Prosodic issues in synthesising Thadou, a Tibeto-Burman tone language

Dafydd Gibbon[1], *Pramod Pandey*[2], *D. Mary Kim Haokip*[3], *Jolanta Bachan*[4]

[1] Universität Bielefeld, Bielefeld, Germany, [2]Jawaharlal Nehru University, New Delhi, India
[3] Assam University, Silchar, India, [4]Adam Mickiewicz University, Poznań, Poland

gibbon@uni-bielefeld.de, pkspandey@yahoo.com,
marykimhaokip@gmail.com, jolabachan@gmail.com

Abstract

The objective of the present analysis is to present linguistic constraints on the phonetic realisation of lexical tone which are relevant for the choice of a speech synthesis development strategy for a specific type of tone language. The selected case is Thadou (Tibeto-Burman), which has lexical and morphosyntactic tone as well as phonetic tone displacement. The last two constraint types differ from those in more well-known tone languages such as Mandarin, and present problems for mainstream corpus-based speech synthesis techniques. Linguistic and phonetic models and a 'microvoice' for rule-based tone generation are developed.

Index Terms: speech synthesis, lexical tone, morphosyntactic tone, Tibeto-Burman

1. Objectives and background

The main objective of the present analysis is to examine linguistic and phonetic constraints on the phonetic realisation of lexical tone which are necessary requirements for the choice of a speech synthesis development strategy for a specific type of tone language. The language chosen is the Tibeto-Burman minority language Thadou, which has both lexical and morphosyntactic tone (which is more typical of Niger-Congo languages [1]) and phonetic tone displacement rules (also more typical of Niger-Congo languages). Lexical tone as such is not an obstacle for speech synthesis, but the specific contextual constraints on morphosyntactic tone and on tone displacement require non-trivial solutions. The speech synthesis scenario determines the choice of a single speaker data set; planning for a more comprehensive, documentation oriented data set is based on the results of the present study..

Thadou grammar has been previously researched by [2] and [3], and Thadou phonology has been described in detail by [4]. There is so far no detailed quantitative phonetic study of Thadou tone or indeed of tone patterns in any Tibeto-Burman language, though there are several non-quantitative studies of a number of languages, including Thadou [5], and many fundamental contributions by [6].

The initial goal in the present context is to clarify the typologically interesting prosodic properties of the language sufficiently explicitly that computational linguistic and computational phonetic methods can be used in the selection of an appropriate rule-based, corpus-based or hybrid approach to speech synthesis. These prosodic properties are of interest not only for basic research on prosodic typology but also for the operational modelling of the sound systems for speech synthesis. One motivation for addressing this issue is the engineering goal of creating a usable speech synthesis system for the language for use in regional information services for a mainly rural agricultural community with an essentially oral culture and non-standardised orthography. However, this goal is at the present stage not the primary goal: basic research questions have to be clarified prior to the creation of operational models, but operational models are also needed for testing the consistency and validity of linguistic insights and descriptions. We consider operational modelling to be a necessary though not sufficient step in validating phonological models, in the formal sense of 'model' as a mapping of a theory into a domain structure, and we develop a 'microvoice' based on selected data for this purpose. Operational modelling requires close interdisciplinary cooperation between linguists, phoneticians, and native speaker experts.

The analysis of non-intonational, tonal properties of languages and the application of these analyses in operational speech synthesis applications is still in its infancy. Further, although expert opinion considers about half the world's languages to be tone languages, they are underrepresented in the spectrum of well-described languages. One reason is politico-economic: the most-described languages are either national languages of prosperous nations, or the languages of large minorities within these nations, all of which, with the notable exception of Mandarin Chinese, are 'intonation languages' (i.e. languages where the major functions of pitch patterning are phrasal, textual and discoursal) with word prosodic systems consisting either of stress-accented words, with abstract stress positions realised by variable pitch accents, as in English and many other Indo-European languages, or pitch-accented words with a stress position realised by fixed pitch patterns, as in Japanese [7], [8].

After discussion of the objectives and background we briefly discuss the situation of the Thadou language in the second section. The third section reports on the phonological and word prosodic properties of Thadou (consonants, vowels and tones) from a linguistic perspective, with a critique of previous work and discussion of problematic aspects for speech synthesis. The fourth section examines quantitative phonetic correlates of the Thadou tonal system based on an initial analysis of the canonical tonal patterns of isolated contrasting words. In the fifth section, phonological and phonetic differences between tonal patterns of canonical isolated tonal patterns and of paired sequences are examined. In the sixth and concluding section, the prospects for fulfilling further requirements for speech synthesis and for extending the model to other functional properties of Thadou tone are considered, and generalisation of the method to modelling other Tibeto-Burman tone languages is briefly discussed.

2. Thadou data

Thadou[1] (ISO 639-3: TCZ) is a Kuki-Chin language of the Kuki-Chin Naga subgroup of the Tibeto-Burman group of Sino-Tibetan languages, and is spoken in the India-Myanmar borderland, with a speaker population of about 200,000. Closely related Kuki-Chin languages are Ralte, Paite and Zo. The nomenclature of the language in the literature is quite diverse and requires mention in order to avoid confusion in

[1]We are highly indebted to the graduate students of the Tibeto-Burman study group at JNU for detailed discussion of prosodic issues in Tibeto-Burman languages. The selection of data for the recordings owes much to an article on the phonology and tonology of Thadou by Hyman [4]. Most background information was provided by the first hand experience of the authors, but some was obtained from the online *Ethnologue* atlas and the online version of the *World Atlas of Language Structures* (WALS).

language identification. Various spellings exist: Thado, Thadou, Thadow, Thaadow, Thaadou. The name also occurs in combination with names for groups of closely related languages, Kuki (in India) and Chin (in Myanmar), such as Thadou-Kuki, Kuki-Thadou, Chin-Thado, etc. Other names such as Kongsai and Thado-Ubiphei are also found. Varieties of Thadou are Baite, Changsen (Jangshen), Haokip (Kaokeep), Khongzai, Kipgen, Langiung, Sairang, Shithlou, Singson (Shingsol) and Thangngen. Thadou spelling is not standardised; local newspapers use a variety of orthographies which reflect phonological variation, particularly with regard to the close back vowel /o/, which appears as [o], [ou] or [ow] (reflected in the spellings 'Thado', 'Thadou' and 'Thadow').

With 200,000 speakers, Thadou is not, strictly speaking, an acutely endangered language, but intergenerational transmission is declining under pressure from neighbouring major languages (Hindi, Burmese), and there are other sociolinguistic influences such as language contact between multiple similar languages. Both of these influences result in extensive borrowing and code-switching as well as in more subtle sociophonetic effects on the realisation of vowels, consonants and tones.

The data for this study were recorded from a female linguistics graduate, a native speaker of the Thadou core area variety in the state of Manipur, India (WALS coordinates: 24° 25′ N, 93° 55′ E). The equipment used was an Edirol R09 solid state digital recorder. External microphones were not available so the device's internal stereo microphones were used. Recordings were made in a quiet room with almost complete sound-absorbent wall covering. The recorded data were preprocessed in the field under Linux with the Audacity audio editor (cutting, amplitude normalisation, mono track selection) and annotated for phonetic analysis with Praat [9]. Pitch data extraction and MBROLA microvoice construction were done with custom Praat and Python scripts; close copy resynthesis was performed with existing software [10].

3. Phonological and word prosodic properties of Thadou

Thadou (ISO 639-3 tcz) is a lexical tone language, like many other Sino-Tibetan languages, the most prominent of which is Mandarin Chinese. A typical Thadou tonal minimal triplet is:

sá	(H tone)	'animal'
sǎ	(LH tone)	'hot'
sà	(L tone)	'build'.

However, Thadou is not only a lexical tone language but also has morphosyntactic tone like most Niger-Congo African languages. Typical functionalities of morphosyntactic tone in Niger-Congo languages are to convey distinctions of person, gender, case, number, tense, aspect, and verb valency. Thadou has similar morphosyntactic tonal patterns, in pronominal proclitics, genitive (associative) noun compounding constructions, subject marking, and verb stem marking [4]. Similar observations have been made of other Tibeto-Burman languages [11], [12]. The grammatical details (though also a non-trivial problem for speech synthesis) are beyond the scope of the present contribution, which concentrates on lexical tone and its phonetic reflexes.

The segmental phonology of Thadou has a number of interesting properties. The consonant system (see Table 1) has an aspirated-nonaspirated contrast in unvoiced stops; labial, alveolar and velar nasals in all positions; glottal stops in final position; h in initial position; a contrast between voiced and unvoiced laterals /l/ and /ɬ/ /(with allotones [l̥], [l̥] and [lh]).

The [x] free variant allophone of /kʰ/ is the main reflex of /kʰ/ in the Thadou variety of the recorded speaker. The glottal stop occurs phonemically, but also as a realisation of /k/ syllable-finally, yielding a neutralisation of /k/ and /ʔ/. There is also dialect specific variation between [j] and [z] as the realisation of /j/. The voiceless lateral fricative /ɬ/ is not a fricative in non-word-initial onsets and may be represented there as [l̥]; it may also resyllabify as /l.h/.

Table 1: *Thadou consonant phoneme system.*

	bilabial	*labio-dental*	*alveolar*	*palatal*	*velar*	*laryn-geal*
plosive	p pʰ b		t tʰ d		k kʰ g	ʔ
nasal	m		n		ŋ	
fricative		v	s z		[x]	h
apical affricate			ts			
lateral fricative			ɬ			
lateral approx			l			
approx.	w			j		

The vowel system of Thadou (Table 2) is phonemically simple, but phonetically intricate. In several Thadou dialects, the mid vowels are realised by the diphthongs [ie] and [ou]. Phonetically Thadou also shows complex interactions at syllable level between vowel length, tone and coda consonants, leading to complex rhythms. This is a separate issue, and does not play a role in the present discussion.

Table 2*: Thadou vowel phoneme system.*

	Front	*Central*	*Back*
Close	i		u
Mid	e	ə	o
Open		a	

Thadou syllable structure is straightforwardly $C_1V(C_2)$, though previous analyses also have CVV. Our analysis shows that with very few exceptions, possibly only in loan words, length is conditioned by coda consonant and tone. In onset consonant position C_1 any consonant except /ʔ/ can occur. There are restrictions on coda consonants at C_2 and hence on diphone combinations: no /h/; final devoicing and deaspiration leading to neutralisations; no affricates; no fricatives, thus also no voiceless lateral fricative /ɬ/; /ʔ/ as the neutralising realisation of /k/.

There is some controversy about the Thadou tone inventory and the effects of tonal context in tone sequences. Hyman [4] gives a detailed analysis which covers the facts about Thadou tone and tone sequencing, but which the present authors consider to be too far abstracted from the phonetic reality of pitch patterning, and from a phonetic perspective too complex, with constraints on downstep, high tone spreading and low tone spreading.

Two different analyses of the Thadou lexical tone inventory, by the authors of this study and by Hyman [4,] are shown in Table 3.

Table 3: *Thadou tone inventories.*

Example	*This study*	*Hyman*	*Gloss*
sá	H	HL	meat, animal
sǎ	LH	LH (later: H)	hot
sà	L	L	thick, to build

The Hyman analysis would make it difficult to use corpus-based speech synthesis techniques in any interesting way, as the combinatorial properties of these rules would yield a prohibitively large corpus (despite the modestly sized syllabic system of the language).

Our simpler analysis adopts only high tone shift (spreading) as a tonological rule, and all else falls out from the choice of underlying tones on the basis of their phonetic correlates and the interpolation of downdrift values. Detailed

discussion of the evidence is beyond the scope of the present contribution, but is summarised in the following section on phonetic correlates of Thadou tone. Even this simpler and more phonetically based description introduces a magnitude of syntagmatic complexity of a kind which has previously been identified as a source of difficulty for Niger-Congo speech synthesis system development [1].

4. Phonetic correlates of Thadou tone

Since the decisive criteria for developing an operational synthesis model combine phonological with phonetic analysis, the phonetic correlates of Thadou tone and tone sequencing are dealt with in some detail.

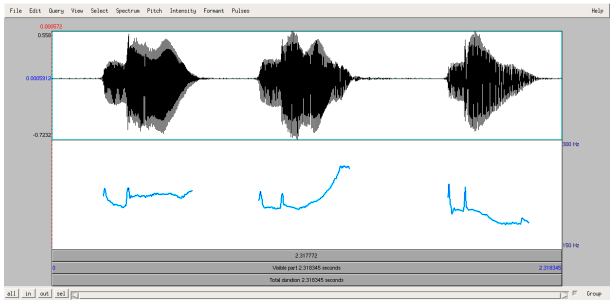

Figure 1: *Thadou tones: lów (H) 'field', lŏw (LH) 'medicine', lòw (L) 'negative marker'.*

Figure 1 illustrates the three lexical tones as a minimal triplet in canonical citation contexts: lów (H) 'field', lŏw (LH) 'medicine', lòw (L) 'negative marker'. The pitch contours also reveal microprosodic features: an initial peak, due to the [l] phone, and a peak between the [l] phone and the vowel-glide combination [ow] which is due to a sudden, tap-like release of the onset [l]. Further analysis of the microprosodic features is not within the scope of this contribution, however.

The phonetic correlates for the tones were analysed in a production experiment with canonical citation contexts, in which for each of the three tones 18 examples of monosyllabic words, each spoken 3 times, were examined. The vowel sections of each of the words were segmented and for each vowel segment the means of 16 equally spaced frequency values were extracted. Using this procedure, the interval lengths were by definition normalised for the duration of the segment and therefore rendered approximately comparable.

Table 4: *Phonetic correlates of Thadou tones (isolated citation forms). Descriptive statistics are over averages of 16 pitch samples per vowel associated with each tone (values over all measurements per tone in parentheses).*

Tone	*N*	*min*	*max*	*mean*	*sd*	*offset*	*slope*
H	18 (864)	200 (220)	244 (222)	221	0.29	221	-0.03
LH	17 (816)	215 (198)	237 (268)	220	7.07	209	1.3
L	18 (864)	192 (178)	213 (227)	203	6.3	215	-1.31

For present purposes it was not necessary to create detailed functions to model the frequency patterns, and simple numerical measures (min, max, mean, standard deviation and linear regression) were used in order to check the linguistic analyses. These results for the three tones are sufficiently expressive for present purposes (Table 4). Clearly this is a simplified approximation, particularly in respect of linear regression, and a more fine-grained modelling procedure will be necessary for a more exact and complete account which takes tonal shapes on vowels, consonantal pitch perturbations, boundary effects and mutual influences between tones into account.

The most interesting value is nevertheless the slope, which shows a clear rise on the LH tone, but also a comparably large fall on the L tone, while H is very flat. Since no utterance effect can be detected on the isolated H and LH tones, it may be provisionally assumed that there is no utterance effect on the falling L tone; however, this requires further investigation.

Correlation measures were not considered at this stage, but checks with paired t-tests on the tone data sets showed that the difference between H and L sets, and between LH and L sets is highly significant, $p<0.001$, but not so between the H and LH sets. This is to be expected from the similarity of the means between H and LH. The data sets have a sequential structure, so the t-test is only of limited discriminating value. However, both offset and slope of the H and LH sets clearly differ, providing the necessary distinguishing criteria. The visualisation of typical samples of the three tones in Figure 1 illustrates these properties.

5. Thadou tones in context

One complication for the modelling of Thadou tones is tonal modification in context (allotones). An examination of the three lexical tones in sequential pairs was performed, yielding 3*3 tones with three utterances of each sequence, i.e. 27 samples. The phonetic properties of tones in context turned out to be somewhat different from those of tones in isolation; in particular the first tone in the sequence, if L or LH, showed a narrower bandwidth than in isolation, and the difference in mean pitch heights of the neighbouring tones was smaller than would be expected from simply concatenating isolated tones.

A sequence effect which at first glance cannot be ascribed to phonetic operations of the kind just discussed, is the transformation of LH+L → L+H and LH+LH → H+H , i.e. H tone spreading [4]. As Hyman points out, this kind of tone implementation is more characteristic of African Niger-Congo tone languages than of Sino-Tibetan languages such as Mandarin. This right-shifting of the H component of the LH tone is illustrated in (Figure 2). It is not clear why these changes are restricted to LH in first position. This requires more investigation in longer sequences and utterance contexts.

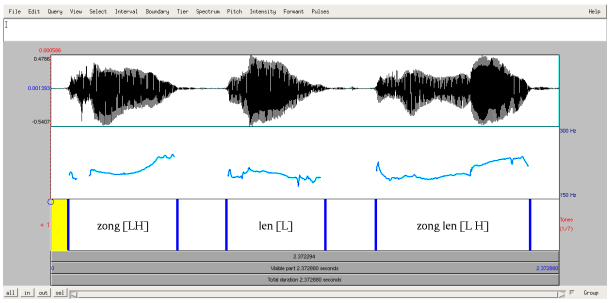

Figure 2: *LH and L tones in isolation, and L+H tone sequence with H tone shift, for* zong len *'eight monkeys'.*

The pitch relations in tone pair sequences are shown in Table 5. A comparison with Table 4 shows in general that tone properties are maintained, with the exception of the previously discussed LH+L sequence (values underlined in the table), but in addition to the narrower bandwidths there are also small downdrift effects between the tones, and sequence-initial and sequence-final effects whose significance for utterance intonation remains to be investigated.

The consequences of these phonological (LH+L→L+H) and phonetic (downdrift, bandwidth narrowing) sequence effects for corpus-based tone synthesis are clear: there is no direct mapping from pitch pattern to tone, and the H tone right-shift rule needs to be taken into account either in corpus labelling or in tone generation rules for speech synthesis.

Table 5: *Summary of pitch relations in tone sequences (mean diff: tone$_2$ – tone$_1$).*

Tones	*mean 1*	*mean 2*	*mean diff*	*sd 1*	*sd 2*	*slope 1*	*slope 2*
L+H	211	223	12	4.08	2.69	-0.84	0.11
L+LH	207	203	-4	3.35	8.12	-0.66	1.58
L+L	207	198	-9	8.12	3.58	-0.73	-0.73
H+H	225	215	-10	2.54	2.91	0.18	-0.13
H+LH	240	217	-23	3.45	2.43	0.66	0.47
H+L	234	201	-33	1.52	5.83	0.15	-1.21
LH+H	222	221	-1	2.28	2.23	0.41	0.23
LH+LH	224	222	-2	0.97	1.56	0.01	0
LH+L	201	227	26	2.58	4	0.38	0.79

It is evident that morphosyntactic contexts for tone placement require a detail of corpus markup which goes far beyond the detail used in current mainstream speech synthesis systems, even taking intonation markup into account: the granularity at the morphosyntactic level is much higher than for markup of stress or pitch accents and phrasal contours.

6. Tone synthesis: conclusion and outlook

The goal of this contribution is to discuss linguistic and phonetic prerequisites for the synthesis of prosody in the Tibeto-Burman tone language Thadou, from the point of view of prosodic typology, based on a single speaker corpus as required by the synthesis scenario. The contribution has presented the typologically unusual prosodic features of Thadou, which (like many other Tibeto-Burman languages) differs from more well-known Sino-Tibetan languages such as Mandarin in having not only phonemic tone but also morphosyntactic tone, features which are more typical of African Niger-Congo languages. A phonological tone-shift rule with, for example, LH+L sequences realised as L+H sequences, which is also characteristic of African tone languages, provides a problem for prosodic synthesis which is not shared by the Mandarin type tone language. In addition to the properties of phonemic and morphosyntactic tone, sequence-level initial, final and downdrift effects were found, indicating that Thadou is not only a tone language per se but may have intonational prosodic features; these are not in the focus of the present study and require further investigation.

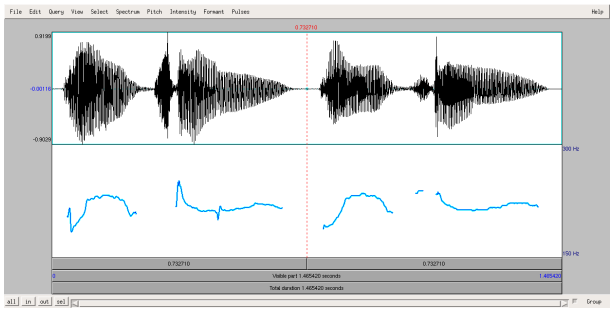

Figure 3: *Original recording and MBROLA re-synthesis with Thadou microvoice, with pitch tracks.*

As a first step towards implementing the results of this requirements study, a 'microvoice' was implemented based on diphones from selected data, using the MBROLA speech engine and voice (diphone database) construction[2] software [13], and used for close copy synthesis of the recorded data (Figure 3). However, without immediate access to native speakers it will not be possible to evaluate the synthesis in the usual way in the short term; this is planned, however, as the next step. The MBROLA diphone synthesiser is a more conservative synthesis concept than current mainstream synthesisers, but it has a clear advantage over these in having a well-defined prosodic interface for integrating rule-based prosodic patterns into the phonetised input to the synthesis engine. In the long term, the results obtained in this way will be available for use in other systems, for instance in signal processing based prosodic modification of the synthesised output of other speech synthesis methods.

The question arises as to whether the results are generalisable. Study of the linguistic literature on Tibeto-Burman shows that Thadou appears to be a rather typical representative of the Tibeto-Burman language group. The present investigation represents a first step in detailed quantitative phonetic investigation of tone systems in Tibeto-Burman languages, and offers new results which are not only of linguistic interest but are relevant to speech synthesis system design.

The next steps in work on Thadou are to complete analyses of tone sequence and consonantal perturbation effects, to evaluate empirically motivated rule-based pitch pattern generation operationally using the 'microvoice' technique described above, to create a full voice, and to collate a more comprehensive, documentation oriented corpus based on the results of the present study.

[2] Under licence from Faculté Polytechnique de Mons.

7. References

[1] Gibbon, Dafydd and Eno-Abasi Urua. "Morphotonology for TTS in Niger-Congo languages." In: Proceedings of 2rd International Conference on Speech Prosody. Dresden: TUD Press. 2006.

[2] Krishan, Shree. Thadou, a Grammatical Sketch. Calcutta: Anthropological Survey of India, Government of India. 1980.

[3] Abbi, Anvita. Reduplication in South Asian Languages. An Areal, Typological and Historical Study. New Delhi: Allied Publishers Ltd. 1992.

[4] Hyman, Larry M. "Kuki-Thaadow: An African Tone System in Southeast Asia." In Franck Floricic (ed.), Mélanges offerts à Denis Creissels. Les Presses de l'Ecole Normale Supérieure. In press.

[5] Thirumulai, M. S. Thaadou Phonetic Reader. Mysore: Central Institute of Indian Languages. 1972.

[6] Matisoff, James. Handbook of Proto-Tibeto-Burman: system and philosophy of Sino-Tibetan reconstruction. Berkeley and Los Angeles: University of California Press. 2003.

[7] Hirst, Daniel and Albert Di Cristo, eds. Intonation Systems. A Survey of Twenty Languages. Cambridge: Cambridge University Press. 1998.

[8] Hyman, Larry M.. "How (not) to do Phonological Typology: The Case of Pitch-Accent." UC Berkeley Phonology Lab Annual Report. 2007.

[9] Boersma, Paul & Weenink, David. "PRAAT, a system for doing phonetics by computer." Glot International 5 (9/10): 341-345. 2001.

[10] Bachan, Jolanta. "Automatic Close Copy Speech Synthesis." In: Speech and Language Technology. Volume 9/10:107-121. Poznań: Polish Phonetic Association. 2007.

[11] Evans, Jonathan P. 'African' tone in the Sinosphere. Language and Linguistics 9.3:463-490. 2008.

[12] Wiersma, Grace. "Yunnan Bai." In Thurgood, Graham and Randy J LaPolla, eds. The Sino-Tibetan Languages. Routledge, pp. 651-673. 2003.

[13] Dutoit, Thierry. An Introduction to Text-To-Speech Synthesis. Dordrecht: Kluwer. 1997.