



HAL
open science

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)

Alexis Michaud, Martine Mazaudon

► **To cite this version:**

Alexis Michaud, Martine Mazaudon. 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). 2006, pp. 819-822. hal-00134939

HAL Id: hal-00134939

<https://hal.science/hal-00134939>

Submitted on 5 Mar 2007

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Reference Type: Conference Proceedings

Author: Michaud, Alexis; Mazaudon, Martine

Year of Conference: 2006

Title: 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)

Conference Name: Speech Prosody 2006

Conference Location: Dresden

Pages: 823-826

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)

Alexis Michaud¹ and Martine Mazaudon²

¹Laboratoire de Phonétique et Phonologie (UMR 7018) CNRS/ Paris 3 Sorbonne Nouvelle

²LACITO, UMR 7107 CNRS/ Paris 3 & 4

alexis.michaud@univ-paris3.fr, martine.mazaudon@vjf.cnrs.fr

Abstract

The tones of Tamang (Sino-Tibetan family) involve both F_0 and voice quality characteristics: two of the four tones (tones 3 and 4) were reported to be breathy in studies from the 1970s. For the present research (thirty years later), audio and electroglottographic data were collected from 5 speakers of the Risiangku dialect in their 30s or 40s. Voice quality is estimated by computing the glottal open quotient. The present results bear on 788 syllables (from a corpus of 6,500). They show that in the speech of three speakers (M2, M3, M5), tones 3 and 4 have a higher open quotient (which provides an indirect cue to the degree of breathiness) than tones 1 and 2, with tone 3 more clearly so than tone 4, especially for speaker M2. The difference in open quotient between the four tones for the other two speakers is negligible or inconsistent.

The Tamang data are compared with similar data from Naxi, which possesses level tones, and from Vietnamese, which possesses pitch-plus-voice-quality tones. The comparison brings out the great variability of Tamang tones in terms of F_0 , as well as in terms of open quotient. The present results appear to confirm that Tamang tones possess several correlates; they offer an insight on ongoing change in the prosodic system of Tamang.

1. Introduction

The tones of the Bodic language group (Sino-Tibetan family), and of the Tamang language in particular, raise a challenge for theories of tone, and for prosodic typology: (i) their domain is the prosodic word, not the mora or the syllable; (ii) their phonetic realisation is highly variable—across speakers, and within the speech of the same speaker—, and involves several features [4, 5]. The present research focuses on the latter issue, and, more specifically, on F_0 and voice quality (the *mode of vibration* of the vocal folds).

Tamang possesses four tones, hereafter referred to as T1, T2, T3 and T4 (see figures 1, 4 and 5, on the last page). In studies from the 1970s, breathiness was described (in two Tamang dialects, Risiangku and Sahugaon, and in other languages of the Tamangish branch: Gurung and Thakali) as one of the features of T3 and T4, the tones that are lowest in terms of F_0 [4, 5 and references therein]. As already pointed out in [4], it appears useful to explore Tamang tones experimentally in order to gain new insights into these complex tones.

As nearly two generations have elapsed since the investigations mentioned above, the present experiment (conducted in 2005) also offers an opportunity to reflect on language change. Variability in the realisation of breathy voice quality is well documented across languages [1], both

diachronically (breathy voice quality frequently transphonologises to tonal contrasts or vowel contrasts) and synchronically (cross-dialect and cross-speaker variation in the realisation of the feature of breathiness is known to be considerable, as compared to other phonological features).

Data from two other languages (also recorded first-hand, using a similar experimental setup) are used as points of reference in the interpretation of the results: these languages are Naxi, which possesses level tones [9], and Vietnamese, which possesses pitch-plus-voice-quality tones [7, 8].

2. Method

The investigation relies on electroglottography (EGG), combined with audio recordings and spectrographic analysis.

2.1. Corpus and speakers

The 136 target morphemes are monosyllabic. They are divided into two sets: on the one hand 49 nouns and grammatical words (namely the particles ⁴ca/ ‘as for’ and ⁴ce/ ‘only’; the deictic ²cu/; and the 3rd person pronoun /¹t^he/), on the other hand 87 verbs, including stative verbs (e.g. ‘[to be] pretty’).

The nouns and grammatical words were placed inside the following carrier sentence:

²cu-ri ____ . “Here [is] ____.”
here target noun

e.g. ²cu-ri ²c^hi “Here [is] grease” (tone is indicated as a superscript figure before the first syllable of the phonological word that carries it: ²c^hi/, ‘grease’, carries T2).

The verbs were placed inside another carrier sentence:

²cu-ri ____ -pa.
here target verb infinitive suffix

e.g. ²cu-ri ³su-pa “Here, [he is] planting” (³su-pa/ means ‘to plant’).

Tamang morphemes have a complex syllabic structure; the present investigation is limited to CV syllables made up of a nonaspirated stop /p, t, k, c, s/ followed by a nondiphthongised vowel /a, a, i, i, u, u, o, o, e, e/. Aspirated stops (which in Tamang are only associated with T1 and T2 syllables) were not included because they may increase the airflow on the following vowel: to take an example, if the voice quality of the syllable rhyme of /¹t^ha/ is breathier than that of /¹ta/, it will thus be less distinct from that of ³ta/, detracting from the accuracy of the comparison of voice quality across tones.

The use of nonsense words did not appear as a valid option, because Tamang does not have a written tradition. The words used were selected from a Tamang dictionary

(currently being finalised by the second author). The distribution of tones in the Tamang lexicon is uneven (T1: 38.5%, T2: 26%, T3: 21%, T4: 13.5% [4, page 84]). ‘Minimal quadruplets’ (such as: ¹ku:pa ‘to wear [a hat]’, ²ku:pa ‘to train [oxen]’, ³ku:pa ‘to draw towards oneself’, ⁴ku:pa ‘to lie in ambush’) are exceptional in the language, so that a phonemically balanced corpus could not be constructed.

Five male native speakers took part in the audio and electroglottographic recordings, conducted in Nepal. The speakers were born in the same village (Risiangku); M1 and M2 are brothers, as are M3 and M4; all five are in their 30s or 40s. These speakers were selected with a view to controlling for the following variables: gender, age, and dialect, thus ensuring a high degree of homogeneity of the speaker group.

The Nepali translation of the target words was used as a prompt (Nepali, the national language of Nepal, is spoken by the Tamang as a second language). The speakers were instructed to say the word inside a carrier sentence, which was repeated twice. The EGG signal (from an EG2-PC two-channel electroglottograph [11]) was recorded simultaneously with the audio, using a laptop computer with a high-fidelity external sound card.

Due to fieldwork conditions, not all items were recorded by all informants; conversely, some parts of the corpus were recorded twice by some speakers. The recordings were annotated manually, labelling the beginning and endpoint of each phone (target words and carrier sentence) on the basis of inspection of the audio signal supplemented by auditory impression. The items which appeared not to have been recorded correctly (e.g. because of an hesitation on the part of the informant, or conversely because of a strongly emphatic rendition of the word) were discarded. The data reported here are a subset of 788 syllables. They are compared with data from 5 Naxi speakers (1200 syllables) and 4 Vietnamese speakers (1008 syllables). The Naxi data [9] and Vietnamese data [7, 8] were recorded with the same equipment as the Tamang data, under a similar experimental setup.

2.2. Measurements

Numerous methods have been proposed to evaluate voice quality from the audio, but they have inherent limitations [12]. The present investigation relies on electroglottography: vocal fold contact area is monitored by means of a pair of electrodes placed on either side of the larynx. The analysis method chosen rests on the analysis of the derivative of the EGG signal, obtained by computing the difference between successive samples [on the interpretation of this derivative signal, see 2 and references therein]. This method allows for the calculation of the fundamental frequency as well as of the open quotient (hereafter O_q), which reflects the degree of vocal fold adduction [10]. This measurement is especially effective to study glottalisation [7]; it provides a less direct indication in the case of breathy voice: glottal configuration is only part of the story, as breathiness—as its name indicates—is a matter of airflow, which depends on subglottal pressure and vocal tract impedance as well as on glottal configuration.

In breathy voice there is no ‘closed phase’ strictly speaking inside each glottal cycle, as some air leaks out constantly. Despite the fact that the airflow baseline is not at zero in breathy voice, each glottal cycle can still be divided into a (so-called) ‘closed’ and an ‘open’ phase.

The detection of the glottal opening and closing peaks was conducted using software developed in the MATLAB

computing environment. For speaker M4, double opening peaks are consistently observed (this phenomenon is not uncommon: for a discussion see [2, page 1324-1327, 3]); O_q was nonetheless calculated, using a barycentre method (about this and other methods, see <http://voiceresearch.free.fr/egg/>, where our software is available for download, with some documentation).

3. Results

3.1. Comparison of the degree of variability with that observed in Naxi and Vietnamese

Figure 1 shows average F_0 and O_q curves for the four tones as realised on monosyllables by speaker M3. (All the rhymes represented come from the first repetition of the carrier sentence, in order to control for the effect of the position of the utterance within the series of two.) The error bars show a considerable standard deviation for F_0 and O_q : the ranges of variation of the four tones overlap to some extent. This is especially clear by comparison with Naxi data (figure 2): Naxi is a textbook example of level tones (High, Mid and Low, plus a rising, Low-to-High tone), that are not specified in terms of voice quality (a fact which is straightforwardly reflected in the higher variation coefficient for O_q than for F_0). Three Vietnamese tones are plotted in figure 3: they are much less variable than Tamang tones, both in terms of F_0 and of O_q .

For the sake of visual clarity, the curves in figures 4 and 5 are plotted with only **half their actual standard deviation**, for F_0 and for O_q .

3.2. Description of the four tones

Tones 1, 2, 4 (from highest to lowest) are decreasing over the syllable. T3 is rising. As for O_q , M5 has a comparable mean O_q on T3 and T4, higher than on T1 and T2; M2 has a clearly higher O_q on T3, with T4 also slightly above T1 and T2. The range is markedly different across speakers: overall, M5 has a much higher mean O_q than M2, i.e. a more breathy voice; this confirms that tone-specific variation in voice quality is relative, not absolute. M3 exhibits the same O_q pattern as M2; the differences are of smaller amplitude, however.

M1 and M4 do not show any variation in O_q across tones. Correlatively, M1 has less variability in F_0 curves than other speakers, but this is not true of M4.

From figure 1, it appears that differences in O_q across tones tend to be strongest on the first half of the syllable; from figure 4, it further appears that this difference (like the difference in F_0) is much stronger on the first than on the second syllable of disyllables (recall that, from a phonological point of view, the domain of tone in Tamang is the phonological word, which may be mono- or polysyllabic). This explains in part why the O_q values shown in table 1 and in graph 1, which are averaged over entire syllables, do not bring out a clear contrast between the four tones: the *time course* of O_q curves appears no less important than that of F_0 curves. (The values in table 1 and graph 1 do not include the second syllable of disyllables, in view of the fact that O_q differences on the second syllable are generally much smaller than on the first.)

Table 1. Mean O_q values for the four tones by the five speakers. In brackets: standard deviation. Number of syllables: 64 for M1, 220 for M2, 178 for M3, 120 for M4, 206 for M5.

	T1	T2	T3	T4
M1	55.0 (4.91)	54.5 (5.07)	57.4 (4.45)	56.8 (5.96)
M2	51.3 (6.53)	49.2 (7.47)	60.8 (8.09)	55.9 (7.04)
M3	54.0 (5.85)	54.5 (5.61)	59.3 (6.37)	57.1 (6.85)
M4	58.8 (6.72)	55.7 (5.39)	60.1 (6.10)	56.9 (7.07)
M5	68.1 (5.33)	67.6 (4.56)	70.5 (5.70)	70.4 (4.81)

A statistical analysis of the data is currently under way, in an effort to assess the extent to which the following variables contribute to the observed variance in F_0 and O_q : speaker; tone; segments (especially vowel length); and F_0 of the preceding syllables (i.e., realisation of the carrier sentence).

4. Discussion

4.1. Cues to tones rather than tonal features?

Speaker-specific variability is observed in all three languages. In the face of such variability, it seemed advisable to look for *patterns of least variability*, e.g. in Naxi there is some variability in F_0 within each tonal category, but F_0 varies less than O_q (reflecting the fact that voice quality plays no distinctive role in Naxi); in Vietnamese, O_q varies less than F_0 , reflecting a strict control on voice quality, which plays a distinctive role. The comparison brings out the great variability of Tamang tones in terms of F_0 , as well as in terms of O_q . The present results appear to confirm that Tamang tones are highly variable in terms of pitch, and possess several correlates. They also suggest differences between the speech of the language consultants of this study and those who participated in earlier studies [4, 5], some two generations ago. This tends to confirm the transitory nature of breathiness as a linguistic trait (typically, in the Sino-Tibetan domain, an intermediate stage in the transphonologisation of consonantal contrasts to pitch contrasts).

Concerning the contrast between T1 and T2, which has been described as ‘ballistic’ (T1) vs. ‘controlled’ (T2), T1 indeed tends to be shorter than T2, and T2 tends to have a slightly more level curve, though here again, the results are far from transparent.

This raises issues in the description of Tamang tones in terms of *features*. No obvious hierarchy among features emerges for speakers M2, M3 and M5, in whose realisations both of the parameters measured here (F_0 and O_q) show consistent patterns of tone-specific variation. Speakers M1 and M4 apparently rely on pitch, not on voice quality. There appears to be multiple cues, rather than a hierarchy of *features*, the one primary, the others secondary.

4.2. On the hypothesis of relations of compensation

Synchronically, it has been hypothesised [6] that a *relation of compensation* may obtain between pitch and breathiness for Risiangku Tamang T4, such that when a token is pronounced at the bottom of T4’s F_0 range, breathiness is often absent. While it is marginally verified that M1, who uses no breathiness cues, has the largest separation of tones in terms of F_0 (i.e. has least F_0 variability), no relation of compensation between F_0 and O_q emerges from the present data: the T3 and T4 tokens with lowest O_q (least breathiness) are not those with

lowest F_0 . If the relation of compensation at issue exists in the language, it did not come into play under the experimental setup chosen—or perhaps the cues to ‘breathiness’ partly lie in parameters other than O_q ; this leads to the issue of the enlargement of the scope of research, addressed in the ‘Perspectives’ below.

5. Conclusion and perspectives

Within a fairly homogeneous group of Tamang speakers, individual strategies in the realisation of tone differ sharply, two of the five informants appearing not to make consistent use of voice quality differences to contrast tones. From a synchronic point of view, these data appear to call for *speaker-dependent modelling*, and for modelling in terms of a *bundle of multiple acoustic correlates*: F_0 , length, voice quality, and allophonic variation of consonants (not studied here). These results confirm and complement description by ear: (i) T3 and T4 show clear hints of breathiness (in the speech of three speakers), more clearly so for the former than for the latter; and (ii) the cross-linguistic evidence adduced shows that the degree of variability observed (in F_0 and O_q) is higher than would be expected on the basis of speaker-specific and phoneme-conditioned variability alone, i.e. this variability is a characteristic of Tamang tones.

From a diachronic point of view, it might be that Tamang is evolving towards an all-pitch tone system; neighbouring languages certainly have a bearing on this evolution, however: the influence of Nepali, the (non-tonal) national language of Nepal, is towards a simplification of the tone system.

With a view to contributing to the study of language change, and more specifically of the evolution of tonal systems, our ongoing research is now directed to the issue of the other cues to tone, in particular the variation in the realisation of syllable-initial consonants in relation to the phonemic tones. Lastly, further experiments are called for to verify the hypothesis of a *relation of compensation among cues to tones*.

6. Hyperlink

Supplementary materials are available at:

<http://ed268.univ-paris3.fr/lpp/pages/EQUIPE/michaud/TAMANG/>

7. References

- [1] Abramson, A.S.; Thongkum, T.L.; et al., 2004. Voice register in Suai (Kuai): An analysis of perceptual and acoustic data. *Phonetica*, 61(2-3), 147-171.
- [2] Henrich, N.; d’Alessandro, C.; et al., 2004. 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.
- [3] Henrich, N.; Gendrot, C.; et al., 2004. Characterization of features observed on the derivative of EGG signal by the use of high speed cinematography. *International Conference on Voice Physiology and Biomechanics*. 2004. Marseille, France.
- [4] Mazaudon, M., 1973. *Phonologie tamang: étude phonologique du dialecte de Risiangkul*. Paris: SELAF.

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

[6] Mazaudon, M., in press. On tone of Tamang and neighbouring languages. *Proceedings of the symposium "Cross-linguistic studies of tonal phenomena"*. Tokyo: Tokyo University of Foreign Studies.

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

[8] Michaud, A.; Vu-Ngoc T., 2004. Glottalized and Nonglottalized Tones under Emphasis. *Speech Prosody 2004*. Nara, Japan, 745-748.

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

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

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

[12] Shrivastav, R.; Sapienza, C.M., 2003. Objective measures of breathy voice quality obtained using an auditory model. *Journal of the Acoustical Society of America*, 114(4), 2217-2224.

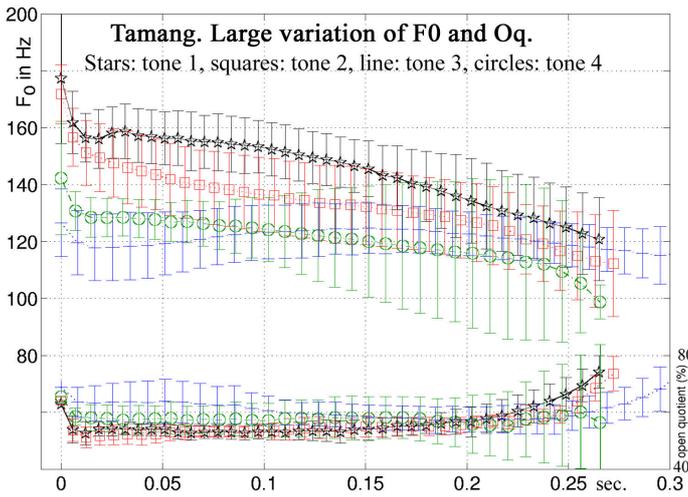


Fig. 1. The 4 tones on monosyllables by speaker M3, initial position, averaged over 36 syllables, nonaspirated initials.

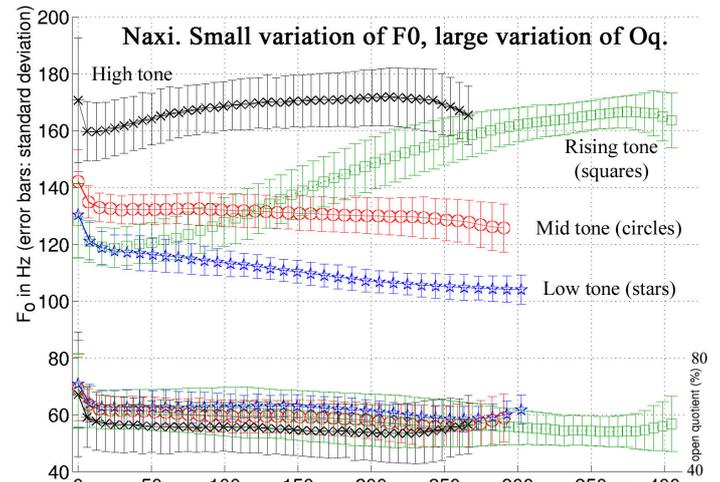


Fig. 2. The four tones of Naxi. Averaged over 160 syll. in carrier sentence. 1 male speaker.

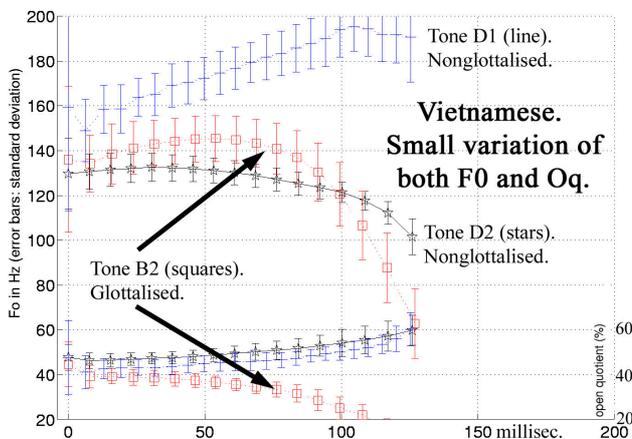


Fig. 3. Tones B2, D1 and D2 of Hanoi Vietnamese. One male speaker. 126 syll.

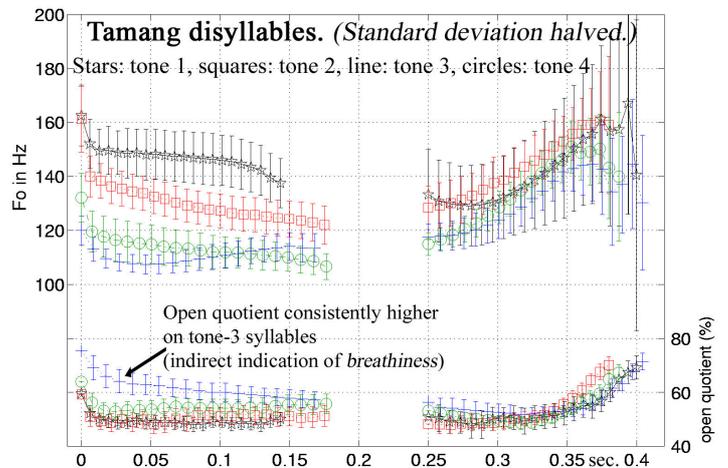
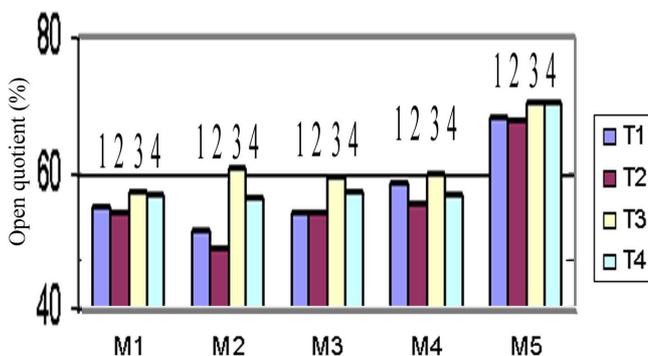


Figure 4. The 4 tones on disyllables by speaker M2, initial position, averaged over 132 syllables, nonaspirated initials.



Graph 1. Open quotient values for the four Tamang tones, averaged over 3,361 syllables with nonaspirated initials. These mean values do not bring out a clear contrast between the four tones, suggesting the necessity of studying the time course of open quotient values.

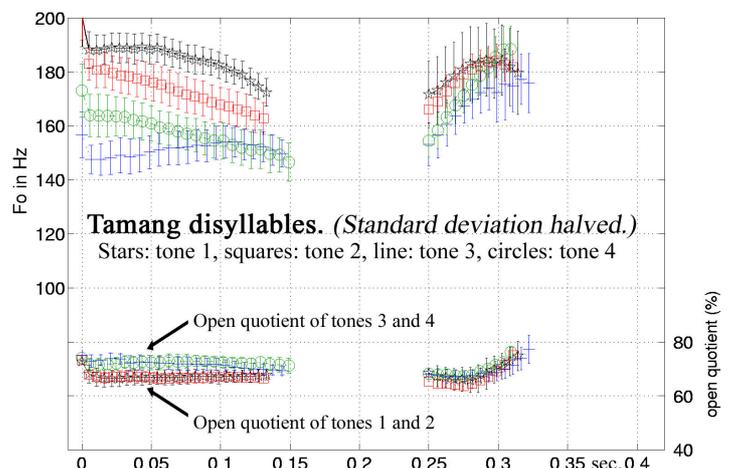


Fig. 5. The 4 tones on disyllables by speaker M5 initial position, averaged over 134 syllables, nonaspirated initials.