurtosis) and center of gravity. The voicing contrasts and place of articulation (PoA) interactions significantly affect these features in fricatives.

The affricates in Angami have three voicing distinctions: voiceless unaspirated, voiceless aspirated, and voiced unaspirated. The affricates occur in three places of articulation: labio-dental, alveolar, and palatal. The study show that acoustic phonetic properties of Affricates in Angami are only the spectral moments. The voicing contrasts and place of articulation (PoA) interactions significantly affect skewness and kurtosis in Angami affricates.

10.1.4 Acoustic properties of Angami Approximants

There are eight approximants in Angami which are produced in three PoA, namely, bilabial, alveolar and palatal. The approximants in Angami two voicing contrasts:

voiceless and voiceless aspirated approximant. The study show a clear differentiation between voiced and voiceless approximants. The voiced and voiceless laterals are distinct for the first and the second formants (F1 & F2). The bilabial approximant, the first and the second formants (F1 & F2) distinguishes voicing distinction. Voiced bilabial approximant exhibits lower formant values, while voiceless counterparts demonstrate higher ones. In the case of alveolar approximants, the voicing contrasts is distinguish in all the first four formants (F1-F4) of the approximants. In the case of palatal approximants, the voicing contrasts is significant only for the first formant (F1). The HNR in the voiced and voiceless approximants show significant differences in voicing. The voiceless approximants have lower HNR values, while the voiced approximants have higher HNR values. Whereas, the intensity of approximant shows that voiceless approximants have lower intensity than their voiced counterpart. Hence, significant acoustic-phonetic properties in distinguishing voicing include formants (F1-F2), Harmonics to Noise Ratio (HNR), and the amount of voicing.

10.1.5 Summary of Angami consonants

In this dissertation, each sound segment is studied in detailed, first by reviewing the existing literature and addressing the research gaps. This is followed by analyzing the phonemic status of the consonants through minimal pairs. Finally, the acoustic-phonetic study of each segment is investigated as discussed above. Hence, this dissertation show that Angami has thirty-nine consonant phonemes as shown in Table 10.1. The consonants in Angami has seven places of articulation, namely, bilabial, labio-dental, Dental, Alveolar, Palatal, Velar and Glottal and six manner of articulations, namely, Plosives, Nasals, Fricatives, Affricates, Lateral and Approximant.

Table 10.1: Phonemic inventory of consonants in Angami.

	Bilabial		Labio-dental		Denti-alveolar		Alveolar	Palatal	Velar	Glottal
Plosives	p p ^h	b	t t ^h	d				k k ^h	g	
Nasals	m ^h	m			n ^h	n		ɲ ^h	ɲ	
Fricative			f	v	s	z		ʃ	ʒ	h
Affricate			pf pf ^h				ts ts ^h	dʒ tʃ ^h	ʒ dʒ	
Lateral							l ^h	l		
Approximant	w ^h	w					ɹ ^h	ɹ	j ^h	j

10.1.6 Acoustic properties of Angami vowels

This study addresses the issues in the discrepancies in the inventory of vowels in Angami. As discussed in Chapter 6, the existing literature on Angami vowel postulate six to seven monophthongs. The issue arises when a few authors (Kuolie, 2006; Suokhrie, 2015) distinguish the mid-central vowel into two categories: [ə] and [ü] based on the position in which the vowel occurs. In this study, the mid-central vowel was separated into two categories, one that occurs only at the onset position and as a single lexical unit, while the other vowel occur with all CV and CVCV combinations. The result showed overlapping of the formants (F1 and F2) for the two central vowel which indicates they occupy the same vowel space. The duration of the two tones were not significant. Hence, this study reveals that there are six monophthongs in the language, as shown in Table 10.2. The acoustic-phonetic properties of the monophthong in Angami is the first three vowel formant (F1-F3) and the duration of the vowel. The study suggest that gender, tone, and vowel type significantly influence vowel formants.

Further, the study also addresses the varying description of diphthongs in Angami. The study found that there are two diphthongs, namely, /ie/ and /uo/ in the Angami

language. The result of the acoustic analyses show that diphthong /ie/ occupy the same vowel space as /i/ and /e/, whereas diphthong /uo/ occupy the same vowel space as /u/ and /o/. For diphthong, the acoustic properties are discrete cosine transform (DCT) of F1 and F2 and duration of the diphthong vowels.

Table 10.2: Phonemic inventory of vowels in Angami.

	Front	Central	Back
Close	i		u
Close-mid		ə	o
Open-mid	e		
Open		a	

10.1.7 Acoustics properties of Angami tones

In this dissertation, extensive studies were carried out to determine the size and types of tones in Angami. As discussed in Chapter 7, the existing work on Angami tones have reported that the language have four or five tones. Further, in a few description, the tones are also described as a combination of contour and level tones (Giridhar, 1980; Kuolie, 2006; Liezietsu et al., 2019) and some described it as only level tones (Suokhrie, 2015; Meyase, 2021). While the high (T1) and low (T5) are distinct, there is an overlap in the intermediary tones. In both lexical and nonce database, the F_0 of intermediary tones T2 and T3 are overlapping, and they are statistically not significant. Hence, this study reveals that Angami has four register tones. The primary acoustic feature of the tones is the relative F_0 height. T1 represent high tone with high F_0 , while T2 represent mid-high tone and T3 represent the low-mid tone and finally, T4 represent the low tones.

10.1.8 Interaction of tones and segments

The tones in Angami were examined based on their interactions with segments and supra-segments. The analysis of the interaction of tones and consonants showed

consonantal perturbations on the F_0 of the following tones. The result show that VOT effect on the tones and tones effect on VOT. The voiced onsets induce F_0 lowering of the following tone, while, voiceless onsets induce raising in the following F_0 . The study showed that voiceless consonantal perturbations were less as tone categories got higher. Conversely, voiced stop consonantal perturbations were less as tone categories got lower. The study also notes a systematic relationship between VOT duration and tone height, where high tones induced shorter VOT, while low tones induced longer VOT.

The study further explores the effect of F_0 on the vowel. The acoustic analysis of the intrinsic F_0 reveal a positive interaction between vowels and tone. The result show that the vowel effect on the tones is prominent at the point of initiation of the tones. The high tone T1 shows the largest vowel F_0 effect. The F_0 effects of vowels emerge where vowels /u/, /i/, and /e/ induce higher F_0 , and vowels /a/ and /o/ induce lower F_0 .

The acoustic analysis of this study show that the tones in Angami have bidirectional tonal coarticulation effects: anticipatory and carryover effects. In terms of anticipatory effects, it was noticed that only the high tone (T1) and mid tone (T3) in the first syllable has dissimilatory effects from the second syllable. However, no other tone in the first syllable showed any systematic effects of the following tones. Regarding carryover effects, the results show that only the initial 25% of the tone contour of the second syllable is affected by the tone in the first syllable. However, both the anticipatory and carryover effects depend on the height of the tone.

10.1.9 Classification of tones in Angami

Another significant contribution of this dissertation is classifying and investigating the acoustic correlates of F_0 in Angami tones. Cross-linguistically, tonal languages with only register tones are rare, and much of the available tone classification is reported

from languages that have both register tonal and contour tones. The classification of Angami tones shows higher accuracy when tones T2 and T3 are merged. Regarding acoustic features for tone identification, it was concluded that F_0 mean, F_0 height, F_0 max, and F_0 min contributed most towards the correct classification of tones.

The dissertation also sheds light on the perception of Angami tones. The two perception tests show that Angami listeners perceive only four tones. The confusion of T2 and T3 is seen in natural and synthesized speech stimuli. In the extensive studies of Angami, results from tone production, perception, and classification show that there are four contrastive tones in Angami.

10.2 Directions for future research

This thesis is a small step in understanding the phonetics of Angami speech sounds. As this research is one of the first detailed study on the speech sounds, this study paves the way for multitude of future studies:

The spoken and written form of Angami (Tenyidie) has a major influence of the variety spoken in the Northern Angami region, in particular, the one spoken in the Kohima village. As such, this work primarily focuses on speech data from Kohima village. It does not represent the inter and intra dialectal variations. Moreover, this study does not talk about the potential khel specific variations in the Angami dialects. Hence, as a future step, systematic speech corpus creation can be initiated for the plethora of varieties Angami has. As the current study reported speech data collected in controlled environment with read speech, in future spontaneous speech data can be collected which may capture finer nuances of the dialectal variations of the language. Each Angami village is said to have its own dialect, even if the geographical distance between the villages is less than 10 kilometers. While the varieties of Angami spoken in the villages have varying degrees of mutual intelligibility, they have village specific

phonetic variations, that can be captured using phonetic features in the future.

There are multiple Naga languages and dialects which are closely related to Angami, as described in the phylogenetic classification of Angami language (Section 1.2.1). Some of the closely related languages to Angami are Chokri, Khezha, Lotha, Mao. These languages share similar sound segments and phonological inventory with Angami, for example, like Angami, Lhotha, Chokri, Mao and Sema are reported to have voiceless nasals in their languages. The methodologies' used in this current study could be employed to study the voiceless nasals in these languages. Moreover, all Naga languages are tonal, however, no acoustic analysis of the tones is available except for the Ao and Sumi language. Hence, a cross-linguistic study of the tones of Naga languages is a future direction.

Although not reported in this dissertation, in the course of the Ph.D. work, a preliminary Automatic Speech Recognition (ASR) speech system in Angami was built (Terhijja et al., 2019). Recording of speech data, transcription and a pronunciation dictionary was created and an automatic speech recognition system using Kaldi, a public domain toolkit was built. This was the first speech recognition system developed for any language indigenous to the state of Nagaland. In order to promote and preserve the indigenous languages, it is important that such language technology tools are built. The International Year of Indigenous Languages – IYIL 2019, highlights the importance of language technologies (LT) and states that of more than 7,500 languages spoken over the world “less than 100, i.e. less than 2%, have some LT with various levels of quality, less than 10 are relatively covered, and only English is well covered, while the remaining 9%, referred to as under-resourced, are not-supported. This situation puts the native speakers of those under-resourced languages in a disadvantageous situation, creating a digital divide, and places their languages in danger of digital extinction, if not complete extinction.”

As we know, North-East India hosts a myriad of languages that belong to three

different language families and that have salient phonological inventories. Officially, of the 99 non-scheduled languages, 60 are spoken in North-East India. Among these 60, 34 languages have more than 50,000 speakers. However, most of the languages are under-studied and still lack comprehensive linguistic phonetic description. Systematic study into these languages may uncover many details that may be of interest to the larger academic community.



Appendix A

Datalist



Table 3: List of Nasal datalist

Sl. no.	Word	Tone	Meaning	Sl. no.	Word	Tone	Meaning
1	ma	T4	price	41	ɲə	T1	grains separation
2	ma	T3	glue	42	ɲə	T4	laugh
3	ma	T5	add	43	ɲə	T5	touch
4	me	T1	mouth	44	m ^h a	T3	something
5	me	T4	really	45	m ^h a	T5	suffocate
6	me	T3	ripe	46	m ^h a	T5	consume
7	mi	T1	fire	47	m ^h a	T3	hurry
8	mi	T3	spirit	48	m ^h e	T5	blow
9	mi	T5	leftovers	49	m ^h i	T5	eye
10	mo	T1	innocent	50	m ^h o	T1	dream
11	mo	T3	body	51	m ^h o	T3	on top
12	mo	T5	negation	50	m ^h o	T1	dream
13	mu	T2	and	51	m ^h o	T3	above
14	na	T1	budding	52	m ^h o	T4	gas
15	na	T5	rituals	53	m ^h u	T3	soiled
16	na	T5	behavior	54	n ^h a	T1	plants
17	ne	T1	push	55	n ^h a	T3	diminished
18	ne	T3	type	56	n ^h a	T3	cloudy
19	ne	T3	click picture	57	n ^h a	T5	mucus
20	ni	T1	happy	58	n ^h a	T5	plants
21	ni	T3	wrap-around	59	n ^h e	T2	ligaments
22	ni	T3	sister-in-law	60	n ^h e	T5	destroy
23	no	T3	you	61	n ^h e	T1	push
24	no	T3	near	62	n ^h e	T5	blow
25	nu	T4	thoughts	63	n ^h o	T1	addition
26	nu	T2	amongst	64	n ^h o	T5	similar
27	nu	T1	Grammatical cat.	65	n ^h o	T5	smear
28	nu	T4	in (direction/place)	66	n ^h o	T3	paste
29	ɲa	T1	crazy	67	ɲ ^h a	T1	weary
30	ɲa	T4	belongings	68	ɲ ^h ə	T1	pull in plenty*
31	ɲa	T5	son-in-law	69	ɲ ^h ə	T4	marriage
32	ɲe	T1	messy hair	70	ɲ ^h ə	T4	snake
33	ɲi	T1	rich	71	ɲ ^h ə	T4	nose
34	ɲi	T3	festive/celebrate	72	ɲ ^h ə	T4	stick
35	ɲi	T3	sick chicken*				
36	ɲo	T2	manure				
37	ɲo	T3	mud irrigation*				
38	ɲo	T5	iron				
39	ɲo	T4	run over				
40	ɲu	T5	feeding				

* words that do not have a direct translation in English.

Table 4: List of fricatives, affricates, laterals and approximants datalist.

Sl. no.	Word	Tone	Meaning	Sl. no.	Word	Tone	Meaning
1	sa	T3	including	31	pf ^h ə	T1	visitors
2	si	T3	know	32	ṭsa	T1	less
3	so	T1	guest	33	ṭse	T1	break
4	sə	T1	gourd cup	34	ṭsi	T2	all
5	se	T2	three	35	ṭso	T3	reach
6	su	T4	grain sapling	36	ṭsu	T3	went*
7	za	T4	distribute	37	ṭsə	T3	sprout
8	zi	T4	early	38	ṭs ^h a	T5	cotton
9	zo	T5	sweep	39	ṭs ^h e	T3	praise
10	zə	T2	top	40	ṭs ^h o	T3	wash (a)
11	ze	T3	follow	41	ṭs ^h u	T1	wash (b)
12	zu	T4	wine	42	ṭs ^h ə	T3	corn cobb
13	ʃa	T5	interrupt	43	ḍza	T5	hang
14	ʃi	T3	question p.	44	ḍze	T1	story
15	ʃo	T4	scrap	45	ḍzi	T3	wool roll
16	ʃə	T3	wings	46	ḍzo	T2	cheek
17	ʃe	T3	draw attention	47	ḍzə	T4	tubber*
18	ʃu	T1	pay	48	ṭʃa	T1	meaning
19	ʒa	T1	big	49	ṭʃe	T4	an expression
20	ʒo	T5	style	50	ṭʃi	T2	all
21	ʒə	T1	sleep	51	ṭʃo	T2	upright
22	ʒe	T4	cut	52	ṭʃu	T3	call
23	ʒu	T1	swept	53	ṭʃə	T1	small
24	p̣fe	T5	stick	54	ṭʃ ^h a	T1	cook
25	p̣fə	T5	carry	55	ṭʃ ^h e	T5	weave
26	p̣fi	T1	monkey	56	ṭʃ ^h i	T2	dislike
27	p̣fu	T5	father	57	ṭʃ ^h o	T2	straight
28	p̣fə	T1	female marker	58	ṭʃ ^h u	T2	crumbs
29	p̣f ^h e	T1	stack	59	ṭʃ ^h ə	T4	pain
30	p̣f ^h i	T4	bee	60	ḍʒa	T1	blessing

* words that do not have a direct translation in English.

Table 4: List of fricatives, affricates, laterals and approximants datalist.

Sl. no.	Word	Tone	Meaning	Sl. no.	Word	Tone	Meaning
61	dʒe	T2	pose	81	ɪa	T4	place
62	dʒo	T3	hit	82	ɪe	T3	smell
63	dʒə	T4	laid back	83	ɪi	T2	but
64	la	T1	because of that	84	ɪo	T4	if
65	le	T1	peel	85	ɪə	T4	aiming
66	li	T1	pot	86	ɪu	T4	bone
67	lo	T2	wool	87	ɪ ^h a	T4	redeem
68	lu	T2	there (distance)	88	ɪ ^h e	T1	thresh
69	l ^h a	T4	grain	89	ɪ ^h i	T4	click
70	l ^h e	T3	healthy	90	ɪ ^h o	T3	alert
71	l ^h i	T4	business	91	ɪ ^h ə	T2	handle
72	l ^h o	T5	would not	92	ɪ ^h u	T5	dirt
73	l ^h u	T4	live	93	ja	T1	cover
74	wa	T4	past tense	94	je	T4	constitution
75	we	T1	plural	95	ji	T3	g. marker
76	wi	T3	mithun	96	jo	T2	g. particle
77	wo	T2	a call	97	ju	T4	jokes
78	w ^h e	T1	hover	98	j ^h a	T1	lift
79	w ^h i	T3	round	99	j ^h o	T4	robust
80	w ^h uo	T4	chase	100	j ^h u	T3	desiccate

Table 5: List of lexical tones datalist.

Sl. no.	Word	Tone	Meaning	Sl. no.	Word	Tone	Meaning
1	pe	T1	lean	21	t̥sə	T1	process of thinning
2	pe	T2	bridge	22	t̥sə	T2	pruning
3	pe	T3	fat	23	t̥sə	T3	head
4	pe	T4	shiver	24	t̥sə	T4	germinate
5	pe	T5	shoot	25	t̥sə	T5	gave
6	se	T4	with	26	su	T1	dispersed*
7	se	T4	three	27	su	T2	traditional game
8	se	T1	erected	28	su	T3	replacement
9	se	T4	thunder	29	su	T4	graing sapling
10	se	T3	fool	30	su	T5	deep
11	t̥sa	T1	less	31	zu	T1	thumb rule
12	t̥sa	T2	stern	32	zu	T2	mountain top
13	t̥sa	T3	forest	33	zu	T3	swept
14	t̥sa	T4	old	34	zu	T4	drain
15	t̥sa	T5	broken	35	zu	T5	uproot
16	la	T1	because	36	do	T1	remnant
17	la	T2	again	37	do	T2	leg
18	la	T3	group	38	do	T3	carve
19	la	T4	discomfort	39	do	T4	ideas
20	la	T5	sow	40	do	T5	weave

* words that do not have a direct translation in English.

Table 6: List of vowels that have the same phoneme and tone but different orthography.

Sl. no.	Word	IPA	Tone	Meaning
1	kei	ki	T4	put together
2	ki	ki	T4	dried vegetables
3	mei	mi	T1	tail
4	mi	mi	T1	fire
5	bou	bu	T1	grab*
6	bu	bu	T1	g. particle
7	nou	nu	T4	depraved
8	nu	nu	T4	in (direction)
9	tou	tu	T4	adjacent
10	tu	tu	T4	well (tears)

Appendix B



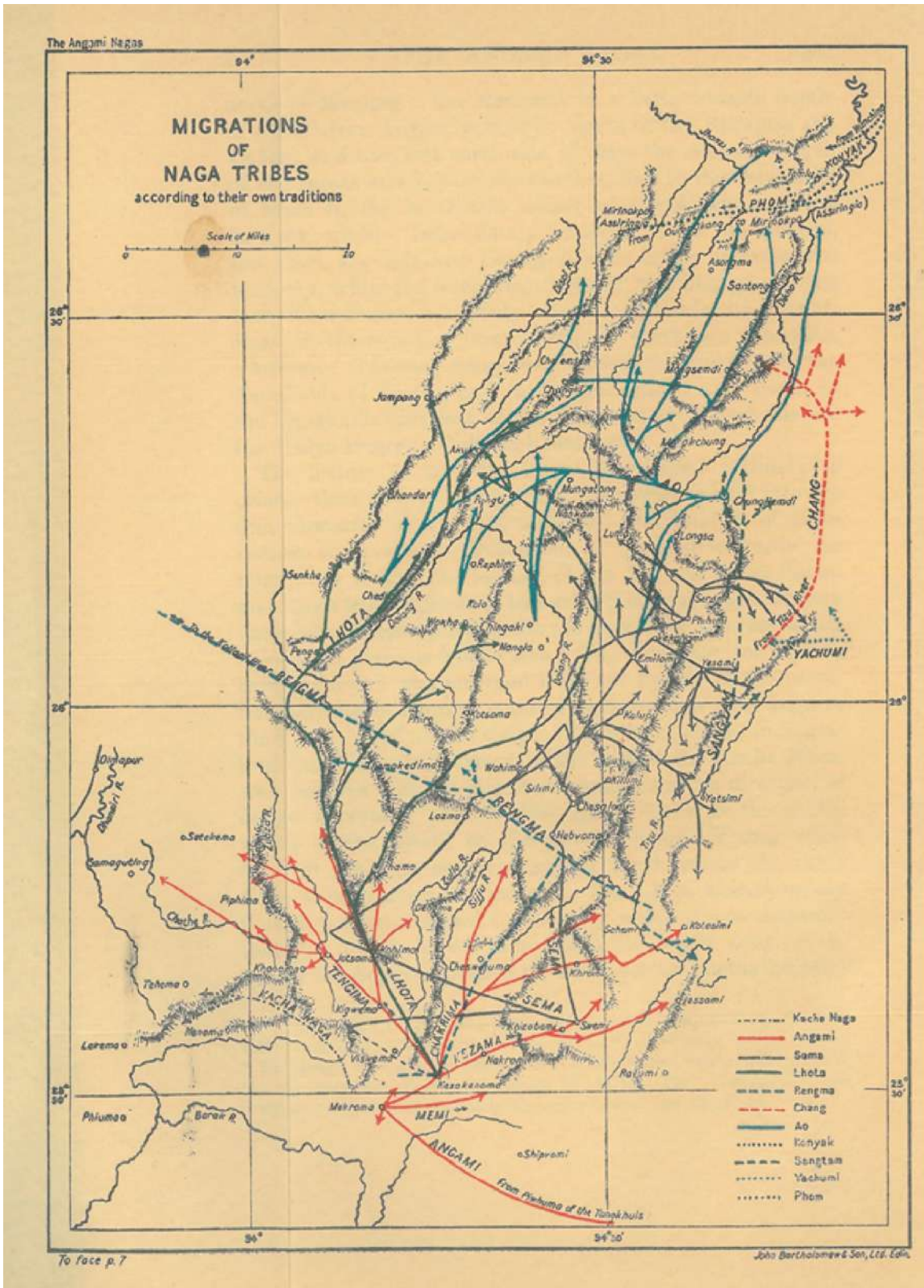


Figure -1: Map of migration of the Tengima (Tenyimia) adapted from Hutton (1921)

Table 7: List of the Angami villages adapted from District Census Handbook Kohima, 2011. The sub-groupings and the division of the villages is done based on common knowledge and cross-checking with native Angami. (subjected to correction)

Angami			
Northern	Southern	Western	Chakro
Kohima Village	Pfuchama	Jotsoma	Signal Angami
Viphoma	Phesama	Phezhu (Jotsoma)	J.V.Perhereilie
Viphoma Basa	Mima	Mezoma	Bade
Phekerkrie Basa	Kigwema	Mezo Basa	Urura
Phekerkrie Bawe	Kigwe Town	Khonoma	Diezephe
Ziezou	Jakhama Hq	Khonoma Basa	Khuovaru
Nachama	Jakhama Vill.	Dzuleke	Sovima
Ciechama	Viswema	Sechu (Zubza)	Vidima
Zhadima	Sweba(Viswema)	Hydro Project Station	Khriezephe
Chiephobozou	Khuzama	Thekrejuma	Sodzulhou
Nerhe Model	Sakhabama	Sechuma	Seithekema (New)
Nerhema	Kezo Town		Tenyiphe I
Phezha	Mitelephe		Tenyiphe II
Tsiese Basa	Kezo Basa		Tenyiphe III (Virazouma)
Tsiese Bawe	Kezoma		Seithekema 'A'
Thizama	Kezocha Hq		Seithekema (Old)
Meriema	Kidima		Seithekema 'C'
4Th NAP Batt. Thiza			Razhapse
Rusoma			Razhapse Basa
Tsiemekhuma Basa			Tsiepama
Tsiemekhuma Bawe			Pherima Vill.
Seiyhama			Medziphema
Seiyha Phesa			Pherima
Botsa			Khanakhuru
Tuophema			Piphema 'A'
Tuophe Phezou			Piphema (Old)
Teichuma			Piphema (New)
Gareiphema			Seirhima
Gareiphe Basa			
Kijumetouma			
Kijumetou Basa			
Dihoma			
Chiede Model Vill.			
Chiedema			
Mengujuma			
Peducha			
Sirhi Angami			
Kiruphema Bawe			
Kiruphema Basa			

Appendix C

Metadata



Table 8: Information of participants in this thesis (2018-2022)

Name	Gender	Age	Data type	Education	Language known
AK	F	31	Act2, Amt	MA	Tenyidie, English, Nagamese
AM	F	34	Avt, Apt1	Ph.D	Tenyidie, English, Nagamese
AL	F	58	Anl2	Graduate	Tenyidie, English, Nagamese
AS	F	53	AstiP	Teacher	Tenyidie, English, Nagamese
AP	M	66	Apt1	HSLC	Tenyidie, English
AZ	F	40	Anl1, Apt2	Teacher	Tenyidie, English, Nagamese
AT	M	38	Aaf, Apx	Graduate	Tenyidie, English, Nagamese
AU	M	34	Apt1	NA	Tenyidie, English, Nagamese
AV	F	27	Aaf, Apx	MA	Tenyidie, English, Nagamese
AO	F	66	Apt1	HSLC	Tenyidie, English
DZ	M	16	Apt1	HSLC	Tenyidie, English, Nagamese
DS	M	29	Avot, Anl2, Anaso, Aaf, Apx	LLB	Tenyidie, English, Nagamese
DK	M	36	Act2, Amt	MA	Tenyidie, English, Nagamese
DV	M	24	Apt1	Graduate	Tenyidie, English, Nagamese
JC	M	34	Apt2	Entrepreneur	Tenyidie, English, Nagamese
KK	F	27	Act1	Graduate	Tenyidie, English, Nagamese
KB	M	43	AstiP	Teacher	Tenyidie, English, Nagamese
KC	F	34	Act2, Aaf, Apx, Amt	MA	Tenyidie, English, Nagamese
KW	F	29	Act1, Avot, Anl, Aaf, Apx, Apt1	MA	Tenyidie, English, Nagamese
KN	F	24	Act1	Graduate	Tenyidie, English, Nagamese
KL	M	22	Apt1	Graduate	Tenyidie, English, Nagamese
KI	F	17	Apt1	HSSLC	Tenyidie, English, Nagamese
KH	F	29	Apt1	MA	Tenyidie, English, Nagamese
KO	F	21	Apt1	Graduate	Tenyidie, English, Nagamese
KT	M	52	Aaf, Apx, Apt2	M.A	English, Tenyidie, Nagamese
KM	M	25	Apt1	Graduate	Tenyidie, English, Nagamese
KP	M	54	Act2, Amt	MA	Tenyidie, English, Nagamese
KL	F	26	Avdt	MA	Tenyidie, English, Nagamese
KV	F	26	Act1, Anl1	Ph.D	Tenyidie, English, Nagamese
KS	F	33	Act2, Amt	MA	Tenyidie, English, Nagamese
KR	M	23	Amt	MA	Tenyidie, English, Nagamese
KG	F	27	Apt2	Entrepreneur	Tenyidie, English, Nagamese
KE	M	31	Apt1, Avot	MA	Tenyidie, English, Nagamese
KY	M	25	Avt, Avot, Apt1	Graduate	Tenyidie, English, Nagamese
MD	M	23	Anl1	Graduate	Tenyidie, English, Nagamese
MT	M	48	Act2, Amt	Ph.D	Tenyidie, English, Nagamese
MB	F	17	Apt1	HSLC	Tenyidie, English, Nagamese
MS	M	31	Act2, Amt	MA	Tenyidie, English, Nagamese
MM	F	26	ALL	MA	Tenyidie, English, Nagamese
MN	F	20	Apt1	MA	Tenyidie, English, Nagamese
ME	M	34	Apt2	Entrepreneur	Tenyidie, English, Nagamese
NK	M	61	Act1, Anl1, Anl2	Graduate	Tenyidie, English, Nagamese
NG	F	29	Act2, Amt	MA	Tenyidie, English, Nagamese
NR	F	27	Apt1	Graduate	Tenyidie, English, Nagamese
NS	F	60	Apt1, Anl2	Graduate	Tenyidie, English, Nagamese
NI	F	53	Anl2	Graduate	Tenyidie, English, Nagamese
NO	M	24	Aaf, Apx	Graduate	English, Tenyidie, Nagamese

Table 8: Information of participants in this thesis (2018-2022) (continued)

NK	M	20	Apt1	Graduate	Tenyidie, English, Nagamese
NS	F	28	Avot	Doctor	Tenyidie, English, Nagamese
NZ	F	40	Act2, Amt	Ph.D	Tenyidie, English, Nagamese
PS	M	25	Avt, Apt1	NA	Tenyidie, English, Nagamese
RK	M	38	Act2, Amt, Apt2	Ph.D	Tenyidie, English, Nagamese
RP	F	24	Avdt	Ph.D	Tenyidie, English, Nagamese
RS	F	24	Avt, Avot, Apt1	MA	Tenyidie, English, Nagamese
RO	F	23	Anaso	LLB	Tenyidie, English, Nagamese
RZ	F	27	Apt1	Graduate	Tenyidie, English, Nagamese
RU	F	17	Apt1	HSLC	Tenyidie, English, Nagamese
SK	M	21	Avt	BA	Tenyidie, English, Nagamese
SS	M	30	Avot	Graduate	Tenyidie, English, Nagamese
TM	F	32	Act2, Amt	MA	Tenyidie, English, Nagamese
TP	M	38	Anaso	Professional	Tenyidie, English, Nagamese, Hindi
TS	F	34	Apt1	MA	Tenyidie, English, Nagamese
VS	M	19	Apt1	Graduate	Tenyidie, English, Nagamese
VK	M	21	Apt1	Graduate	Tenyidie, English, Nagamese
VL	M	32	Act1	Graduate	Tenyidie, English, Nagamese
VP	M	37	Act2, Amt	MA	Tenyidie, English, Nagamese
VV	M	30	Act1	Graduate	Tenyidie, English, Nagamese
VT	M	30	Anl1	Graduate	Tenyidie, English, Nagamese
VW	F	32	AstiP	MA	Tenyidie, English, Nagamese
VY	F	35	Act2, Aaf, Amt	MA	Tenyidie, English, Nagamese
VK	M	29	Anl2	MA	Tenyidie, English, Nagamese
VB	F	20	Apt1	Graduate	Tenyidie, English, Nagamese
VT	F	31	Act1, Avot, Anaso, Apt2	Ph.D	Tenyidie, English, Nagamese
ZZ	M	23	Aaf	B.A	Tenyidie, English, Nagamese, Hindi
ZG	F	28	Act1, Act2, Amt	Ph.D	Tenyidie, English, Nagamese

List of Publications

Journals (peer-reviewed)

1. **Viyazonuo Terhijja**, and Priyankoo Sarmah. “Voice Onset Time in Angami”, Linguistics of the Tibeto-Burman Area, John Benjamins Publishing Company, vol. Volume 47, ISBN.0731-3500, pp.32-56, [2024].
2. **Viyazonuo Terhijja**, and Priyankoo Sarmah. “Acoustics of Voiceless Nasals in Angami”, Journal of Asian and African Studies, Supplement, Research Institute of Languages and cultures of Asia and Africa, ISBN.2758-0326, pp.129-144, [2024].
3. **Viyazonuo Terhijja**, and Priyankoo Sarmah. “Acoustics of vowels in Angami”, Journal of Himalayan Linguistics (under review)

Conference Proceedings (peer-reviewed)

1. **Viyazonuo Terhijja**, and Priyankoo Sarmah. “Voiced and voiceless laterals in Angami”. In Proceedings of Interspeech 2024, pp. 3704-3708, doi: 10.21437/Interspeech.2024-2361
2. **Viyazonuo Terhijja**, Zhonei I Gwirie, and Priyankoo Sarmah. “Tonal coarticulation in Angami level tones”. Proceedings of the Second International Conference on Tone and Intonation. Chinese and Oriental Languages Information Processing Society. November 2023.
3. **Viyazonuo Terhijja**, and Priyankoo Sarmah. “Aspiration in voiceless nasals in Angami”. In Proceedings of Meetings on Acoustics 179 ASA (Vol. 42, No. 1, p. 060008). Acoustical Society of America. December 2020. doi: 10.1121/2.0001403

4. Wendy Lalhminghlui, **Viyazonuo Terhiija**, and Priyankoo Sarmah. “Vowel-Tone Interaction in Two Tibeto-Burman Languages”. In Proceedings of INTERSPEECH 2019, (pp. 3970-3974),2019. doi: 10.21437/Interspeech.2019-2808

Conferences (other than thesis work)

1. **Viyazonuo Terhiija**, and Priyankoo Sarmah. “Vowels in the Khels of Two Southern Angami Villages”. 12th International Conference of the North East Indian Linguistics Society (NEILS12), Guwahati, India. January 2023.
2. **Viyazonuo Terhiija**, Priyankoo Sarmah, and Samudra Vijaya. “Development of Speech Corpus and Automatic Speech Recognition of Angami”. 22nd Oriental-COCOSDA, Cebu, Philippines. October 2019
3. **Viyazonuo Terhiija**, Priyankoo Sarmah, and Samudra Vijaya. “Spoken Language Technology for North-East Indian Languages”. In Proceedings of the Language Technologies for All (LT4All), pages 182–185. European Language Resources Association (ELRA), licenced under CC-BY-NC. 2019.
4. **Viyazonuo Terhiija**, Priyankoo Sarmah, and Samudra Vijaya. “Speech Technology of three Tonal Languages in North-East India”. Poster presentation at Language Technologies for All (LT4All), at UNESCO Headquarters, Paris, 5-6 December 2019.
5. **Viyazonuo Terhiija**, Priyankoo Sarmah, and Samudra Vijaya. “Acoustic Analysis of Vowels in Two Southern Angami Dialects”. In Oriental COCOSDA (pp. 7-8).2018.
6. **Viyazonuo Terhiija**, and Luke Horro. “A sociolinguistic comparison between two pidgin languages of North East India”. Presented at North East Indian Lin-

guistics Society 10th Conference, Assam university, Silchar on 29th-31st January 2018.



Bibliography

- Abramson, A. S. (1977). Noncategorical perception of tone categories in Thai. *The Journal of the Acoustical Society of America*, 61(S1):S66–S66.
- Abramson, A. S. (1979). The coarticulation of tones: An acoustic study of Thai. *Studies in Tai and Mon-Khmer phonetics and phonology in honour of Eugenie JA Henderson*, pages 1–9.
- Abramson, A. S. and Whalen, D. H. (2017). Voice Onset Time (VOT) at 50: Theoretical and practical issues in measuring voicing distinctions. *Journal of phonetics*, 63:75–86.
- Adank, P., Smits, R., and 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.
- Asu, E. L., Nolan, F., and Schötz, S. (2015). Comparative study of Estonian Swedish voiceless laterals: Are voiceless approximants fricatives? In *the 18th International Congress of Phonetic Sciences (ICPhS XVIII)*. University of Glasgow.
- Baart, J. L. (1997). The sounds and tones of Kalam Kohistani. *National institute of Pakistan studies Quaid-i-Azam University and Summer Institute of Linguistics. P*, pages 1–9.
- Bahl, K. C. (1957). Tones in Punjabi. *Indian Linguistics*, 17:139–147.
- Bates, D., Maechler, M., Bolker, B., and Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, 67(1):1–48.
- Bhaskararao, P. and Ladefoged, P. (1991). Two types of voiceless nasals. *Journal of the International Phonetic Association*, 21(2):80–88.
- Bhatia, T. K. (1957). The evolution of tones in Punjabi. *Studies in the Linguistic Sciences*, 5:12–24.
- Bhatia, T. K. (1993). Punjabi (Descriptive Grammars).
- Bielenberg, B. and Nienu, Z. (2001). Chokri (Phek dialect): phonetics and phonology. *Linguistics of the Tibeto-Burman Area*, 24(2):85–122.

- Blankenship, B., Ladefoged, P., Bhaskararao, P., and Chase, N. (1993). Phonetic structures of Khonoma Angami. *UCLA working papers in phonetics*, 84:127–142.
- Blevins, J. (1993). Klamath laryngeal phonology. *International journal of American linguistics*, 59(3):237–279.
- Boersma, Paul and Weenink, David (2020). Praat: doing phonetics by computer [computer program]. *Version*, 6:1.
- Boucher, V. J. (2002). Timing relations in speech and the identification of voice-onset times: A stable perceptual boundary for voicing categories across speaking rates. *Perception & Psychophysics*, 64:121–130.
- Brown, N. (1851). Specimens of the Naga Language of Asam. *Journal of the American Oriental Society*, 2:157–165.
- Bruhn, D. W. (2014). *A phonological reconstruction of Proto-Central Naga*. PhD thesis, University of California.
- Brunelle, M. (2009a). Northern and Southern Vietnamese tone coarticulation: A comparative case study. *Journal of Southeast Asian Linguistics*, 1(1):49–62.
- Brunelle, M. (2009b). Tone perception in Northern and Southern Vietnamese. *Journal of Phonetics*, 37(1):79–96.
- Burling, R. (1960). Angami Naga phonemics and word list. *Indian Linguistics*, 21:51–60.
- Burling, R. (2003). The Tibeto-Burman languages of Northeastern India. *The Sino-Tibetan Languages*, 3:169–191.
- Burling, R. (2013). The ‘sixth’ vowel in the Boro-Garo languages. *North east Indian linguistics*, 5:271–279.
- Chao, K.-Y. and Chen, L.-m. (2008). A cross-linguistic study of voice onset time in stop consonant productions. In *International Journal of Computational Linguistics & Chinese Language Processing, Volume 13, Number 2, June 2008*, pages 215–232.
- Chao, Y.-R. (1930). A system of tone letters. *Le maître phonétique*.
- Chase, N. (1992). *A descriptive analysis of the Khwunomia dialect of Angami*. PhD thesis, Ph.D. dissertation, University of Poona.
- Chen, M. Y. (1997). Acoustic correlates of English and French nasalized vowels. *The Journal of the Acoustical Society of America*, 102(4):2360–2370.
- Chen, Y. (2011). How does phonology guide phonetics in segment-f₀ interaction? *Journal of Phonetics*, 39(4):612–625.

- Chirkova, K., Basset, P., and Amelot, A. (2019). Voiceless nasal sounds in three Tibeto-Burman languages. *Journal of the International Phonetic Association*.
- Cho, T. and Ladefoged, P. (1999). Variation and universals in VOT: evidence from 18 languages. *Journal of phonetics*, 27(2):207–229.
- Connell, B. (2002). Tone Languages and the Universality of Intrinsic F0: Evidence from Africa. *Journal of Phonetics*, 30(1):101–129.
- Coupe, A. R. et al. (2003). *A phonetic and phonological description of Ao: a Tibeto-Burman language of Nagaland, North-East India*. Pacific Linguistics, Research School of Pacific and Asian Studies.
- Dantsuji, M. (1984). A Study on Voiceless Nasals in Burmese. *Studia Phonologica*, 18:1–14.
- Dantsuji, M. (1986). Some acoustic observations on the distinction of place of articulation for voiceless nasals in Burmese. *Studia phonologica*, 20:1–11.
- Delvaux, V., Metens, T., and Soquet, A. (2002). French nasal vowels: acoustic and articulatory properties. In *INTERSPEECH*.
- Diehl, R. L. and Kluender, K. R. (1989). On The Objects of Speech Perception. *Ecological Psychology*, 1(2):121–144.
- Directorate of Census Operations, Nagaland (2011). *District Census Handbook, Kohima. Series 14 Part XII A. Census of India 2011 - Nagaland*.
- Dixit, R. (1975). *Neuromuscular aspects of laryngeal control: with special reference to Hindi*. Univ. of Texas, Austin. PhD thesis, doctoral dissertation.
- Eberhard, D. M., Simons, G. F., and Fennig, C. D. (2024). *Ethnologue: Languages of the World*. SIL International, Dallas, Texas, USA, 27 edition.
- Erickson, D., Baer, T., and Harris, K. S. (1982). The role of the strap muscles in pitch lowering. *Haskins Laboratories Status Report on Speech Research SR-70*, pages 275–284.
- Ewan, W. G. and Kronen, R. (1974). Measuring larynx movement using the thyro-mbrometer. *Journal of Phonetics*, 2(4):327–335.
- Ezung, M. K. (2014). A Typological Overview of Tenyiedie (Angami). *Nagaland University Research Journal (A Multi Disciplinary Approach)*, 7:116–131.
- Ezung, M. K. (2018). *Relative Clause Formation in Tenyidie*. Heritage Publishing House, Dimapur, Nagaland. India, 1st edition.
- Fanai, L. (1992). *Some aspects of the lexical phonology of Mizo and English: an autosegmental approach*. PhD thesis, The Central Institute of English and Foreign Languages.

- Fernandes, J., Teixeira, F., Guedes, V., Junior, A., and Teixeira, J. P. (2018). Harmonic to noise ratio measurement-selection of window and length. *Procedia computer science*, 138:280–285.
- Flege, J. E. and Eefting, W. (1986). Linguistic and developmental effects on the production and perception of stop consonants. *Phonetica*, 43(4):155–171.
- Flemming, E. (2009). The phonetics of schwa vowels. *Phonological weakness in English*, 493:78–95.
- Flemming, E. and Johnson, S. (2007). Rosa’s roses: Reduced vowels in American English. *Journal of the International Phonetic Association*, 37(1):83–96.
- Ford, C. (2016). *The status of voiceless nasals in Ikema Ryukyuan*. Bachelor’s thesis, University of Alberta, Edmonton, Canada.
- Fox, J. and Weisberg, S. (2019). *An R Companion to Applied Regression, Third Edition*. Thousand Oaks CA:Sage. URL.
- Fujimura, O. and Lindqvist, J. (1971). Sweep-tone measurements of vocal-tract characteristics. *The Journal of the Acoustical Society of America*, 49(2B):541–558.
- Gandour, J. (1974). Consonant types and tone in Siamese. *Journal of phonetics*, 2(4):337–350.
- Gandour, J., Potisuk, S., and Dechongkit, S. (1994). Tonal coarticulation in Thai. *Journal of Phonetics*, 22(4):477–492.
- Gandour, J. T. (1978). II - The Perception of Tone. In Fromkin, V. A., editor, *Tone*, pages 41–76. Academic Press.
- Gandour, J. T. and Harshman, R. A. (1978). Crosslanguage differences in tone perception: A multidimensional scaling investigation. *Language and speech*, 21(1):1–33.
- Geissler, C. (2019). Tonal and laryngeal contrasts in Diaspora Tibetan. *S. Calhoun*, pages 2421–2424.
- Gill, H. S. and Gleason, H. A. (1969). *A reference grammar of Punjabi*. Department of Linguistics, Punjabi University Patiala, India.
- Giridhar, P. P. (1980). *Angami grammar*, volume 6. Central Institute of Indian languages.
- Glover, W. W. (1970). Gurung tone and higher levels. *Occasional papers of the Wolfenden Society on Tibeto-Burman linguistics*, 3:52–73.
- Gogoi, P., Dey, A., Lalhminghlu, W., Sarmah, P., and Prasanna, S. M. (2020a). Lexical tone recognition in Mizo using acoustic-prosodic features. In *Proceedings of the Twelfth Language Resources and Evaluation Conference*, pages 6458–6461.

- Gogoi, P., Tzudir, M., Sarmah, P., and Prasanna, S. (2020b). Automatic Tone Recognition of Ao Language. In *Proceedings of the 10th International Conference on Speech Prosody 2020*, pages 1005–1008.
- Goldsmith, J. A. (1976). *Autosegmental phonology*, volume 159. Indiana University Linguistics Club Bloomington.
- Gonzales, A. (2009). Intrinsic F0 in Shona Vowels: A Descriptive Study. In *Selected Proceedings of the 39th Annual Conference on African Linguistics*, ed. Akinloye Ojo and Lioba Moshi, pages 145–155.
- Gordon, M., Barthmaier, P., and Sands, K. (2002). A cross-linguistic acoustic study of voiceless fricatives. *Journal of the International Phonetic Association*, 32(2):141–174.
- Grierson, G. A. (1903). Tibeto-Burman Family: Specimens of the Bodo, Naga, and Kachin Groups, Volume III Part II of Linguistic Survey of India.
- Hajek, J. (2009). Labiodental Nasal'mj(M)' in the Angami Area. *Linguistics of the Tibeto-Burman Area*, 32(2):107–117.
- Halle, M. and Stevens, K. N. (1971). A note on laryngeal features. *Quarterly progress report*, 101:198–213.
- Hammond, M. (2019). Voiceless Nasals in Welsh. *Journal of Celtic Linguistics*, 20(1):31–60.
- Han, M. S. and Kim, K.-O. (1974). Phonetic variation of Vietnamese tones in disyllabic utterances. *Journal of Phonetics*, 2(3):223–232.
- Haudricourt, A.-G. (1954). *De l'origine des tons en Vietnamien*. Imprimerie Nationale.
- Haudricourt, A. G. (1972). *Tones in Punjabi*.
- Hawks, J. W. and Miller, J. D. (1995). A formant bandwidth estimation procedure for vowel synthesis. *The Journal of the Acoustical Society of America*, 97(2):1343–1344.
- Hirose, H. (1997). *Investigating the physiology of laryngeal structures*. Oxford: Basil Blackwell.
- Hombert, J.-M. (1975). Phonetic explanation of the development of tones from pre-vocalic consonants. *UCLA Working Papers in Phonetics*, 33:23–39.
- Hombert, J.-M. (1976). Consonant types, vowel height and tone in Yoruba. *UCLA Working Papers in Phonetics*, 33:40–54.
- Hombert, J.-M. (1978). Consonant Types, Vowel Quality. *Tone: A linguistic survey*, page 77.

- Hombert, J.-M., Ohala, J. J., and Ewan, W. G. (1979a). Phonetic Explanations for the Development of Tones. *Language*, 55(1):37–58.
- Hombert, J.-M., Ohala, J. J., and Ewan, W. G. (1979b). Phonetic explanations for the development of tones. *Language*, pages 37–58.
- Hoole, P. and Hu, F. (2004). Tone-Vowel Interaction in Standard Chinese. In *International Symposium on Tonal Aspects of Languages: With Emphasis on Tone Languages*.
- Hussain, Q. (2018). A typological study of Voice Onset Time (VOT) in Indo-Iranian languages. *Journal of Phonetics*, 71:284–305.
- Hutton, J. H. (1969). *The Angami Nagas. With Some notes on Neighbouring Tribes*. Oxford University Press, Bombay, India, 2nd edition.
- Hyman, L. M. (2006). Word-prosodic typology. *Phonology*. University of Berkeley.
- Jacewicz, E. and Fox, R. A. (2015). Intrinsic Fundamental Frequency of Vowels is Moderated by Regional Dialect. *The Journal of the Acoustical Society of America*, 138(4):EL405–EL410.
- Jessen, M. and Pétursson, M. (1998). Voiceless nasal phonemes in Icelandic. *Journal of the International Phonetic Association*, 28(1-2):43–53.
- Jongman, A., Wayland, R., and Wong, S. (2000). Acoustic characteristics of English fricatives. *The Journal of the Acoustical Society of America*, 108(3):1252–1263.
- Kenstowicz, M. J. (1971). *Lithuanian phonology*. University of Illinois at Urbana-Champaign.
- Kessinger, R. H. and Blumstein, S. E. (1997). Effects of speaking rate on voice-onset time in Thai, French, and English. *Journal of phonetics*, 25(2):143–168.
- Khouw, E. and Ciocca, V. (2007). Perceptual correlates of Cantonese tones. *Journal of Phonetics*, 35(1):104–117.
- Kirby, J. and Hyslop, G. (2019). Acoustic analysis of onset voicing in Dzongkha obstruents. *dimensions*, 15:16.
- Kirby, J. P. (2018). Onset pitch perturbations and the cross-linguistic implementation of voicing: Evidence from tonal and non-tonal languages. *Journal of Phonetics*, 71:326–354.
- Kirby, J. P. and Ladd, D. R. (2016). Effects of obstruent voicing on vowel F0: Evidence from “true voicing” languages. *The Journal of the Acoustical Society of America*, 140(4):2400–2411.
- Klatt, D. H. (1975). Voice onset time, frication, and aspiration in word-initial consonant clusters. *Journal of speech and hearing research*, 18(4):686–706.

- Kochetov, A. and Arsenault, P. (2020). Kalasha affricates: An acoustic analysis of place contrasts. *Journal of South Asian Languages and Linguistics*, 6(2):245–280.
- Koopmans-van Beinum, F. J. (1994). What’s in a schwa? *Phonetica*, 51(1-3):68–79.
- Kuang, J. (2013). The tonal space of contrastive five level tones. *Phonetica*, 70(1-2):1–23.
- Kuhn, M. (2021). *caret: Classification and Regression Training*. R package version 6.0-88.
- Kumar, D., Meghwani, S. S., and Thakur, M. (2016). Proximal support vector machine based hybrid prediction models for trend forecasting in financial markets. *Journal of Computational Science*, 17:1–13.
- Kuolie, D. (2006). *Structural description of Tenyidie: a Tibeto-Burman language of Nagaland*. Ura Academy, Publication Division.
- Kuznetsova, A., Brockhoff, P. B., and Christensen, R. H. (2017a). lmerTest package: tests in linear mixed effects models. *Journal of statistical software*, 82:1–26.
- Kuznetsova, A., Brockhoff, P. B., and Christensen, R. H. B. (2017b). lmerTest Package: Tests in Linear Mixed Effects Models. *Journal of Statistical Software*, 82(13):1–26.
- Ladefoged, P. (1964). Some Possibilities in Speech Synthesis. *Language and Speech*, 7(4):205–214.
- Ladefoged, P. (1967). *Three Areas of Experimental Phonetics: Stress and Respiration Activity, the Nature of Vowel Quality, Units in the Perception and Production of Speech*. Oxford University Press.
- Ladefoged, P. and Disner, S. F. (2012). *Vowels and consonants*. John Wiley & Sons.
- Ladefoged, P., Johnson, K., and Ladefoged, P. (2006). *A course in phonetics*, volume 3. Thomson Wadsworth Boston, MA.
- Ladefoged, P. and Maddieson, I. (1996). *The sounds of the world’s languages*. Blackwell, Oxford, UK.
- Lai, Y. and Jongman, A. (2004). The effect of aspiration on fundamental frequency in Taiwanese syllables. *Unpublished MA Thesis. University of Kansas*.
- Lalhminghlu, W. and Sarmah, P. (2020). Interaction of Tone and Voicing in Mizo. *Proc. Interspeech 2020*, pages 1903–1907.
- Lalhminghlu, W. and Sarmah, P. (2021). Characterizing Voiced and Voiceless Nasals in Mizo. In *Interspeech*, pages 426–430.

- Lalhminghlu, W., Terhijja, V., and Sarmah, P. (2019). Vowel-Tone Interaction in Two Tibeto-Burman Languages. *Proc. Interspeech 2019*, pages 3970–3974.
- Leben, W. R. (1973). *Suprasegmental phonology*. PhD thesis, Massachusetts Institute of Technology.
- Lee, S., Kawahara, S., Guillemot, C., and Monou, T. (2019). The acoustic correlates of the four-way laryngeal contrast in Drenjongke stops. In *Proceedings of the 19th International Congress of Phonetic Sciences, Melbourne, Australia*, pages 1445–1449.
- Lehiste, I. (1964). *Acoustical characteristics of selected English consonants*, volume 30. Indiana University Bloomington, IN.
- Lehiste, I. (1970). *Suprasegmentals*. Massachusetts Institute of Technology Press.
- Lenth, R. (2019). emmeans: Estimated Marginal Means, aka Least-Squares Means. *R package version*, 1:4.
- Li, Y. (2016). English and Thai Speakers' Perception of Mandarin Tones. *English Language Teaching*, 9(1):122–132.
- Liaw, A. and Wiener, M. (2002). Classification and Regression by randomForest. *R News*, 2(3):18–22.
- Lieberman, A. M., Harris, K. S., Hoffman, H. S., and Griffith, B. C. (1957). The discrimination of speech sounds within and across phoneme boundaries. *Journal of experimental psychology*, 54(5):358.
- Lieberman, P. (1970). A Study of Prosodic Features. *Current Trends in Linguistics*, 12(Part 10):2419–2450.
- Liezietsu, M., K.Kesiezie, and S.Liezietsu (2019). *MKS Dieda, 2nd eds*. Ura Academy.
- Lisker, L. and Abramson, A. S. (1964). A cross-language study of voicing in initial stops: Acoustical measurements. *Word*, 20(3):384–422.
- Lisker, L. and Abramson, A. S. (1967). Some effects of context on voice onset time in English stops. *Language and speech*, 10(1):1–28.
- Lisker, L. and Abramson, A. S. (1970). The voicing dimension: Some experiments in comparative phonetics. In *Proceedings of the 6th international congress of phonetic sciences*, volume 563, pages 563–567. Academia Prague Prague, Czech Republic.
- Lobanov, B. M. (1971). Classification of Russian vowels spoken by different speakers. *The Journal of the Acoustical Society of America*, 49(2B):606–608.
- Maddieson, I. (1978a). Tone effects on consonants. *Journal of Phonetics*, 6(4):327–343.

- Maddieson, I. (1978b). Universals of tone. *Universals of human language*, 2:335–365.
- Maddieson, I. (1984a). *Patterns of Sounds*. Cambridge Studies in Speech Science and Communication. Cambridge University Press.
- Maddieson, I. (1984b). *Patterns of Sounds*. Cambridge University Press, Cambridge.
- Maddieson, I. (2005). Bilabial and labio-dental fricatives in Ewe. *UC Berkeley Phon-Lab Annual Report*, 1(1).
- Maddieson, I. (2013a). Tone. *The world atlas of language structures online*.
- Maddieson, I. (2013b). Voicing in plosives and fricatives (v2020.3). In Dryer, M. S. and Haspelmath, M., editors, *The World Atlas of Language Structures Online*. Zenodo.
- Maddieson, I. (2013c). Vowel quality inventories. *The world atlas of language structures online*.
- Maddieson, I. and Emmorey, K. (1984). Is there a valid distinction between voiceless lateral approximants and fricatives? *Phonetica*, 41(4):181–190.
- Maeda, S. (1993). Acoustics of vowel nasalization and articulatory shifts in French nasal vowels. In *Nasals, nasalization, and the velum*, pages 147–167. Elsevier.
- Mahapatra, B. P. and McConnell Grant D., P. P. . V. V. (1989). Book 2: Non-Constitutional Languages. In McConnell, H. K. . G. D., editor, *The Written Languages of the World: A Survey of the Degree and Modes of Use (2)*, pages 611–1359. Office of the Registrar General, India and the International Centre of Research on Bilingualism, Laval University, Canada.
- Mahapatra, B. P., Padmanabha, P., McConnell, G. D., Verma, V., of the Registrar General, I. O., and the International Centre of Research on Bilingualism, Laval University, C. (1989). *Written languages of the world: a survey of the degree and modes of use; volume 2, India; book 2, non-constitutional languages*. Presses de l'université Laval, Québec, QC, CA.
- Malécot, A. (1960). Vowel nasality as a distinctive feature in American English. *Language*, pages 222–229.
- Marrison, G. E. (1967). *The classification of the Naga languages of Northeast India*. PhD thesis, School of Oriental and African Studies, University of London, London.
- Matisoff, J. A. (1970). Glottal dissimilation and the Lahu high-rising tone: a tonogenetic case-study. *Journal of the American Oriental Society*, pages 13–44.
- Matisoff, J. A. (1973). Tonogenesis in Southeast Asia. *Consonant types and tone*, 1.
- Matisoff, J. A. (2003). *Handbook of Proto-Tibeto-Burman: system and philosophy of Sino-Tibetan reconstruction*. Univ of California Press.

- Mays, C. and Beckman, M. (2008). An acoustic study of affricates in the Songyuan dialect of Mandarin Chinese. In *Poster presented at the 22nd Annual CIC SROP Research Conference, Michigan State University*, pages 24–26.
- McCabe, R. B. (1887). *Outline Grammar of the Angāmi Nāgā Language: With a Vocabulary and Illustrative Sentences*. Superintendent of Government Printing.
- Meyase, S. (2019). *A Morphological Investigation of the Tonal Phonology of Tenyidie*. Unpublished Ph.D. Dissertation, The English and Foreign Languages University, Hyderabad.
- Meyase, S. M. (2016). *A morphological investigation of the tonal phonology of Tenyidie*. PhD thesis, The English and Foreign Languages University.
- Meyase, S. M. (2021). Polarity in a four-level tone language: tone features in Tenyidie. *Phonology*, 38(1):123–146.
- Meyer, D., Dimitriadou, E., Hornik, K., Weingessel, A., and Leisch, F. (2021). *e1071: Misc Functions of the Department of Statistics, Probability Theory Group (Formerly: E1071), TU Wien*. R package version 1.7-8.
- Mok, P. P.-K. and Wong, P. W.-Y. (2010). Perception of the merging tones in Hong Kong Cantonese: preliminary data on monosyllables. In *Speech Prosody 2010-Fifth International Conference*.
- Moran, S. and McCloy, D., editors (2019). *PHOIBLE 2.0*. Max Planck Institute for the Science of Human History, Jena.
- Murty, K. S. R. and Yegnanarayana, B. (2008). Epoch extraction from speech signals. *IEEE Transactions on Audio, Speech, and Language Processing*, 16(8):1602–1613.
- Narayan, C. and Bowden, M. (2013). Pitch affects voice onset time (VOT): A cross-linguistic study. In *Proceedings of Meetings on Acoustics*, volume 19. AIP Publishing.
- Nirgianaki, E. (2014). Acoustic characteristics of Greek fricatives. *The Journal of the Acoustical Society of America*, 135(5):2964–2976.
- Nuh, V. (2002). *The origin of Naga. The Khezhakeno legendary stone*. Vision press.
- Office of the Registrar General & Census Commissioner of India (2011). *Statement-1: Abstract of Speakers' Strength of Languages and Mother Tongues - 2011*.
- Ohala, J. J. (1972). How is pitch lowered? *The Journal of the Acoustical Society of America*, 52(1A_Supplement):124–124.
- Ohala, J. J. (1974). *Experimental historical phonology*. publisher not identified.
- Ohala, J. J. (1993). Sound change as nature's speech perception experiment. *Speech Communication*, 13(1-2):155–161.

- Ohala, J. J. and Eukel, B. W. (1987). Explaining the intrinsic pitch of vowels. *In honor of Ilse Lehiste*, pages 207–215.
- Ohala, M. (1983). *Aspects of Hindi phonology*. Motilal Banarsidass Publishers Pvt. Limited.
- Ohala, M. and Ohala, J. (1972). The problem of aspiration in Hindi phonetics. *Project on Linguistic Analysis Reports*, pages 63–70.
- Ohala, M. and Ohala, M. (1975). Nasals and nasalization in Hindi. *Nasalfest*, pages 317–32.
- Peterson, G. E. (1961). Parameters of vowel quality. *Journal of Speech and Hearing Research*, 4(1):10–29.
- Peterson, G. E. and 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.
- Purnell, H. C. (1965). *Phonology of a Yao Dialect*. Hartford Seminary Foundation, Hartford.
- R Core Team (2022). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.
- Ravindran, N. (1974). *Angami phonetic reader*, volume 10. Central Institute of Indian Languages.
- Recasens, D. (2022a). Acoustic characteristics and placement within vowel space of full schwa in the world's languages: A survey. *Journal of the International Phonetic Association*, 52(1):59–94.
- Recasens, D. (2022b). Acoustic characteristics and placement within vowel space of full schwa in the world's languages: A survey. *Journal of the International Phonetic Association*, 52(1):59–94.
- Rose, P. (1987). Considerations in the normalisation of the fundamental frequency of linguistic tone. *Speech communication*, 6(4):343–352.
- Sanyu, V. (1996). *A history of Nagas and Nagaland: dynamics of oral tradition in village formation*. Commonwealth Publishers.
- Sarma, B. D., Sarmah, P., Lalhminghlu, W., and Prasanna, S. M. (2015). Detection of Mizo tones. In *Sixteenth Annual Conference of the International Speech Communication Association*.
- Sarmah, P., Dihingia, L., and Choudhury, D. (2015a). An acoustic study of Bodo vowels. *Himalayan Linguistics*, 14(1).

- Sarmah, P., Dihingia, L., and Lalhminghlui, W. (2015b). Contextual variation of tones in Mizo. In *Sixteenth Annual Conference of the International Speech Communication Association*.
- Schwartz, J.-L., Boë, L.-J., Vallée, N., and Abry, C. (1997a). Major trends in vowel system inventories. *Journal of phonetics*, 25(3):233–253.
- Schwartz, J.-L., Boë, L.-J., Vallée, N., and Abry, C. (1997b). Major trends in vowel system inventories. *Journal of Phonetics*, 25(3):233–253.
- Shafer, R. (1955). Classification of the Sino-Tibetan languages. *Word*, 11(1):94–111.
- Shrem, Y., Goldrick, M., and Keshet, J. (2019). Dr. VOT: Measuring positive and negative voice onset time in the wild. *arXiv preprint arXiv:1910.13255*.
- Simonet, M., Casillas, J. V., and Díaz, Y. (2014). The effects of stress/accent on VOT depend on language (English, Spanish), consonant (/d/,/t/) and linguistic experience (monolinguals, bilinguals). In *Speech Prosody 7: Proceedings of the 7th International Conference on Speech Prosody: Social and Linguistic Speech Prosody. Trinity College. ISSN*, pages 2333–2042.
- Smith, K. D. (1968). Laryngealization and de-laryngealization in Sedang phonemics. *Linguistics*, 6(38):52–69.
- Smith, W. F. (1973). *A formant study of whispered vowels*. PhD thesis, The University of Oklahoma.
- Sreedhar, M. (1976). *Sema phonetic reader*. CIIL, Mysore, India, phonetic reader series, 15. edition.
- Steed, W. and Hardie, P. (2004). Acoustic properties of the Kuman voiceless velar lateral fricative. In *Proceedings of the 10th Australasian International Conference on Speech Science and Technology*, pages 346–351. Citeseer.
- Stevens, K. N. and Blumstein, S. E. (1975). Quantal aspects of consonant production and perception: A study of retroflex stop consonants. *Journal of Phonetics*, 3(4):215–233.
- Styler, W. (2013). Using Praat for linguistic research. *University of Colorado at Boulder Phonetics Lab*.
- Styler, W. (2015). *On the acoustical and perceptual features of vowel nasality*. PhD thesis, University of Colorado at Boulder.
- Styler, W. (2017). On the acoustical features of vowel nasality in English and French. *The Journal of the Acoustical Society of America*, 142(4):2469–2482.
- Subbarao, K. and Kevichüsa, M. (1999). Internal relative clauses in Tenyidie (Angami): a case of hierarchical precedence vs. linear precedence. *Linguistics of the Tibeto-Burman Area*, 22(1):149181.

- Sun, K.-C. and Huang, T. (2012). A cross-linguistic study of Taiwanese tone perception by Taiwanese and English listeners. *Journal of East Asian Linguistics*, 21:305–327.
- Suokhrie, K. (2015). Internal variation in Angami: A case study of Kohima village. *Unpublished M. Phil. dissertation, University of Delhi*.
- Suokhrie, K. (2017). Clans and clanlectal contact. *Asia-Pacific Language Variation*, 2(2):188–214.
- Svantesson, J.-O. (1986). Acoustic analysis of Chinese fricatives and affricates. *Working papers/Lund University, Department of Linguistics and Phonetics*, 25.
- Svantesson, J.-O. and House, D. (2006). Tone production, tone perception and Kammu tonogenesis. *Phonology*, 23(2):309–333.
- Tabain, M., Butcher, A., Breen, G., and Beare, R. (2014). A spectral analysis of laterals in three Central Australian languages. *Proceedings of the 15th Australasian Speech Science and Technology Association*, pages 98–101.
- Tabain, M., Butcher, A., Breen, G., and Beare, R. (2016a). An acoustic study of multiple lateral consonants in three Central Australian languages. *The Journal of the Acoustical Society of America*, 139(1):361–372.
- Tabain, M., Butcher, A., Breen, G., and Beare, R. (2016b). An acoustic study of nasal consonants in three Central Australian languages. *The Journal of the Acoustical Society of America*, 139(2):890–903.
- Teo, A. B. et al. (2014). *A phonological and phonetic description of Sumi, a Tibeto-Burman language of Nagaland*. Asia-Pacific Linguistics.
- Terhiiija, V. and Sarmah, P. (2020). Aspiration in voiceless nasals in Angami. In *Proceedings of Meetings on Acoustics 179ASA*, volume 42, page 060008. Acoustical Society of America.
- Terhiiija, V., Sarmah, P., and Vijaya, S. (2019). Development of Speech Corpus and Automatic Speech Recognition of Angami. *22nd Oriental-COCOSDA, Cebu, Philippines*.
- Thurgood, E. (2009). Coronal contrasts in Anong. *Journal of the International Phonetic Association*, 39(1):53–66.
- Tse, H. (2005). *The phonetics of VOT and tone interaction in Cantonese*. PhD thesis, University of Chicago.
- Tzudir, M., Sarmah, P., and Prasanna, S. M. (2021). Analysis and modeling of dialect information in Ao, a low resource language. *The Journal of the Acoustical Society of America*, 149(5):2976–2987.

- Van Dam, M. (2003). VOT of American English stops with prosodic correlates. *The Journal of the Acoustical Society of America*, 113(4):2328–2328.
- Van Hoof, S. and Verhoeven, J. (2011). Intrinsic Vowel F0, the Size of Vowel Inventories and Second Language Acquisition. *Journal of Phonetics*, 39(2):168–177.
- Wang, W. S.-Y. (1976). Language change. *Annals of the New York Academy of Sciences*, 280(1):61–72.
- Whalen, D. H. and Levitt, A. G. (1995). The Universality of Intrinsic F0 of Vowels. *Journal of phonetics*, 23(3):349–366.
- Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York.
- Wong, Y. W. (2006). Contextual tonal variations and pitch targets in Cantonese. In *Proceedings of speech prosody*, pages 317–320. Citeseer.
- Xu, C. X. and Xu, Y. (2003). Effects of consonant aspiration on Mandarin tones. *Journal of the International Phonetic Association*, 33(2):165–181.
- Xu, Y. (1994a). Asymmetry in contextual tonal variation in Mandarin. *Advances in the study of Chinese language processing*, 1:383–396.
- Xu, Y. (1994b). Production and perception of coarticulated tones. *The Journal of the Acoustical Society of America*, 95(4):2240–2253.
- Xu, Y., Gandour, J. T., and Francis, A. L. (2006). Effects of language experience and stimulus complexity on the categorical perception of pitch direction. *The Journal of the Acoustical Society of America*, 120(2):1063–1074.
- Yip, M. (2002). *Tone*. Cambridge University Press.
- Yu, K. M. (2009). Contextual tonal variation in level tone languages. *The Journal of the Acoustical Society of America*, 125(4):2571–2571.
- Zakaria, M. (2020). The segmental phonetics and phonology of Hyow, A Tibeto-Burman language of Bangladesh. *Linguistics of the Tibeto-Burman Area*, 43(2):309–346.
- Zec, D. (1999). Footed tones and tonal feet: rhythmic constituency in a pitch-accent language. *Phonology*, 16(2):225–264.