INTERSPEECH 2019
September 15–19, 2019, Graz, Austria

Vowel-Tone Interaction in Two Tibeto-Burman Languages

Wendy Lalhminghlui[1], *Viyazonuo Terhiija*[1], *Priyankoo Sarmah*[1,2]

[1]Department of Humanities and Social Sciences
[2]Center for Linguistic Science and Technology
Indian Institute of Technology Guwahati, Assam, India
{wendy,viyazonuo,priyankoo}@iitg.ac.in

Abstract

Intrinsic F0 (IF0) has been considered a phonetic phenomenon that has a physiological basis. However, considering cross linguistic variation in IF0, it is also assumed that there is an amount of speaker intended control on IF0. This work looks into the two tone languages spoken in North East India and confirms the evidence of IF0 in the languages. However, it also shows that as soon as speakers exert control over F0 for tone production, IF0 differences diminish. As previously reported, in this study too, IF0 differences were noticed to be more pronounced in the higher F0 regions.

Index Terms: Intrinsic F0 (IF0), tone, Angami, Mizo, vowel tone interaction

1. Introduction

In this work, we look into the interaction of tones and vowels in two Tibeto-Burman languages, namely, Angami and Mizo, spoken in the North-East of India. The specific aim of this work is to identify and account for the changes in tone characteristics due to the influence of the vowels. The positive correlation of vowel height and F0 has been referred to as intrinsic F0 (IF0) in the literature and we investigate how IF0 affects the pitch of different tones in Angami and Mizo.

According to Lieberman, IF0 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 [1]. 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 production of high vowels, the tongue pulls on the larynx, giving rise to high tension of the vocal folds resulting in higher F0 values [2, 3]. Ohala and Eukel provided substantial evidence for the tongue pull hypothesis through their bite block experiment, making it a dominant explanation for IF0 in the years to follow [4]. However, the amount of control a speaker exerts on IF0 is not quite clear.

There is also an enhancement hypothesis that posits that IF0 variation is deliberate and is aimed at improving auditory salience [5]. It argues that IF0 is used as an additional feature in enhancing the perception of vowel contrasts. In other words, languages that have large vowel inventory may need IF0 to enhance the perception of different vowel categories. While earlier studies did not find any correlation between vowel inventory size and IF0 differences [6], a later study confirms that there is a direct relationship between the two [7]. Contrary to that, a later work investigating IF0 differences in American English varieties found that the sources of the control of F0 use may not be the size of vowel inventory but may be socio-cultural, reflecting regional variation in speech [8]. 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 case of tone languages, a few interesting observations are made in the literature [6]. 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, induce higher tones [9] as in non-tone languages. Clear evidence for IF0 in three African register tone languages namely, Ibibio, Kunama, Dschang; are found in a previous work [10]. 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 [11].

While there are studies that have investigated the evidence of IF0 in Chinese and African languages, there are not many studies that 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 aims at addressing this gap in the literature by investigating the interaction of vowel and F0 in two tone languages of the Tibeto-Burman languages family, namely, Angami and Mizo. This study tries to answer the following questions:

1. Is there any evidence of vowel effect on the F0 of tones in Angami and Mizo?
2. Do vowel effects manifest equally along the duration of the tone?
3. Considering the differences in the tone inventory of Angami and Mizo, is there a difference in the vowel effect on F0 of tones?

The rest of the paper is arranged in the following manner: Section 2 describes the languages and the methodology adopted in this study. Section 3 presents the results and finally Section 4 discusses the findings and concludes the study.

2. Methodology

2.1. Angami and Mizo Language

Angami and Mizo are under-resourced languages of the Tibeto-Burman language family, spoken in the North-East of India. According to the Census of India, 2011, Angami (also referred to as Tenyidie, ISO $639-3$: njm) is spoken by $152,796$ people in Nagaland and Mizo (ISO 639-3 code: lus) is spoken by $8,30,846$ people mainly in Mizoram and its neighbouring states of India. Both Angami and Mizo are tone languages of the Tibeto-Burman language family [12, 13]. Mizo has five vowels,

http://dx.doi.org/10.21437/Interspeech.2019-2808

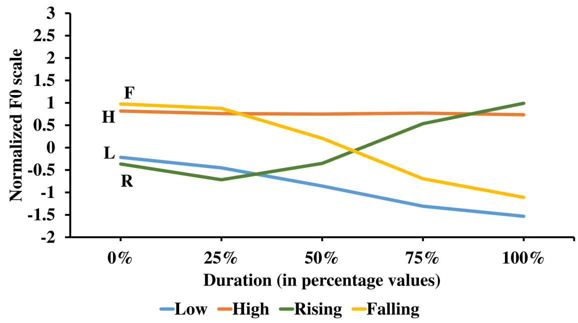

Figure 1: *Average F0 contours of four Mizo tones.*

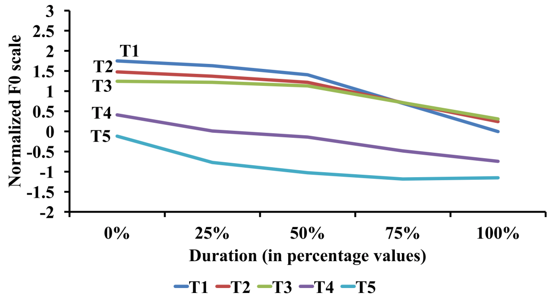

Figure 2: *Average F0 contours of five Angami tones.*

/a,ɛ,i,ɔ,u/ and four lexical tones, namely, low, high, rising and falling [14, 15, 16, 17, 18, 19] as shown in Figure 1. The salient characteristics of Mizo tones also aid automatic speech recognition in noisy conditions [20]. Apart from the four phonological tones, in a previous study, evidence of another salient Mizo tone resulting from tone sandhi is also mentioned [21].

Standard Angami, as spoken in the Kohima area, has six vowels, /a, i, ɛ, u, o, ə/. However, previous literature is divided on the number of tones in Standard Angami and report either four or five tones in the language [22, 23, 24, 25, 26, 27]. We conducted a pilot study before collecting the data from the participants in this study and found that there is no acoustic difference between the T2 and T3 produced by the Standard Angami speakers. The previous study on two non-standard varieties of Angami also yielded similar results on tones [28]. Hence, in this study T1 stands for the highest tone, T5 for the lowest tone and T2 and T3, even though these are reported as two separate tones in some of the previous literature, they are considered as a single tone category (T2/T3 hereafter) in this work as shown in Figure 2.

2.2. Speakers

For Mizo, 8 native speakers of Mizo (4 male, 4 female) participated in the study. All the speakers were born and brought up in Mizoram with Mizo as their first language, they also speak and understand English. Their age was between 20 and 40 years at the time of data collection. No one reported any speech and hearing impairment. As for Angami, 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.

Table 1: *Distribution of vowel tokens in Mizo.*

	High	Low	Rising	Falling	Total
/a/	23	24	23	22	92
/ɛ/	23	24	23	19	89
/i/	24	24	23	23	94
/ɔ/	24	18	22	23	87
/u/	23	23	24	24	94
Total	117	113	115	111	456

Table 2: *Distribution of vowel tokens in Angami.*

	T1	T2 / T3	T4	T5	Total
/a/	20	26	16	7	71
/ɛ/	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

2.3. Materials

The data for Mizo consists of 5 Mizo long vowels /a, i, ɛ, u, ɔ/ produced in isolation with each of the 4 tones in Mizo, namely, high (H), low (L), rising (R) and falling (F) tone resulting in 20 unique tokens. To replicate the Mizo four tones on the vowels, a meaningful Mizo word, /bɛl/, was given to the participants as cue. The word /bɛl/ with four different meanings for the four Mizo tones, i.e, high - 'pot', low - 'to rely', rising - 'thorough' and falling - 'to stick' along with the target vowel were displayed on a computer screen. Each token was displayed randomly on the screen to avoid listing effect. The participants were requested to produce the vowel by using the same tone of /bɛl/ shown on the screen with all the five vowels. Out of the 20 tokens, only 5 could be associated with real lexical words in Mizo, viz. /a/ with rising tone for 'mad', /o/ with falling tone for 'yes', /e/ with low tone for 'defecate', /u/ with a rising tone for 'howl' and /u/ with a high tone for 'elder brother/ sister'. However, the speakers did not seem to access the lexical meaning of the words as they were focused on replicating the tones in the speaking task and seemed to have considered all 20 of the tokens devoid of any meaning. Each unique vowel was produced three times by each participant resulting in 480 tokens. However, only 456 tokens were considered for analysis as 24 token were mispronounced. The number of occurrences of vowels in each tone category for the present study is presented in Table 1.

The data for Angami consists of 6 Angami vowels /a, i, ɛ, 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 /pɛ/, 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 listing effect. The participants were requested to mimic the tone of /pɛ/ syllable and produce the vowels. Each token was produced five times by each participant resulting in 900 tokens. However, 585 tokens are considered as a large number of tokens were produced with creaky phonation, making pitch estimation impossible and many others were mispronounced. The tokens considered for analysis in this study are presented in Table 2.

2.4. Acoustic and statistical analysis

The speech recordings were transferred to a computer for analysis. The sound files were segmented and annotated by the native speakers of Mizo and Angami by means of listening and visual examination of the waveforms and spectrograms using Praat 6.0.43 [29]. The tone boundaries are marked at the initiation of the F0 till the point of termination. A script was used to automatically extract F0 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 F0 values were normalized using the z-score normalization method [30]. The averaged normalized F0 values for each vowels were plotted for visual examination. Statistical analysis, such as the Linear Mixed Effects (LME) test, Analysis of Variance (ANOVA) and post-hoc tests were conducted to support the results of the analysis. All statistical analyses were conducted using R [31] and for LME, the *lme4* package was used [32]. For the LME test, separate analysis were conducted with F0 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.

3. Results

3.1. Mizo

A visual examination of the plots of the Mizo tones with different vowel categories (Figure 3, Figure 4, Figure 5 and Figure 6) show that there is a significant effect of vowel type at the beginning of the pitch contours (0%). That is, the high vowels /i/ and /u/ induce higher F0 and low vowels do the reverse. As the pitch contour progresses, the contour characteristics change according to tone types. In case of low and falling tones (Figure 3 and 6), vowel height specific effects are observed at 25% of the pitch contour. However, as the pitch falls, the vowel category induced differences diminish. In case of the high tone, as the pitch contour remains consistently high throughout the total duration of the pitch, the vowel category specific differences are maintained throughout the total duration of the pitch contour. In case of the rising tone, the vowel category specific differences in the pitch contour reduced in the 25% of the total duration. However, as the pitch contour rises to higher F0 levels in 75% of the total duration, the vowel specific differences show more prominently.

The mid vowels /ɛ/ and /ɔ/ are in the middle of the two extreme F0 values in all the four tones, except for /ɔ/ in the falling tone context. In low and rising tones, the mid vowels show similar patterns where both overlap with each other from the initial until the final point of the F0 contour. In falling tones, the mid vowels start with different F0 and then overlap after 50% till the end. Low vowel /a/ and /ɔ/ overlap in falling tones throughout the total duration. High vowels and mid vowels show almost similar F0 contours in most of the F0 contour plots.

The results of the LME tests, as seen in Table 3, on Mizo tones showed that vowel categories interact significantly with F0 values for all the tones in Mizo. While the effect is significant from the beginning till 75% of the total duration, no significant effect of vowel categories is noticed at the end of the pitch contour (100%).

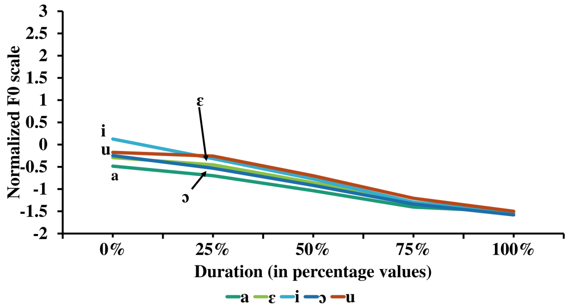

Figure 3: *F*0 *contours of low tones with five Mizo vowels.*

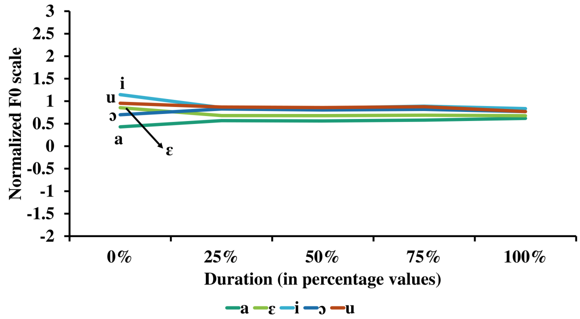

Figure 4: *F*0 *contours of high tones with five Mizo vowels.*

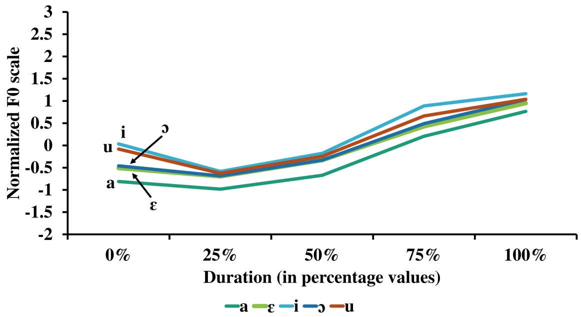

Figure 5: *F*0 *contours of rising tones with five Mizo vowels.*

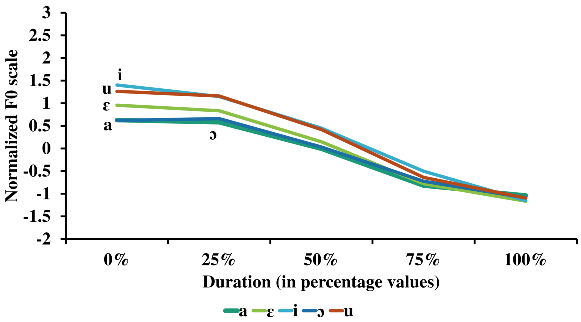

Figure 6: *F*0 *contours of falling tone with five Mizo vowels.*

Table 3: *Result of LME test in Mizo tones*

	0%	**25%**	**50%**	**75%**	**100%**
$\chi^2(4)$	58.72	53.29	43.30	26.23	1.34
p	< 0.001	< 0.001	< 0.001	< 0.001	0.86

3.2. Angami

In the present study, all the register tones in Angami tend to fall when produced in isolation. The average F0 contours of the five Angami tones are provided in Figure 2. As seen from Figure 7 till Figure 10, 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 F0 levels, the vowel specific differences diminish. T1, the highest tone in Angami shows largest vowel specific F0 differences until the 50% of the pitch contour.

The height of the vowel /u/ seems to correspond with higher F0 in T1, T4 and T5, indicating robust vowel and tone interaction. At around 25%, high vowel /i/ and mid vowel /ɛ/ are overlapping 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, vowel /u/, /i/ and /ɛ/ induce higher F0 and vowels /a/ and /o/ induce lower F0. The vowel /ə/ stands out as in most of the tones, it reaches its tone target only at the 50% of duration of the tone.

The results of the LME test, reported in table 4, showed that vowels interact significantly with the F0 at all points of the F0 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.

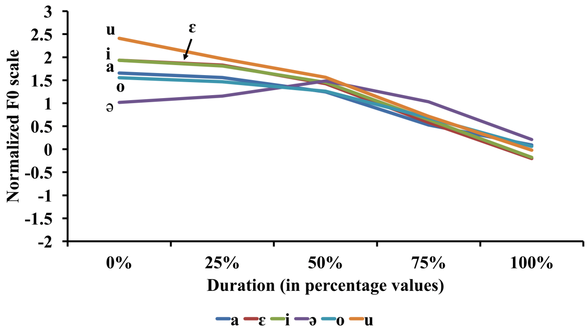

Figure 7: F0 *contour plot of T1 with all the six Angami vowels.*

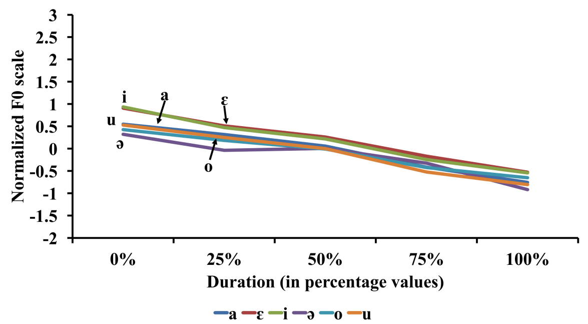

Figure 8: F0 *contours of T2/T3 with all Angami vowels.*

Table 4: *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

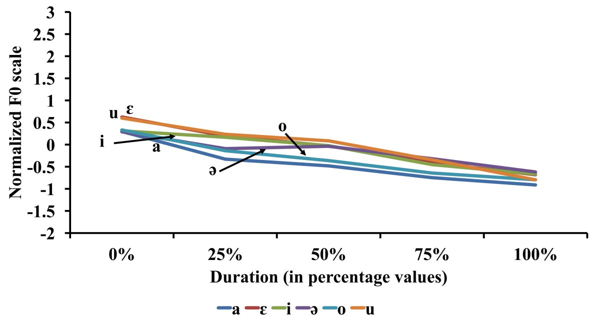

Figure 9: F0 *contour plot of T4 with all the six Angami vowels.*

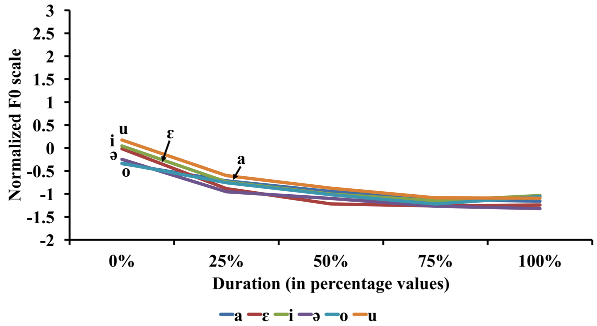

Figure 10: F0 *contour plot of T5 with all the six Angami vowels.*

4. Discussion and conclusions

This study suggests that there is a clear correlation between F0 and vowel quality. High vowels induce higher F0 and low vowels induce F0 lowering in Mizo and Angami. The effect of vowels on tones is prominent in both languages at the initial part of the F0 contour. At the initial points (0%), the speakers of the two languages do not control the tone specific pitch in the utterance and hence, IF0 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 F0 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 IF0. However, as speakers try to reach the tone targets, the IF0 differences are subdued by tone specific values. Hence, the effect of vowels may not be manifested equally along the duration of the tone. While Mizo tones has contour tones and Angami has register tones, no significant vowel effect difference was observed on F0 based on the salient pitch contour dynamics of the two languages.

5. Acknowledgements

The speech data used in this study came from two projects, “Acoustic and Tonal Features Based Analysis of Mizo”, funded by the Ministry of Human Resource Development (MHRD) & Ministry of Information Technology (MeITy), Govt. of India and “Sociolinguistic Study of Phonetic Variations among the Clans and Khels of Two Southern Angami Villages”, by the Indian Council of Social Science Research.

6. References

[1] P. Lieberman, "A Study of Prosodic Features," *Current Trends in Linguistics*, vol. 12, no. Part 10, pp. 2419–2450, 1970.

[2] I. Lehiste, *Suprasegmentals*. Massachusetts Institute of Technology Press, 1970.

[3] P. Ladefoged, "Some Possibilities in Speech Synthesis," *Language and Speech*, vol. 7, no. 4, pp. 205–214, 1964.

[4] J. J. Ohala and B. W. Eukel, "Explaining the intrinsic pitch of vowels," *In honor of Ilse Lehiste*, pp. 207–215, 1987.

[5] R. L. Diehl and K. R. Kluender, "On The Objects of Speech Perception," *Ecological Psychology*, vol. 1, no. 2, pp. 121–144, 1989.

[6] D. H. Whalen and A. G. Levitt, "The Universality of Intrinsic F0 of Vowels," *Journal of phonetics*, vol. 23, no. 3, pp. 349–366, 1995.

[7] S. Van Hoof and J. Verhoeven, "Intrinsic Vowel F0, the Size of Vowel Inventories and Second Language Acquisition," *Journal of Phonetics*, vol. 39, no. 2, pp. 168–177, 2011.

[8] E. Jacewicz and R. A. Fox, "Intrinsic Fundamental Frequency of Vowels is Moderated by Regional Dialect," *The Journal of the Acoustical Society of America*, vol. 138, no. 4, pp. EL405–EL410, 2015.

[9] P. Hoole and F. Hu, "Tone-Vowel Interaction in Standard Chinese," in *International Symposium on Tonal Aspects of Languages: With Emphasis on Tone Languages*, 2004.

[10] B. Connell, "Tone Languages and the Universality of Intrinsic F0: Evidence from Africa," *Journal of Phonetics*, vol. 30, no. 1, pp. 101–129, 2002.

[11] A. Gonzales, "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*, 2009, pp. 145–155.

[12] R. Burling, "The Tibeto-Burman Languages of Northeastern India," *The Sino-Tibetan Languages*, vol. 3, p. 169, 2003.

[13] G. E. Marrison, "The Classification of the Naga Languages of Northeast India,," Ph.D. dissertation, U. London, SOAS, vol. 1, 292p., vol. 2,, 1967.

[14] A. Weidert, *Componential Analysis of Lushai Phonology*. John Benjamins Publishing, 1975, vol. 2.

[15] L. Fanai, "Some Aspects of the Lexical Phonology of Mizo and English: An Autosegmental Approach," Ph.D. dissertation, CIEFL, Hyderabad, India, 1992.

[16] L. Chhangte, "Mizo Syntax," Ph.D. dissertation, University of Oregon, Eugene, U.S.A, 1993.

[17] P. Sarmah and C. R. Wiltshire, "A Preliminary Acoustic Study of Mizo Vowels and Tones," *Journal of the Acoustical Society of India*, vol. 37, no. 3, pp. 121–129, 2010.

[18] P. Sarmah, L. Dihingia, and W. Lalhminghlui, "Contextual Variation of Tones in Mizo," in *Sixteenth Annual Conference of the International Speech Communication Association*, 2015, pp. 983–986.

[19] B. D. Sarma, P. Sarmah, W. Lalhminghlui, and S. M. Prasanna, "Detection of Mizo Tones," in *Sixteenth Annual Conference of the International Speech Communication Association*, 2015.

[20] B. D. Sarma, A. Dey, W. Lalhminghlui, P. Gogoi, P. Sarmah, and S. M. Prasanna, "Robust Mizo Digit Recognition Using Data Augmentation and Tonal Information," in *Proc. 9th International Conference on Speech Prosody*, 2018, pp. 621–625.

[21] W. Lalhminghlui and P. Sarmah, "Production and Perception of Rising Tone Sandhi in Mizo," *in Proceedings of Tonal Aspects of Languages*, 2018.

[22] R. Burling, "Angami Naga Phonemics and Word List," *Indian Linguistics*, vol. 21, pp. 51–60, 1960.

[23] N. Ravindran, *Angami Phonetic Reader*. Central Institute of Indian Languages, 1974, vol. 10.

[24] P. P. Giridhar, *Angami Grammar*. Central Institute of Indian languages, 1980, vol. 6.

[25] D. Kuolie, *Structural Description of Tenyidie: A Tibeto-Burman Language of Nagaland*. Ura Academy, Publication Division, 2006.

[26] N. Chase, "A Descriptive Analysis of the Khwunomia Dialect of Angami," Ph.D. dissertation, University of Poona, Maharashtra, India, 1992.

[27] K. Suokhrie, "Clans and Clanlectal Contact," *Asia-Pacific Language Variation*, vol. 2, no. 2, pp. 188–214, 2017.

[28] V. Terhiija, P. Sarmah, and S. Vijaya, "Acoustic Analysis of Vowels in Two Southern Angami Dialects." *Presented at Oriental COCOSDA 2018, 7-8 May, Miyazaki, Japan*, 2018.

[29] P. Boersma *et al.*, "Praat, A System for Doing Phonetics by Computer," *Glot International*, vol. 5, 2002.

[30] P. Rose, "Considerations in the Normalisation of the Fundamental Frequency of Linguistic Tone," *Speech Communication*, vol. 6, no. 4, pp. 343–352, 1987.

[31] R Core Team, *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria, 2018. [Online]. Available: https://www.R-project.org/

[32] D. Bates, D. Sarkar, M. D. Bates, and L. Matrix, "The lme4 Package," *R Package Version*, vol. 2, no. 1, p. 74, 2007.