Phonetic insights into a simple level-tone system: ‘careful’ vs. ‘impatient’ realizations of Naxi High, Mid and Low tones

Alexis Michaud, Jacqueline Vaissière, Minh-Châu Nguyễn

PHONETIC INSIGHTS INTO A SIMPLE LEVEL-TONE SYSTEM: 'CAREFUL' VS. 'IMPATIENT' REALIZATIONS OF NAXI HIGH, MID AND LOW TONES

Alexis Michaud, Jacqueline Vaissière** & Minh-Châu Nguyễn****

*International Research Institute MICA, HUST-CNRS-Grenoble INP
**Laboratoire de Phonétique et Phonologie, CNRS-Paris 3 ***Vietnam National University, Hanoi
alexis.michaud@mica.edu.vn; jacqueline.vaissiere@univ-paris3.fr; minhchau.ntm@gmail.com

ABSTRACT

The Naxi language has three level tones: H, M and L (plus a marginal Rising tone). The present study aims to offer phonetic insights into this simple system through examination of production data from three male speakers and one female speaker: realizations of the three level tones on CV syllables, under two reading conditions, labelled as 'CAREFUL' and 'IMPATIENT'. Fundamental frequency (F_0), glottal open quotient (O_q), and formant frequency characteristics are estimated. The three level tones span about 8 semitones under 'CAREFUL' reading and 11 semitones under 'IMPATIENT' reading. The average distance separating H from M is on the same order as that separating M from L. Under 'IMPATIENT' reading, F_0 register is higher. O_q follows speaker-specific patterns. No clear pattern of influence of tone or reading condition on vowel articulation was found. These findings (along with the original data, made available in full) offer a basis for cross-linguistic comparison.

Keywords: Naxi, level tones, H tone, M tone, L tone, intonation, glottal open quotient.

1. INTRODUCTION

1.1. The tone system of Naxi

The Naxi language, spoken in Yunnan, China, has H, M and L tones, plus a marginal Rising tone. This is an 'easy' tone system. Tones are lexically attached to syllables and there are extremely few processes of tonal change [20]. Phonetically, the tones are readily identifiable; there are no oppositions in the language that could complicate their production and perception, such as vowel quantity [24], phonation types, distinctive stress, or glottalized consonants [6]. Syllable structure is (C)(G)V+Tone. There is full agreement about tone across Naxi dictionaries [11], [23]; the notation of tones is correct even in amateur work by a botanist-explorer [26], contrasting with the difficulties encountered by that author in the transcription of Naxi vowels and consonants [18]. This testifies to the ease with which Naxi tones can be recognized by ear.

The present study aims to offer phonetic insights into this system through examination of production data.

1.2. State of the art: phonetic studies of level tones

1.2.1. The relative scarcity of phonetic studies

Much phonetic research about tone focuses on phonation-type-rich systems, which offer special challenges for experimental investigation (e.g. [27], [2], [14]). Level-tone systems (i.e. systems in which the distinctive units are relative tone heights, not detailed templates of the time course of F_0) are relatively less studied. It could seem as if the typological range were adequately covered nonetheless, since instrumental analyses are available for (i) two-tone systems, e.g. Dinka [25] and Sotho [30], (ii) three-tone systems [4:100-106], [15], [28:48-65], and (iii) four-tone systems, e.g. Mambila [3]. But the number of levels is only one aspect of these systems' diversity. Level tone systems differ considerably in the amount of combinations of levels: e.g. G|ui (Khoe-San) has H, M and L levels that combine into six tonemes, H, M, L, HM, HL, and LM [21]; the tonal space is unlikely to be similar to that of a language where there are fewer tonemes. Seen in this light, a phonetic study of Naxi tone does not appear redundant.

1.2.2. A potential side interest of phonetic studies of lexical tone: metaphoric carry-over to intonation

A paradoxical result of the transfer of the notion of level tone to intonation studies is that there is currently a more extensive phonetic literature on level tones in languages where they are posited as part of the intonation system (e.g. English: [10]) than on *bona fide* level tones: lexical (and/or morphosyntactic) tones. One of the uses of phonetic studies of level-tone systems is as an empirical basis for assessing the real degree of similarity with the "intonational tones" used in some analyses of English and other non-tonal languages.

1.3. Specific goals

The present research's main goals are:

- to ascertain the phonetic distance in terms of F_0 between the three level tones, H, M and L
- to examine other parameters, such as fine detail in the articulation of vowels
- to progress towards a dynamic picture of tones, through comparison of two reading conditions.

2. METHOD

2.1. Read materials

Forty minimal sets such as /lɑ́/ 'to strike', /lɑ̄/ 'tiger', /lɑ̀/ 'hand' were selected from a dictionary [8] (i.e. a total of 120 syllables). As monosyllables only constitute a fraction of the lexicon, it was not possible to be picky in terms of part of speech (N, V and ADJ had to be included) or phonemic composition. The set of initials is /**pʰ tʰ kʰ tsʰ ʈʂʰ**, **p k ts**, **g**, **l m n**, **s ʂ h**/ plus the empty-onset filler /**ʝ**/; vowels are /**i y ɯ u e ɤ o ɚ æ ɑ wɑ v̩**/.

The consultants were shown the target items in writing, as part of a randomized list on paper. The carrier sentence was as shown in (1):

(1)	ʈʰɯ̄	_	ʈʂʰɯ̄	ndzỳ	pɚ́	nè
	3SG		DEM	sign	write	ASP.ongoing

'(S)he is writing the character __ .'

The imagined context was that people are watching a calligrapher. The target words were recorded under two reading conditions, both tending towards hyperarticulation [16]. The context for the first reading, referred to below as CAREFUL, is that a child asks which character is being written, and the consultant provides the answer. Under the second condition, referred to as IMPATIENT, the child is naughty, asking for the umpteenth time without paying attention to answers; the consultant says the answer again, but in a way that will clarify to the child that (s)he is being rude and should not speak up again. A final affirmative particle /**mɤ̀**/, reduplicated to /**mɤ̌~mɤ̀**/ for emphasis, is added to the carrier sentence to facilitate the rendering of this attitude.

2.2. Consultants and recordings

The recordings took place in Lijiang. The four consultants are native speakers of Naxi, bilingual in Mandarin Chinese; they are members of the tiny fraction of the Naxi population that has learnt Naxi Pinyin, a romanized writing system that was encouraged during the 1980s (China's "Reforms and Opening up" decade [31]).

The recording took place in a quiet room; some household linen was hung against the walls to decrease reverberation. Audio and electroglottographic (hereafter EGG; see [7], [1]) signals were collected. The entire data set is available online from the Pangloss Collection [17], together with explanations about slight adjustments to the carrier sentence that were made in consultation with speakers (in particular, use of the topic marker /ʂɤ́/ after the target item is mandatory in some dialects); these adjustments detract somewhat from the symmetry of the data, but they helped consultants feel fully comfortable with the task, and they do not affect the study's central topic: comparison across tones and across reading conditions.

The speaker codes are F2, M5, M7 and M9, following the numbering used for the Naxi data in the Pangloss Collection [17] (current address: http://lacito.vjf.cnrs.fr/pangloss/languages/Naxi_en.htm).

2.3. Measurements

F_0 and the glottal open quotient O_q were estimated from the derivative of the EGG signal [12] using PEAKDET, a script available from the COVAREP GitHub repository [5].

F_0 values were converted to semitones using the formula shown in (2), where (i) F_{ABS} is the absolute value in Hz; (ii) F_{MEDIAN} is the speaker's median F_0, calculated as the median of all F_0 measurements for the target syllables; and (iii) F_{REL}, "relative F_0", is a value in semitones.

$$(2)\quad F_{REL} = 12 \times \frac{\log\left(\frac{F_{ABS}}{F_{MEDIAN}}\right)}{\log(2)}$$

Conversion to a logarithmic scale has been argued to correspond best to speakers' intuitions [22]; it also facilitates cross-speaker comparison.

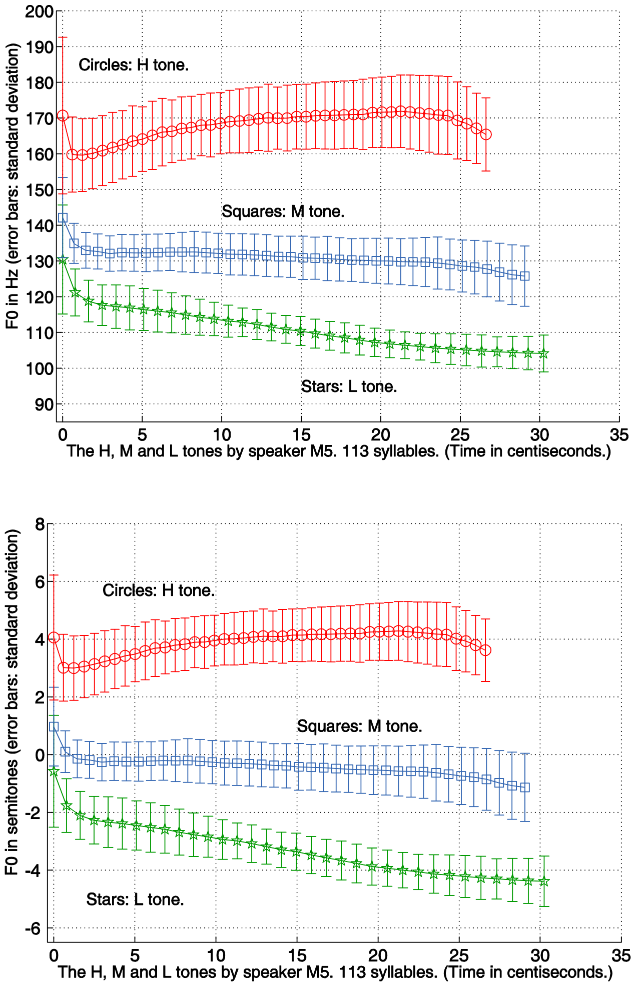
3. RESULTS

3.1. Comparison across tones

3.1.1. Fundamental frequency

Figure 1 shows the time course of F_0 for the H, M and L tones by speaker M5, in the CAREFUL reading condition. On the Hertz scale (top of Fig. 1), standard deviation is noticeably higher for the H tone than the M tone, with least variability for the L tone. This is partly an artifact of the Hertz scale, however: in the bottom part of Fig. 1, which shows the same data in semitones, the difference is slightly smaller.

Figure 1: F_0 curves for the three tones, speaker M5. Top: Hz scale; bottom: semitones.



As a rule of thumb of the distance between the three tones, one value was selected at a point where visual inspection of the curves for the various speakers (full set available in [19:470-474]) suggested that values were most stable: at 75% of vowel duration for tones H and M, and 80% for tone L (because the L target tends to be reached later). This value does not encapsulate all the information about the time course of F_0, e.g. the amount of F_0 movement, known to affect the perception of pitch [29]; but it allows for cross-speaker comparison. Table 1 shows the results in semitones, taking the value for M tone as a reference: for instance, the value 3.6 for speaker M5's L tone means that it is 3.6 semitones lower than his M tone (for the reading condition at issue). The 'total' column is the sum of the previous two, i.e. the distance between H and L tone.

The tonal space as viewed through Table 1 is relatively symmetrical: the average distance between H and M is almost the same as between M and L, with clear cross-speaker differences.

Table 1: Average distance from M tone, in semitones.

speaker	reading	H tone	L tone	total
F2	Careful	3.0	3.2	6.2
	Impatient	4.0	3.8	7.8
M5	Careful	4.8	3.6	8.4
	Impatient	6.4	4.6	11
M7	Careful	5.0	6.2	11.2
	Impatient	4.7	6.8	11.5
M9	Careful	6.8	4.3	11.1
	Impatient	5.2	5.0	10.2
average		**5.0**	**4.7**	**9.7**
std. dev.		1.22	1.26	1.96

3.1.2. Glottal open quotient (O_q)

Glottal open quotient (O_q) values associated with the H, M and L tones follow speaker-specific patterns. For M5, O_q is inversely correlated with tone height: the respective average O_q values for H, M and L are 57%, 60% and 64%. For F2 and M7, the correlation is positive (F2: {57, 53, 49}; M7: {58, 56, 50}). No difference was detected in M9's data.

3.2. Comparison across reading conditions

3.1.1. Fundamental frequency

Values are compared below: e.g. the value under 'H tone' indicates the difference in semitones between realizations of the H tone under the 'CAREFUL' and 'IMPATIENT' reading conditions (in semitones). The values set in normal (non-bold) font are based on one value, as explained above (at 75% of syllable duration for H and M, and 80% for L); the values in bold are based on average F_0 over the entire vowel. The sign '=' is used where differences are less than 1/3rd of a semitone.

Table 2: Difference between the 'Careful' and 'Impatient' reading conditions, in semitones.

speaker	H tone	M tone	L tone	mean
F2	+2.3	+1.3	+0.7	+1.4
	+2	**+1.5**	**+1/2**	**+1 1/3**
M5	+1.3	=	+1.2	+0.8
	+1	**=**	**=**	**+1/3**
M7	+2	+2.5	+1.9	+2.1
	+2	**+2**	**+2**	**+2**
M9	+2.4	+4	+3.3	+3.2
	+2	**+4**	**+2 2/3**	**+2.9**
mean (std. dev.)				+1.75 (1)

Values averaged over the entire vowel tend to be slightly less sensitive, bringing out less difference between the two reading conditions. The general mean and standard deviation values in the last lines of Table 2 take into account both measurements (selected value, and F_0 averaged over the syllable).

3.1.2. Glottal open quotient (O_q)

Glottal open quotient (O_q) values associated with the two reading conditions follow speaker-specific patterns. O_q is significantly higher under 'IMPATIENT' reading for speaker M5, for all three tones; it is significantly lower for speaker M7, a tendency which t-tests show to be significant for tones H and L. For F2, differences are not statistically significant. For M9, the only significant difference is a lower O_q for the L tone under 'IMPATIENT' reading.

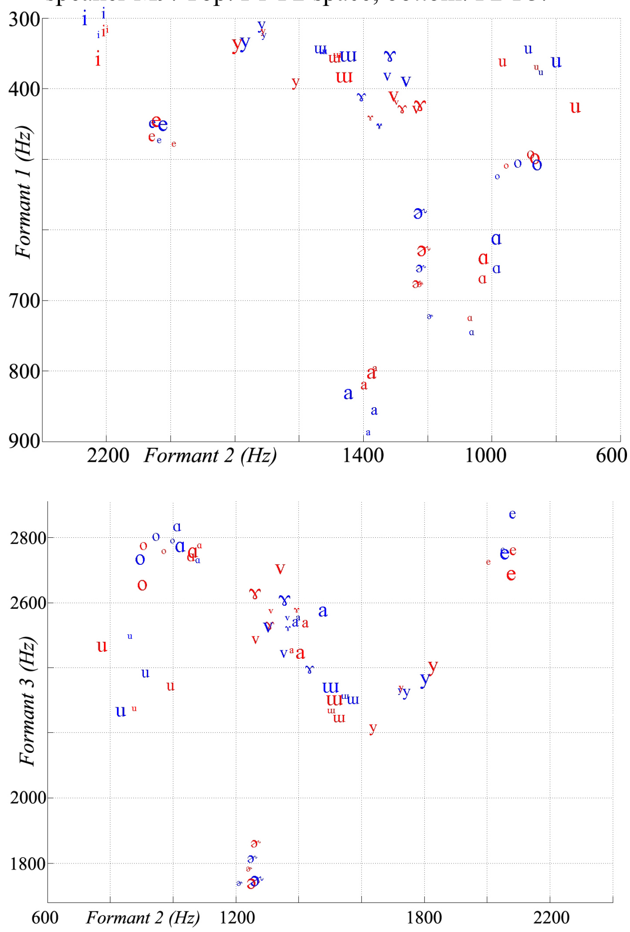
3.1.3. Acoustic intensity (mean RMS amplitude)

Measurements of acoustic intensity for this experiment are not precise, as the all-important parameters of distance to microphone and head orientation were not controlled for: the microphone was placed on a table, not head-mounted. It nonetheless appears relevant to report that, averaging across tones and speakers, intensity was 2.5 dB higher under 'IMPATIENT' reading.

3.3. Formant frequency measurements

Figure 2 shows formant frequency measurements averaged over the syllable for the various vowels carrying the three tones (from largest to smallest font size: tones H, M and L), under both reading conditions: 'Careful' in blue (appearing darker on a grayscale printout), and 'IMPATIENT' in red (lighter in grayscale). Due to the small number of items, no statistical processing was attempted. The data shown are from speaker M9; data from all speakers are available from [19:475-479]. No systematic influence of tone or reading condition on vowel articulation was found.

Figure 2: estimation of F1, F2 and F3 frequencies, speaker M9. Top: F1-F2 space; bottom: F2-F3.



4. DISCUSSION AND CONCLUSION

The average distance between lexical tones is on the order of five semitones between L and M, and between M and H. The Naxi tonal space thus appears relatively symmetrical.

All tones are realized with higher F_0 under 'IMPATIENT' reading, confirming that lexical tone is by no means incompatible with intonational malleability [13]. The average distance between 'CAREFUL' and 'IMPATIENT' realizations is 1.75 semitones, a figure that remains relatively small in comparison with the difference across lexical tones.

The Mid and Low tones are slightly longer, in keeping with a cross-linguistic tendency [9]. The hypothesis (based on auditory impression) that the Mid tone was the longest is disproved by the production data.

These findings, however modest, offer a basis for cross-linguistic comparison; availability of the full set of original data allows for verification and for further experiments, e.g. perceptual.

5. ACKNOWLEDGMENTS

Many thanks to the Naxi language consultants and friends. Support from Agence Nationale de la Recherche (HimalCo project, ANR-12-CORP-0006, and LabEx "Empirical Foundations of Linguistics", ANR-10-LABX-0083) is gratefully acknowledged.

6. REFERENCES

[1] Abberton, E., Fourcin, A. J. 1984. Electroglottography. In: Code, C. (ed.), *Experimental clinical phonetics*. Worcester: Billing & Sons, 62–78.

[2] Brunelle, M., Nguyễn Khắc Hùng, Nguyễn Duy Dương. 2010. A laryngographic and laryngoscopic study of Northern Vietnamese tones. *Phonetica 67*(3), 147–169.

[3] Connell, B. 2003. Pitch realization and the four tones of Mambila. In: S. Kaji (ed.), *Proceedings of "Cross-linguistic studies of tonal phenomena"*. Tokyo: Tokyo University of Foreign Studies, Institute for the Study of Languages and Cultures of Asia and Africa, 181–198.

[4] Coupe, A. *A phonetic and phonological description of Ao: a Tibeto-Burman language of Nagaland, north-east India*, Pacific Linguistics. Canberra: Australian National University, 2003.

[5] Degottex, G. 2012. COVAREP: A Cooperative Voice Analysis Repository for Speech Technologies. https://github.com/covarep/covarep

[6] DiCanio, C. 2012. Coarticulation between tone and glottal consonants in Itunyoso Trique. *Journal of Phonetics 40*, 162–176.

[7] Fabre, P. 1957. Un procédé électrique percutané d'inscription de l'accolement glottique au cours de la phonation: glottographie de haute fréquence. *Bulletin de l'Académie Nationale de Médecine*, *141*, 66–69.

[8] Fang Guoyu, He Zhiwu. 1995. *A dictionary of Naxi pictographic characters.* (In Chinese: 《纳西象形文字谱》.) Kunming: Yunnan Renmin Chubanshe.

[9] Faytak, M., Yu, A. C. L. 2011. A typological study of the interaction between level tones and duration. In: *Proceedings of the ICPhS XVII*, 659–663. Hong Kong.

[10] Goldsmith, J. 1981. English as a tone language. In: D. L. Goyvaerts (ed.), *Phonology in the 1980s*. Ghent: Story-Scientia, 287–308.

[11] He Jiren. 2011. *A dictionary of colloquial Naxi*. (In Chinese:《纳西语常用词汇》.) Kunming: Yunnan Minzu Chubanshe.

[12] Henrich, N., d'Alessandro, C., Castellengo, M., Doval, B. 2004. On the use of the derivative of electroglottographic signals for characterization of nonpathological phonation. *Journal of the Acoustical Society of America 115*(3), 1321–1332.

[13] Kochanski, G. P., Shih Chilin. 2003. Prosody modelling with soft templates. *Speech Communication 39*(3-4), 311–352.

[14] Kuang, J. 2013. The tonal space of contrastive five level tones. *Phonetica 70*(1-2), 1–23.

[15] Laniran, Y. O., Clements, N. 2003. Downstep and high raising: interacting factors in Yorùbá tone production. *Journal of Phonetics 31*, 203–250.

[16] Lindblom, B. 1990. Explaining phonetic variation: a sketch of the H&H theory. In: W. J. Hardcastle, A. Marchal (eds), *Speech production and speech modelling*, 403–439. Dordrecht: Kluwer.

[17] Michailovsky, B., Mazaudon, M., Michaud, A., Guillaume, S., François, A., Adamou, E. 2014. Documenting and researching endangered languages: the Pangloss Collection. *Language Documentation and Conservation 8*, 119–135.

[18] Michailovsky, B., Michaud, A. 2006. Syllabic inventory of a Western Naxi dialect, and correspondence with Joseph F. Rock's transcriptions. *Cahiers de Linguistique - Asie Orientale 35*(1), 3–21.

[19] Michaud, A. 2005. *Prosodie de langues à tons (naxi et vietnamien), prosodie de l'anglais : éclairages croisés*. Ph. D., Université Paris 3-Sorbonne Nouvelle.

[20] Michaud, A., He Xueguang. 2007. Reassociated tones and coalescent syllables in Naxi (Tibeto-Burman. *Journal of the International Phonetic Association 37*(3), 237–255.

[21] Nakagawa, H. 2009. A preliminary phonetic investigation of the lexical tones of G|ui. *Corpus-Based Linguistics and Language Education 3*, 45–52.

[22] Nolan, F. 2003. Intonational equivalence: an experimental evaluation of pitch scales. In: *Proceedings of ICPhS XV*, 771–774. Barcelona.

[23] Pinson, T. M. 2012. *A Naxi-Chinese-English Dictionary.* Kunming: Yunnan Minorities Publishing House.

[24] Remijsen, B. 2014. Investigating tone in languages with a quantity contrast. *Language Documentation and Conservation 8*, 672–689.

[25] Remijsen, B., Ladd, R. 2008. The tone system of the Luanyjang dialect of Dinka. *Journal of African Languages and Linguistics 29*(2), 173–213.

[26] Rock, J. 1963. *A Na-Khi - English encyclopedic dictionary*. Roma: Instituto Italiano per il Medio ed Estremo Oriente.

[27] Rose, P. 1990. Acoustics and phonology of complex tone sandhi: An analysis of disyllabic lexical tone sandhi in the Zhenhai variety of Wu Chinese. *Phonetica 47*, 1–35.

[28] Teo, A. 2014. *A phonological and phonetic description of Sumi, a Tibeto-Burman language of Nagaland*. Asia-Pacific Linguistics Open Access Monographs. Canberra : Asia-Pacific Linguistics.

[29] Vaissière, J. 2004. The perception of intonation. In: D. B. Pisoni, R. E. Remez (eds), *Handbook of Speech Perception*. Oxford: Blackwell, 236–263.

[30] Zerbian, S.; Barnard, E. 2010. Realisations of a single high tone in Northern Sotho. *Southern African Linguistics and Applied Language Studies 27* (4), 357–380.

[31] Zhu, G. 2014. The right to minority language instruction in schools: negociating competing claims. *Human Rights Quarterly 36*(4), 691–721.