Towards the tonal system of an unknown language from south-east Asia: a deeper insight

Geneviève Caelen-Haumont

MICA Institute, HUST - CNRS/UMI 2954 - Grenoble INP
Hanoi University of Science and Technology, 1 Dai Co Viet St., Hanoi, Vietnam
genevieve.caelen@mica.edu.vn

Abstract
This paper is focused on melodic and tonal analyses of a language without a writing system, the Mo Piu one, from an endangered ethnic minority of the south-east Asia in North Vietnam. The Mo Piu language is a branch still unknown of the Hmong-Mien family. Based on a previous experience, we try to get a deeper insight into the tonal system of this language, getting support with a dedicated tool, MISTRAL+, recently upgraded with new functionalities under Praat.

Index Terms: tonal system, unknown language, ethnic minority, Hmong-Mien family.

1. Introduction
The present study is developed in the context of the “Au Co” Project which started in 2008 in MICA Institute, which aims to help saving endangered languages and cultures in Vietnam. This project is based on the collaboration of a wide variety of experts like French and Vietnamese specialists of ethnic languages, French specialists of speech in the domains of phonetics, phonology, prosody, Vietnamese specialists of ethnic groups in Northern Vietnam (Province of Lao Cai), and computer scientists.

First the project Au Co focused on an ethnic minority called Mo Piu. This minority is located in the mountains of North Vietnam not so far from the Chinese border. The Mo Piu village situated in a sort of circus on the side of a hill, is named Nam Tu Thuong, meaning the “the stream river spring up” in Tày language. Though the village name is Tày, the whole population is strictly Mo Piu.

According to preliminary studies [1, 2], the Mo Piu language is really an endangered language because it is uncharted, undocumented, unwritten and spoken only by 237 people in 2011. Moreover, the 7 or 8 ethnic groups in their surrounding do not understand this language. The Mo Piu language being not documented at all, it is urgent to study it before the fusion with the dominant culture, the Vietnamese one.

In fact this study is multifold: one devoted to the traditional linguistic investigation in speech (mainly at the phonetic, and tonal levels), and the other one, to the automatic process of phonetic labeling of an under-resourced language [3]. This paper is concerned with the tonal domain of the Mo Piu language.

Since a long time, many works have been published in the domain of the Hmong tonal system [4], [5], [6, 7, 8], [9], [10, 11]. Among the different Hmong languages, the White Hmong is likely the most studied. According to Niederer [11], this variety presents 50 consonants, 13 vowels (oral, nasal), and 7 tones. If we consider now the Hmong language on the whole according to the different varieties of the Hmong languages, the tones may vary from 3 to 11.

2. Corpus
Three field trips were undertaken in 2009, 2010 and 2011. On the whole, our sound and video corpus is composed of 36h for films, 35h for speech (French / Vietnamese / Mo Piu), fluent speech and lists of words included, 1h for songs, 2350 images, 335 video-clips, and more than 2000 sound files and more than 2000 video ones.

The continuous speech is composed of a large set of cultural inquiries, the domains of the Mo Piu life being split up in about 50 domains of interests (questions / answers), tales, life stories, and drawings or video comments. The lists of words used are “500 words”, Calmsrea”, “Vo”. Just a part of them (plants, parts of body, animals, directions, natural phenomena, numbers…) was until now registered but with several repetitions (3 at least) by 20 male and female speakers.

In the context of this present study, we used the translation of 44 French words from the Calmsrea list (Parts of body), issuing in 209 Mo Piu lexical units (x 2 speakers, thus 418 tokens): these units are complex ones made of possible determinants, classifiers and compound words. For instance while the male speaker VAP01 supplies these 3 words /s zæ/, /s phɛt/, /s mbjɛ/ respectively for urine, sweat and snot, where /s/ probably means aqueous humor or (body) liquid, the female speaker uses before the compound words, the determinant /pʊt(ɛ)/. It seems until now that most of the lexical units are monosyllabic.

3. Methodology

3.1. Previous study
A previous study [1] laid the foundations of the present one. This previous one was based on continuous Mo Piu speech. As nothing at this time was known about this language, and thus no cue was available, we first settled a zone of confidence, i.e. the syllable (even though some uncertainty was remaining about the right syllable boundaries). After the segmentation and labeling task at the lexical and phonetic levels, on a specific Praat TextGrid tier, we put some labels describing the melodic variation (rise, fall, plateau, and F0 levels). The main results showed that 1° the language was mainly monosyllabic, 2° it was tonal, 3° there were existing many tones in the shape of a plateau. Moreover some tones could also have been specified in the frame of the study. All these findings have been confirmed when the following year (2010), we could register for the first time the isolated words.

A second study [2] was based on the same data that the present one, but considered a single speaker. Thus the present one is grounded on twice as many as of the previous data, passing then from 260 to 583 vowels and 2 speakers.
3.2. Present study

Our aim is now to go thoroughly into the tonal system after having got a better insight of the phonetic level. So these 418 lexical items have been first studied at the phonetic and phonologic levels.

3.2.1. Phonetic level

This language being still uncharted, this first step has to be driven very carefully. So 4 expert phoneticians have separately segmented and labeled these 418 lexical items, and at the end of this task, all the results have been merged, in order to reach a settlement between us for each phonetic unit. For this aim, a confusion matrix has been build up for the vowels and for the consonants [12].

3.2.2. MISTRAL+ and the tonal analysis

The second part of the study is concerning the tonal system of these lexical items from the Calmsea list, and especially of their vocalic nucleus. For this part, we used a specific tool (MISTRAL+ [13]). This tool running as a script under Praat [14] built in 2004 [15], was regularly improved until 2011. In 2011 and 2012, a brand new version has been written, developing new functionalities to study all kinds of languages (accented, tonal or not, standard, expressive or emotional speech), allowing many (semi-) automatic tasks in order to help linguists in their hard and boring task of segmentation and labeling.

MISTRAL+ is composed of two different modules: MISTRAL Praat and MISTRAL xls.

The first module of MISTRAL+ is a plugin integrated to the tool Praat. MISTRAL Praat has been designed in order to facilitate the linguistic tasks of studying any language and to save time by the simplification and automation of the maximum of the tasks.

MISTRAL Praat is divided into three main functionalities: 1- extraction and approximation of the fundamental frequency F0, 2- synchronization between segmentation boundaries and target points, and 3- processing of melodic annotation based on the segmentation done by the linguist. To each of these functionalities, a dedicate Praat script is written.

The first step of the language study is done manually. From the speech file, the user segments it into words and/or phonetic units and annotates it in a textgrid. MISTRAL Praat supports the International Phonetic Alphabet. The user may add other tiers in order to help him in his study (words translation, tier for phonetic units, tier for comments...).

Once the annotation has been completed, the user uses the first functionality of MISTRAL Praat. It enables him to automatically extract and approximate F0 for the speech file he is studying.

The praat script extracts the F0 values of the signal every 10 ms using an accurate autocorrelation method supplied by Praat. The user defines the floor and ceiling values based on the data he has already collected on this speaker or he can use default values: 75Hz and 600Hz. From the values extracted and the user parameters, the MOMEL algorithm [16] approximates the F0 variation of the signal. Several ‘target points’ are created which represent the points where the F0 slope is changing its direction. The resulting curve is a sequence of entirely sonorant segments and it constitutes the macro-prosodic component. From these calculations a manipulation file is resulting, which can be used, modified and studied under Praat (see below Figure 1 the part above).

The approximation done by the system aims to fit the best the original signal but at this point a manual step is needed: the file needs to be validated. The user modifies and adapts the target points of the F0 approximation in time and frequency till he perceives it (by vision and audition) in the same way than the original sound file.

To each boundary he creates in his textgrid, a new target point in the F0 approximation of the manipulation file is created by MISTRAL Praat. By defining such new targets in accordance to the description and the segmentations he had done in the textgrid, the user will obtain the information he need to study the language in his way.

Besides, as the annotation and modification of the manipulation file are manually done, the user may have created unwanted results that will impact the melodic and/or duration information. While the user is modifying the frequency height of a wrong target point, or adding a new one, he may add accidently a time delay between the target and the corresponding boundary. The user has thus the possibility via MISTRAL Praat to synchronize the targets points with the boundaries of his textgrid. Depending on the accuracy of the segmentation, he can choose the dedicate precision: 5 ms, 10 ms, or 20 ms. The praat script will then modify all of the target points according to the user choice. Only the time position of the target point is modified, its frequency level is kept in order not to modify the perception.

At this point, the speech signal has been approximated as a manipulation file and has been annotated in a textgrid (see below Figure 1 the part below). The third functionality of MISTRAL Praat bases its calculations on these two files. It computes melodic information in Hz and semi-tones, duration, and derivative enabling the linguist to study the language according to the segmentation he has done.

As MISTRAL+ is dedicated to a great F0 precision (under-resourced languages, and expressive/emotional speech as well), it allows the user to define between 2 and 9 different levels according to the language (for instance tonal, accent or standard), the kind of speech he is studying (expressive, emotional, standard), the precision he wishes. The user can choose between using numeric symbols {1, 2, 3, 4, 5, 6, 7, 8, 9} or alphabetic symbols {A, S, H, E, m, c, b, i, g} standing for Acute, Supra-high, High, elevated, middle, centered, bottom, infra-grave, grave. In the frame of this study, we follow the main trend concerning the tonal languages using 5 levels with a numeric scale. These levels are automatically computed.

The user can also define the frequency range (Fmin, Fmax) and the mean frequency, if he has already collected data on his speaker or choose default values. Once he has defined the number of F0 levels, and the frequency values to use, the user launches the praat script. The script not only performs calculations based on the F0 variation i.e. based on the target points, but also on the segmentation the user did in the textgrid. Three different files result from these calculations: a text file, a textgrid, and an excel file.

The text file contains the description of the thresholds and of the F0 levels in Hz and semi-tones, and the frequency values (minimum, mean, maximum).

The textgrid is based on the file made by the user during the annotation and segmentation step. Several tiers may be added (depending on the choice of the user). Some automatic tiers are giving the level value of each target points contained in the manipulation file, and the resulting dynamic range between two target points. The value in Hz (F0) of each boundary of the textgrid made beforehand by the user, is
converted in ‘Semi-Tones’ \( (ST) \) using \( F_{mid} \) (F0 mean) as shown in Equation 1 below:

\[
ST_i = 12 \times \log_{10} \frac{F_i}{F_{mid}}
\]

The tiers ‘Duration’ and ‘Derived’ give respectively the duration and the value of the derivative of F0 between two targets points.

The third file created by MISTRAL+ is an excel file: it is the export of all the information contained in the textgrid described above. The linguist will be able then to study, use and do the calculation he wants using excel.

MISTRAL xls is an excel VBA module. It enables to create easily macros on the data contained in the xls file exported from the module MISTRAL praat. The user may save a lot of time: he can create the macros he needs and obtains the relevant data in the format he wants. The user defines which files he wants to use and which reports he wants to obtain from the selected files. Once the macro is done, he can then apply it quickly and massively on several files.

First, the user defines the format and order of the results he wants to obtain in an excel sheet. Once the sheet structures of the report are defined, the user can run the VBA module by clicking on the MICA logo. As Praat saves Unicode in Big Endean order which cannot be read properly in Excel, if needed, the procedure converts the exported result from UTF-16BE to UTF-16LE.

### 3.2.3. Intra/inter speaker phonetic and tonal variations

Thus concerning this present work, the first step consists in extracting the 583 vowels from their context. These vowels are simple or diphthongs, and their nature are oral, nasal, creaky and even glottalized.

Then following step requires to sort all the vowel data according first to the number of the tonal parts in the vowels, i.e., one or two parts for a more complex tone, and secondly according to the vowel nature, a simple one or a diphthong, and finally in alphabetic order.

As the phonologic system is not yet established, but under way, it is not possible to determine the phonemes among their variants. The phonetic variation is operating along two axes, at the intra- and inter-speaker ones. In this domain, the most striking variation among the female and male speakers concerns the use of the creaky voice. The female speaker uses it often while the male one doesn’t. As we know that in Mo Piu language the creaky voice may differentiate two tones, we inspected thoroughly these vowels. But in our sample of words, the creaky vowels are not consistent neither between the same lexical item from the same speaker, neither between the two speakers. We thus can conclude that in our limited sample of words, the feature creaky / not creaky is useless.

Moreover there is existing also another kind of variation: the speakers do not always agree on the word due to possible mistakes, but also to the difference between the French concepts, their translation in Vietnamese (our pivot language) and then in Mo Piu ones. For instance when the French /ongle/ and English /nail/ are using one lexical item, the Vietnamese as well the Mo Piu are using two, differentiating the nails between the hand (Mo Piu, female speaker, \( /\text{ng} \text{ t\text{e}ng/}\) and the foot (Mo Piu, female speaker, \( /\text{ng} \text{ qu\text{a}t/}\).

### 3.2.4. Reducing the tonal variation

Another task consists in reducing the tonal variation. First we built a two-way table with all the tone levels of our present corpus (abscissa) and the phonetic items (ordinate) ranked according to their Calmsea list number. The first operation consisted in gathering all the intra-speaker variations of the same item per speaker: the phonetic variation does not concern the tonal one.

In these conditions, now we only deal with the tonal variation. First we have to precise that there exists a strong invariant concerning the vowel F0 slope. It is quite amazing to observe that whoever the speaker could be, the automatically computed F0 slopes are monotonous, shaping mostly a plateau or a falling slope without any modulation (Figure 1 below).

![Figure 1: In the above window, stands the Manipulation file showing 2 plateaux corresponding to vowels (V, third tier), and below, the final Praat-MISTRAL is displayed, showing from top to bottom the manual tiers (tiers 1-4), and the automatic ones (tiers 5-11).](image-url)

Concerning now the tonal variation, an important task is to reduce the deviant tones which are produced by a laugh (all are thus suppressed), an emotion (induced for instance by some word belonging to the intimate parts of the body), or because an emphasis, often produced at the third word repetition by the male speaker. Emotion and emphasis tends to make the tones more contrastive, resulting mostly in raising the F0 maximum and sometimes lowering the F0 minimum. Contrary to laugh, an emotion and an emphasis do not destructure the phonetic and/or the tonal pattern. So these deviations are considered as simple variations, and are easy to reduce. The simple observation of the other repetitions of the same words by the same speaker, -and if necessary by the other speaker-, enables mostly to settle the question.

Another matter of variation is concerning the problem of one or two part in the tonal shape. It is not so easy to take a decision when one can observe for instance a falling double tone where the external boundaries (for instance /43-32/) correspond to the values of a single tone (for instance /42/). Is an internal modulation of a single tone or a double tone?

The first way of solving this problem is to check whether the double tones are produced by both speakers. If so, the hypothesis of a double tone is stronger. Another cue is to take into account the variation. A weak variation at the F0 level among the repetitions of the same items (intra- and inter-speaker levels), and/or a small number of variants, strengthen the hypothesis, whatever it may be.
4. Results

4.1. Tones heights

The figure 2 below presents the distribution of the 492 vowels presenting a unique slope. One can observe that 3 zones of tones are emerging: 3 plateaux, levels /22, 33, 44/ and 2 falling tones /32, 43/. Due to their weak population, the other possible tones may be considered as variations of the former ones.

There are only 6 examples presenting a twofold tone, and with the 3 repetitions, 36 items have to be reported. Only one is well represented in the data: four lexical units are using the shape /33-32/. The other shapes are not clear neither well represented.

4.2. Duration

In the previous study stemming from continuous speech [1], we suggested that duration might play a role to differentiate between the tones. We undertook to check this hypothesis in the context of the present data.

In order to focus on the core of the tones, we selected the most frequent tones (/32, 33, 43, 44/) of the simple vowels (oral, nasal but no diphthongs), with a single slope, and not followed by another vowel. In such conditions, we gathered 300 vowels, sorted by each speaker (a man and a woman). The aim was to compare the same lexical units (at the phonetic level but not at the conceptual one) spoken with the same tone and thus observe if the duration could differentiate the tones. The data present often the same lexical units (possibly with phonetic variation), but unfortunately in each case they correspond to the same concept: this lexical item forms either a word or a part of a compound word. So our discussion can only concern a general perspective.

Table 1: The main duration features of the 300 vowels/tones (tones /32, 33, 43, 44/), sorted by speaker

<table>
<thead>
<tr>
<th>Speakers</th>
<th>Nb</th>
<th>Min</th>
<th>Max</th>
<th>Average</th>
<th>Standard dev.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Man</td>
<td>147</td>
<td>105</td>
<td>533</td>
<td>334</td>
<td>87</td>
</tr>
<tr>
<td>Woman</td>
<td>153</td>
<td>59</td>
<td>691</td>
<td>300</td>
<td>169</td>
</tr>
</tbody>
</table>

From the table 1 above and the Figure 3 below, we see that the durations between the two speakers are a matter of opposition: as the number of data is about the same in both cases (147 vs. 153), the woman gets her duration extrema (minimum / maximum) beyond the man’s ones. In other words, the woman shows both smaller tonal durations and greater ones than the man. We also observe a break in the woman duration around 200 ms while the man one is continuous.

We only notice one regularity among vowels duration, and concerning only the woman data because most of the time, she is using what seem to be determinants (same token frequently used before a lexical item) and also vowels as vocalic support before a complex cluster (especially when the lexical item begins with a prenasalized occlusive) at the beginning of the word, while the man does not: on one hand the determinants and such support vowels are always put just before the lexical entity, and on the other one they correspond to the smallest duration. It simply means that Mo Piu language behaviors as many other languages: function units get smaller duration. Nevertheless some lexical items present also small durations. Finally all these findings strengthen the hypothesis that duration plays a role for the tonal distinction, but until now, we have no means to prove this strong hypothesis.

4.3. Derivative

Besides the duration, the MISTRAL+ tool supplies the derivative. It is computed from the tones. For this analysis, we used the same method as for duration, i.e. the vowels of the core of the study presenting the tones the most frequent (/32, 33, 43, 44/). As in 4.2 above, they all correspond to the simple vowels (oral, nasal no diphthongs), with a single slope, and not followed by another vowel. Our aim was once more to try to put to light some interesting features concerning the tones. The study of the derivative supplies some interesting results about the data. This analysis could be also made from the tones levels, but the derivative is more precise. So let us first consider all the vowels of the present study, that is to say 583 vowels (simple, diphthongs from all the tones). Over these 583 vowels, only 81 are presenting a positive derivative, which means that only 81 vowels (14%) present a rising tone (part of a tone or the whole tone).
If we now consider the positive derivatives, 52 (64%) over 81 are corresponding to a plateau. The remaining 26% of these 81 derivatives, or in other words 9% of the 583 vowels, present a rising tone, and 2% belong to a twofold tone. Now coming back to the restricted data (tones /32, 33, 43, 44/) in the same way as we did for duration, we notice in the Table 2 above and the Figure 4 below, that the curves of the derivative are almost parallel between the woman and man.

Table 2: The main derivative features of the 300 vowels/tones (tones /32, 33, 43, 44/), sorted by speaker.

<table>
<thead>
<tr>
<th>Speakers</th>
<th>Nb</th>
<th>Min</th>
<th>Max</th>
<th>Average</th>
<th>Standard dev.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Man</td>
<td>147</td>
<td>-27.56</td>
<td>2.37</td>
<td>-8.63</td>
<td>4.47</td>
</tr>
<tr>
<td>Woman</td>
<td>153</td>
<td>-26.92</td>
<td>31.42</td>
<td>-2.44</td>
<td>7.56</td>
</tr>
</tbody>
</table>

However, the woman, like duration, shows a greater range between her derivative extrema, particularly due to the positive values. Though these last ones are not frequent at all, nevertheless 25 vowels (i.e. tones) present positive values from the woman, while only one is present in the man’s ones.

Figure 4: Comparison of the tones derivative ranked in ascending order between the man (M, solid line, 147 vowels / tones) and woman (W, dashed line, 153 vowels / tones). The abscissa presents the derivative values, and the ordinate corresponds to the number of the vowels / tones.  

5. Conclusion

This study focused on the tonal investigation in the context of an unknown branch of a Hmong language, the Mo Piu one which is strongly endangered. After some considerations about the phonetic domain of this language, then the presentation of the tool MISTRAL+ making it possible to study the Mo Piu phonetics and tones, we presented the main findings concerning the tonal domain. On one hand over the 17 tones appearing in our speech samples (528 vowels) in the context of a single slope, 2 falling tones /32, 43/, and 2 plateaux /33, 44/, and on the other one, a twofold falling tone (/33-32/) are statistically clearly emerging. Very rare rising patterns were noticed. More data are required in order to test whether duration, creaky and breathy voice are relevant cues for better differentiating the tones.

6. Acknowledgements

We warmly thank Benoît Weber, research engineer in MICA Institute, who makes it possible to write the new version of MISTRAL+ just in time for the phonetic and tonal analysis in the context of this paper.

7. References


