Manual and automatic tone annotation: the case of an endangered language from North Vietnam "Mo Piu"

Geneviève Caelen-Haumont[1], *Katarina Bartkova*[2]

[1] MICA Institute, HUST – CNRS/UMI2954 - Grenoble INP, Hanoi University of Sciences and Technology, Hanoi, Vietnam
[2]ATILF-UMR 7118, Université de Lorraine, France
genevieve.caelen@mica.edu.vn, katarina.bartkova@atilf.fr

Abstract

The goal of our study is twofold. First, the results of two approaches in a tone annotation task are presented: a manual approach provided by the system MISTRAL+ and an automatic approach provided by the system PROSOTRAN. Second, an attempt is made here to determine the number and the patterns of the Mo Piu language tones with respect to the pitch slope and the tone level. It is found that both approaches of the tone annotation yield similar results on the speech material studied: tone patterns detected by the two systems are mainly falling (68% in the manual detection and 55% in the automatic detection) and flat tones (30% in manual detection and 43% in automatic detection) while rising tones are detected very seldom by the two systems (only 2% of the patterns are considered as rising ones). The agreement between the two systems is quite high: 70% of the detected tones have the same pitch movement (rising or flat) and only 4% of the tones detected by the two systems are completely different (neither the slope direction nor the tone levels are identical). The main tone patterns detected in our study are /41, 43, 54/ as falling slopes and /33/ as a plateau.

Index Terms: tone pattern annotation, tonal system, Mo Piu language

1. Introduction

The present study is part of the "Au Co" Project started in 2008 in MICA Institute, aiming to help saving endangered languages in Vietnam. It is based on the collaboration of a wide variety of experts, French and Vietnamese specialists of ethnic languages and computer scientists.

The Au Co project allowed to study an ethnic minority called *Mo Piu*. This minority is located in the mountains of North Vietnam not far from the Chinese border.
According to our previous studies [1], the Mo Piu language is an endangered language. Indeed it is unwritten and spoken only by 237 people in 2011, and until our study, this language was uncharted and undocumented.

This paper is the continuation of our previous studies [1, 2], and it is concerned with the tonal domain of the Mo Piu language. In fact the goal of this study is twofold: 1° to compare 2 methods of tonal descriptions, i.e. a manual/semi-automatic and an entirely automatic one, and 2° to define more precisely the number and the patterns of the Mo Piu tones. For this second goal, a prototone approach is used. These prototones are scientifically built from the observed tones of the same family of languages, -here the Hmong-Mien one-, and represent a common heritage before the evolution of each language. Based on these findings, a list of 8 prototones is elaborated [3], supplying 8 supposed different prototones thus leading possibly to specific Mo Piu tones.

Although, many works have been published in the domain of the Hmong tonal system [4, 5, 6, 7, 8, 9, 10, 11], however, among the different Hmong languages, the White Hmong is the most studied. The Hmong languages supply a great diversity and studies report that tone numbers may vary from 3 to 11. Therefore great uncertainties remain about the Mo Piu tonal system.

In the following sections of our study we give first a description of the semi-automatic and automatic approaches used, then the number and the patterns of the Mo Piu tones are discussed, and finally the study focuses on the core of the prototone list, enabling to extract the best information concerning tone levels, their patterns and durations. In the context of this study, neither the voice quality of the tones, nor the tone sandhi (contextual tonal influence) are taken into account.

2. Description of the corpus and data

In order to collect Mo Piu linguistic data, four field trips were undertaken from 2009 to 2012. Our sound and video corpora recorded during these trips supplies about 45 hours of video and speech material (French/Vietnamese/Mo Piu), composed of continuous speech and lists of isolated words (for further details, see [2]). The present study uses specific recordings made in 2012, and is based on a list of prototones [3]. This list just encompasses 8 lists of words without any reconstruction of tonal patterns. Our two best Mo Piu speakers recorded this prototone list in the MICA Institute recording room. The aims of using such a prototone list based originally on the Calmsea list, are: 1° to better detect the number and patterns of the Mo Piu tonal system, 2° to contribute to the knowledge of the Hmong-Mien family languages. In this pilot study on tone annotation and analyses, only data recorded by one of our speakers are studied.

While recording the prototone list, some items appeared to the speakers to be semantically linked to others and were therefore recorded as well (31 ones). The prototone corpus of the present study is made of 206 (often compound) words repeated 3 times by 1 speaker and labeled by the two systems used in the present study. Thus our data are composed of 3868 tokens (consonants, vowels, pauses) issued from the Praat TextGrids, 1752 tonal sequences (one or more per lexical item), corresponding to the 818 vowels (or tones) of the prototone list including new words added by our speakers. In these data, tones are annotated using our automatic and manual tone annotation systems. Finally aiming at checking the prototone list against the Mo Piu data, our analysis is restricted (see further details in § 4.2. below) to the strict prototone list, i.e. without the extra words added by our speakers. From these items, as explained below, 525 vowels are finally extracted and their tone patterns thoroughly studied.

3. Comparison of the automatic and manual tools

Our main concern in this study is to describe the most reliably possible the Mo Piu tonal system. To meet this aim, the performances of two tools of tone annotation (PROSOTRAN and MISTRAL+) are evaluated and compared. The two systems use the same manual segmentation provided by an expert phonetician.

3.1. A semi-automatic approach: MISTRAL+

MISTRAL+ [12] is a tool running under Praat enabling to compute the speaker's tonal range (Hz or semi-tones, 10 ms sampling) in a number of levels defined by the user (5 levels are generally required for tones studies). Moreover MISTRAL+ allows also 1° under the menu "Manipulations task", to synchronize boundary labels with the melodic alignment (i.e. one melodic segment for one phonetic item), 2° to compute all the data issuing from the different tiers and labels: segment tones at phonetic and word levels, duration at different levels, derivative… 3° to create an xls file with all the manual and computed data, and the IPA codings. In its new version, MISTRAL+ is operating a pre-alignment of the melodic curve at the boundaries chosen by the user. But at this step, the automatic stylization of the melodic curve in MISTRAL+ remains raw, thus the user has to operate a final alignment with a level of stylization of his choice. This part of the manipulation is manually made, matching at the best the sound with its F0 line. The tonal forms are evaluated on a 5 level scale (for instance /43/, /32/, /44/…), and an important task remains to detect, for a tonal segment, its tone prototype pattern and its variants. Thus another task is to check all the deviant patterns in order to verify whether deviant tonal values are on the edge of a tone threshold (in most cases, it is the cause of variations). These deviant values are then easily corrected.

3.2. Automatic approach: PROSOTRAN

Prosotran [13] is a system enabling to annotate tone patterns automatically. At the acoustico-phonetic level, it uses F_0values in semi-tones calculated from the speech signal every 10 ms by the acoustic analysis Aurora [14]. The input of the system requires also the segmentation of the speech signal, which in this study is carried out by hand (by an expert phonetician).

Prosotran uses 5 tone levels for the tone annotation. In order to define tone levels, all the speech material of a speaker is used to build up a histogram of the distribution of the F0 values. Once the extreme (seldom) F0 values are discarded, the remaining pitch range is then divided into 5 tone levels.

Regarding the tone detection, the system provides three different types of information for each vowel: the level and the orientation of the pitch movement (rising, falling, etc.) and the importance of the slope. The tone level is calculated per vowel on 3 distinct points that correspond to the F0 values measured at the beginning, at the end of the vowel and to a third value which is considered as the most pertinent value reflecting the pitch movement between the two extreme (first and last) values. The third pitch point corresponds thus to the turning point of the pitch slope and is obtained by determining two consecutive segments that approximate the best a sequence of pitch values (see Figure 1). The automatic approach indicates also the pertinence of the slope change: whether it is perceptible or not. Therefore an additional label is related to the steepness of the F_0 slope itself. The F_0 slope value is compared to the glissando threshold ($0.16/T^2$) [15, 16] and is annotated symbolically between $Vowel_{Slope}$++++ (very strong slope, rising or falling) and $Vowel_{Slope}$- (flat). This symbolic annotation is useful in adjusting the tone level detection: when the F0 slope is steep (higher than the glissando threshold) and the tone levels detected for the different F0 values are the same, then the tone level detection is post-processed in order to obtain an annotation that would reflect the slope movement.

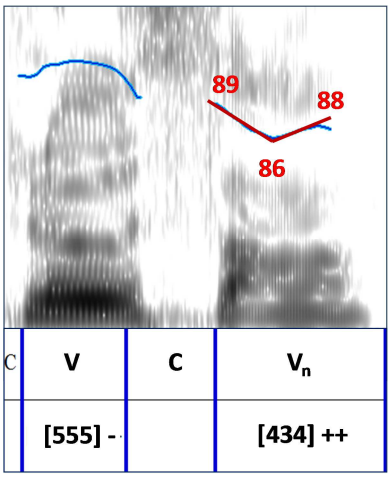

Figure 1: *Prosotran annotation: F0 turning value detection (pitch in thin (blue) line and regression lines in (thicker) red). Automatic slope annotation (/555/ and /434/) in 3 points.*

4. Results

The tone detection is obtained systematically on three points per vowel by the automatic system, however the number of pertinent tone points is variable in our manual detection (ranging from 2 to 4 points), therefore the tone pattern comparison between the two system performances is made using only two tone points, at tones boundaries (first and the last point).

4.1. Occurrences of tone patterns

4.1.1. Tone detection yielded by MISTRAL+

The results are presented from a general up to the most detailed perspective. This part of the study concerns all the data (660 vowels, from the strict prototone list, including simple and compound words as well).

Confirmed also by the PROSOTRAN analysis (see below § 4.1.2), the rising tone patterns are very seldom used. In the manual task described in paragraph 4.2, these patterns are thoroughly analyzed.

In our previous study [2] based on 44 compound words in 3 repetitions made by 2 speakers, the tone patterns /32,33,43,44/ are detected the most frequently. Hopefully, our previous findings will be completed and confirmed by this new study of a new speech material containing a new speaker, a new selection of words, more items, and a more detailed manual procedure. Figure 2 presents the results issued from our new data.

Although the tone patterns /32,33,43,44/ are still detected in our new data, however there are major differences as far as their occurrences are concerned. But as the number of words is

different between the 2 corpora (only ¼ of the words is common between them), these word differences can partly account for the differences of the tone patterns. The patterns /32/ and /33/ are the most frequent in our first study, although in the present one, these patterns are seldom, while the pattern /43/ is far the most frequent

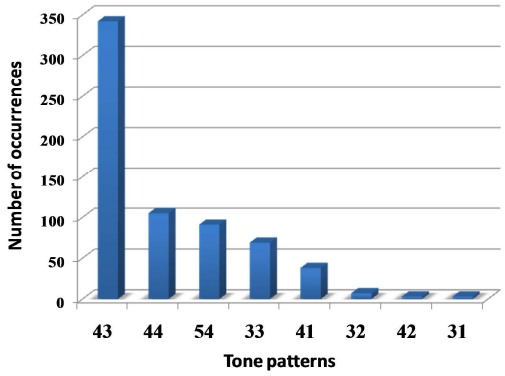

Figure 2: *Occurrences of tone patterns annotated by MISTRAL+*

4.1.2. Tone detection yielded by PROSOTRAN

The most frequently detected tone patterns by our automatic approach are the falling ones /43,54,53/ and the plateaux /55, 33, 44/. Rising tones are very seldom detected in the speech material by the PROSOTRAN system.

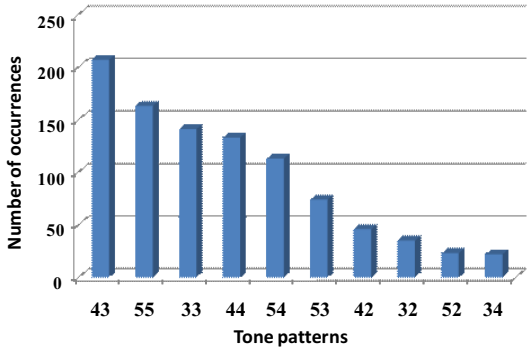

Figure 3: *Occurrences of tone patterns annotated by PROSOTRAN*

No clear relation is observed in our data between context and tone pattern: falling and flat tones occur indifferently in consonant and pause contexts (which are the main vowel contexts in our data).

The PROSOTRAN analysis comprises the whole data of the prototones and that accounts for the differences between the number of vowels annotated by the two systems and can explain part of the existing differences between the two annotations.

4.1.3. MISTRAL & PROSOTRAN: tone detection comparison

The two tone detection systems work in quite a similar way: they detect mostly falling and flat tones while rising tones are very seldom detected by either of them (only 2% of rising tones are detected by the two systems). As far as rising and falling tones are concerned, less flat tones (30%) and more falling tones (68%) are detected by our manual approach than by the automatic system used (43% flat tone patterns and 55% falling tone patterns). Moreover, the two systems agree in the most frequently detected tone pattern which is the pattern /43/ for both detections. Agreement between the two approaches when 3 points (F0 value at the beginning, at the central pivot point and at the end of the vowel) of the tone detection are considered is about 60%. But when only the first and the last points are considered (the middle point is left out from the comparison), the agreement reaches 70 % between the systems. Furthermore for 26% of the tone patterns, there is at least one tone level shared by the two (manual and automatic) annotations and only 4 % of the detected tone patterns are mismatching between the two systems: neither the slope direction nor the tone levels are identical.

4.2. Prototones and tones

In the following paragraphs the prototone list, reduced to its core, is studied using an analysis yielded by MISTRAL+. The speech material used here contains (after that extra words are discarded) 660 words/vowels/tones. Our goal is to put to light a structure, if any, for each of the 8 prototones. 175 lexical items are used here uttered in 3 repetitions, yielding 525 items of single and compound-reduced words (only one item from the compound words is selected). Moreover in case of a discrepancy among the 3 repetitions of a same lexical item, F0 values are checked in order to correct a possible threshold effect: ± 5 Hz around the threshold are considered as tolerable.

The use of some F0 values can be considered as a speaker effect: before tuning the prototypical target, the speaker can undershoot the F0 values, or he can produce tone pattern modulations provoking a change of the tone level (at most ± 1 level) deviating thus from the main tone pattern (falling slope or plateau). Furthermore a clear fact seems to emerge from the data: no matter what the tone middle value is, the last portion of the tone seems to carry ultimately its discriminative function. For instance, the tones /54, 44, 43/ generate respectively the modulations /545*54*, 45*44*, 43*43*/. The few rising tones annotated in our data are thus also due to the speaker effect discussed here above and they do not carry a discriminative function. Therefore, all the annotated tones can be thus considered as either falling or flat (plateau).

Nb	P	43	54	44	33	41	32
141	**1**	19%	***65%***	12%	2%	0%	2%
75	**2**	***59%***	8%	9%	8%	16%	0%
93	**3**	***84%***	0%	10%	3%	3%	0%
45	**4**	**33%**	0%	0%	***60%***	7%	0%
93	**5**	***84%***	0%	6%	3%	6%	0%
24	**6**	***50%***	0%	0%	0%	***50%***	0%
33	**7**	***91%***	0%	9%	0%	0%	0%
21	**8**	*71%*	14%	14%	0%	0%	0%
525		**299**	**100**	**45**	**42**	**36**	**3**

Table 1. *Occurrences (in absolute (Nb) and relative (%) values) of lexical items/vowels/tones per prototone type (P)*

As it appears from Table 1, lexical items are unevenly distributed through the different prototones (see column Nb), especially through the prototones 6, 7 and 8.

4.2.1. *Distribution of the tones across the prototones*

Whatever the number of examples of each prototone can be, a stable trait is striking in Table 1: the tone /43/ accounts for more than the half of the data occurrences (299/525): in fact, this pattern is present in all the prototones, and ranks as the first pattern in the prototones 2, 3, 5, (6), 7, 8. Of course, the prototone list used is not a Mo Piu tone list. From these limited data restricted to only one speaker and containing an unbalanced number of prototones, we can only formulate some new hypotheses: 1° some prototones (2, 3, 5, 7, 8) seem to have merged into the tone /43/; 2° the prototone 6 is distributed equally into tones /43,41/ but its data are not numerous enough to draw valid conclusions; 3° though one third of the tones belongs to tone /43/, the prototone 4 seems to converge towards the plateau /33/, but again the lack of data prevents further conclusions; 4° the prototone 1 seems to have mostly converged towards the tone /54/. From this perspective, the tones /44, 32/ seem to be overshot or undershot targets of the other tones. However, it is still possible that the tone patterns /44, 32/ exist in Mo Piu words not included in our list, and also that other tones exist in the Mo Piu language than those presented in the prototone list.

4.2.2. *Tonal patterns*

Percentages have been calculated for the tones presenting the biggest population (/43,54,41,33/) among the prototones, which can be considered as Mo Piu tone candidates. Each tone supplies phonetically (i.e. *tonetically*) modulations which are not significant at the meaning level but are so at the physiological, expressive, emotional or attitudinal level. The task consists then in extracting the right pattern of the tone. In other words, one has to decide whether a pattern such as /433/ is a modulation of the tone /43/ or the true prototype pattern /433/?

Tones	1 direction	2 directions	> 2 directions	Ranked 1st
54	**4%**	***75%***	**21%**	**544**
43	**15%**	***53%***	**32%**	**433**
41	***50%***	**28%**	**22%**	**41**
33	***100%***	**0%**	**0%**	**33**

Table 2. *Tones occurrences according to the number of their orientations (slopes or plateaux), and preferred patterns*

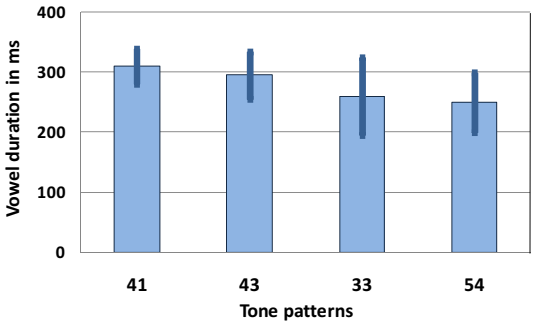

Figure 4: *Mean durations of the main tones (525 vowels / tones) and their standard-deviation*

The answer to this question can be drawn from two embedded cues: the difference of occurrences between tones with 1, 2 or more than 2 orientations (falling slopes or plateaux), and the most frequently observed pattern.

According to these cues, the tones /54,43/ are the most numerous among the two direction tones (their main patterns are respectively /544/ and /433/), while the tones /41,33/ present one direction. The other patterns (same number of slopes or more) are thus considered as simple variants of these main patterns. On the other hand, the "one orientation" variants for tones /54,43/ can be considered on the basis of their weak population, as carrying slighter modulations than the perceptual threshold (during the tonal alignment in the Praat Manipulations step), and therefore not taken into account.

4.2.3. *Duration*

The mean vowel/tone duration is 293 ms with a standard-deviation of 72 ms. Figure 4 above presents the durations of the 525 vowels/tones and their standard deviations. There is no clear tendency enabling to make further oppositions between the tones or their orientation (falling slopes, plateaux) and the measured vowel durations.

5. Conclusion

The goal of our study is not only to give a description of tone patterns of the Mo Piu language but also to evaluate the detection and the annotation of the tone patterns. In order to obtain a reliable annotation of the tone patterns, this task is achieved here by two approaches, by a semi-automatic one allowing manual adjustments (MISTRAL+) and an automatic one (PROSOTRAN).

In the tone pattern detection, both approaches use F0 values in semi-tones and a manual segmentation of the speech signal made by an expert phonetician. The two systems yield similar results. They detected with the highest occurrence the same falling pattern /43/ and the overall agreement between the two systems as far as the F0 slope direction is concerned, is about 70%. The main difference between the two systems is in a higher ratio of plateau pattern detection provided by the automatic system, probably due to the differences in the pitch range calculation that influences the limits of the tone levels.

Although our data are limited, the main results concerning tone levels confirm findings of previous studies [1, 2]. Our study allow to reduce the number of different tone patterns and to give more precision about them, putting to light a bidirectional pattern for falling tones /54/ (/544/) and /43/ (/433/), and a one direction pattern for /41/ (and of course for the plateau /33/).

The adopted perspective can be considered as a step toward tone phonology, supplying a pre-phonological scope about the Mo Piu tonal system.

Concerning the relation between Mo Piu tones and the 8 prototones, in the limits of our data, it is only possible to put forward that the 8 prototones do not lead to 8 Mo Piu tones. A tone restructuring is made in Mo Piu under the pressure of history, internal reorganizationand loan words. Nowadays in Mo Piu, the main tone patterns are falling slopes and plateaux. Some prototones 2, 3, 5, 8 seem to have merged into the same tone /43/, while the prototone 1 seems to mostly correspond to the tone /54/. As for the other prototones, other data are needed to confirm whether 6 is split up in two tones /43, 41/, and whether 4 leads to the plateau /33/.

6. References

[1] Caelen-Haumont, G., Cortial B., Culas C., Hong T.D., Lê Thi .X., Nguyen T.N., Pannier E., Salmon J.-P., Vittrant A., Vuong H.T., Song L.A., "Mo Piu minority language: data base, first steps and first experiments", Second International Workshop on Spoken Languages Proc., Technologies for under-resourced Languages, SLTU'10, Penang, Malaysia, 42-50, 2010.

[2] Caelen-Haumont, G., "Towards the tonal system of an unknown language from south-east Asia: a deeper insight", the Tonal Aspects of Languages Proc., TAL 2012, Nanjing, China, Section 01, 60, 1-5, 2012.

[3] Ratliff, M., "Hmong–Mien Language history", Canberra, Australia: Pacific Linguistics, 2010.

[4] Haudricourt, A.G., "Introduction à la phonologie historique des langues miao-yao", Bulletin de l'Ecole française d'Extrême-Orient, 44, 2, 555-576, 1951.

[5] Downer G.B., "Tone-change and Tone-Shift in White Lao", Bulletin of the School of Oriental and African Studies, 30/3, 589-599, 1990.

[6] Jarkey, N., "Vowel Phonemes in White Hmong", Ms, 1985.

[7] Ratliff, M., "An Analysis of Some Tonally-Differentiated Doublets in Whiet Hmong (Mioa)", in Linguistics of the Tibeto-burman Area, 9:2,1-35, 1986.

[8] Ratliff, M., "Tone Sandhi Compounding in White Hmong", in Linguistics of the Tibeto-burman Area, 10:2, 71-105, 1987.

[9] Ratliff, M., "Meaningful Tone: A Study of Tonal Morphology in Compounds, Form Classes, and Expressive Phrases in White Hmong", Dekalb, Illinois: Center for Southeast Asian Studies, Northern Illinois University, 1992.

[10] Niederer, B., "Les langues Hmong-Mjen (Miáo-Yáo), Phonologiehistorique", Lincom Studies in Asian Linguistics 07, Munich, Lincom Europa, 1998.

[11] Niederer, B., " La langue Hmong", Amerindia, 26-27: 347-381, 2001-2002.

[12] Weber, B., Caelen-Haumont, G., Pham, B.H., Tran, D.D, 2012, "Mistral: A *M*elody *I*ntonation *S*peaker *T*onal *R*ange semi-automatic *A*nalysis using variable *L*evels", LREC 2012 Proc., Istanbul, Turkey. On line:http://www.lrec-conf.org/proceedings /lrec2012/index.html, 2012.

[13] Bartkova K., Delais-Roussarie E., Santiago-Vargas F., "PROSOTRAN: a tool to annotate prosodically non-standard data", Speech Prosody 2012 Proc.,Shanghai: Chine (2012)

[14] ETSI ES 202 212 V1.1.1, STQ, "Distributed speech recognition; Extended advanced front-end feature extraction", 2005.

[15] Smith, J.O. and Abel, J.S., "Bark and ERB Bilinear Transforms", IEEE Trans. Speech and Audio Proc., 7(6):697-708, 1999.

[16] Rossi, M., "Le seuil de glissando ou seuil de perception des variations tonales pour les sons de la parole ", Phonetica 23, 1-33, 1981.