Tonal contrasts and initial consonants: a case study of Tamang, a 'missing link' in tonogenesis

Martine Mazaudon, Alexis Michaud

Tonal contrasts and initial consonants: a case study of Tamang, a 'missing link' in tonogenesis

Martine Mazaudon and Alexis Michaud LACITO, CNRS mazaudon@vjf.cnrs.fr

Preprint version of article published in *Phonetica*, vol. 65, issue 4, 2008.

The figures are placed on separate pages, after the text page where their caption is provided.

Abstract:

Tamang (Bodic division of Tibeto-Burman) is spoken at the edge of the East Asian "tone-prone" zone, next to the almost tone-free Indian linguistic area, and is, chronologically, at the late end of the tone multiplication wave which has swept through East Asia in the course of the last two millenia. It can be regarded as a 'missing link' in tonogenesis: following the loss of voicing contrasts on syllable-initial consonants, Tamang has four tonal categories instead of its earlier two-tone system; the present state of the prosodic system is typologically transitional, in that these four tonal categories are realised by several cues which include fundamental frequency (F_0), phonation type, and allophonic variation in the realisation of consonants. Acoustic and electroglottographic recordings of 131 words in two carrier sentences by five speakers were conducted (total number of target syllables analysed: 1651). They allow for a description in terms of F_0, glottal open quotient, duration, and realisation of consonants. The results confirm the diversity of cues to the four tonal categories, and show evidence of laxness on tones 3 and 4, i.e. on the two tones which originate diachronically in voiced initials. The discussion hinges on the phonological definition of tone.

1. Introduction: the phonetic complexity of Tamang tones

The Tamang language, spoken in Nepal, is a member of the TGTM group, an acronym for the names of the four main languages of the group: Tamang-Gurung-Thakali-Manangke. TGTM is a branch of the Bodish section of the Bodic division in Shafer's classification of Sino-Tibetan languages [Shafer 1955]. The tonal systems of TGTM languages have several interesting characteristics. The first, which will not be studied here, is that the domain of tone is the phonological word, composed of a lexical morpheme, which is tonally specified, and a sequence of tonally neutral suffixes. A single tonal melody spreads over the whole word. A second characteristic is the use of multiple cues to tones.

Tamang has four tones – numbered 1 to 4, roughly from highest to lowest – which are phonetically complex. To the unaided ear of the linguist, these four tonal categories are distinguished by a combination of cues involving pitch level, pitch contour, and phonation type. Tones 3 and 4 have been described as characterised by a general laxness, evidenced by breathy/whispery voice and by occasional voicing of initial stops – stops are phonologically unspecified for voicing and are generally realised as voiceless in initial position [Mazaudon 1973:66]. Preliminary experimental results have already been obtained concerning F_0 [Mazaudon 2005] and phonation type [Michaud and Mazaudon 2006]; the present study aims to provide a more comprehensive picture of Tamang tone, building on the analysis of a broader range of parameters. Two major hypotheses are (i) that tones 3 and 4 are realised with nonmodal phonation, more specifically with whispery voice, and (ii) that non-contrastive partial voicing is occasionally found on the initial stops of syllables with tones 3 and 4, never with tones 1 and 2.

In addition to its synchronic interest, the experimental study of Tamang tones can contribute to a better understanding of tonogenetic processes. The present state of the Tamang prosodic system is typologically transitional; it can shed light on the stages that follow the loss of voicing contrasts on syllable-initial consonants – a process of great historical importance that took place in numerous Far Eastern languages [Maspéro 1912; Haudricourt 1954, 1961]. Haudricourt recognised the present state of Tamang as the missing link between the stage where a language has two tones and three series of initial consonants (voiced, voiceless, aspirates) and the later stage where there are four tones and only two series of initials [Haudricourt 1975].

2. Method

2.1. Items recorded

About one half of the lexical morphemes of Tamang are monosyllabic, the other half are polysyllabic. A word is composed of one lexical morpheme followed by one to three grammatical morphemes which are toneless monosyllabic affixes. The tone of the whole word is determined by that of the lexeme. All lexical morphemes, whether monosyllabic or polysyllabic, carry one and only one of the four tones of the system. The tonal characteristics spread over all the syllables, though they are most salient over the first syllable. The corpus used in the present study is restricted to monosyllabic lexemes. It is made up of 127 open-syllable monosyllabic roots without initial clusters, with initial /pʰ/, /tʰ/, /kʰ/, /tsʰ/, /p/, /t/, /k/, /ts/ or /s/ and vowel /aː/, /a/, /iː/, /i/, /uː/, /u/, /oː/, /o/, /eː/ or /e/. It includes morphemes of diverse grammatical nature and frequency. Sixteen other morphemes of the desired phonemic composition were discarded because they were unfamiliar to the speakers or unsuitable for use within the carrier sentences chosen.

The target morphemes are divided into two sets: 41 nouns (including 2 tone-carrying particles, /^{4}tsa/ 'as for' and /^{4}tse/ 'only', and the deictic /^{2}tsu/) on the one hand, and 86 verbs, including stative verbs such as /^{4}tseː/ '[to be] pretty', on the other. Tone is indicated as a superscript figure before the first syllable of the phonological word that carries it: e.g. /^{4}tseː/ carries tone 4.

2.2. Carrier sentences

The nouns were placed inside the two following carrier sentences:

(1)	^{2}tsu-ri	___	-ka	-tsim.	
	here	target noun	FOC	be	(FOC = focus particle)

e.g. ^{2}tsu-ri ^{4}tsi-ka-cim "Hey, there is beer here" (/^{4}tsi/ means 'beer').

(2)	^{2}tsu-ri	___ .
	here	target noun

e.g. ^{2}tsu-ri ^{2}ts^{h}i "Here [is] grease."

The verbs were placed inside another carrier sentence:

(3)	^{2}tsu-ri	___	-pa.
	here	target verb	infinitive suffix

e.g. ^{2}tsu-ri ^{3}suː-pa "Here, [he is] planting" (/^{3}suː-pa/ means 'to plant').

2.3. Speakers

Five male native speakers of Tamang (in their 30s or 40s) from the village of Risiangku in Central-Eastern Nepal participated in the recordings. The Nepali equivalent of the target morpheme was provided orally as a prompt; like most speakers of Tamang, the subjects speak Nepali as a second language. They were instructed to repeat the item within the carrier sentence twice, e.g. /^{2}tsu-ri ^{3}suː-pa ‖ ^{2}tsu-ri ^{3}suː-pa/. It was originally intended to go through the entire recording procedure twice with each speaker on different days, but for practical reasons this was not possible for all five speakers (some details are provided in the Results section). The audio and the electroglottographic signal were recorded simultaneously (sampling rate: 44,100 Hz). Concerning electroglottography, see section 2.5. The equipment used was a two-channel electroglottograph [Rothenberg 1992]. The audio and electroglottographic signals were time-aligned after recording, to compensate for the time lag of about 1.5 ms due to the distance between the microphone and the voice source.

2.4. Identification test

The distinctiveness of the four tonal categories has been well established over many years of research with several generations of speakers of the Risiangku dialect. It was nonetheless considered useful to perform an identification test to evaluate the distinctiveness of the tokens that had been recorded in the experiment. Due to the constraints of fieldwork, this identification test could be performed only by one speaker (M2) and only over the data that he had recorded. Each item was played within its carrier sentence through loudspeakers, and the subject was instructed to provide a translation of the target word in Nepali.

2.5. Electroglottographic analyses

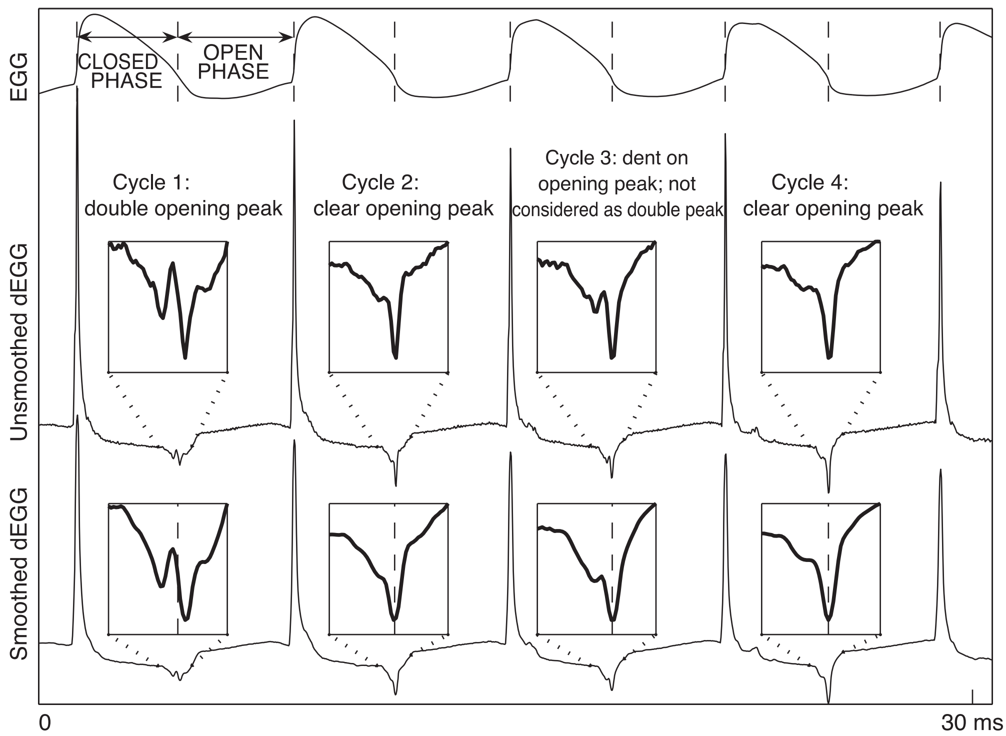
The initial hypothesis concerning the phonation type of Tamang tones is that tones 3 and 4 are realised with more breathy/whispery voice than tones 1 and 2. The present study relies on electroglottography to investigate phonation type. Electroglottography is a technique for monitoring vocal fold contact area by means of electrodes placed on either side of the Adam's apple (the thyroid cartilage). On the derivative of the EGG signal (hereafter dEGG), obtained by calculating the slope of the EGG curve at each point, alternating positive and negative peaks are typically observed, as seen on figures 1 and 2. The interpretation of these peaks was established in light of the comparison of EGG and dEGG signals with simultaneously recorded images of the glottis [Childers et al. 1983; Baer et al. 1983; Anastaplo and Karnell 1988; Karnell 1989; Hess and Ludwigs 2000]: the positive peak corresponds to the instant when the glottis closes over its full length; the negative peak usually marks the onset of glottal opening along the superior surface of the vocal folds. It is therefore assumed here that the positive and negative peaks on the dEGG signal are indicators of glottis closing and opening, respectively. They allow for the measurement of the open quotient (hereafter O_q), defined as the length of the open phase – from one opening instant to the following closing instant – divided by the length of the glottal cycle, itself defined as the time between two closing instants: see top of figure 1. The open quotient allows for the monitoring of vocal fold adduction: a low open quotient is an indicator of a tight/pressed voice. The open quotient is known to correlate positively with airflow [Rothenberg and Mahshie 1988], though it must be cautioned that the open quotient only provides an indirect cue to airflow, and that the relationship between open quotient and airflow is not linear.

The MATLAB routines used in this study were devised specifically for the investigation of linguistic uses of phonation types.[1] The shape of the negative portion of the dEGG signal within each cycle is appraised individually: a clear opening peak will be included in the results even though it is flanked by cycles that do not have a clear opening peak. Measurements are semi-automatic. The software displays the open quotient values obtained by two methods: (i) detection of highest peak, and (ii) barycentre of the peaks whose amplitude reaches 50% of the highest peak. Both methods are run twice: once without smoothing of the dEGG signal, once with smoothing. The smoothing step is selected by the user: from 0.07 ms to 0.16 ms, depending on the clarity of the original EGG signal. Figure 1 shows the dEGG signal before and after smoothing with a step of 0.113 ms. The user verifies the results in view of the shape of the EGG and dEGG signals. It is well established that unclear or double peaks are not artefacts, but are related to the way in which the glottis closes [see discussion in Henrich et al. 2004:1324-1327]. Exclusion of all the cases where there is no unique and well-defined opening peak may appear as the safest option. In the present study, the barycentre method was nonetheless used in a few instances; figure 1 shows an example from speaker M1, the speaker for whom the barycentre method was used most often (9.1% of cases).

Figure 1. Illustration of the barycentre method for estimation of glottal opening. Speaker M1, vowel /a/, first syllable of /^{3}ka-pa/ "to please, to satisfy taste".

The portions of dEGG signal corresponding to opening peaks are enlarged; the vertical dashed line indicates the estimated position of the glottis-opening-instant. The first cycle has a double opening peak; the barycentre calculated from these two peaks, being weighted as a function of peak amplitude, is closer to the higher of the two. The second and fourth cycles have simple opening peaks. The dent in the opening peak of the third cycle is overlooked because its amplitude is less than half that of the main peak. The barycentre method and the detection of the local minimum in the dEGG signal yield the same result for cycles 2, 3 and 4, whereas for cycle 1 the barycentre method gives a slightly higher estimation: 53.6%, as against 52.8%. The proportion of excluded values for each speaker, and of values calculated by means of the barycentre method, is provided in the Results section (table 1).

[1] The software is available for download from: http://voiceresearch.free.fr/egg/


EGG
CLOSED PHASE
OPEN PHASE
Unsmoothed dEGG
Smoothed dEGG
Cycle 1: double opening peak
Cycle 2: clear opening peak
Cycle 3: dent on opening peak; not considered as double peak
Cycle 4: clear opening peak
0
30 ms

2.6. Analysis of consonants

Several pieces of information concerning the realisation of the consonants were obtained by inspecting the audio and electroglottographic signals item by item, as shown in figure 2.

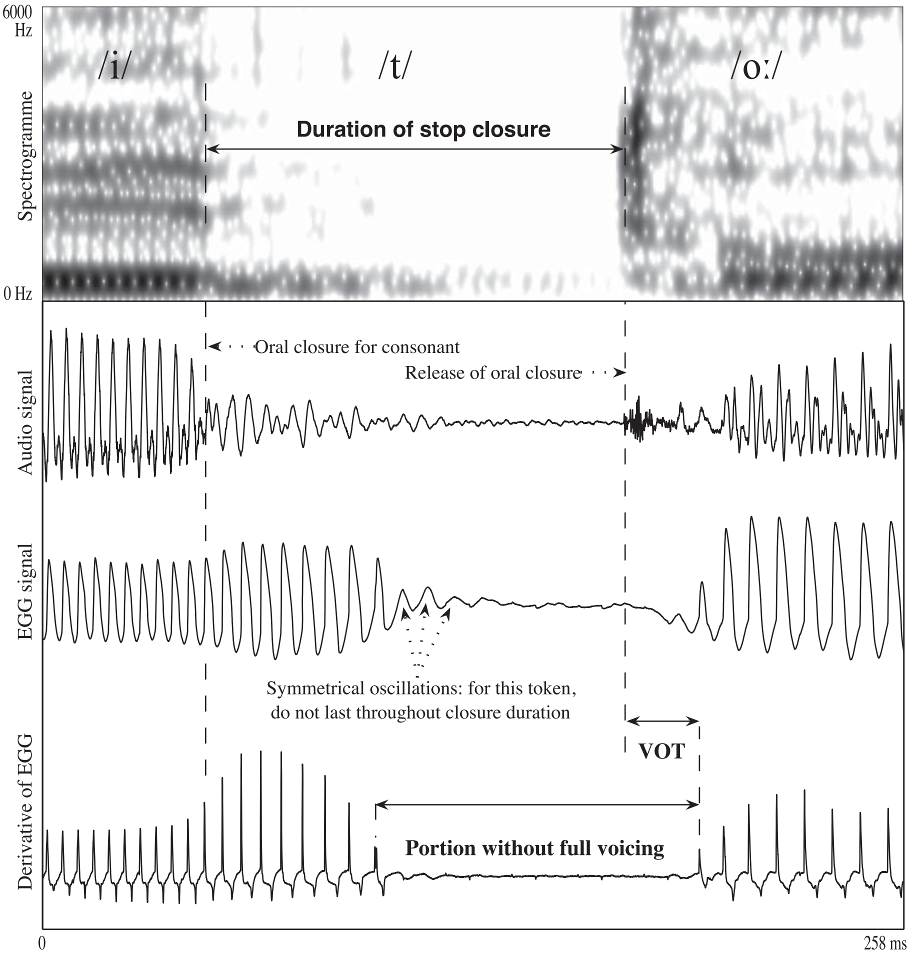
Figure 2. Illustration of the measurements conducted over consonants on the basis of spectrograms and dEGG signals. Speaker M3, first syllable of /^{3}to:-pa/ "to reach, to attain", in carrier sentence /^{2}tsu-ri ^{3}to:-pa/.

(i) Information on the state of the glottis during the initial consonant of the target morpheme. Recall that this consonant is phonemically unspecified for voice in all cases – unaspirated stops are voiced intervocalically, and voiceless initially and in final position – but is hypothesised to undergo tone-conditioned allophonic variation. The three categories that were retained in the description of voicing are set out in the Results section.

(ii) VOT: Voice Onset Time, "the temporal relation between the onset of glottal pulsing and the release of the initial stop consonant" [Abramson 1977:296; see also Lisker and Abramson 1964]. Since the target items of the present study always follow a vowel, VOT as measured here coincides with voicing lag: it is the duration between the burst of the obstruent, observed on the acoustic signal and on a spectrogram, and the next positive peak on the derivative of the electroglottographic signal, indicative of glottis closure. The VOT cannot be measured when the consonant does not have a clear burst, nor when it is voiced throughout: in these cases, the item was excluded from the calculation of VOT.

(iii) Duration of portion without full voicing between the end of the previous vowel (vowel /i/ of the carrier sentence, /^{2}tsu-ri …/) and the onset of voicing for the target morpheme. This measurement was defined solely in terms of the positive peaks on the derivative of the electroglottographic signal: from the last positive peak at the end of the preceding syllable (/-ri/) to the first positive peak of the target syllable. When the consonant was voiced throughout, this measurement was set at zero. This measurement was conducted as part of the attempt to verify experimentally the auditory impression that the same stop initials are clearly unvoiced in syllables bearing tones 1 and 2 and sometimes voiced in syllables bearing tones 3 and 4.

(iv) Duration of oral closure for the initial consonant, measured from a spectrogram: from closure, where the formants (F2 and above) of the vowel of the preceding syllable cease to



6000 Hz
0 Hz
Spectrogramme
/i/
/t/
/oː/
Duration of stop closure
Audio signal
Oral closure for consonant
Release of oral closure
EGG signal
Symmetrical oscillations: for this token, do not last throughout closure duration
VOT
Derivative of EGG
Portion without full voicing
0
258 ms

be visible, to release. In cases where these events could not be reliably detected, closure duration was not measured.

A limitation of these measurements is that the speaking rate was not controlled during the recording sessions. Rather than normalise for duration – recalculating each value relative to the duration of the sentence – it appeared preferable to exclude the cases where the speaker had made a pause before the target word, and consider the speaking rate of the remaining data as sufficiently homogeneous for statistical analyses.

3. RESULTS

3.1. Result of the identification test

The test was conducted three years after the recordings, which means that the speaker had to identify the words from his complete lexicon, and not only among the test words. The answers pertaining to the 127 words fall into one of the following four sets:

(i) In most cases, the word was identified as that elicited, or with an error that did not involve tone. These 'tonally correct' mistaken identifications (14 items) consist in: giving a homophonous verb instead of a noun (implying mishearing a consonant in the frame: in the carrier sentence, nouns are followed by /ka/ and verbs by /pa/); a different initial stop consonant: /p/ heard as /k/, /k/ heard as /p/, /t/ heard as /p/ or /ʈ/; a confusion between long and short vowel; and one error on vowel timbre.

(ii) Lack of identification occurred with rare words which the speaker understands but which are not part of his active vocabulary, or which are unlikely to appear in the carrier sentences chosen in this study. In these cases, after hesitating, the speaker supplied a word that had some phonetic resemblance, and in 8 items the tone was different.

(iii) Two common items were identified as words having another tone: /^{4}koː.pa/ 'to sing' heard as /^{2}koː.pa/ 'to contaminate', and /^{4}ciː.pa/ 'to remember' heard as /^{3}ciː.pa/ 'to pinch'.

(iv) One item, /^{3}coː/ 'point', was first identified as /^{1}c^{h}oː/ 'rope' then corrected to 'point'.

3.2. Measurements of fundamental frequency, open quotient, and duration

Table 1 provides an overview of the data obtained by quantitative analysis of the electroglottographic signal.

Table 1. Overview of the data obtained by quantitative analysis of the electroglottographic signal.

Speaker	M1	M2	M3	M4	M5	total
Total number of target syllables analysed	415	306	367	200	363	1,651
Total number of syllable rhymes analysed (including carrier sentences)	1,508	1,054	1,314	720	1,325	5,921
Total number of glottal cycles	31,198	21,137	29,229	11,830	24,145	117,539
Proportion of cycles for which the barycentre method was used	9.1%	0.11%	0.68%	0.94%	2.9%	3.4%
Proportion of cycles for which O_q could not be estimated	3.47%	26.6%	3.81%	32.8%	6.54%	14.6%
Mean F_0 and median of F_0	167 (166)	131 (130)	145 (143)	124 (124)	175 (174)	
Mean O_q and median of O_q	56.8 (56.0)	55.6 (55.0)	55.2 (55.31)	59.71 (59.5)	68.5 (68.6)	

The unequal number of tokens recorded by each speaker is due to the fact that some recording sessions had to be shortened or cancelled for practical reasons, so that the entire recording procedure could not be completed twice with each speaker as originally intended. The proportion of cycles for which O_q could not be estimated varies from 3.5% to 33% across speakers, confirming the existence of considerable cross-speaker differences in the proportion of well-defined opening peaks. (Henrich et al. 2004:1325 provide figures on this phenomenon for 18 subjects performing singing and speaking tasks.)

The analysis programme was run over the rhymes of all the syllables in the corpus, including those of the carrier sentences. The average value of O_q calculated over all the syllable rhymes produced by a given speaker is used below as a rule-of-thumb reference for inter-speaker normalisation of the results. Mean O_q is within 0.8% of median O_q for all speakers, and mean F_0 within 2 Hz of median F_0.

3.2.1 The case for normalisation

If measured in absolute value, extreme variability across items is observed for both F_0 and O_q, with some overlap across tones for F_0, and more overlap for O_q. This result is in keeping with the hypothesis that Tamang tones are not realised by F_0 alone, but by a bundle of characteristics. This variability is not an effect of phonemic composition – vowel quantity, vowel height, stop or fricative character of the initial consonant –: data plots for phonemically homogeneous data subsets still show a comparable degree of variability. The averaged F_0 curves are also close in some cases: for instance, the F_0 curves of tone 1 and tone 2 as realised by speaker M4 on syllables with an unaspirated initial are almost indistinguishable. However, taking the F_0 curve of the frame into account, it appears that the closeness of tones 1 and 2 in M4's data is an artefact of the measurement in absolute values: M4 actually anticipates the target syllable, realising both syllables of the frame (/^{2}tsu-ri/) higher before tone 2. Once the data are recalculated relative to the frame, the average curves for tones 1 and 2 do not coincide. This raises the issue of which value to use for normalisation: the F_0 value over the first few cycles of the utterance, under the hypothesis that it 'sets the tone' for all that follows? An average over the entire frame, or over the two vowels of /^{2}tsu-ri/? A midpoint value, or a final value, within /^{2}tsu-ri/? The decision was made on the basis of a study of the influence of the tone of the target syllable on the F_0 curves of the carrier sentence: the value selected was an average over the first half of the /i/ in /^{2}tsu-ri/, because that is where the F_0 differences conditioned by the tone of the target syllable were found to be strongest. The formula used for obtaining relative F_0 values (in semitones) is the following, where F_{REL} is the relative value (in semitones), F_{TARGET} the measurement on the target syllable (in Hz) and F_{FRAME} the measurement over the frame (in Hz):

$$F_{REL} = 12 \times \frac{\log\left(\frac{F_{TARGET}}{F_{FRAME}}\right)}{\log(2)}$$

An important result is that, even after the correction relative to the F_0 of the frame, there is some overlap between categories: some tone-2 syllables are realised in the same F_0 range as tone 1, others in the same range as tone 4. This confirms the auditory impression that F_0 differences across tones are not considerable, and are unlikely to suffice for the correct identification of the tonal categories.

The speakers repeated each item within the carrier sentence twice at a go, e.g. /[2]cu-ri [3]suː-pa || [2]cu-ri [3]suː-pa/ "Here, he is planting. Here, he is planting." The repetition produced a degree of intonational variation: some speakers preplanned both repetitions and introduced an intonational marking of continuation (realised in Tamang as a rise) at the end of the first repetition. Furthermore, some tokens are followed by one or two syllables (/-pa/ or /-ka-cim/) whereas others are utterance-final. Visual inspection of the data for individual tokens suggests that the data are nonetheless sufficiently homogeneous to be pooled together in the statistical calculations. The measurements reported in what follows correspond to the target syllables only, i.e. excluding the suffixes, /-pa/ and /-ka-cim/.

Open quotient values were also recalculated, relative to a mean value for the speaker. In view of the strong cross-speaker differences in mean open quotient, shown in table 1, O_q values were converted using the following formula, where $O_{q\ \mathrm{TARGET}}$ is the measurement for the glottal cycle at issue, and $O_{q\ \mathrm{MEAN}}$ the speaker's mean O_q value, obtained by averaging across all the syllable rhymes in the corpus, carrier sentences included.

$$O_{q\ \mathrm{REL}} = 100 \times \left(\frac{O_{q\ \mathrm{TARGET}}}{O_{q\ \mathrm{MEAN}}} - 1\right)$$

For example, an O_q of 71.4% for a speaker whose mean O_q is 68% will be translated to the same value as an O_q of 57.8% for a speaker whose mean O_q is 55%: both correspond to +5% above the reference.

3.2.2. Averaged curves and statistical analyses

Since we are looking for fine-grained differences in phonation type in correlation with tone, it seemed advisable to consider separately syllables with an aspirated initial in the statistical treatments, so as to distinguish the glottal effects of aspiration from those of whispery voice. So, in computing the results of the analyses, we distinguish six categories of syllables, rather than four categories of tones. Words under tones 1 and 2, as we have seen, have an opposition between aspirated and unaspirated initial stops, whereas words under tones 3 and 4 only have one series of initial stops (unaspirated). The six syllable categories are: tone 1 with an aspirated initial (1_{asp}); tone 1 with an unaspirated initial (1_{unasp}); tone 2 with an aspirated initial (2_{asp}); tone 2 with an unaspirated initial (2_{unasp}); tone 3 (3); and tone 4 (4).

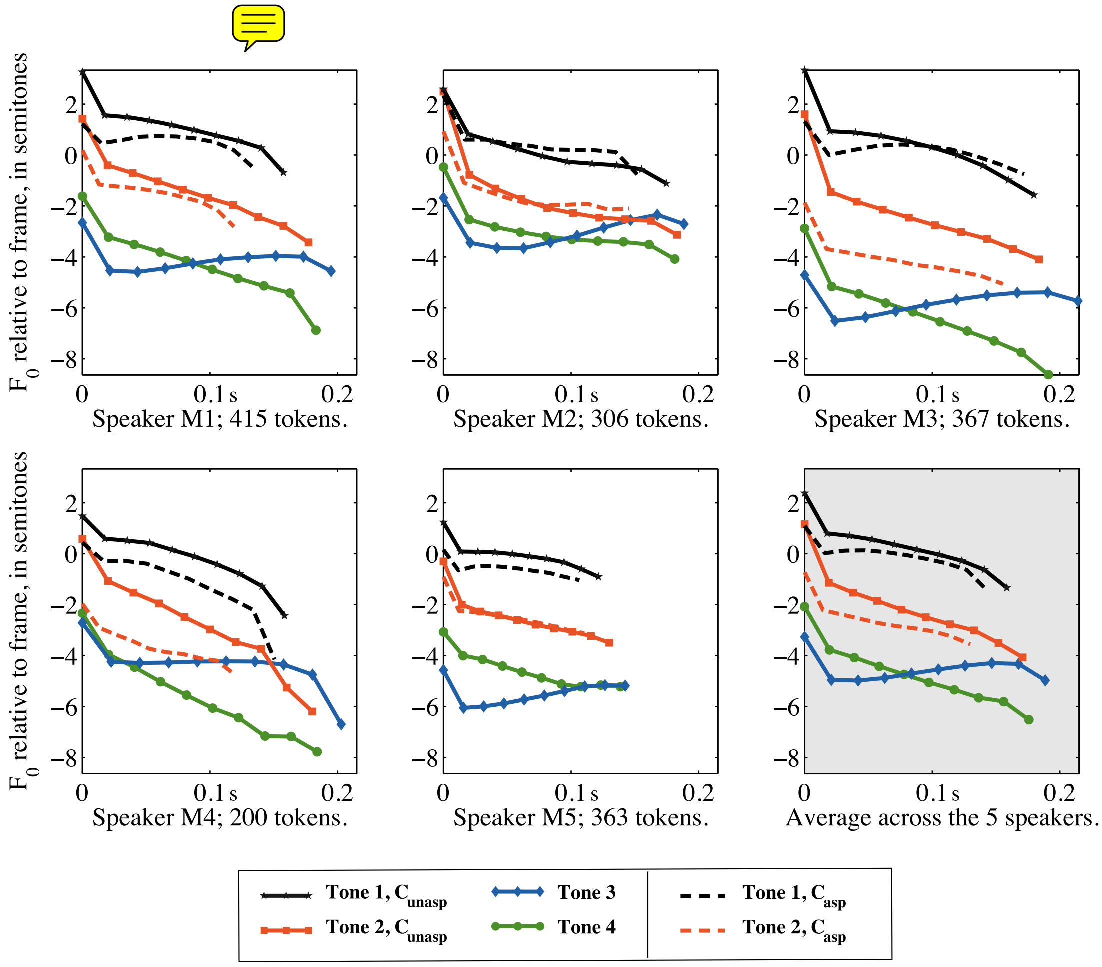
Figure 3. Averaged curves of fundamental frequency (relative to the F0 of the preceding syllable), plotted against average duration.

Figure 4. Averaged curves of open quotient (relative to the mean Oq value of each speaker), plotted against average duration.

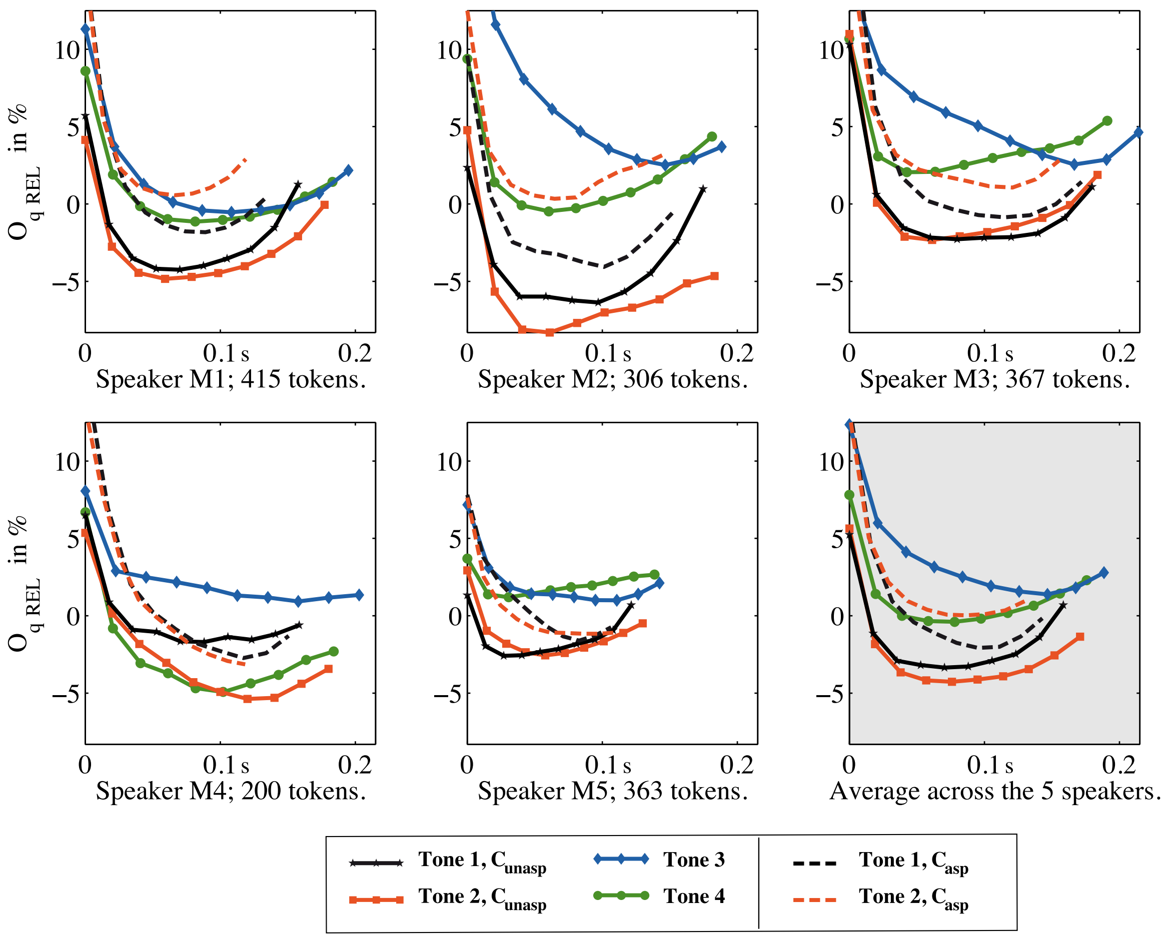
Averaged curves of fundamental frequency for each of the five speakers are shown in figure 3, and open quotient in figure 4. Curves averaged over the five speakers are also provided on the lower right-hand corner of these figures. No error bars are plotted because standard deviations are high, and plotting them would make the figures quite difficult to read. The time courses of tones 1, 2 and 4 are roughly parallel, whereas tone 3 stands out by its rise. For speakers M1, M2 and M5, syllables 1_{asp} and 2_{asp} are close to 1_{unasp} and 2_{unasp}, respectively, in terms of F_0. On the other hand, no such proximity emerges from the data of speakers M3 and M4, where the F_0 curves of 1_{asp} and 2_{asp} appear to occupy a larger part of the tonal space than 1_{unasp} and 2_{unasp}, as if dividing the tonal space into two halves: the upper half for 1_{asp}, the lower half for 2_{asp}. The relative brevity of the F_0 curve on syllables with an aspirated initial in comparison with syllables with an unaspirated initial does not reflect the duration of the syllable as a whole; it is due to the fact that the measurement of F_0 and O_q bears on the voiced portion of the syllable, excluding the VOT.

The average differences in open quotient are small. For the human voice, O_q ranges from about 40 – and even less in cases of glottalisation – to about 75; the window chosen for figure 4 covers no more than two thirds of this range, and the curves only occupy part of this window. The tones of Tamang do not involve such very clear oppositions in phonation type as that between *extremely pressed* and *modal-to-whispery* phonation types in Vietnamese tones [documented by Michaud 2004, Michaud et al. 2006, Brunelle forthcoming and references therein]. It may be useful to recall at this point that the correlation between open quotient and airflow is not linear: a difference of 4% in open quotient between two tones suggests a difference in airflow probably much higher than 4%.

Tones 1 and 2 share the same range of O_q. Tones 3 and 4 also share in part the same range, that of higher O_q values, indicative of more relaxed/whispery phonation. A sizeable proportion of tone-3 tokens show a steep rise in O_q at some point in the first half of the syllable, resulting in a telltale bell-like shape of the O_q curve; this phenomenon, which is levelled out in the averaged curves, suggests a sharp and brief increase of airflow occurring within the first half of the vowel. A rise in O_q in the course of the vowel is also found in some tone-4 tokens, whereas it is rare for tones 1 and 2.



F_0 relative to frame, in semitones
Speaker M1; 415 tokens.
Speaker M2; 306 tokens.
Speaker M3; 367 tokens.
Speaker M4; 200 tokens.
Speaker M5; 363 tokens.
Average across the 5 speakers.
0.1 s
Tone 1, C_{unasp}
Tone 2, C_{unasp}
Tone 3
Tone 4
Tone 1, C_{asp}
Tone 2, C_{asp}



$O_{q\,REL}$ in %
10
5
0
−5
0
0.1 s
0.2
Speaker M1; 415 tokens.
Speaker M2; 306 tokens.
Speaker M3; 367 tokens.
Speaker M4; 200 tokens.
Speaker M5; 363 tokens.
Average across the 5 speakers.
Tone 1, C_{unasp}
Tone 2, C_{unasp}
Tone 3
Tone 4
Tone 1, C_{asp}
Tone 2, C_{asp}

A multivariate analysis of variance (MANOVA) was performed over the entire set of data; the total number of values, excluding the data points where O_q values were missing, was 14,772. The dependent variables were O_q and F_0, and the independent, nominal variables were the six tone/syllable categories (1_{unasp}, 1_{asp}, 2_{unasp}, 2_{asp}, 3 and 4), the ten points in time, and the five speakers. The values of O_q and F_0 were normalised as explained in §3.2.1. The software used was StatView 5.0. A global significant effect was observed for all three variables, at $p < 10^{-4}$. This highly positive overall result legitimates the use of more focused statistical analysis. One-factorial ANOVAs (analyses of variance) were specifically targeted at the O_q data at $4/10^{ths}$ of vowel duration, a time point where visual inspection of the curves in figure 4 suggested that the O_q differences were at their clearest. This test was conducted separately on the data of the five speakers, with the six tone/syllable categories (1_{unasp}, 1_{asp}, 2_{unasp}, 2_{asp}, 3 and 4) as the dependent variable. Table 2 shows the results.

Table 2. Results of an ANOVA over the open quotient values at $4/10^{ths}$ of vowel duration. Same data as in figure 4. Degrees of freedom: 5.

speaker	F value	p value
M1	25.0	<0.0001
M2	31.6	<0.0001
M3	41.6	<0.0001
M4	2.96	0.014
M5	7.65	<0.0001

Table 3. Results of a post-hoc Scheffé test following the ANOVAs reported in table 2. The value provided in each cell is the p value. Blank cells indicate that p is higher than 0.05.

speak -er	1_{unasp} /2_{unasp}	1_{unasp} /3	1_{unasp} /4	2_{unasp} /3	2_{unasp} /4	3 /4	1_{unasp} /1_{asp}	1_{unasp} /2_{asp}	2_{unasp} /1_{asp}	2_{unasp} /2_{asp}	3 /1_{asp}	3 /2_{asp}	4 /1_{asp}	4 /2_{asp}	1_{asp} /2_{asp}
M1		$<10^{-4}$	0.0007	$<10^{-4}$	$<10^{-4}$		$<10^{-4}$	$<10^{-4}$	$<10^{-4}$	$<10^{-4}$					
M2		$<10^{-4}$	0.040	$<10^{-4}$	0.0006	$<10^{-4}$		0.011		10^{-4}	$<10^{-4}$	0.008			
M3		$<10^{-4}$	$<10^{-4}$	$<10^{-4}$	$<10^{-4}$	$<10^{-4}$		10^{-4}		0.0002	$<10^{-4}$	0.0007			
M4															
M5		10^{-3}	0.0075	0.0058	0.023		0.012		0.036						

The effect of the dependent variable is significant ($p < 0.05$) for all speakers. Table 3 provides the results of a post-hoc Scheffé test. For speakers M1 and M5, among words with unaspirated initials, all pairs are significantly different for O_q, with two exceptions: $1_{unasp}/2_{unasp}$ and 3/4. This result brings out the hypothesised difference between two subsets: on the one hand, tones 1 and 2 with unaspirated initials, and on the other hand, tones 3 and 4, which have a higher open quotient. In the data of speakers M2 and M3, tone 3 emerges as the one with highest open quotient (indicative of whispery voice): its open quotient is even significantly higher than that of tone 4. Again, 1_{unasp} and 2_{unasp} do not have statistically different O_q values from each other. For M4, no single pair differs significantly. Recall from table 1 that there is a smaller amount of data for speaker M4 than for the four other speakers: this has a bearing on statistical analyses. Tone 4 does not appear to pattern like tone 3 for this speaker: it shows no signs of whispery phonation. It would seem that speaker M4 only uses whispery phonation type in association with tone 3; he realises tone 4 markedly lower than the other speakers, and without any traces of whispery phonation, as if F_0 were low enough to render additional cues to tone unnecessary.

As for syllables with aspirated initials, we observe on figure 4 that O_q is globally higher on 1_{asp} and 2_{asp} than on 1_{unasp} and 2_{unasp}. In terms of open quotient, there is no difference between 1_{asp} and 2_{asp} in the data of speakers M4 and M5, whereas M1, M2 and M3 have lower O_q for 1_{asp} than for 2_{asp}. This trend is not found for 1_{unasp} and 2_{unasp}. Across speakers, there does not appear to be any relationship, either direct or indirect, between the distance in terms of O_q and the distance in terms of F_0 for 1_{asp} and 2_{asp}: speaker M3, who makes the largest difference between 1_{asp} and 2_{asp} in terms of F_0, also produces them with some differences in O_q. In the data of speaker M5, 1_{asp} and 2_{asp} are relatively close in terms of F_0, and very close indeed in terms of O_q. For speaker M2, 1_{asp} and 2_{asp} are relatively close in terms of F_0 and the furthest apart in terms of O_q.

Now turning to a comparison of the 'whispery' tones (3 and 4) with the tones found after aspirated initials (1_{asp} and 2_{asp}), it appears that values of the open quotient at the very onset are even slightly higher for 1_{asp} and 2_{asp} than for 3 and 4. The average values across speakers are: 73% for {1_{asp}, 2_{asp}}, 69% for {3, 4}, and 65% for {1, 2}. This reflects the presence of high airflow up until the onset of voicing after aspirated initials. For tones 1_{asp} and 2_{asp}, O_q then decreases rapidly, getting close to the speaker's value of reference for O_q (value zero in figure 4) within about 50 ms after the onset of voicing. On the other hand, for tone 3, while the O_q value is always much higher for the first glottal cycle than for the following cycles, an increase in O_q in the first half of the vowel is not infrequent. On average, the slope of the curve is gently decreasing, with a final rise also found in the other tones, corresponding to the offset of voicing. As for tone 4, the decrease in O_q in the first half of the syllable is steeper than for tone 3. There is only one speaker, M1, for whom the shape of the O_q curves for tones {1_{asp}, 2_{asp}} is not well differentiated from that of tones {3, 4}; at 4/10^{ths} of the syllable, the O_q of the tones found after aspirated initials (1_{asp} and 2_{asp}) is not statistically different from that of tones 3 and 4: see table 3. To sum up in a nutshell: the whispery phonation type which the data show to be associated with tone 3 (and, less saliently, with tone 4 in the data of four out of the five speakers) is manifested by O_q curves that differ from those found after aspirated initials.

3.3. Observations on the realisation of initial consonants

Some tone-linked allophonic variation of initial stops was observed. It is described below in terms of voice onset time (hereafter VOT), of stop closure duration, and of voicing of the initial consonant: presence or absence of voicing, and, if the consonant is not voiced throughout, duration of the interval without full voicing. The sibilant /s/ is consistently voiceless throughout.

3.3.1. Voice onset time (VOT)

Table 4 presents the results for VOT. The total number of tokens for which the stop release could be reliably detected from a spectrogram, and hence the VOT estimated, is 1251.

Table 4. Results for VOT. For each cell: average, in milliseconds; standard deviation; number of tokens.

category	1_{unasp}			2_{unasp}			3			4			1_{asp}			2_{asp}		
M1	42	22	58	39	33	46	32	21	80	41	27	39	100	28	52	101	28	38
M2	57	39	40	43	33	37	63	33	57	69	48	27	130	33	37	111	30	20
M3	38	34	52	36	31	40	48	38	75	49	34	45	133	40	61	136	47	37
M4	35	14	15	30	22	22	33	28	54	43	31	27	102	27	19	112	41	16
M5	41	25	54	39	33	36	36	28	68	35	23	34	97	29	41	101	26	24
average, in ms	43			37			42			47			112			112		

In terms of voice onset time, two distinct sets emerge from the data: VOT only differentiates between syllables with aspirated initials and syllables with unaspirated initials. In syllables without aspirated initials, no difference in VOT emerges across tones 1 to 4; standard deviation is considerable; mean values are close. This is confirmed by an ANOVA: for all five speakers, a significant effect is found, but it is only due to the difference between unaspirated and aspirated initials. A Scheffé post-hoc test shows that all the comparisons involving a set of aspirated initials and a set of unaspirated initials conclude to a significant difference (at $p < 10^{-4}$), whereas none of the other pairs differ significantly.

3.3.2. Closure duration

Table 5. Results for closure duration. For each cell: average in milliseconds; standard deviation; number of tokens.

	1_{unasp}			2_{unasp}			3			4			1_{asp}			2_{asp}		
M1	87	26	60	83	35	46	77	19	77	82	20	35	66	23	51	75	24	36
M2	91	24	39	88	28	31	78	33	54	76	19	25	68	20	31	84	25	17
M3	132	40	40	129	48	30	105	29	69	107	32	37	90	36	55	97	36	33
M4	89	19	15	99	31	22	83	17	53	90	46	27	62	13	17	69	18	14
M5	90	23	46	90	37	33	66	21	53	67	21	27	61	28	35	63	24	18
average, in ms	98			98			82			84			69			78		

Table 5 presents the results for closure duration. The average closure duration is slightly shorter for tones 3 and 4 than for tones 1_{unasp} and 2_{unasp}; it is shortest for 1_{asp} and 2_{asp}. The correlation of syllable category (tone + aspiration) with closure duration is significant for all speakers. Table 6 presents the results of an analysis of variance and the pairs that differ significantly according to a post-hoc Scheffé test. For M3, tone 3 differs from all the others. For M4, no significant difference is brought out. For M5, two groups differ from each other and have no consistent internal differences: 1_{unasp} and 2_{unasp} on the one hand, 3, 4, 1_{asp} and 2_{asp} on the other.

Table 6. Results of ANOVA tests over the closure duration measurements for the five speakers. The number of degrees of freedom, df, is 5 in all cases.

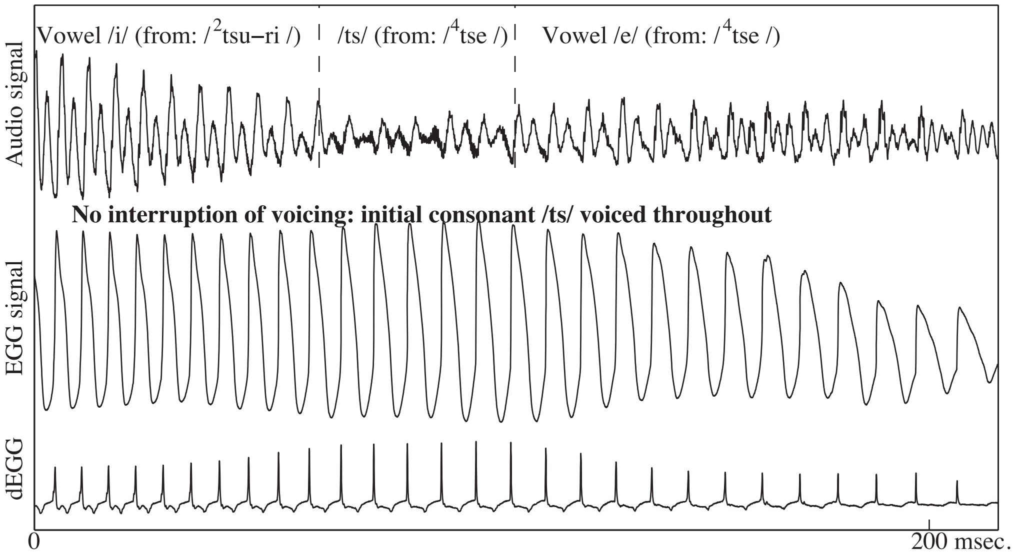
speaker	F	p	pairs that differ significantly, according to a post-hoc Scheffé test
M1	4.768	3×10^{-4}	$1_{unasp}/1_{asp}$, $1_{unasp}/2_{asp}$
M2	3.316	6.8×10^{-3}	$1_{unasp}/1_{asp}$
M3	9.05	$< 10^{-4}$	$1_{unasp}/3$, $1_{unasp}/1_{asp}$, $1_{unasp}/2_{asp}$, $2_{unasp}/1_{asp}$, $2_{unasp}/2_{asp}$
M4	4.890	4×10^{-4}	$2_{unasp}/1_{asp}$
M5	9.997	$< 10^{-4}$	$1_{unasp}/3$, $1_{unasp}/4$, $1_{unasp}/1_{asp}$, $1_{unasp}/2_{asp}$, $2_{unasp}/3$, $2_{unasp}/4$, $2_{unasp}/1_{asp}$, $2_{unasp}/2_{asp}$

3.3.3. Voicing of the initial consonant

Not a single case of voicing of initial fricatives was observed for any of the five speakers. As for stops and affricates, a variety of realisations were observed, from those with a clear interruption of voicing, as in figure 2, to uninterrupted voicing, as in figure 5.

Figure 5. A case of fully voiced realisation of the initial consonant of the target syllable. Speaker M3, consonant /ts/ in sequence /i.tse/, from sentence /^{2}tsu-ri ^{4}tse/.

For the sake of classifying the observed realisations, three cases were distinguished: (i) the consonant was labelled as *fully voiced* when there was no interruption of full voicing between the preceding syllable and the target syllable (e.g. in figure 5); (ii) the consonant was labelled as *partly voiced* when quasi-periodic fluctuations of vocal fold contact area


Vowel /i/ (from: /^{2}tsu–ri /)
/ts/ (from: /^{4}tse /)
Vowel /e/ (from: /^{4}tse /)
Audio signal
No interruption of voicing: initial consonant /ts/ voiced throughout
EGG signal
dEGG
0
200 msec.

were observable on the electroglottographic signal throughout the consonant but some or all of the glottal cycles had no closing peak (symmetrical oscillations of small amplitude, as in figure 6); (iii) the consonant was labelled as *unvoiced* if laryngeal vibrations were entirely interrupted during part or all of the consonant, as in figure 2.

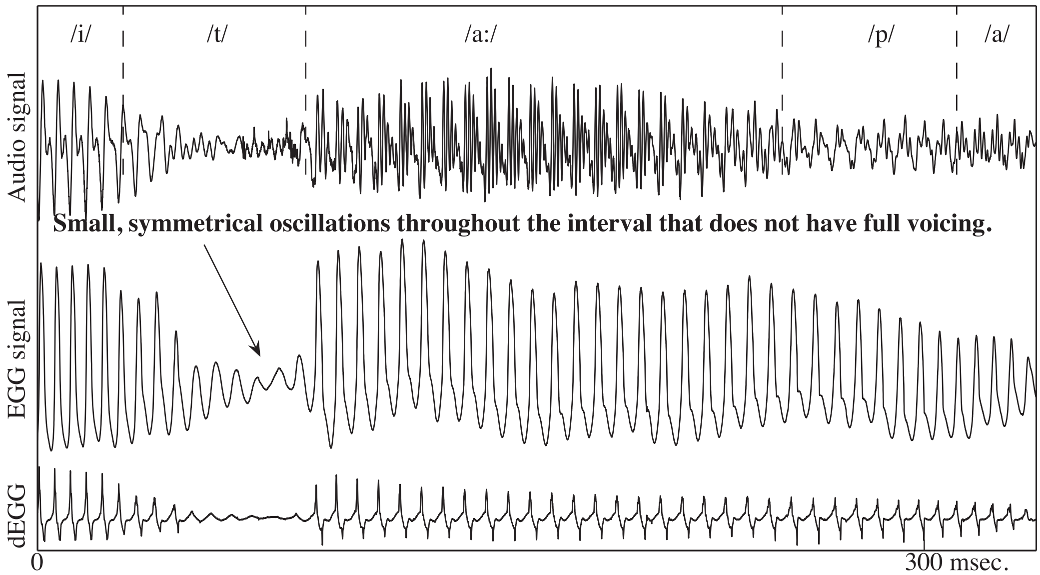
Figure 6. A case where continuous symmetrical oscillations are observed on the electroglottographic signal throughout the initial consonant. Speaker M5, consonant /t/ in sequence /i.taː/, from sentence /^{2}tsu-ri ^{3}taː-pa/.

Category (ii) calls for some explanations. Symmetrical oscillations of small amplitude on the electroglottographic signal indicate that, as vocal fold adduction decreases, the glottis does not have a sharp closure anymore (i.e. no *glottis-closure-instant*); fluctuations in vocal fold contact area become very small while remaining quasi-periodic. A few such oscillations are generally observed during a 'soft' (nonglottalised) offset of voicing. Figures showing such oscillations during voice decay in Vietnamese are provided by Michaud 2004:126-129. Images of the glottis by high-speed cinematography (English and French data) show that these small oscillations correspond to a state where the glottis does not actually close, but the vocal folds still oscillate quasi-periodically, making contact at the anterior part of the glottis [Kitzing 1982, and Cédric Gendrot, p.c.]. Whether an initial consonant has small symmetrical oscillations or not correlates with the duration of the portion without full voicing: the longer the portion without full voicing, the more likely it is that the periodic oscillation of the vocal folds will be interrupted altogether – or, at least, will become so small as to be undetectable from the EGG signal. The categories proposed here are based on the electroglottographic signal; it should be borne in mind that the intermediate category, (ii), does not correspond to a distinct mode of vibration (a voice register in the sense of Roubeau et al. 1987), but simply to a transitional state. A visual representation of the results for unaspirated initials is offered in figure 7.

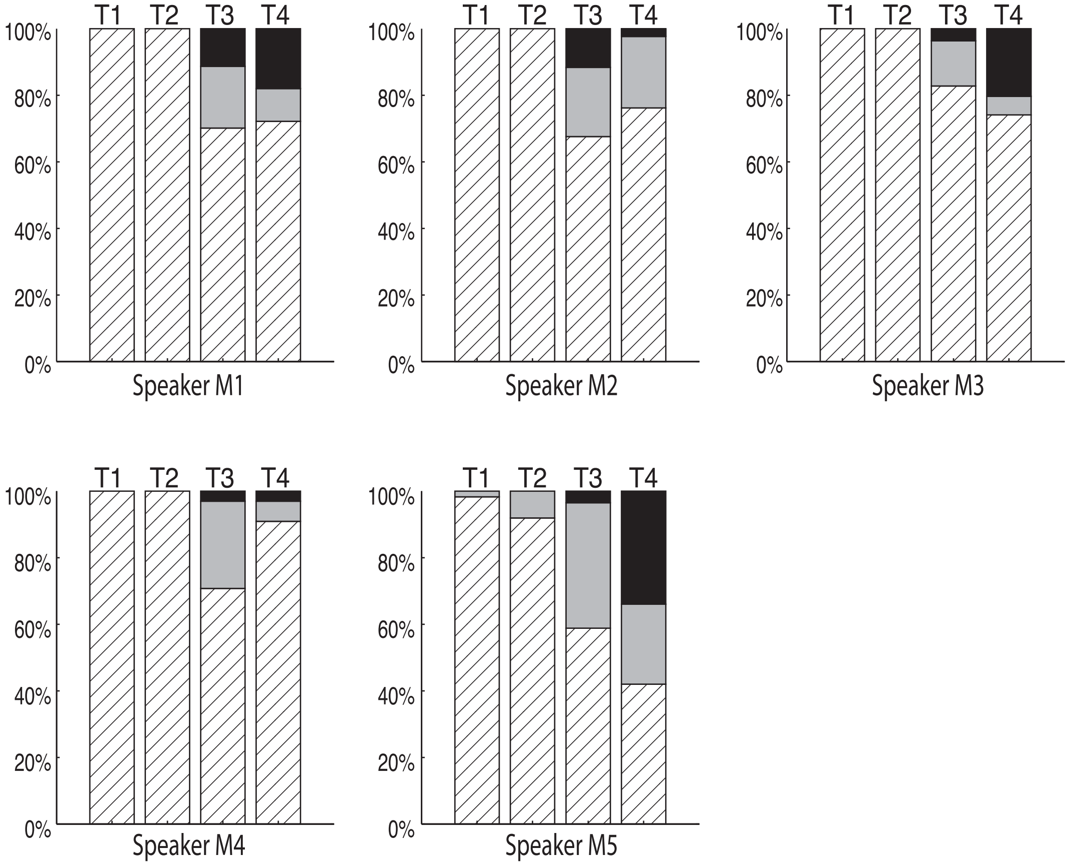
Figure 7. Voicing of the initial consonant of the target syllable in relation to tone (aspirated initials excluded). Stripes = interruption of voicing, grey = symmetrical oscillations, black = fully voiced.

The results are clear: full voicing of the initial consonant is never found in tone-1 and tone-2 syllables, whereas it is not uncommon in tone-3 and tone-4 syllables. A χ^2 test was conducted for each speaker, excluding syllables with aspirated initials; it confirmed that the distributions in Table 12 are not random (M1, M2, M3, M5: $p < 10^{-4}$; M4: $p = 4.6 \times 10^{-3}$).

On the fricative /s/ (total for the 5 speakers: 135 tokens), no single case of voicing was observed: neither full voicing, nor symmetrical oscillations.



/i/
/t/
/a:/
/p/
/a/
Audio signal
Small, symmetrical oscillations throughout the interval that does not have full voicing.
EGG signal
dEGG
0
300 msec.


T1 T2 T3 T4
100%
80%
60%
40%
20%
0%
Speaker M1
Speaker M2
Speaker M3
Speaker M4
Speaker M5

3.3.4. Duration of the portion without full voicing

Table 7. Duration of the portion without full voicing found at the beginning of the target syllable. For each cell: average duration of the portion without full voicing, in milliseconds; standard deviation; number of tokens.

category	1_{unasp}			2_{unasp}			3			4			1_{asp}			2_{asp}		
M1	112	33	64	107	28	49	70	39	97	68	49	61	154	33	60	154	44	44
M2	139	39	47	110	52	39	113	61	77	105	55	42	161	66	49	175	60	29
M3	117	71	52	109	72	40	105	49	81	80	71	54	186	67	68	189	90	39
M4	125	19	19	120	25	23	93	32	65	117	48	33	165	23	40	148	47	18
M5	124	34	57	120	47	37	79	34	85	58	51	50	154	38	61	154	35	36
average	123			113			92			87			164			164		

Table 7 reports on the duration of the portion without full voicing found at the beginning of the target syllable. This is defined as the portion between the last positive peak on the dEGG during the consonant, and the first positive peak during the vowel, as shown on figure 2. This temporal measurement aims at providing more fine-grained results than the three-way classification applied in figure 7. When 'full' voicing (characterised by an uninterrupted sequence of opening and closing peaks on the derivative of the EGG signal) was continuous throughout the consonant, this parameter was set at zero; these cases are taken into account in the statistical calculations. To get an idea of cross-speaker regularities, equal weight was given to each speaker, rather than proportionally to the number of tokens. This amounts to the assumption that the data given for each speaker are representative. For fricative-initial syllables, no significant tone-correlated differences emerge concerning the duration of the portion without full voicing, so these syllables are not included in the statistical analysis. An ANOVA test was applied separately for each speaker, followed by a post-hoc Scheffé test.

- For M1 and M5, all pairs are significantly different except $1_{unasp}/2_{unasp}$, 3/4 and $1_{asp}/2_{asp}$, i.e. the measurement of the portion without full voicing successfully reflects the hypothesised difference between tones 3 and 4 on the one hand, and tones 1 and 2 on the other hand (aspirated-initial syllables making up a distinct set).

- For M2, the only statistically significant differences that emerge are those between items with aspirated initials and with unaspirated initials. On average, 1_{unasp} has a relatively long portion without full voicing; as a result, it is statistically different neither from 1_{asp} and 2_{asp} nor from the other unaspirated syllables.

- For M3, two sets emerge: unaspirated 1 to 4 on the one hand, 1_{asp} and 2_{asp} on the other.

- For M4, the average duration of the interval without full voicing is lower for tone 3 than for the other tones; tone 4 does not appear to pattern together with tone 3, any more than it does in terms of O_q. This suggests that tone-conditioned allophonic variation of the initial consonant goes hand in hand with whispery phonation: it is not expected that a speaker will exploit the one to characterise a certain tone (say, tone 3) and the other to characterise another (say, tone 4).

Overall, the data in table 7 differentiate three sets, from longest to shortest period without full voicing: items with tones 1 and 2 and aspirated initials > items with tones 1 and 2 and unaspirated initials > items with tones 3 and 4 (which we recall never have aspirated initials). Since we observed in §3.3.1 that VOT is markedly longer for aspirated initials, it seemed interesting to compute the duration of the portion without full voicing before stop release, i.e. excluding VOT. Table 8 lists these values. Categories 1_{unasp} and 2_{unasp} stand out as having the longest unvoiced portion before stop release. This is consistent with the hypothesis that they are cleary voiceless, as opposed to 3 and 4.

Table 8. Average duration of the portion without full voicing before the stop release, i.e. the values in table 7 minus VOT.

category	1_{unasp}	2_{unasp}	3	4	1_{asp}	2_{asp}
total duration without full voicing	123	113	92	87	164	164
VOT	43	37	42	47	112	112
duration of closure without full voicing	80	76	50	40	52	52

3.4. Realisation of the medial consonant

The initial consonant of the suffix is realised in intervocalic position, a position in which unaspirated stops are reported to be commonly voiced, and sometimes spirantised. We checked if any tone-correlated variation could be detected also in the realisation of the initial consonant of the suffix. The observations were conducted in the same way as for word-initial consonants: see §3.3.3. In addition, the consonant was labelled as either 'stop' (for /k/: realisation as [k] or [g]; for /p/: [p] or [b]) or 'approximant' (for /k/: [ɣ]; for /p/: [ß]), on the basis of auditory impression. In cases where the stop was unvoiced, it was always perceived as a stop, never as an approximant, i.e. there were no cases of [x] or [ɸ].

3.4.1 Voicing of the initial of the suffix

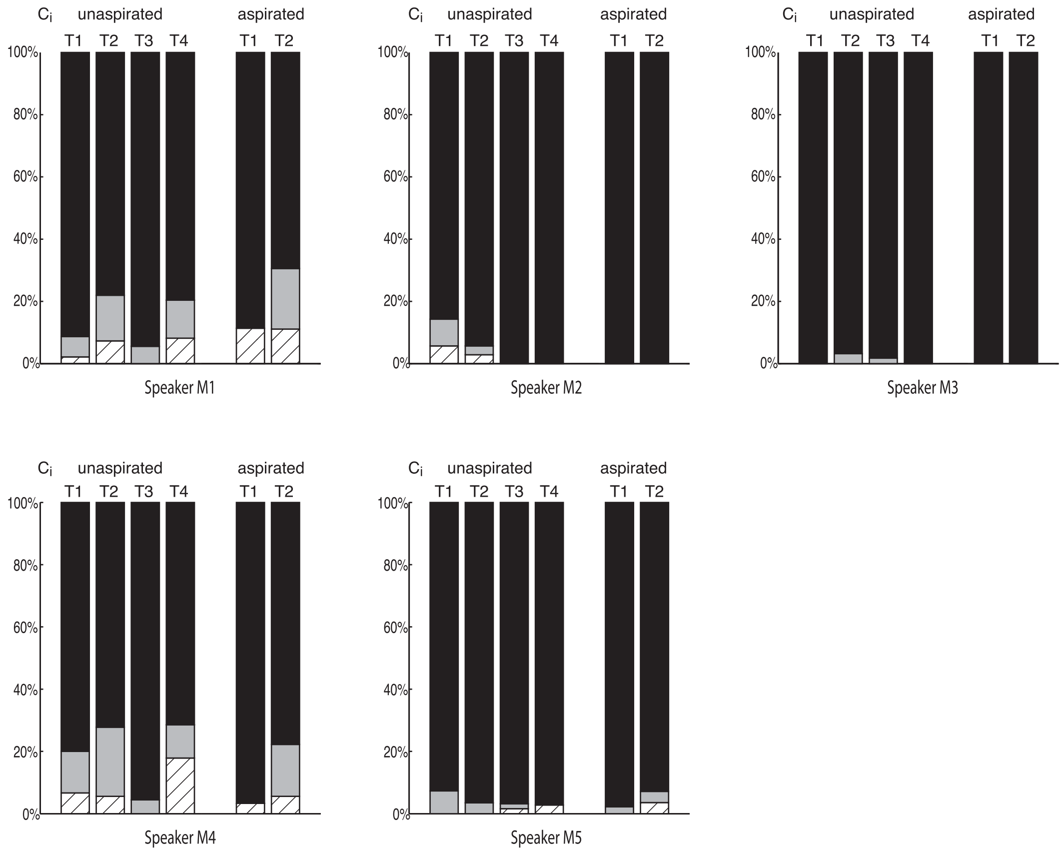
Figure 8 provides a visual representation of the results concerning the state of the glottis during the initial of the suffix. For M1, M3 and M5, no difference emerges across tones. Slight tendencies are observed for speakers M2 and M4. For M2, some cases of full interruption of vocal fold oscillations (as observed by EGG) are found for initials of suffixes following tones 1 and 2, whereas no such case is found for tones 3 and 4. It is clear that this trend, if present at all, is of a different order of magnitude than that observed on the initial of the target syllable. For M4, tone 3 has the fewest cases of full interruption of vocal fold oscillations on the initial of suffixes of any of the categories.

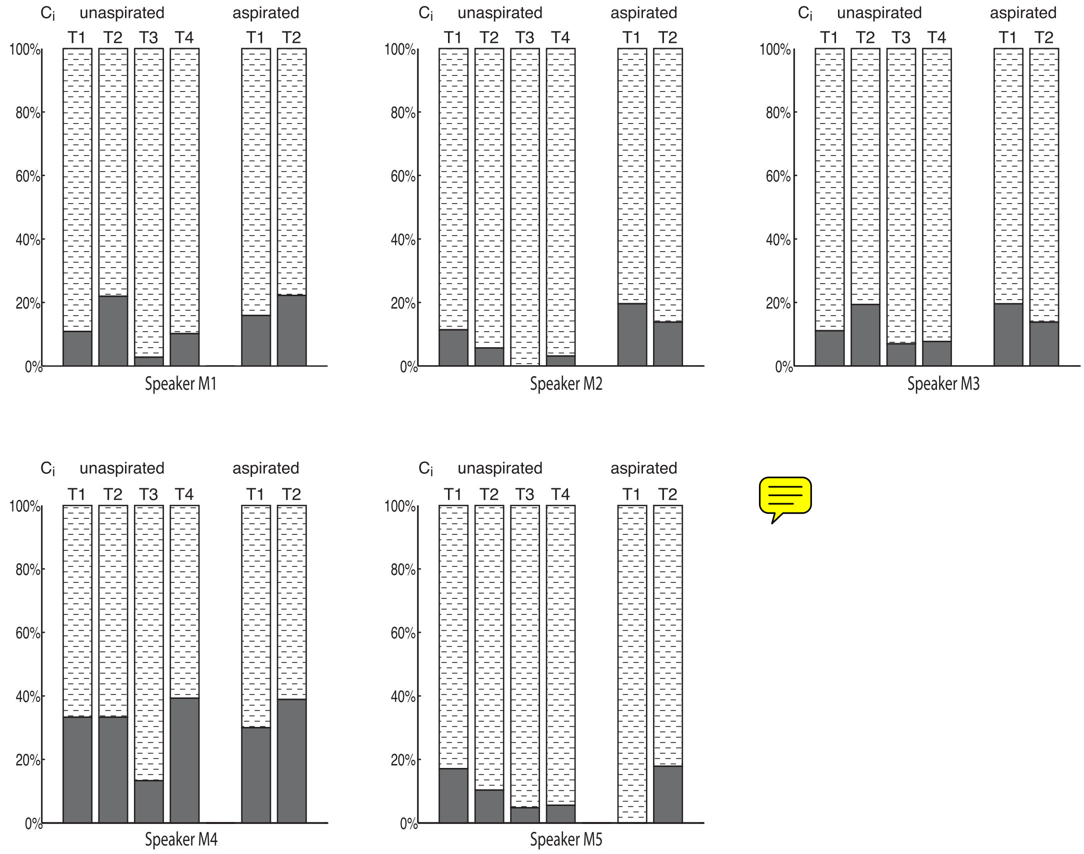
Figure 8. Voicing of the initial consonant of the suffix in relation to tone and mode of articulation of the initial consonant of the target syllable. Stripes = interruption of voicing, grey = symmetrical oscillations, black = fully voiced.

3.4.2 Lenition of the initial of the suffix

Under the hypothesis that tones 3 and 4 may be realised with a general "laxness", it was thought plausible that lenition of the intervocalic stop could be more frequent with these tones. Figure 9 sums up the results.

Figure 9. Spirantisation of the initial consonant of the suffix in relation to tone and mode of articulation of the initial consonant of the target syllable. Grey = occlusive, dashed lines = spirant.


C_i unaspirated aspirated
T1 T2 T3 T4 T1 T2
100% 80% 60% 40% 20% 0%
Speaker M1
Speaker M2
Speaker M3
Speaker M4
Speaker M5



C_i
unaspirated
aspirated
T1
T2
T3
T4
100%
80%
60%
40%
20%
0%
Speaker M1
Speaker M2
Speaker M3
Speaker M4
Speaker M5

The general picture that emerges is that the suffix's consonant was perceived as a stop in only 5% of cases for tone 3, as against 12% to 18% for the other tonal categories. A statistical treatment by χ^2 was applied to these data (1 degree of freedom). For M1, tone 3 has more cases of approximant realisations of the initial of the suffix; tone 4 does not stand out clearly ($\chi^2 = 13.5$, $p = 0.019$). For M2 and M3, tones 3 and 4 have more cases of approximant realisations; differences are statistically significant for M2 ($\chi^2 = 16.2$, $p = 6.4 \times 10^{-3}$), not for M3 ($\chi^2 = 6.06$, $p = 0.30$). For M4 as for M1, tone 3 has more cases of approximant realisations, but the trend is not significant ($\chi^2 = 8$, $p = 0.15$). For M5, no clear tendencies emerge.

For words with an initial fricative in the target syllable, averaging across speakers, the proportion of realisations of the suffix's consonant as stop amounts to 22% for tone 1 (total: 40 tokens), 5% for tone 2 (40 tokens), 0% for tone 3 (24 tokens). For reasons of lexical distribution, no single fricative-initial with tone 4 is present in the list of items recorded. These data, though very limited, reach statistical significance ($\chi^2 = 10.2$, $p = 6.2 \times 10^{-3}$), suggesting that there may be a small correlation of tone with the (subphonemic) spirantisation of the consonant of the suffix after syllable-initial fricatives. To sum up, there is some evidence to substantiate the claim that the realisation of the initial of the suffix is influenced by the tone of the target word, but this influence is limited, and is not consistent across speakers.

4. Discussion

It is customary in phonological typology to distinguish between tonal languages and voice-register languages, yet it seems that the boundary between them is fuzzy. The present experiment confirms (i) that tones 3 and 4 have a significantly higher glottal open quotient, indicative of higher airflow than tones 1 and 2, and (ii) that there is some tone-correlated allophony of word-initial consonants. We will try to bring out the complex way in which the four tones of Tamang are differentiated in speech production by various combinations of fundamental frequency, phonation type and allophony of consonants (§4.1). After some diachronic reminders (§4.2), we will propose a critical evaluation of phonological analyses that have been hitherto proposed for Tamang tones (§4.3).

4.1. Reflections on the respective importance of the cues to tone in Tamang

The higher open quotient found for tones 3 and 4 cannot be put down to a phonetic universal whereby fundamental frequency and open quotient would be correlated either directly or inversely. For instance, in Naxi, where the lexical tones are simply specified in terms of pitch, divergences on open quotient are found across speakers: in an acoustic and electroglottographic study of five speakers [Michaud 2005:128-143], it appeared that there was a statistical correlation of F_0 and O_q for one of the speakers, and an inverse correlation for another speaker. In Tamang, the two tones that are most dissimilar in terms of O_q are tones 2 and 3, which are closest in terms of mean F_0.

In the present experiment, the somewhat whispery phonation type of tones 3 and 4 was compared with that of aspirated-initial syllables (syllables coded 1_{asp} and 2_{asp}). The results are reminiscent of those found by Blankenship 2002, who compared the effect of initial aspiration in Swedish and English [as reported by Löfqvist 1992] with the effect of whispery/breathy phonation type over vowels in Mazatec (Oto-Manguean, Popolocan); the latter extends for roughly twice as long as the high-airflow phonation found at the onset of voicing after the aspirated initials of Swedish and English. Likewise, in Tamang, it was found that the whispery phonation type associated with tone 3 – and, less saliently, with tone 4, in the data of four out of the five speakers – is manifested by O_q curves that differ from those found after aspirated initials: the high O_q after aspirated initials generally lasts only for a few cycles after the onset of voicing.

Concerning the presence of a tone-correlated allophonic variation of word-initial consonants, a salient finding is that this is not a difference in VOT. No consistent tone-correlated VOT differences are observed among unaspirated initial syllables, whereas slight differences emerge in terms of closure duration, and clearer differences in terms of the degree of voicing of the initial consonant (as reported in §3.3.2-3, and in figure 7).

A perceptual study of the relative importance of each cue to tone was not attempted here. It would not be easy to conduct with the unschooled villagers who are the best speakers of the language. The relative importance of each cue is thus evaluated from the production data only. In terms of overall F_0 register, it is apparent from figure 3 that the tones are layered from highest to lowest (from 1 to 4), taking into account the fact that tone 3 is rising, which, from a perceptual point of view, gives the overall impression of higher pitch than tone 4. However, as hypothesised on the basis of auditory impressions, the phonetic distance across tones in terms of overall F_0 (F_0 register) is not considerable in Tamang. The shape of the contour of the tone is an important cue. The rise of tone 3, even though moderate, contrasts with the fall of tone 4. Tones 1 and 2 are both falling in the data reported here. To the ear, tone 1 often sounds more falling than tone 2. This could be related to the fact that tone 1 tends to be somewhat shorter than any of the others. Perceptually, it gives the impression of a more "energetic" tone. Mazaudon 1973, after Pike, called it "ballistic". Disyllabic words were out of the scope of our study, but Mazaudon as well as Weidert noted than on second syllables of tone-1 words, the fall continued on the second syllable, whereas on the second syllable of tone-2 words it did not. As mentioned in the introduction, non-initial syllables of words, whether they be a suffix or part of a single morpheme, never carry their own tone, so that their F_0 curve can be considered an expression of the tone lexically carried by the initial lexeme, which is allowed to unfold over the available space (the entire phonological word). The same is true of tone 4, which, on a polysyllabic morpheme, rises on the second syllable, before falling again on the third if there is one, whereas tone 3 evens out on following syllables.

The second obvious cue to tone in Tamang is phonation type, which is used to a different extent by the different speakers, as evidenced by the cross-speaker differences in the range of O_q in figure 4. For instance, data from speaker M2 show a small F_0 range, and conversely a large range of O_q. We are led to conclude that each tone has an individual prototype in which the fluctuating equilibrium of the cues is different from that of all the other tones. Variability appears to be part of the nature of the Tamang tones, in that none of the cues to tone is sufficient by itself to identify the tone with certainty. There is no obvious evidence for proposing a hierarchy of cues. As a summary, the tones on monosyllabic items could be characterised as in table 9, keeping in mind that the extent to which each cue is realised may vary.

Table 9. An attempt at a summary of the phonetic properties distinguishing the four tones of Risiangku Tamang.

<table>
<tr><td>tone</td><td>overall F0</td><td>course of F0</td><td>phonation type</td><td>voicing of initial stops</td></tr>
<tr><td>1</td><td>highest</td><td>"ballistic"</td><td rowspan="2">modal</td><td rowspan="2">no</td></tr>
<tr><td>2</td><td>second highest</td><td></td></tr>
<tr><td>3</td><td>low</td><td>rising</td><td>whispery</td><td rowspan="2">occasional</td></tr>
<tr><td>4</td><td>lowest</td><td>somewhat "ballistic"</td><td>somewhat whispery</td></tr>
</table>

Tone 1 is the highest; it tends to be shorter, and is falling; it gives the perceptual impression of a short, "ballistic" tone. Tone 2 lacks any salient characteristic apart from being the second highest; it is not whispery. Voicing of the initial stop, when present, identifies with certainty a tone as being either 3 or 4. Tone 3 is whispery and rising, within an overall low register. Tone 4 is low, somewhat whispery, and its falling contour is somewhat "ballistic", though less so than tone 1.

4.2. Diachronic reminders

The TGTM languages provide a text-book case of the two-way tonal split which spread across Asia and Southeast Asia in the last millenium and a half [Haudricourt 1961, 1975]. Proto-TGTM is reconstructed as having had a two-tone system, tones A and B – whose phonetic nature remains unknown –, three series of initial stops (p, p^{h}, b, t, t^{h}, d, ʈ, ʈ^{h}, ɖ, ts, ts^{h}, dz, k, k^{h}, g) and two series of sonorants and sibilants (m, m^{h}, n, n^{h}, ŋ, ŋ^{h}, l, l^{h}, r, r^{h}, j, j^{h}, w, w^{h}, s, z, h). In all the languages of the TGTM group there occurred a merger of the voiced and voiceless unaspirated series, leading to the phonemicisation of four tones instead of two [Mazaudon 1978]. This accounts for a number of synchronic idiosyncrasies found in languages of the TGTM group. For instance, since there was no voicing contrast in intervocalic position when tone developed from the loss of that contrast, tone did not develop on non-initial syllables. As a consequence, modern TGTM languages, like Tibetan languages, have word-tone systems [Sun 1997, 2003, Mazaudon 2005; for data on various TGTM languages: Hale and Pike 1970; Hale and Watters 1973; Mazaudon 1973, 1978, 1996; Noonan 2003; Hildebrandt 2003]. The topic of the present study, namely the realisation of tone by a bundle of characteristics, can also be understood in this light, since tone 3 and tone 4 originate in words with distinctive voicing of the initial, as opposed to tones 1 and 2.

4.3. Phonological modelling of Tamang tones

4.3.1. Analyses in terms of two tones plus an orthogonal feature

The diachronic facts summarised in the previous section help place the present-day situation of Tamang tones in perspective; however, they do not entail a synchronic phonological model of this tone system. An etymologising description is admittedly possible in theory: the four tonal categories of Tamang could be described in terms of two tones plus a /voice/ feature of the initial consonants, the observed variability being attributed to the diversity of the phonetic correlates of this /voice/ feature. This is precisely the analysis proposed by Kjellin 1975 for Tibetan. The reconstruction reflected in this analysis is based on abundant and converging evidence; however, it was observed in §3.3.3 that the whispery phonation type associated with tones 3 and 4 is more consistently present than the voicing of word-initial stop consonants also associated with these two tones. This provides compelling evidence that whispery voice cannot be treated, in present-day Tamang, as a phonetic effect of a consonantal difference. We see traces, in the present-day system of Tamang, of what its ancestor may have been, and we use these traces as powerful indicators towards reconstruction, but we are fully aware that these traces cannot be granted the synchronic status of distinctive features.

Another proposal to reduce the four-way contrast to a two-tone contrast plus an orthogonal feature was put forward by Ian Maddieson. Risiangku Tamang, as described in Mazaudon 1973, constitutes language 507 in Ian Maddieson's UPSID database; Maddieson's representation of its prosodic system has served as a source for other authors, e.g. Silverman 1997. Maddieson presents Risiangku Tamang as having only two tones, and a set of "breathy vowels". His argument for rejecting Mazaudon's analysis of "breathiness as an inherent part of a set of contrasting tone" and attributing it to the vowels is, he claims, that "the two 'breathy tones' have the same pitch shapes as the two plain tones" [Maddieson 1984:132]. The results of the present experiment certainly do not support pairing the tones according to their 'pitch shape' as Maddieson proposes. Reducing the four-way prosodic contrast of Tamang to two tones plus a non-tonal feature does not appear synchronically acceptable, as it would amount to excluding altogether from the tonal system the register contrast originating diachronically in the voicing contrast over initials.

4.3.2. Analysis in terms of two tones plus two registers

Moira Yip proposes a phonological analysis of Tamang tones based on Weidert's description of a dialect of Tamang which is close to the Risiangku dialect [Weidert 1987:262]. She follows Duanmu 1992 in "identifying low register with obstruent voicing", making the voicing of Tamang initials on LOW tones a "straightforward assimilation rule, spreading low register/obstruent voice leftwards" [Yip 1995:486]. Yip's analysis amounts to flipping around the diachronic conditioning of whispery phonation and lower F_0 by initial voicing into a synchronic phonological dependency of voicing on tone: voicing, which used to be the conditioning factor, is now conditioned. This is a much more insightful account than an etymologising analysis under which consonant voicing would be considered as distinctive. But it is not basically different from an analysis of the four tones into tonal features, for which several options were considered by Hale and Pike 1970 and Mazaudon 1973. There are inherent uncertainties in a synchronic subdivision of the tone system: should one pair 1 and 3, versus 2 and 4, arguing that the first set is 'relatively higher' than the second, each within their own register (HIGH vs LOW)? Or should one pair 1 and 4, versus 2 and 3, arguing that the first set is "ballistic" and characterised by salient contours, the second set more level? Both analyses were rejected by Mazaudon 1973 as unconvincing.

4.3.3. Tamang tones as defined by a bundle of cues

The next step will consist in proposing a phonological model that captures the growing asymmetry in the present-day state of the system, where there remains only slender evidence for a neat bipartition of the tonal system into a higher register and a lower register. Panchronic phonology [Hagège and Haudricourt 1978, Mazaudon and Michailovsky 2007] aims to describe how systems evolve from one type into one or more new types. The tone system of Risiangku Tamang is not only a missing link of tonal bipartition, illustrating the evolution from two tones plus a voicing contrast to four tones. It also sheds light on the stage where the system begins to lose its etymological symmetry, and the four tones, breaking their last ties with the correlation of voicing, become free to evolve away from their original phonetic register. "Once constituted, the tonal system evolves without remembering its origins" [Haudricourt 1961:286; English translation: Haudricourt 1972]. This is indeed what happened in three TGTM languages, where one tone has moved through the tonal space away from its etymological partner: in the Tamang dialect of Taglung and in Marphali, tone 4 has now become high, as has tone 3 in Manangke [Mazaudon 1978, 2005, Hildebrandt 2003], and Manangke has lost phonation type differences altogether, so that the analysis in terms of two registers does not apply anymore.

5. Concluding remarks: on the importance of studying transitional states

The present experiment confirmed that Tamang tones are phonetically complex, presenting whispery phonation type for two of the tones, as well as some tone-correlated allophony of word-initial consonants. The organisation in two registers, inherited from the etymological voicing contrast on initials, is showing signs of weakening, so that a phonological account in terms of two tones plus two registers leaves aside important aspects of the present-day state of the system. There appears to be much to be learnt from the study of such transitional states. Residual "redundant" features can actually survive for hundreds of years, going through avatars that present special interest for phonetics and phonology. The current, transitional state of the Tamang tone system is to be modelled without either looking back to a stable state in the reconstructed past of the system, or looking forward to a future stabilised state. Tamang tones are defined in terms of a bundle of cues, rather than of hierarchically organised features. It is clear that the conception of tone as pitch – a widespread conception, reflected in the International Phonetic Alphabet symbols for tone – is not adequate for Tamang tones. The present data serve as a reminder that it is necessary to take a "polydimensional approach to tonal investigation contra the prevalent monodimensional stance which ignores from the outset all parameters except F_0/pitch variation" [Rose 1982:48].

ACKNOWLEDGMENTS

Many thanks to Boyd Michailovsky for his help with the Tamang recordings; to Kiyoshi Honda and Shinji Maeda for indications about acoustics; to Cédric Gendrot for discussions about the interpretation of the electroglottographic signal; to Jean-Yves Dommergues and Barbara Kühnert for statistical advice; and to Jean-Michel Roynard for help and advice in the preparation of some of the figures. We gratefully acknowledge the comments offered by Jacqueline Vaissière, Klaus Kohler and two anonymous reviewers. Last but not least, many thanks to the Tamang language consultants. Any errors are our own. The electroglottograph used in the present study was kindly lent by the Laboratoire de Phonétique et Phonologie (CNRS/ Sorbonne Nouvelle).

REFERENCES

Abramson, A.S.: Laryngeal timing in consonant distinctions. Phonetica **34**: 295-303 (1977).

Anastaplo, S.; Karnell, M.P.: Synchronized videostroboscopic and electroglottographic examination of glottal opening. Journal of the Acoustical Society of America **83**(5): 1883-1890 (1988).

Baer, T.; Lofqvist, A.; McGarr, N.S.: Laryngeal vibrations: A comparison between high-speed filming and glottographic techniques. Journal of the Acoustical Society of America **73**(4): 1304-1308 (1983).

Blankenship, B.: The timing of nonmodal phonation in vowels. Journal of Phonetics **30**: 163-191 (2002).

Brunelle, M.: Tone perception in Northern and Southern Vietnamese. Journal of Phonetics (forthcoming).

Childers, D.G.; Naik, J.M.; Larar, J.N.; Krishnamurthy, A.K.; Moore, P.: Electroglottography, speech and ultra-high speed cinematography. Vocal Fold Physiology: Biomechanics, Acoustics and Phonatory Control: 202-220. I. R. Titze and R. C. Scherer. (Denver Center for the Performing Arts, Denver, Colorado 1983).

Duanmu, S.: An autosegmental analysis of tone in four Tibetan languages. Linguistics of the Tibeto-Burman Area **15**(1): 1-27 (1992).

Hagège, C.; Haudricourt, A.-G.: La phonologie panchronique (Presses Universitaires de France, Paris 1978).

Hale, A.; Pike, K., Eds. Tone systems of Tibeto-Burman languages of Nepal. Occasional Papers of the Wolfenden Society on Tibeto-Burman Linguistics 3. (Department of Linguistics, University of Illinois, Urbana, Ill. 1970).

Hale, A.; Watters, D., Eds. Clause, Sentence and Discourse Patterns in selected languages of Nepal. SIL publications in Linguistics and related fields: 40. (SIL, Norman, Oklahoma 1973).

Haudricourt, A.-G.: De l'origine des tons en vietnamien. Journal Asiatique **242**: 69-82 (1954).

Haudricourt, A.-G.: Bipartition et tripartition des systèmes de tons dans quelques langues d'Extrême-Orient. Bulletin de la Société de Linguistique de Paris **56**(1): 163-80 (1961).

Haudricourt, A.-G.: Two-way and three-way splitting of tonal systems in some Far Eastern languages (Translated by Christopher Court). Tai phonetics and phonology: 58-86. J. G. Harris and R. B. Noss. (Central Institute of English Language, Mahidol University, Bangkok 1972).

Haudricourt, A.-G.: Compte-rendu de M. Mazaudon, Phonologie tamang (Langues et Civilisations à Tradition Orale 4), Paris: SELAF, 1973. Bulletin de la Société de Linguistique de Paris **70**(2): 436-437 (1975).

Henrich, N.; d'Alessandro, C.; Castellengo, M.; Doval, B.: On the use of the derivative of electroglottographic signals for characterization of non-pathological voice phonation. Journal of the Acoustical Society of America **115**(3): 1321-1332 (2004).

Hess, M.H.; Ludwigs, M.: Strobophotoglottographic transillumination as a method for the analysis of vocal fold vibration patterns. Journal of Voice **14**(2): 255-71 (2000).

Hildebrandt, K.: Manange tone: scenarios of retention and loss in two communities (PhD, University of California at Santa Barbara, Santa Barbara 2003).

Karnell, M.P.: Synchronized videostroboscopy and electroglottography. Journal of Voice **3**(1): 68-75 (1989).

Kitzing, P.: Photo- and electroglottographical recording of the laryngeal vibratory pattern during different registers. Folia Phoniatrica **34**: 234-41 (1982).

Kjellin, O.: How to Explain the "Tones" in Tibetan. Annual Bulletin of the Research Institute of Logopedics and Phoniatrics, Faculty of Medicine, University of Tokyo **9**: 151-166 (1975).

Lisker, L.; Abramson, A.S.: A cross-language study of voicing in initial stops: Acoustical measurements. Word **20**: 384-422 (1964).

Löfqvist, A.: Acoustic and aerodynamic effects of interarticulator timing in voiceless consonants. Language and spech **35**: 15-28 (1992).

Maddieson, I.: Patterns of Sounds (Cambridge University Press, Cambridge 1984).

Maspéro, H.: Etude sur la phonétique historique de la langue annamite. Les initiales. Bulletin de l'Ecole Française d'Extrême-Orient **12**: 1-127 (1912).

Mazaudon, M.: Phonologie tamang: étude phonologique du dialecte tamang de Risiangku, langue tibéto-birmane du Népal (Société d'études linguistiques et anthropologiques de France, Paris 1973).

Mazaudon, M.: Consonantal mutation and tonal split in the Tamang subfamily of Tibeto-Burman. Kailash **6**(3): 157-79 (1978).

Mazaudon, M.: An outline of the historical phonology of the dialects of Nar-Phu (Nepal). Linguistics of the Tibeto-Burman Area **19**(1): 103-114 (1996).

Mazaudon, M.: On tone in Tamang and neighbouring languages: synchrony and diachrony. Proc. "Cross-linguistic studies of tonal phenomena": 79-96. S. Kaji. (Tokyo University of Foreign Studies, Institute for the Study of Languages and Cultures of Asia and Africa, Tokyo 2005).

Mazaudon, M.; Michailovsky, B.: La phonologie panchronique aujourd'hui: quelques repères. Combats pour les langues du monde: hommage à Claude Hagège: 351-362. J. Fernandez-Vest. (L'Harmattan, Paris 2007).

Michaud, A.: Final consonants and glottalization: new perspectives from Hanoi Vietnamese. Phonetica **61**(2-3): 119-146 (2004).

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

Michaud, A.; Mazaudon, M.: Pitch and voice quality characteristics of the lexical word-tones of Tamang, as compared with level tones (Naxi data) and pitch-plus-voice-quality tones (Vietnamese data). Proc. Speech Prosody 2006, Dresden: 823-826 (2006).

Michaud, A.; Vu-Ngoc, T.; Amelot, A.; Roubeau, B.: Nasal release, nasal finals and tonal contrasts in Hanoi Vietnamese: an aerodynamic experiment. Mon-Khmer Studies **36**: 121-137 (2006).

Noonan, M.: Nar-Phu. The Sino-Tibetan Languages: 336-352. G. Thurgood and R. LaPolla. (Routledge, London and New York 2003).

Rose, P.: Acoustic characteristics of the Shanghai-Zhenhai syllable-types. Papers in Southeast Asian Linguistics n°8: Tonation: 1-53. D. Bradley. (Australian National University Press, Pacific Linguistics, Series A, n°62, Canberra 1982).

Rothenberg, M.: A multichannel electroglottograph. Journal of Voice **6**(1): 36-43 (1992).

Rothenberg, M.; Mahshie, J.J.: Monitoring vocal fold abduction through vocal fold contact area. Journal of Speech and Hearing Research **31**: 338-51 (1988).

Roubeau, B.; Chevrie-Muller, C.; Arabia-Guidet, C.: Electroglottographic study of the changes of voice registers. Folia Phoniatrica **39**: 280-89 (1987).

Shafer, R.: Classification of the Sino-Tibetan languages. Word **11**: 94-111 (1955).

Silverman, D.: Phasing and recoverability (Garland Publishing, New-York and London 1997).

Sun, J.T.-S.: The Typology of Tone in Tibetan. Proc. "Chinese Languages and Linguistics IV: Typological studies of languages in China", Taipei, Taiwan, Symposium Series of the Institute of History and Philology-Academia Sinica: 485-521 (1997).

Sun, J.T.-S.: Variegated tonal developments in Tibetan. Language variation: papers on variation and change in the Sinosphere and the Indosphere in honour of James A. Matisoff: 35-51. D. Bradley, R. LaPolla, B. Michailovsky and G. Thurgood. (Pacific Linguistics, Canberra 2003).

Weidert, A.: Tibeto-Burman tonology (Benjamins, Amsterdam 1987).

Yip, M.: Tone in East Asian Languages. Handbook of Phonological Theory 476-494. J. Goldsmith. (Basil Blackwell, Oxford 1995).