Replicating in Naxi (Tibeto-Burman) an Experiment Designed for Yorùbá: An Approach To ‘Prominence-Sensitive Prosody’ vs. ‘Calculated Prosody’

Alexis Michaud

Laboratoire Phonétique et Phonologie (UMR 7018) CNRS/ Paris 3 Sorbonne Nouvelle alexis.michaud@univ-paris3.fr

Abstract

The starting-point of this study is the hypothesis (suggested by an overview of typologically varied languages) that it may be useful to characterise prosodic systems in terms of the degree to which they rely on the calculation of tone sequences. Each language could be placed at a certain position along a typological continuum between two types of prosodic organisation: (i) ‘calculated prosody’, in languages such as Yorùbá (in which tone serves complex morphophonological functions), whose prosodic structure hinges on the calculation of a tone sequence, by categorical processes such as the association of lexical tones and/or boundary tones, reassociation ‘tone floating’, and downstep; and (ii) ‘prominence-sensitive prosody’, typologically more common, found in languages such as Chinese, which have fewer elements of categorical tonal calculation, and in which intonation appears to reflect phrasing and informational structure in a largely noncategorical way. In an effort to test (and refine) this hypothesis, an experiment used for Yorùbá [8, 9] is adapted to Naxi, a Sino-Tibetan language which, like Yorùbá, has three lexical tones (High, Mid and Low), but which is hypothesised to be closer to the ‘prominence-sensitive prosody’ type, whereas Yorùbá would be closer to the ‘calculated prosody’ type. This pilot study on sentences in which all syllables bear the same tone does bring out differences between the two languages in terms of phenomena of phrasing and of prominence.

1. A typological proposal: ‘prominence-sensitive prosody’ vs. ‘calculated prosody’

1.1. Some definitions

As a preliminary to the proposed typological overview, some indications on the terms used may be useful. Prosody is here understood in its broadest sense: it includes lexically distinctive suprasegmentals, intonation, the expression of attitudes and emotions, and performance factors: rhythm and speech rate [15]. Lexically distinctive suprasegmentals may be, depending on the language, lexical stress, lexical tone, lexical pitch accent... Intonation is an abstract structure, not to be confused with the parameters whereby it manifests itself (in particular, it is a different notion from $F_0$, which is a physical parameter, and pitch, which is a percept). Intonation can be conveniently divided into two components: (i) intonational phrasing, sometimes called syntactic intonation because of its strong (though not bijective) link to syntax, (ii) pragmatic intonation, determined by the information structure of the utterance. The term intonational group is used here loosely as a cover term for the various levels of intonational phrasing.

1.2. An overview of typologically varied tone languages

Typologically, there appears to be considerably more diversity among prosodic systems than labels such as ‘tone languages’, ‘pitch accent languages’ and ‘stress languages’ suggest [7, page 1]. ‘Tone languages’ do not make up a unified typological pole. An overview suggests that some prosodic systems, more than others, go by a calculation of tone sequences. The prosody of Mambila (Niger-Congo, Bantoid) seems to be highly constrained by its four level lexical tones. Ngamambo (Bantoid) [1] also appears highly constrained, as a five-term paradigm is open after a High-tone syllable: High, Mid, Low, downstepped Low (L), or flat Low (L'). On the other hand, in Hausa, an Afro-Asiatic language which only has two tones, speakers apparently play to a larger extent on intonation to convey degrees of informational prominence (A. Riialand, p.c.). It thus seems that the degree of constraint on the margin of freedom left to the speaker depends on the number of level tones and of categorical processes of tone modification.

1.3. Formulation proposed

The proposal put forward here is that prosodic systems could be placed along a typological continuum between (i) prosodic systems that hinge on the calculation of a tone sequence (by categorical processes of association of lexical tones and/or boundary tones, reassociation ‘tone floating’, and downstep), and (ii) prosodic systems in which phonetic realisation is sensitive to the prominence of individual items within the utterance, i.e. where intonation reflects informational structure (with such typical phenomena as contrastive emphasis). ‘Prominence-sensitive prosody’ can be illustrated by languages such as Mandarin Chinese. Despite having four lexical tones, Mandarin allows local intonational modifications which carry a communicative load comparable to the one they have in a language such as English (as has been known since [3]): the phonetic realisation of tones depends on the degree of prominence of individual syllables, and on intonational phrasing. Rather than unambiguously categorical processes such as downstep and tonal reassociation, Mandarin Chinese shows many noncategorical modifications of lexical tones, and few categorical modifications. In the study of Mandarin tones, much attention has been directed to the sandhi process whereby a sequence of two tone-3 syllables yields a tone 2+tone 3 sequence; on the other hand, the many cases of “minor sandhi” that have been proposed for Mandarin (e.g. by Chao Yuen-ren) are not clearly categorical, so much so that “It is not clear whether it is desirable or even possible to segregate tonal coartication [i.e. a noncategorical modification] from tone sandhi proper [i.e. a categorical modification]” [4, page 25].
It seems that ‘calculated-prosody’ systems would lend themselves especially well to the automatic synthesis of \( F_0 \) curves, on the basis of tone sequences alone, whereas for ‘prominence-sensitive’ prosodic systems, other parameters have to be taken into account: for instance, a system of synthesis proposed for Mandarin assigns a coefficient of strength to each individual syllable [14].

Section 2 aims to provide some experimental evidence to help refine these tentative typological reflections.

2. A pilot study comparing two widely different prosodic systems (Naxi and Yɔrùbá) using a similar experimental setup

2.1. Overall presentation of the experiment

The pilot study reported below consists in adapting to Naxi, a Sino-Tibetan language, an experiment originally designed to investigate Yɔrùbá, a language of the Niger-Congo family [8, 9]. Both languages have High, Mid and Low lexical tones; unlike in Naxi, however, tone plays complex morphophonological functions in Yɔrùbá (see, e.g., the discussion by [2] of the subject marker as a floating High tone); and the Mid tone of Yɔrùbá can be considered as the absence of tone [6 and references therein], whereas the Mid tone of Naxi is a tone in its own right. In some respects, the prosodic system of Naxi (investigated first-hand [10, 11]) appears closer to that of neighbouring tone languages (such as Mandarin Chinese [3, 4, 14]) than to that of Yɔrùbá. In the terms proposed in section 1, Yɔrùbá is closer to the ‘calculated prosody’ type, whereas Naxi is closer to the ‘prominence-sensitive prosody’ type (though it possesses some elements of tonal calculation, namely the addition of a High tone for intensification, and some cases of ‘floating’ High tone [11]). Using a similar method for both languages appeared as an interesting means to gain a perspective on the differences between these two types of prosodic organisation.

2.2. Method

The data reported here consist in sentences in which all syllables bear the same lexical tone, as in [9]:

(1) kó sò sè nỳ day after tomorrow COND./TOP. 2nd sg. honorary
kò sò yò yè tʰú object particle say give (be) possible
“The day after tomorrow, [I’ll] be able to tell you.”

(2) tʰú tʰú nỳ̲ kʰ nú̲ s̄i̲ m̄̄̀ nd̄zǔ 2nd sg. today dog meat NEG. eat
“He’s not eating any dog meat today.”

(3) nỳ kʰ nú̲ kʰ kó tʰè bỳ m̄̄̀ s̄è lᵗ sg. inside ASP. ongoing crawl final particle
“I’m crawling into [it].” (The final particle was /s̄è/ or /s̄/ at the language consultant’s choice.)

Five native speakers of (Western) Naxi recorded these sentences (their Chinese equivalent was used as a prompt). (For ease of cross-reference with other publications based on data from the same language consultants, they are referred to using the same codes as in [10]: F2, M4, M5, M7, M9.)

In addition to its H, M and L tones, Naxi possesses a rising LH contour [11]; due to its rarity on lexical items, constructing an utterance such as (4a) is difficult; it appeared useful to record it nonetheless.

(4a) tʰú lỳ nù̲ m̄̄̀ p̄̄̀ ḡi̲j̄̄̄ h̄ɔ̀ 3rd sg. see TOP. 2nd sg. and 1st sg. very correct, adequate
“He thinks that you and I [getting together] is quite right./ He thinks that our getting together is a very good thing.”

Informants F2, M7 and M9 chose to record a variant (4b) that they considered more satisfactory.

(4b) tʰú nù̲ m̄̄̀ p̄̄̀ ḡi̲j̄̄̄ h̄ɔ̀ m̄̄̀ 3rd sg. 2nd sg. and 1st sg. very correct final part.
“He, you and I [getting together] is quite right, isn’t it.”

The recordings took place in a quiet room in Lijiang Prefecture, China. An electroglottographic signal was recorded simultaneously with the audio, in order to obtain a precise measurement of \( F_0 \) and indications on voice quality. The Chinese translation of the sentences was used as a prompt (sentences 1-4 were arranged in random order). The speakers were instructed to say the sentence twice, with a deep breath in-between repetitions. Due to some repetitions, a total of 54 utterances were finally retained (10 by F2, 17 by M4, 11 by M5, 8 by M7, 7 by M9). The beginning and endpoint of each syllable rhyme (i.e. excluding initial consonants) were annotated manually on the basis of inspection of the audio signal supplemented by auditory impression. \( F_0 \) and the open quotient (hereafter \( O_q \); it offers an indication on the degree of vocal fold abduction [5 and references therein]) were estimated using a method set out in [5, 10].

2.3. Results

Figures 1 and 2 (on the last page) show average curves of \( F_0 \) and \( O_q \) for tones H (sentence 1) and LH (sentence 4a) by speaker M5, superimposing the \( F_0 \) curves of all the syllables of each sentence. It was chosen not to average the data across speakers due to differences in overall \( F_0 \) register which might have biased the results. The curves are resampled at 20 equally spaced time points. Due to the small number of repetitions averaged (two on fig. 1, four on fig. 2), it did not appear useful to indicate standard deviation. These figures can be proposed as representative of the behaviour of the four speakers (cross-speakers differences will not be discussed in detail; graph 1 shows the cross-speaker differences in declination over the utterance).

On figure 1, the third syllable, which corresponds to an intonational juncture, is markedly longer than the seven others; this third syllable and the last syllable (the 8th) both have a more strongly decreasing \( F_0 \) than the others (the decrease is on the order of 3 musical tones), and, contrary to all other syllables, have a decreasing \( O_q \), indicating a tensing of the vocal folds, as opposed to the relaxing of the vocal folds at the end of the other six syllables. On figure 2, the longest syllable is, again, the one located at an intonational juncture (syllable 2). The curves in figure 2 do not show a strict layering as in figure 1, where the mean \( F_0 \) of successive syllables decreased from the first to the last: in figure 2, syllables 4 and 7 have a higher mean \( F_0 \) than the syllable that precedes them. The extent of the \( F_0 \) rise also differs across syllables.

These piecemeal observations, by themselves, do not have demonstrative value; they nonetheless illustrate the general fact that, in Naxi, grammatical words are typically weaker (which does not appear to be equally true in Yɔrùbá), e.g. /m̄̄̀ n̄̄́/, ‘and’; the fourth syllable, is realised with flat rather than rising \( F_0 \), in the continuity of the preceding syllable (its starting-point by and large corresponds with the \( F_0 \) endpoint of the preceding syllable). The adverb ‘very’, /ḡi̲j̄̄̄/ (6th
syllable), likewise a grammatical word, is also weak—weaker than the adjective /hò́/ (syll. 7), which is realised slightly higher, and with a more clearly rising $F_0$.

These facts have a bearing on the representation of $F_0$ curves. In the study of Yorùbá that inspired the present experiment, $F_0$ curves over individual syllables are simplified to two data points, “a point towards the beginning of the vowel portion of the syllable and another towards the end” [8, page 60]. This is an ‘impressionistic’ choice, in that it does not rest on evidence from perception tests.) Such an approach does not seem to commend itself for Naxi, because of the effect of intonational junctures, manifested by lengthening and a downward tilt in $F_0$. Also, the data points retained for Yorùbá are plotted at an equal distance, which amounts to normalising for duration, as on figure 3. This stylisation of the $F_0$ curves may be adequate for Yorùbá: two data points per syllable can capture an upward or downward movement, and the literature on this language suggests that this information is sufficient to represent prosodic events; on the other hand, such a stylisation does not appear adequate for the Naxi data.

As shown by figures 1 and 2, considerable differences in length emerge (beyond the thresholds indicated by [13]) at the junctures indicated below by the symbols [1] (optional) and [l] (realised by all five speakers): (1) /kò só sè l ńg’ kò l pè pè t’ə́/ (2) /t’ə́/ /t’ə́/ (3) /t’ə́/ (4) /t’ə́/ t’ə́/ /t’ə́/ /t’ə́/ /t’ə́/ 

In view of these observations, normalising the data for duration does not appear adequate for Naxi. Moreover, it does not appear enlightening to approximate declination over the utterance by comparing the average $F_0$ of the first and last syllables (as was done by [8, 9] for Yorùbá), because this procedure overlooks the division of the utterance into intonational groups. The most apparent downward trend in Naxi appears to be due to group- and utterance-final lowering, rather than to (continuous) declination. With a view to comparison with the Yorùbá data, declination was nonetheless estimated as a ratio of the average $F_0$ of the first and last syllable. The results are shown on graph 1: values on the order of -1.6 musical tone over the utterance for tone H and tone M, -1.75 musical tone for tone L. (The instruction that the speakers should take a deep breath between sentences may have contributed to this relatively steep declination.) $F_0$ decrease in the course of the last syllable is more frequent than downward shift in $F_0$ register. Graph 1 brings out the variability observed across speakers: in Naxi, even under controlled, ‘laboratory’ conditions (uttering sentences that have been prepared beforehand), local modifications are of such amplitude that the declination line cannot be recovered from the observed $F_0$ curve in a straightforward way.

3. General discussion and conclusion

Though very slender indeed (so much so that no statistical treatment was attempted), and open to some criticisms, the experimental results set out in section 2 reflect the considerable structural difference between the prosody of Naxi and of Yorùbá. The following differences emerge:

(i) Final lowering is observed in Yorùbá [8, page 17] as well as in Naxi; in Naxi, however, it is usually accompanied by lengthening.

(ii) The marking of intonational junctures (by lengthening and a final dip in $F_0$) is very salient in Naxi prosody.

(iii) Grammatical words, as opposed to lexical words, are typically realised (all other things being equal) with slightly lower $F_0$ in Naxi, whereas it seems to be a general observation that grammatical words are NOT realised as ‘weaker’ or with lower $F_0$ than lexical words in Yorùbá (and in other languages of the Niger-Congo family; A. Rialland, p.c.).

(iv) Monosyllables elicited in a carrier phrase in Naxi are much longer (average over 600 syllables: 260 ms [10]) than would be expected for Yorùbá; I consider this lengthening as evidence of the prominence-sensitive realisation of Naxi words (a syllable placed inside a carrier sentence is in focal/prominent position), as opposed to the more calculated realisation of tone in Yorùbá, whereby $F_0$ realisations do not usually depart considerably from what can be predicted on the basis of the tone sequence of the utterance.

Unlike in Yorùbá, where it appears possible, in read speech, to elicit sequences of like tones that come out as regular, gradually lowered $F_0$ curves, it appears that in Naxi, under a similar experimental setup, at least two factors disrupt (to a certain extent) the regularity of the tone sequences:

(i) junctures between intonational groups are clearly marked by lengthening and $F_0$ decrease; and (ii) the informational importance (in the broad sense) of individual words influences their prosodic realisation. (These factors are even more salient in continuous/’spontaneous’ speech.)

Concerning perspectives for future research, the typological distinction between ‘calculated prosody’ and ‘prominence-sensitive prosody’ could arguably be extended to languages that do not possess lexical tones: a language such as Wolof, which signals information structure by morphosyntactic means (a verb form may be ‘subject-emphatic’, ‘object-emphatic’ or ‘verb-emphatic’), appears to have a prosodic system very unlike ‘prominence-sensitive’ systems, despite not having any lexical tones [12].
4. Hyperlink
Some recordings (audio and EGG) are available at: http://cd268.univ-paris3.fr/tpp/pages/EQUIPE/michaud/PROSODY/

5. Acknowledgments
Many thanks to Larry Hyman and Annie Rialland for useful discussions, and to Jacqueline Vaissière, Cédric Patin and two anonymous reviewers for suggesting improvements. Any remaining errors are naturally my own responsibility.

6. References