

# Analysis of Manipuri Tones in ManiTo: A Tonal Contrast Database

**Thiyam Susma Devi**

IIT Guwahati

India

thiyam.devi@iitg.ac.in

**Pradip K. Das**

IIT Guwahati

India

pkdas@iitg.ac.in

## Abstract

Manipuri is a low-resource, tonal language spoken mainly in India's northeastern state, Manipur. It has two tones - level and falling tones. For an acceptable Automatic Speech Recognition (ASR) system, integrating tonal cues from a potent Tone Recognition model is essential. ASR research on tonal languages, African, Asian and Indo-European tonal languages such as Thai, Chinese, Vietnamese and Mandarin have been done, but Manipuri is underexplored. This paper focuses on the fundamental analysis of the developed hand-crafted tonal contrast dataset, ManiTo. It is observed that the height and slope of the pitch contour can be used to distinguish the two tones of the Manipuri language.

## 1 Introduction

Automatic Speech Recognition (ASR) generates the transcript of a spoken speech. It plays a significant role in stopping the extinction of endangered languages and is also required for fostering economic growth and benefits. For improving the performance of ASR, tone recognition carries a crucial role in tonal languages where tone distinguishes the meaning of the words. Intensive work has been done on tonal languages like Mandarin, Thai, Vietnamese, Chinese, etc., (Kaur et al., 2020), but no work has been done on the ASR of Manipuri. Manipuri is one of the Indian Tibeto-Burman languages spoken in Manipur, a northeastern state of India. It is a tonal language in which the tone differentiates the word's meaning. So, robust tone recognition can enhance the speech-understanding tasks of Manipuri. It has its script known as Meitei Mayek (Singh et al., 2007). The script has twenty-seven main alphabets (Mapung Mayek), eight un-released characters (Lonsum Mayek), eight vowel signs (Cheitap Mayek), three punctuation marks, including diacritics or tone marker (Khudam Mayek

and Cheising Mayek for the numerals.

Manipuri has two tones (Thoudam, 1980; Khan, 1987; Chelliah, 1992; Devi, 2004; Singh et al., 2007), level and falling tones. Each one of the syllables in Manipuri bears one of the two tones. The level tone is unmarked, while the falling tone is marked with // in English representation and with the falling tone marker or lum mayek, “.” just after the syllable in the Manipuri script. Some of the tonal contrast words pair with their meanings are shown in Figure 1.

Sl. No.	Falling Tone	Meaning	Level Tone	Meaning
1	ꯀꯂꯩ /ùn/	skin	ꯀꯂꯩ /un/	ice, snow
2	ꯀꯂꯩ /in/	push	ꯀꯂꯩ /in/	follow
3	ꯀꯂꯩ /t <sup>h</sup> òŋ/	door	ꯀꯂꯩ /t <sup>h</sup> oŋ/	bridge
4	ꯀꯂꯩ /mi/	man	ꯀꯂꯩ /mi/	spider
5	ꯀꯂꯩ /pùbə/	to borrow	ꯀꯂꯩ /pubə/	to bring
6	ꯀꯂꯩ /cà/	wax	ꯀꯂꯩ /ca/	tea
7	ꯀꯂꯩ /i/	blood	ꯀꯂꯩ /i/	thatch
8	ꯀꯂꯩ /k <sup>h</sup> òi/	navel	ꯀꯂꯩ /k <sup>h</sup> oi/	bee/ fishing hook
9	ꯀꯂꯩ /là/	wide basket	ꯀꯂꯩ /la/	banana leaf
10	ꯀꯂꯩ /siŋ/	firewood	ꯀꯂꯩ /siŋ/	Ginger
11	ꯀꯂꯩ /səm/	hair	ꯀꯂꯩ /səm/	Basket
12	ꯀꯂꯩ /məŋ/	dream	ꯀꯂꯩ /məŋ/	grave

Figure 1: List of tonal contrast word pairs with their meaning.

## 2 Related Works

Internationally, progressive work on tone recognition has been done for tonal languages like Mandarin, Thai, Cantonese, and Vietnamese for the last three decades. (Linkai et al., 2021) recognized lexical tones in Mandarin speech on the 863 corpora using a model that integrates multiple features at a different scale. Experimental results showed that their technique achieved a Tone Error Rate of 10.5%. (Nguyen et al., 2016) proposed an acoustic model using the tone feature to investigate the effect of tone in the Vietnamese Large Vocabulary Continuous Speech. The result improved the perception of phonemes by 19.25% against the non-tonal phonemes system. (Krittakom and Narissara, 2015) analyzed Neuro-fuzzy based approach to recognize Thai speech. Their dataset consisted of eight Thai words recorded in different environments. The result showed that the system was robust to noise and could yield higher recognition than other recognizers. In India, some work has been done on

the tonal languages Punjabi and Mizo. (Gogoi et al., 2020) proposed a technique for Mizo tone recognition using Support Vector Machine (SVM) and Deep Neural Network (DNN) based classifiers. The experiment is performed on a dataset obtained from nineteen speakers. It achieved an accuracy of 73.39% for the SVM model and 74.11% for the DNN model. (Jyoti and Achyuta, 2020) proposed an ASR system based on pitch-dependent features and the probability of voicing estimated features. The results significantly improved the word error rate of the system.

However, there is no work done on the ASR of the Manipuri Language, which motivates us to build a dataset that contains the tonal contrast word pairs of Manipuri and examine the tone information to develop a robust ASR system for Manipuri.

## 3 Manipuri Tonal Contrast Dataset

A Manipuri tonal contrast dataset, ManiTo consisting of 3000 speech samples collected from

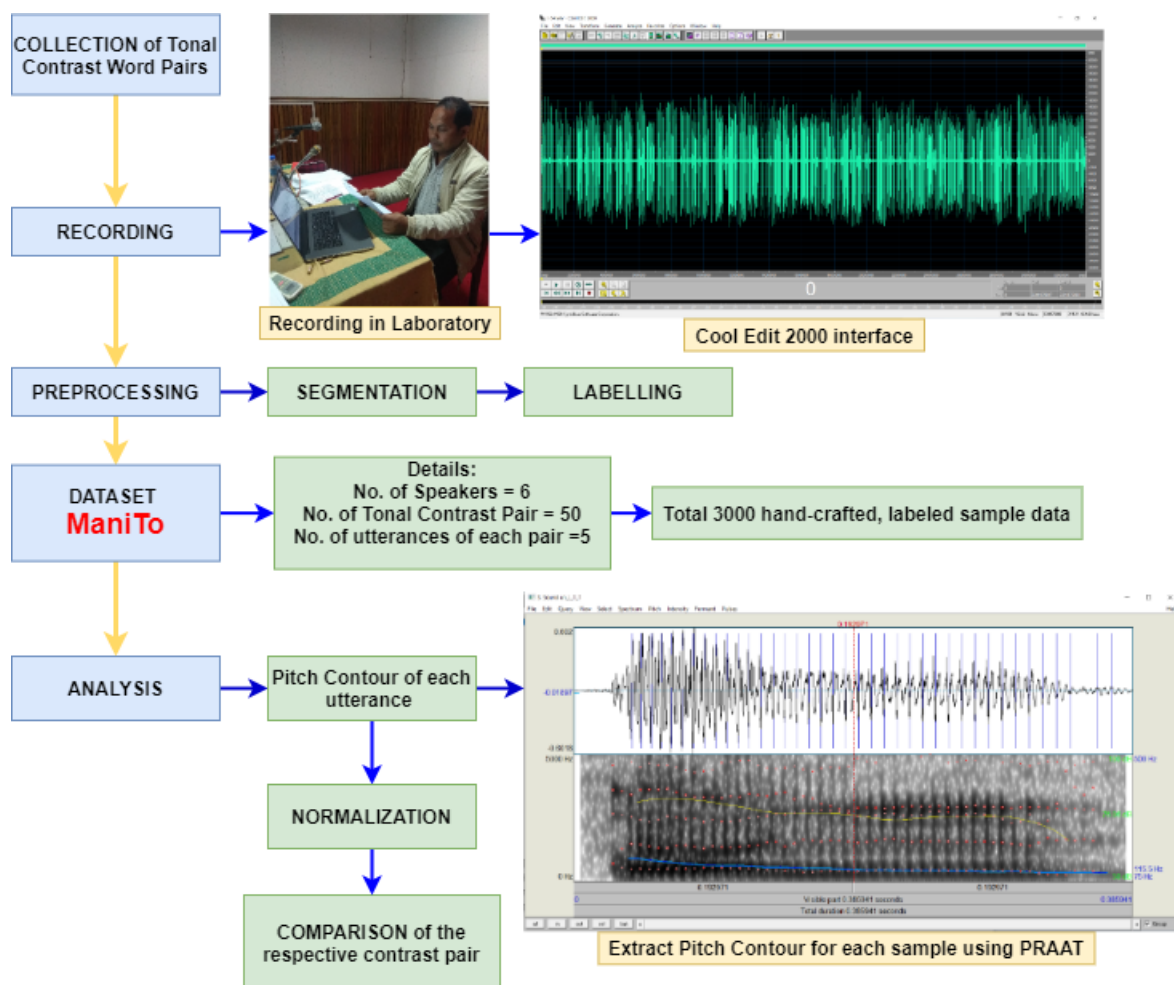


Figure 2: Flowchart depicting the summary of the work done

6 native speakers, 3 males and 3 females aged 20-35, is developed. The upper portion of the flowchart in Figure 2 shows the steps for the development of ManiTo. The corpus contains a total of 50 tonal contrast word pairs collected from different sources (Khan, 1987; Takhellambam, 2014; Thoudam, 1980; Chelliah, 1997; Singh, 2019), five utterances of each pair recorded using Cool Edit 2000 in a laboratory as well as in a quiet office environment. The recorded speech is further examined and manually segmented, keeping 1000 samples of silence at the prefix and suffix of the speech sample and saved as .wav format. Each file has the naming convention: Word-Name\_ToneType\_UtteranceID\_SpeakerID. For example, un\_f\_2\_3.wav means the particular speech sample is the falling tone “un” sound spoken by speaker 3, second utterance. The corpus ManiTo consists of hand-crafted labeled speech data of size 273MB.

## 4 Experimental Analysis

Praat, version 6.1.51 (Boersma and Van Heuven, 2001) tool is used to analyze speech data in ManiTo. The lower portion of the flowchart in Figure 2 shows the steps for analyzing the speech data.

### 4.1 Pitch Extraction

Feature extraction plays a vital role in developing a robust tone recognition system. As the pitch contour carries the salient information regarding tones, the fundamental frequency (F0) is generally used for tone recognition (Chao et al., 2019).

Initially, the pitch listing of a particular tonal contrast pair for a speaker, i.e., ten samples, 5 for falling and 5 for level, are retrieved using Praat. Praat is built with the most precise pitch extraction algorithm (Boersma and Van Heuven, 2001). All pitch values are collected and stored in different files.

### 4.2 Normalization

Normalization is performed on each extracted pitch listing value to compare the speech data efficiently with one another. For particular word pairs of a speaker, it contains 5 utterances of level tone speech and 5 utterances of falling tone speech. The normalization obtained the same length pitch listing values of the specific tone type utterances, i.e., after the normalization, the five utterances will have the same length of pitch listing values. Algorithm 1

shows the normalization method that we employed on the recorded data.

---

#### Algorithm 1: Normalization of a particular speech sample

---

**Input:**

$f0[L]$ : Pitch listing containing L values

$L_{max}$ : Maximum Pitch listing length

**Output:**

$Norm_{f0}[L_{max}]$ : Pitch listing with  $L_{max}$  values

```

1 begin
2   Step 1:  $k \leftarrow 0$   $\triangleright f0$  index
3   Step 2:  $j \leftarrow 0$   $\triangleright$  track insertion point
4   Step 3: if  $L < L_{max}$  then
5      $insert_{Loc} \leftarrow L / (L_{max} - L)$ 
6     for  $i \leftarrow 0$  to  $L_{max}$  do
7        $Norm_{f0}[i] \leftarrow f0[k]$ 
8       if  $j = insert_{Loc}$  then
9          $i \leftarrow i + 1$ 
10         $Norm_{f0}[i] \leftarrow f0[k]$ 
11         $j \leftarrow 0$ 
12      end
13       $k \leftarrow k + 1$ 
14       $j \leftarrow j + 1$ 
15    end
16  end
17 end
```

---

### 4.3 Results

The analysis on the Tonal Contrast word pair\_10, i.e., “sing” based on the pitch contour of each sample, is shown in Figure 3. The speech sounds of three male and two female speakers are analyzed separately as the female pitch is not comparable to the male pitch. The left column shows the analysis of male speech and the right for the female speech data. Figure 3a shows the normalized pitch contour of fifteen utterances of falling tone “sing” of the three male speakers, Figure 3c shows level tone normalized pitch contour, Figure 3e compares the two tones. Similarly, for female speech sounds, Figure 3b shows normalized pitch contour of ten utterances of falling tone “sing” and, Figure 3d shows the level tone and finally Figure 3f compares the two tones. From the plotted graph, we can initially claim that we can use the slope and height of pitch contour to distinguish the two tones, falling and level tones of Manipuri. The pitch of the level tone is lower than that of falling tone.

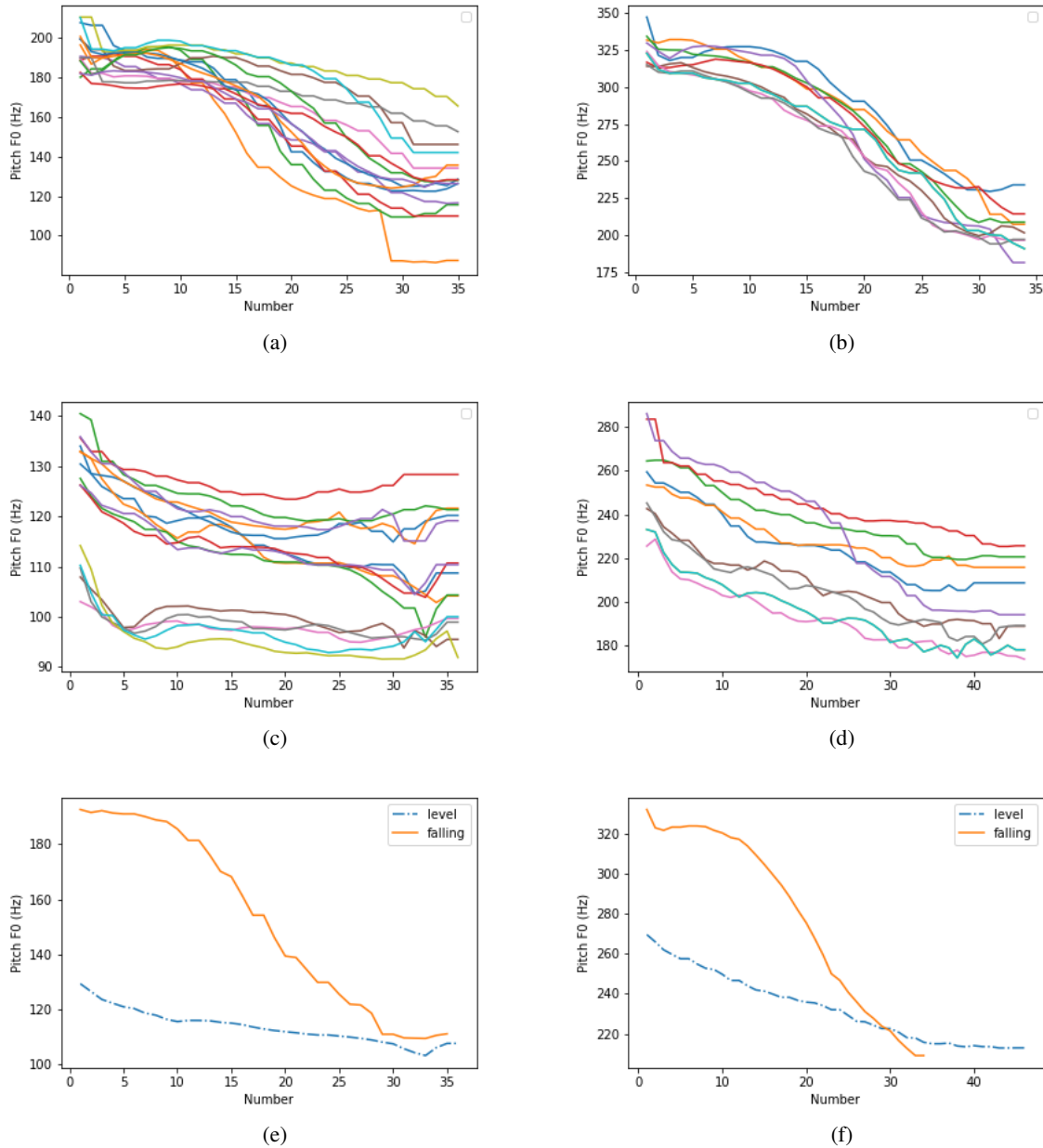


Figure 3: Normalized Pitch Contour of (a) 3 Male Speakers “sing” 15 utterances(falling tone) (b) 2 Female Speaker “sing” 10 utterances(falling tone) (c) 3 Male Speakers “sing” 15 utterances(level tone) (d) 2 Female Speaker “sing” 10 utterances(level tone) (e) Comparison of male average falling and level pitch (f) Comparison of female average falling and level pitch

## 5 Conclusion and Future Work

Initial analysis of the ManiTo, Manipuri Tonal Contrast dataset is performed. It is inferred that we can use the pitch contour to distinguish the two tones, falling and level tones. Further analysis is presently being done to precisely differentiate the tones and build a tone recognition system for Manipuri. Currently the dataset consists of 3000 samples from 6 speakers and unprocessed recordings from 3 more speakers. In the future, we will extend the dataset

for 20 speakers and this will be made available for the speech community.

## Acknowledgments

This work was supported by the Linguistic Department of Manipur University, Imphal. The authors would like to thank Prof. Chungkham Yashawanta Singh, Nameirakpam Amit, Laishram Niranjana and Dr. Yumnam Aboy Singh for their valuable advice.

## References

- Paul Boersma and Vincent Van Heuven. 2001. Speak and unSpeak with PRAAT. *Glott International*, 5.
- Hao Chao, Cheng Song, Bao yun Lu, and Yong li Liu. 2019. [Feature Extraction based on DBN-SVM for Tone Recognition](#). *Journal of Information Processing Systems*, 15(1):91–99.
- Shobhana L. Chelliah. 1992. Tone in Manipuri. In *K. L. Adams and T. J. Hudak (Eds.), Papers from the first annual meeting of the Southeast Asian Linguistics Society*. Tempe, AZ: Arizona State University.
- Shobhana L Chelliah. 1997. A grammar of Meitei. In *Mouton de Gruyter, Berlin, New York*.
- Hajarimayum Subadani Devi. 2004. Loanwords in Manipuri and their impact. In *Linguistics of the Tibeto-Burman Area, Volume 27.1 Spring*.
- Parismita Gogoi, Abhishek Dey, Wendy Lalhminghlui, Priyankoo Sarmah, and S. R. Mahadeva Prasanna. 2020. [Lexical Tone Recognition in Mizo using Acoustic-Prosodic Features](#). In *Proceedings of The 12th Language Resources and Evaluation Conference, LREC 2020, Marseille, France, May 11-16, 2020*, pages 6458–6461. European Language Resources Association.
- Guglani Jyoti and Mishra Achyuta. 2020. [Automatic speech recognition system with pitch dependent features for Punjabi language on KALDI toolkit](#). *Applied Acoustics*, 167:107386.
- Jaspreet Kaur, Amitoj Singh, and Virender Kadyan. 2020. [Automatic Speech Recognition System for Tonal Languages: State-of-the-Art Survey](#). *Archives of Computational Methods in Engineering*, 28.
- Abdul Ghaffar Khan. 1987. A Contrastive Study of Manipuri (Meiteilon) And English Phonology. In *Thesis of Doctor of Philosophy, Guwahati University*.
- Srijiranon Krittakom and Eiamkanitchat Narissara. 2015. [Thai speech recognition using Neuro-fuzzy system](#). In *2015 12th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology (ECTI-CON)*, pages 1–6.
- Peng Linkai, Dai Wang, Ke Dengfeng, and Zhang Jinsong. 2021. [Multi-scale model for mandarin tone recognition](#). In *2021 12th International Symposium on Chinese Spoken Language Processing (ISCSLP)*, pages 1–5.
- Quoc Bao Nguyen, Tat Thang Vub, and Chi Mai Luong. 2016. The Effect of Tone Modeling in Vietnamese LVCSR System. In *Procedia Computer Science*.
- Chungkham Yashwanta Singh. 2019. Manipuri Grammar, 2nd Edition,. In *Textbook, Rajesh Publications*.
- Leihaorambam Sarbajit Singh, Kabita Thaoroijam, and Pradip Kumar Das. 2007. Written Manipuri (Meiteiron) from Phoneme to Grapheme. In *Language in India, Volume 7:6*.
- Meiraba Takhellambam. 2014. Tones in Meiteilon: A Phonetic Description. In *Language in India*.
- Purna Chandra Thoudam. 1980. A Grammatical Sketch of Meiteiron. In *Thesis of Doctor of Philosophy, Jawaharlal Nehru University, New Delhi*.