THE ACOUSTIC AND PERCEPTUAL FEATURES OF TONE IN THE TIBETO-BURMAN LANGUAGE AO NAGA

Alexander Robertson Coupe
Phonetics Laboratory, Department of Linguistics (Arts), Australian National University

ABSTRACT

The tonemes of the Waromung Mongsen dialect of Ao Naga, a Tibeto-Burman of northeast India, are described with respect to their auditory and acoustic features. Even though rather small FO differences are found to separate each contrasting toneme, the results of a perception test nevertheless demonstrate that these small differences are perceptually salient to a native speaker and are readily identifiable.

1. INTRODUCTION

1.1. The Language

Ao Naga belongs to the Northern Naga branch of Kuki-Naga (Benedict 1972) and is spoken in the Mokokchung area of Nagaland, a state of north-east India. There are two principal dialects – Mongsen and Chungli. The latter has a Roman script that was developed by a missionary in the late nineteenth century and is considered the standard dialect. Each of the two main dialects subsumes a number of varieties that demonstrate varying degrees of phonological, morphological and lexical divergence. A preliminary survey suggests that every village speaks its own variety; native speakers report that the unique village-specific characteristics of each variety serve as shibboleths to identify their speakers’ villages of origin. Tonal contrasts, however, appear at the first blush to be quite constant across varieties, at least within the Mongsen dialect. A comparison of two Mongsen varieties (Waromung and Khensa) spoken in villages located on opposite sides of the traditional Ao territory demonstrates an extremely high degree of tonal correspondence, both in auditory impressions of pitch shape and number of tones. Presumably varieties of the Chungli dialect also demonstrate similar degree of tonal uniformity, although this has not yet been personally attested. Data used in this paper were gathered from two native speakers of Waromung Mongsen (henceforth WM), a variety of the Mongsen dialect spoken in Waromung village. Half of the corpus was collected on a fieldwork trip to Nagaland between December 1996 and February 1997, and the other half was collected in Australia between 1996 and 1998.

1.2. Previous Work

Very little is known about the typological characteristics of tonal languages spoken in this region of north-east India. Descriptions of languages done earlier this century tended to ignore phonemic tonal contrasts altogether. More recently a number of grammatical sketches done on Naga languages continue to give tonal phenomena rather short shrift, usually devoting no more than a cursory paragraph to its description. To the best of my knowledge Blakenship et. al. (1993) is hitherto the only article to present an acoustic description of any Naga language; their paper briefly describes the segmental phonemes, voice onset time of stops, and tones of the Khonoma dialect of Angami, with the main focus centred on the characteristics of the voiceless nasal series. This work was preceded by Burling (1960), which provides a brief description of the tone system of Angami based on an auditory analysis. Weidert (1987) gives a account of Tibeto-Burman tonology, presumably based upon his auditory analyses of the tones of Naga languages, as well as those of other Tibeto-Burman languages. This work however is concerned solely with comparative tonology, and consequently gives only fleeting descriptions of tonal systems in the languages examined. Next to nothing is known about the tone system of Ao Naga. This paper presents the first detailed auditory and acoustic description done on the language.

1.3. Data Sources

Data were collected from two tertiary-educated native speakers of WM (CL and MA), both male and in their early thirties. English was used as the contact language. The informants speak the standard Chungli dialect in addition to their native dialect, as well as a creolised variety of Assamese called Nagamese, the lingua franca of Nagaland. CL’s data were recorded in Nagaland on a portable Sony TCM-5000 with an external Sony Electret condenser microphone. MA’s data were recorded in the recording studio of the Phonetics Laboratory of the Department of Linguistics (Arts), at the Australian National University. Approximately two hundred words were uttered in isolation by each speaker, with three elicitations of each word.

2. AUDITORY ANALYSIS

2.1. Procedure

The corpus was transcribed phonetically and each syllable was assigned a tone according to auditory impressions. Initially Chao letters were used to transcribe the tones and each syllable was assigned an integer of 1 to 5, with 1 representing the lowest level. As the tones are heard to be quite consistent in pitch height over their relatively short duration, generally just one integer was found to be sufficient to characterise the pitch shape of each syllable. An exception to this was demonstrated by the one and only conjunction elicited in the corpus of CL, namely [t’ ´ʊ]
3. ACOUSTIC ANALYSIS

3.1. Procedure

The acoustic analysis is based on CL’s data. Monosyllabic citation forms representative of the three contrastive tonemes identified in the auditory analysis were digitised with Kay CSL software using 16 bit digitisation at a sampling rate of 10 KHz. Aligned narrow band spectrograms with a bandwidth of 293 Hz were then made of each digitised token. Because the pitch extraction algorithm of the software was sometimes found to be unreliable or unable to extract pitch over the total duration of the rhyme (probably attributable to the presence of reverb in some of the recorded data), the following methodology was employed. Firstly, comparison of the wave form to the formants of the temporarily aligned wide band spectrogram allowed the rhyme onset and offset to be identified and marked. The rhyme onset was taken to be the first spectrographic evidence of glottal phonation demonstrating a formant structure consistent with the vowel of the citation form, and offset was judged to be the point at which periodicity was lost in both the wave form and the temporarily aligned spectrogram, usually coinciding with a decrease in peak-to-peak amplitude. The view of the wide band spectrogram was then deleted and replaced by a narrow band spectrogram (24Hz), aligned to the marked section of the wave form. The frequency range was adjusted to 0-1000 Hz to allow a clearer view of the first few harmonics of the rhyme. Fundamental frequency (henceforth FO) was then manually measured off the highest harmonic with the best resolution at seven sampling points, being 0%, 5%, 25%, 50%, 75%, 95% and 100% of the duration of the rhyme relative to its onset and offset. As in the auditory analysis, three types of monosyllabic words were identified in preparation for the acoustic analysis: stopped monosyllables e.g. \[\text{ts} \text{\textsc{N}}\] ‘came’, stopped monosyllables with nasal codas e.g. \[\text{ts} \text{\textsc{N}}\] ‘full’, and stopped monosyllables (those with a stop coda) e.g. \[\text{ti}\] ‘horn’. These were then subdivided into three groups according to their specific tones (i.e. high, mid or low, as determined by the auditory analysis), and mean FO at each sampling point and mean rhyme duration were calculated for each group. The mean FO shapes displayed at Figure 1 are based on five to nine tokens of each tone group representing the three syllable types. An exception is the data of the Tone 2 stopped monosyllables, of which only three examples occurred in the corpus.

3.2. Results

Figure 1 below plots mean FO values as a percentage of rhyme duration for low, mid and high tonemes on three types of syllables. The most remarkable features of these citation tones are the minimal FO differences between levels – just 10 to 20 Hz – and the uniformity of their FO height over the durations of their respective rhymes. A comparison of FO
heights for unstopped monosyllables with and without nasal codas demonstrates consistently higher mean FO values for syllables with nasal codas. This could be attributable to the fact that the codas of all but two of the nineteen nasal-final monosyllables used in the acoustic analysis had a velar place of articulation. In an acoustic study of a dialect of Chinese, Rose (1992) found that velar nasal codas had a significant effect on the FO height of syllables with falling contours throughout most of their durations, and that the FO differences decreased with overall FO height.

As expected, the stopped syllables are significantly shorter in duration than the unstopped syllables, while the nasal-final syllables are marginally longer than the unstopped syllables. The tonemes of the stopped syllables are observed to have overall higher mean FO values than their unstopped non-nasal counterparts. FO contours demonstrate little variation over time between and within the tonemes of the three syllable types; indeed they are remarkably similar in shape, suggesting that height of FO is the sole perceptual cue for identification of the three tonemes.

4. PERCEPTION TEST

4.1 Procedure

The preparation for the perception test was as follows. From MA I elicited tokens of target words constituting sets of segmentally identical minimal triplets demonstrating tonal contrasts. These were then dubbed in quasi-random fashion onto another tape, interspersed with randomly selected words from MA’s previously recorded corpus. The end product was a test tape of two hundred and twenty-five words in total, containing nine tokens of each target word. The target words were recorded one month before the perception test was administered, and the informant (MA) was deliberately kept ignorant as to the purpose and nature of the test. The test tape was played and his response to each test word was noted. He was permitted to replay any words as many times as required, and the entire session was carried out in one sitting at a pace of his choosing. In the test four sets of minimal contrasts were scattered throughout the test tape. Because of limitations of space, the following discussion of the results shall be confined to just one of the sets.

4.2 Results

The results of the perception test are presented in the confusion matrix of Table 1 and the FO shapes of the stimuli are provided at Figure 2. The initial syllables of these

Figure 1: Mean FO shapes and rhyme durations for CL’s three citation tones on (a) unstopped monosyllables, (b) unstopped monosyllables with a nasal coda, and (c) stopped monosyllables.
Disyllabic words are observed to have an extremely short duration of just 30 to 60 msec and rarely exceed four or five glottal pulsations in the wave forms and aligned spectrograms examined. The small FO differences and overlapping contours suggest that the FO target is phonologically irrelevant on the first syllable of these words, although as might be expected, the initial syllables do show a dissimilation in height to the high and low tones of their respective second syllables.

The confusion matrix demonstrates a high rate of intelligibility for the words tested, with confusions only occurring between two stimuli having adjacent pitches (i.e. the LM tone of /t' maN/ ‘don’t believe’ was confused for the MM tone of /t’ maN/ ‘corpse’). Interestingly, the four instances of confusion were unidirectional. It is difficult to state with certainty what the significance of this is; one would also expect to find confusions with /t’ maN/ (LM) ‘don’t believe’ in response to the stimulus /t’ maN/ (MM) ‘corpse’, yet there was one hundred percent intelligibility for all the tokens of this word that were tested.

5. CONCLUSION

As far as an individual male’s tonal FO range is concerned, mean differences in FO height between contrasting tonemes are rather small. Nevertheless, the perception test demonstrates that despite these small differences, tones can be readily identified by a native speaker.

6. REFERENCES


