Comparison between two models of language for the automatic phonetic labeling of an undocumented language of the South-Asia: the case of Mo Piu

Geneviève Caelen-Haumont[1], Sethserey Sam[2]
[1]MICA Institute, HUST, Hanoi, Vietnam
[2]Institute of Technology of Cambodia, Pnomh Penh, Cambodia
gcaelen@gmail.com, sam.sethserey@itc.edu.kh

Abstract
This paper aims at assessing the automatic labeling of an undocumented, unknown, unwritten and under-resourced language (Mo Piu) of the North Vietnam, by an expert phonetician. In the previous stage of the work, 7 sets of languages were chosen among Mandarin, Vietnamese, Khmer, English, French, to compete in order to select the best models of languages to be used for the phonetic labeling of Mo Piu isolated words. Two sets of languages (1° Mandarin + French, 2° Vietnamese + French) which got the best scores showed an additional distribution of their results. Our aim is now to study this distribution more precisely and more extensively, in order to statistically select the best models of languages and among them, the best sets of phonetic units which minimize the wrong phonetic automatic labeling.

Keywords
Phonetic labeling, assessment of automatic procedures, under-resourced languages, models of language, Hmong.

1. Introduction

The present study belongs to the "Au Co" Project which started in 2008 in MICA Institute, which aims to help save endangered languages and cultures in Vietnam. This project is based on the collaboration of a wide variety of experts like French and Vietnamese linguists, Vietnamese specialists of minority languages, French specialists of speech in the domains of phonetics, phonology, prosody, French and Vietnamese anthropologists, specialists of minority 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 along the Chinese border. The Mo Piu village situated in a sort of cirque on the side of a hill, is named *Nam Tu Thuong*, meaning the "stream river spring up" in Tày language.

According to preliminary studies [1], the "Mo Piu" language is an endangered language because it is uncharted, undocumented, unwritten and spoken only by 237 people. Moreover, the 7 or 8 minority 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.

Our purpose is then to carry on studying linguistic analysis and develop tools to save the Mo Piu culture and language, in the framework of two international CNRS-ANR-Blanc Projects "PI Language" and "AppSy".

In fact the study of the Mo Piu language is twofold: one devoted to the traditional linguistic investigation in speech (phonetic, tonal, lexical domains…), and the other one, to the automatic process of phonetic labeling of an under-resourced language (in this present case, it was unknown before we started studying it). This paper is concerned with a human assessment of the automatic procedures.

2. Automatic labeling for the Mo Piu unknown language

Concerning the endangered languages, the lack of human language technologies (HLT) systems for these languages accelerates their extinction while on the other hand HLT could help to stop this trend by giving various resources to this people in order to facilitate the access to their language in the next decades and also from now on, to the linguistic analysis.

As no (or only limited) resources are available for such under-resourced languages, we are particularly interested in new techniques and tools for rapid development of spoken language technologies for this kind of languages. One way to document an unstudied language, especially if it is a language without writing, is to use computer tools to transcribe the speech segments in sequence of phonemes or words. As for such languages, no acoustic model, nor language model nor lexical model do exist, we propose for this task to use acoustic models of other languages. In a previous work [2] concerning the speech transcription of a new language for which neither adaptation nor learning of acoustic model was done, the acoustic-phonetic multilingual systems were given preference for this task over the monolingual system.

2.1. Multilingual acoustic-phonetic recognizers

In order to generate the phonetic sequences from the Mo Piu speech, we used different multilingual acoustic-phonetic recognizers. Each multilingual acoustic-phonetic recognizer covers at least two of the five languages: Mandarin-Chinese (MA), English (EN), French (FR), Khmer (KH, the official language of Cambodia), and Vietnamese (VN). In this case, the multilingual acoustic

models are created by combining the existing monolingual acoustic models of MA, EN, FR, KH and VN, trained respectively on several broadcast news corpora [3], [4], [5], [6] and [7]. The combination of acoustic models is simply based on the ML-sep combination method [8]. It means that there is no data to share across languages among the involved monolingual acoustic models (e.g: the acoustic units /a/ of French and Vietnamese are trained by two different HMMs). For instance, there are 75 acoustic units in the multilingual acoustic model of Vietnamese and Mandarin (MultAM-VNMA) constituted respectively by 34 and 41 acoustic units, in Mandarin and Vietnamese acoustic models.

To enable linguistic analyses and evaluation on automatic phonetic transcription, all phonetic items issuing from the recognizer are converted to the TexGrid format readable by the Praat software [9]. Thus the assessment by an expert phonetician can be set in motion [10].

2.2. The Mo Piu Corpus

Our first studies gave evidence that this ethnic minority belongs to the family of Hmong-Mien, but as a little and totally unknown branch. This fact is certainly due to the extremely small size of the community which lives in the remote North Mountains of Vietnam, with no roads reaching their village.

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 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 inquiries (questions / responses), tales, life stories, and drawings or video comments. Several lists of words are used, but in the context of this study, we only refer to the Calmsea list [11, 12]. 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) and by 20 male and female speakers. Since 2010, this corpus has been the matter of many international papers (for instance [1, 10, 13, 14, 15]).

3. Phonetic Assessment

3.1. First experiment

In the previous study [10] as in the present one, our aim is to assess the output of the recognizers against manual annotations by an expert phonetician. Let us say first that as this language is unknown, the task of the phonetician is not easy as she has to check the automatic performances in the same time where she has to determine with a minimal error the phonetic status of the items. But these conditions cannot be improved because they are precisely those of the experiment.

In the previous one, 2 comparisons were at the stake, first that of the 2 performances (computer / expert), and secondly of the group of languages competing each other. In those conditions, 5 languages were chosen in several combinations. The choice of these specific languages relies on geographically close languages (such as Mandarin in China where the Mo Piu are coming from 350 years ago, the Vietnamese because they settle in Vietnam, and Khmer because the phonetic is very rich). As for the English and French languages, their resources are largely represented and well known. Thus we first studied 3 combinations: VN-MA, VN-MA-KH (as geographically close languages), VN-MA-KH-EN-FR (adding languages with computering resources), then as French surprisingly seems to increase the scores results, we added 3 new combinations of languages: VN-MA-KH-FR, MA-FR, VN-FR.

In order to assess the performance, we used only isolated words as such: 1 female speaker x 11 words x 3 repetitions x 6 groups of languages, thus leading on the whole to 198 Praat TextGrids and 198 xls files. These xls files were all manually merged in one single big file composed of 1425 data of phonetic events, split up in 754 vowel events, 671 consonant ones, and 65 events automatically labeled as phonetic event, but which have been actually labeled as a part of a pause by the phonetician.

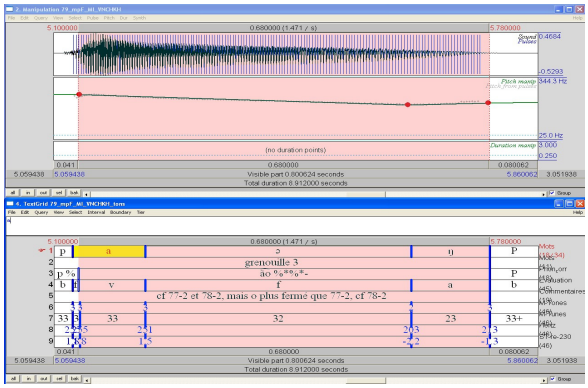

Figure 1: The upper window shows the signal and the F0 slopes of a mo piu word /pão/ or *frog* in English. Below, the Praat TextGrid presents respectively the automatic phonetic labelling (tier 1) using a model of language with three components (Vietnamese, Mandarin, Khmer), then the translation in French of the mo piu word (*grenouille*, tier 2), the expert phonetic manual labelling (tier 3), the assessment of the automatic labeling (tier 4), and finally the comments and addresses of the same word in the data basis (tier 5). The others tiers below (6 to 9) deliver automatic information about the lexical tone and F0 levels using several units (Praat-Momel-MISTRAL+).

So the assessment was concerning the phonetic labeling, in terms 1° of a right identification or not, 2° of right boundaries or not, 3° and then the best combination of languages (i.e. the best models of languages). In such an experiment, we expected of course to find very bad results concerning the phonetic labeling. Nevertheless, our aim was not to get a tool for an automatic labeling, but to build it, and therefore to perform that, to previously select the best models of language.

On the whole, the first results (merging the right and close labels) are better for consonants identification than for vowels ones: vowel identification spreads from 11 to 37%, while the consonant one, from around 10 to 44%. The nasal consonants supply the best recognition (from 30 to 44%).

Concerning the sets of languages, the results of the assessment *for the vowels,* are ranking in such an order: VN-FR (46%) > MA-FR (44%) > VN-MA-KH-FR (42%) > VN-MA-KH-EN-FR = VN-MA (36%) > VN-MA-KH (33%).

Concerning *the consonants*, the scores are greater than for the vowels: MA-FR (58%) > VN-MA-KH-FR (54%) > VN-MA-KH-EN-FR = VN-MA-KH (51%) > VN-MA (48%) > VN-FR (44%). Surprisingly, VN-FR appeared as the best for vowels and the worst for consonant language assessments.

On the whole, *for all the vowels and consonants* included in our corpus, the final order is thus: MA-FR (51%) > VN-MA-KH-FR (48%) > VN-FR (45%) > VN-MA-KH-EN-FR (43%) > VN-MA (41%) > VN-MA-KH (40 %).

3.2. Second experiment

The goal of the second experiment is to enlarge the set of data in order to thoroughly analyze the results concerning the best models of language found as such in the first experiment. According to the previous study, the best models are French + Mandarin (FR-MA), and French + Vietnamese (FR-VN). In this paper we focus on the phonetic level, and we do not taking into account the question of the place of the boundary.

3.2.1. Choice of the new Mo Piu corpus

As we check fewer models of language than for the first experiment, we use a greater corpus. So for this new experiment, we chose 2 speakers, a woman (VTD01) and a man (VAP01). As previously the list of words is extracted from the Calmsea list, and concerns the body parts.

For the second experiment, the number of the words is now: 2 speakers x 44 words x 3 repetitions x 2 models of language = 528 words. In fact these 44 words are French ones, but in the Mo Piu language they generally result in several ones (classifier, determinant, simple or complex content words). These Mo Piu words are generating 6813 phonetic events (consonants, vowels, pauses). From these 6813 items, 5034 have been extracted, split up into 2777 consonants and 2257 vowels, which have been extracted and manually assessed. Putting apart the other symbols previously written on the other Praat TextGrid tiers, a total of *70 983 signs (symbols or letters)* on the whole have been manually written on the different tiers in order to achieve this present assessment.

PI VTD01 VAP01 FR MA VN 1-41 chron [Compatibility Mode]

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
1	Items	Models of	Speaker	Word	French	Mo Piu	Phonetic	Manual	Automatic	Manual	Item	Item	Item	Tone at the	Tones	Tones in	Semi-	Comments
2	number	languages		number			Classes	correction	labeling	assessment	beginning	end	Duration	left boundary	in 5 levels	Hertz	tones	
3	1	FR-MA	VTD01	1	PP	PP	PP	PP %*-%	w	F	14.81	14.83	14832					
4	2	FR-MA	VTD01	1	ventre 1y	i	V	*%- i %*-%	w	A	14.83	14.84	107	3	35	232	0.9	
5	3	FR-MA	VTD01	1	ventre 1y	i	V	*%- i %*-%	ə	F	14.84	14.92	107	3	35	232	0.9	
6	4	FR-MA	VTD01	1	ventre 1y	i	V	*%- i %*-%	ŋ	A	14.92	14.94	107	3	35	232	0.9	
7	5	FR-MA	VTD01	1	ventre 1a	h͡ŋa	C	h %	ŋ	F	14.94	15.04	101	5	54	321	6.5	
8	6	FR-MA	VTD01	1	ventre 1a	h͡ŋa	Cn	ɲ %*-%	ŋ	V	15.04	15.16	180	4	44-	291	4.9	
9	7	FR-MA	VTD01	1	ventre 1a	h͡ŋa	Cn	ɲ %*-%	j	B	15.16	15.22	180	4	44-	291	4.9	
10	8	FR-MA	VTD01	1	ventre 1a	h͡ŋa	V	*%- a %*%*-%	j	A	15.22	15.24	482	4	43	255	2.6	
11	9	FR-MA	VTD01	1	ventre 1a	h͡ŋa	V	*%- a %*%*-%	a	B	15.24	15.52	482	4	43	255	2.6	
12	10	FR-MA	VTD01	1	ventre 1a	h͡ŋa	V	*%- a %*%*-%	ɛ̃	V	15.52	15.64	482	4	43	255	2.6	
13	11	FR-MA	VTD01	1	ventre 1a	h͡ŋa	V	*%- a %*%*-%	ŋ	A	15.64	15.70	482	4	43	255	2.6	
14	12	FR-MA	VTD01	1	PP	PP	PP	*%- PP %*-%	ŋ	F	15.70	15.94	648	3	34	219	-0.1	
15	13	FR-MA	VTD01	1	PP	PP	PP	*%- PP %*-%	ʏ	F	15.94	16.14	648	3	34	219	-0.1	
16	14	FR-MA	VTD01	1	PP	PP	PP	*%- PP %*-%	l	F	16.14	16.35	648	3	34	219	-0.1	
17	15	FR-MA	VTD01	1	ventre 2y	i	V	*%- i %*-%	l	A	16.35	16.38	139	4	45	242	1.7	
18	16	FR-MA	VTD01	1	ventre 2y	i	V	*%- i %*-%	ə	F	16.38	16.46	139	4	45	242	1.7	
19	17	FR-MA	VTD01	1	ventre 2y	i	V	*%- i %*-%	w	F	16.46	16.49	139	4	45	242	1.7	
20	18	FR-MA	VTD01	1	ventre 2a	h͡ŋa	C	*%- h %	w	F	16.49	16.51	122	5	54	306	5.7	
21	19	FR-MA	VTD01	1	ventre 2a	h͡ŋa	C	*%- h %	ŋ	F	16.51	16.61	122	5	54	306	5.7	
22	20	FR-MA	VTD01	1	ventre 2a	h͡ŋa	Cn	ɲ %*-%	ŋ	V	16.61	16.74	198	4	44-	266	3.3	
23	21	FR-MA	VTD01	1	ventre 2a	h͡ŋa	Cn	ɲ %*-%	j	B	16.74	16.81	198	4	44-	266	3.3	
24	22	FR-MA	VTD01	1	ventre 2a	h͡ŋa	V	a %*%*-%	j	B	16.81	17.14	198	4	44-	266	3.3	
25	23	FR-MA	VTD01	1	ventre 2a	h͡ŋa	V	a %*%*-%	ɛ̃	V	17.14	17.27	198	4	44-	266	3.3	
26	24	FR-MA	VTD01	1	ventre 2a	h͡ŋa	V	a %*%*-%	ʏ	F	17.27	17.29	198	4	44-	266	3.3	
27	25	FR-MA	VTD01	1	PP	PP	PP	*%- PP %*-%	ʏ	F	17.29	17.40	444	3	34	220	0	
28	26	FR-MA	VTD01	1	PP	PP	PP	*%- PP %*-%	PP	B	17.40	17.72	444	3	34	220	0	
29	27	FR-MA	VTD01	1	PP	PP	PP	*%- PP %*-%	m	F	17.72	17.74	444	3	34	220	0	
30	28	FR-MA	VTD01	1	ventre 3x	nɔt	Cn	n	m	V	17.74	17.84	103	4	44+	248	2.1	occlusive fi
31	29	FR-MA	VTD01	1	ventre 3x	nɔt	V	ɔ %	m	V	17.84	17.93	103	4	44+	248	2.1	occlusive fi
32	30	FR-MA	VTD01	1	ventre 3a	h͡ŋa	C	*%- h %	m	F	17.93	17.95	15	4	45	277	4	
33	31	FR-MA	VTD01	1	ventre 3a	h͡ŋa	C	*%- h %	n	F	17.95	18.06	24	5	54	322	6.6	
34	32	FR-MA	VTD01	1	ventre 3a	h͡ŋa	Cn	ɲ %*%*-%	n	B	18.06	18.19	189	4	43	266	3.3	
35	33	FR-MA	VTD01	1	ventre 3a	h͡ŋa	Cn	ɲ %*%*-%	j	B	18.19	18.24	189	4	43	266	3.3	

Table 1: Extract of the xls file showing the manual and automatic labels and the automatic values computed by the script Praat-Momel-MISTRAL+

3.2.2. Method of assessment

For the assessment we use the Praat software, including in addition the MISTRAL+ script (updated in 2011 [17, 18 see in the same Proceedings]), which enables to use phonetic IPA symbols, then filling up the xls files from all the TextGrids data. The task initially consists in adding several layers (called *tiers*) in the TextGrid. First concerning the manual segmentation and labeling, the tiers are completed in this following way (see the Figure 2 below):

1° the French word
2° its translation in Mo Piu word
3° the generic classes of each Mo Piu phonetic unit
4° the manual segmentation and labeling, which includes the errors concerning the automatic labeling and segmentation. The phonetic event assessed by the expert is thus preceded and/or followed by a suite of possible symbols, indicating the boundaries and/or phonetic mistakes of the automatic tier (please see below),
5° the phonetician assessment of the automatic phonetic labeling,
6° the last tier (*Commentaires / Comments)* consists in mentioning the address of the same word in the other Praat TextGrids, and sometimes a comment about the nature of the phonetic events. The same phonetic units may vary due to the speaker or the context variation.

Concerning the point 4 above, the error(s) of each phonetic event is gathered in a TextGrid tier which takes into account all the sub-parts of the automatic segmentation and labeling according to the expert phonetician assessment.
For instance the symbols used to point out the mistakes performed by the automatic performance are: %, *, - and +, indicating respectively an event concerning a boundary (e.g. suppression %- or adding %+), an event concerning a phonetic item (suppression *-, adding *+). As a suppression of a boundary generally occurs with that of a phonetic event, the suite is reduced to %*-, %*+ when they occur after a phonetic item assessed by the expert phonetician, or conversely *%-, *%+ when it precedes it. If several boundaries therefore phonetic events are concerned, the format is the same, just adding the number of symbols suppressed/added, such as %*%*%*-, etc.
The Figure 2 below shows a TextGrid extract after the MISTRAL+ procedure has been achieved, adding the last 7 automatic tiers below the 6 previous manual ones and to the automatic Mo Piu phonetic labeling (tire 4).
Now regarding the phonetician assessment (point 5° above), 5 labels are used: *F* = "faux" (wrong or *F/W*), *B* = "bon" (right or *B/R*), *V* = "voisin" (phonetically close, one feature difference on the articulatory or acoustic level, or *V/C*), *A* = "acoustic". This label applies to a section of a phonetic unit, and not on the whole of it as conversely *F/W, V/C, B/R* do. For instance, it is often used in the final transition of the phonetic units. Thus the labels *F, B, V* are concerning the phonetic unit while *A* is labeling a sub-part of it.

In the same way as *A,* a new label "*#*" was added comparatively to the first experiment (see 4.1. above), in order to increase the precision. It is used when the boundaries between the automatic tier and the manual one in the Praat TextGrid is practically the same, spreading from 2 to 6 ms. Below the limit of 2 ms, it is considered as the same boundary.

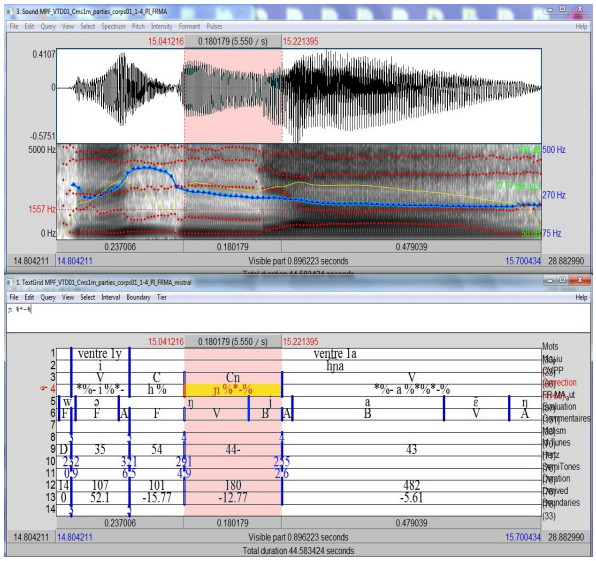

Figure 2: Example of a Praat TextGrid after the MISTRAL+ processing.

In fact this symbol "#" means that the place of the boundary is not considered as wrong, neither the corresponding phonetic item between the automatic and manual boundaries. Nevertheless let us precise that this tiny chunk of speech may be considered as wrong when for instance the automatic procedure writes a vocalic item while a consonant burst occurs, and in this case this tiny phonetic share is labeled as *F* (wrong).

3.2.3. Results

All the automatic and manual annotations from the Praat TextGrids are automatically feeding an xls file (see Table 1 above), enabling to sort the data along several criteria: rank of the phonetic item, file name, speaker, French word, Mo Piu word, phonetic classes (C = general consonant, Cn = nasal consonant, G = glide; V = oral vowel, Vn = nasal vowel, Vd = diphthong), phonetic item with the symbol errors, automatic label, address of its beginning / end, assessment, duration of each speech portion assessed, comments.

3.2.3.1. Global perspective

The main interest of the second study is to compare which model of language perform best, either FR-MA or FR-VN.
The Figure 3 below presents the comparison between the results of each model of languages.
On the whole it appears that FR-MA ranks above FR-VN. Putting apart the symbol # which presents a great stability

around of 4% in all the conditions, the difference (25/33%) concerning the wrong results *(F/W)* between the two models of language is mainly explained by the better scores for the right labeling (*B/R,* 20/16%) and the acoustic one (*A,* 22/18%), whereas the close one (*C,* 29/28%) is nearly the same in both cases.

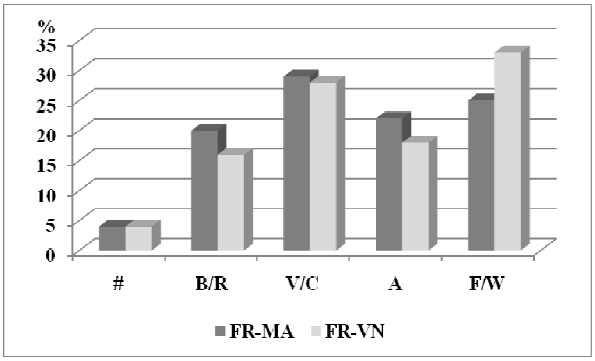

Figure 3: Comparison between the two models of language FR-MA (2548 phonetic items) and FR-VN (2486 items)

The next point to check is concerning the results between consonants and vowels according to the two models of languages. The question is thus: are these previous results for these two models of language remaining stable for consonants and vowels?

3.2.3.2. Consonant assessment

Concerning the 2777 consonant items checked, split up between 1369 for FR-MA and 1408 for FR-VN, the Figure 4 below shows that the results globally present the same trend than that of the global perspective (Figure 3 above).

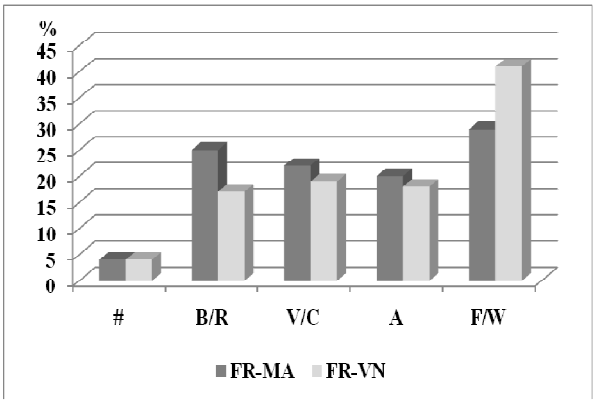

Figure 4: Comparison of the two models of language (2777 consonantic items)

Nevertheless the difference between FR-MA and FR-VN is strengthened: the distance between the wrong results is increasing *(F/W,* 29/41%) while all the other FR-MA labels are benefiting from this difference (*B/R* 25/17%, *V/C* 22/19%, *A* 20/18%).

3.2.3.3. Vowel assessment

For this evaluation, 1179 vocalic items were labeled for FR-MA and 1078 for FR-VN, so on the whole, 2257 items were assessed (Figure 5 below).

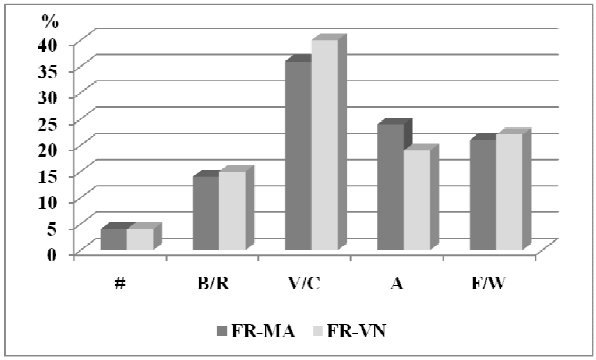

Figure 5: Comparison of the two models of language (2257 vocalic items)

The Figure 5 presents a trend which is a little bit less in favour to FR-MA than previously. Whereas on the one hand the wrong results are indeed stable (respectively FR-MA, FR/VN, *F/W,* 21/22%), the right ones also (respectively FR-MA, FR/VN, *B/R* 14/15%), and the acoustic labels turn to be better for FR-MA (*A,* 24/19%), conversely on the other hand the close phonetic targets are benefiting to FR-VN (*V/C*, 36/40%). So nothing can be clearly establish for the model FR-VN, as neither the right results nor the wrong ones definitively settle the question.

3.2.3.4. The assessment of the phonetic cores

The first point concerns the global distribution of the percentages. The wrong results (consonantic shares respectively FR-MA, FR-VN 29/41%; vocalic ones, respectively FR-MA, FR-VN 21/22%) are no so great though the automatic procedure is labeling an unknown language with resources borrowed from another languages (French, Mandarin, Vietnamese). But conversely the right results are still less great (consonantic shares respectively FR-MA, FR-VN 25/17%; vocalic ones, respectively FR-MA, FR-VN 14/15%).

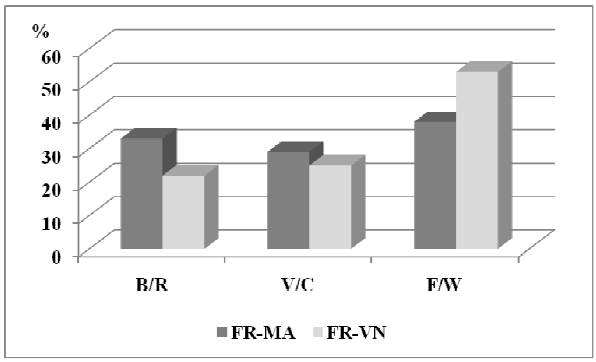

Figure 6: Comparison of the two models of language (2133 consonants)

Until now the study was concerning the consonantic and vocalic items, whatever they could be, whole phonetic items or part(s) of them. We now focus on the phonetic items cores. As explained above, they are addressed by 3 labels, *B/R, V/C,* and *F/W*.
The figure 6 above presents the comparison between the two models of language (2133 consonants: FR-MA, 1042 items, FR-VN, 1091).

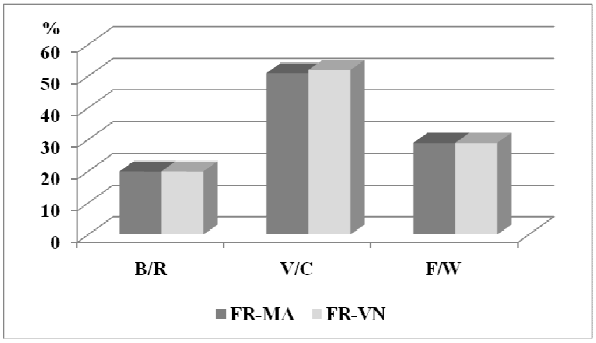

Figure 7 : Comparison of the two models of language (1680 vowels)

Concerning the consonants cores, FR-MA appears as the most efficient model of language: greater right results (respectively FR-MA, FR-VN, 33/22%), greater close results (respectively FR-MA, FR-VN, 29/25%), fewer wrong results (respectively FR-MA, FR-VN, 38/53%).
The Figure 7 above shows the results concerning the vowels. This Figure presents an outstanding stability of the results between the two models of language (respectively FR-MA and FR-VN, *B/R* 20/20%, *V/C* 51/52%, *F/W* 29/29%).
Thus the conclusion appears clearly as shown by the Figure 8 below.

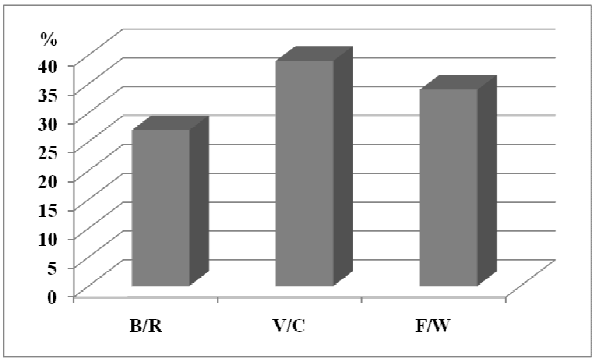

Figure 8: Final results of the best model of language FR-MA for labeling the unknown Mo Piu phonetic units (1883 consonants and vowels)

The model of language FR-MA is the more efficient for labeling the unknown Mo Piu consonants and vowels. Though the wrong labels *(F/W)* are reaching 34%, the close ones *(V/C)* perform 39%, and the right ones *(B/R)* 27%. Thus concerning the improvement of the automatic procedure, the main task –and perhaps at the same time the easiest one-, will consist in tipping up the number of the close labels towards the right ones, and of course in trying to reduce the number of wrong labels.

4. Discussion

To address the comparison between the first and the second study, some questions may be put.
On the whole are still the results better for the consonant labeling than for the vowel ones?
If we merged the right results with the close ones *(B/R + V/C)* as it has be done for the first study, the response is negative: FR-MA, FR/VN consonants are respectively reaching 47/36%, and vowels 50/55%. But if we consider only the right labels *(B/R),* the consonants results remain the better (25/17%) in comparison with the vowels ones (14/15%). In fact the close labels are much more numerous for the vowels than for the consonants.
Concerning the vowels, is the model of language FR-VN remaining first above the FR-MA?
In the second study, if we merge as above the results *(B/R + V/C)*, FR-VN is first (55/50%) as for the previous study.
If we consider only the core of the phonetic units, the results are strictly equivalent between FR-VN and FR-MA (see Figure 7).
Concerning the consonants, is the model of language FR-MA remaining first above the FR-VN?
The FR-MA remains the most efficient model of languages, whatever the conditions, ranking first for the right labels *(B/R* 25/17%), the close ones (*V/C* 22/19%), and the core results *(B/R* 33/22%; *V/C* 29/25%).
And in the whole, comparing all the results merging vowels and consonants, FR-MA is the best model of language for labeling the unknown Mo Piu phonetic items: wrong labels (*F/W* 34%), close ones (*V/C* 39%), right ones (B/R 27%).

5. Conclusion

The second study encompassing many more data (from 11 words to 44, from one speaker to two ones), is in tune with the previous one. It supplies more precision to the assessment. Finally the model of language FR-MA is confirmed as the most efficient one for the task of the automatic labeling of an under-resourced language such as the Mo Piu one.
The French phonetic items contribute in increasing the results because in both cases of Mo Piu and French, there are surprisingly many nasal or nasalized vowels.
The next task will consist in reaching a consensus between 4 phonetician having labeled the same Mo Piu words, which will be soon achieved. This agreement will supply the first list of the Mo Piu phonemes. Then these phonemes will be included in the automatic procedure in order to improve the results.

6. Acknowledgements

This research is granted thanks to 2 Projects ANR-CNRS Blanc "*Langues PI"* and *"AppSy"*.

7. References

[1] Caelen-Haumont, G., Cortial, B., Culas C., Hong T. D., Lê T. X., Salmon J.-P., Vittrant A., Vuong H. T., Song L. A. (2010). Mo Piu minority language: data base, first steps and first experiments. In *Proceedings of the Second International Workshop on Spoken Languages*, SLTU'10, Penang, Malaysia, pp. 42--50.

[2] Schultz, T., Kirchhoff, K. (2006). *Multilingual Speech Processing,* Academic Press, Elsevier.

[3] CCC resources (2005). Online, Chinese Corpus Consrtium Retrieved from http://www.dear.com/ CCC/resources.htm.

[4] Paul, D., Baker, J. (1992). The design for the Wall Street Journal-based CSR corpus. In *Proceedings of the workshop on Speech and Natural Language*, pp. 357--362.

[5] Lamel L., Gauvain, J., Eskenazi, M. (1991). BREF, a large vocabulary spoken corpus for French. In *Proceedings of Eurospeech*, pp. 24--26.

[6] Seng, S., Sam, S., Besacier, L., Bigi, B., Castelli, E. (2008). First Broadcast News Transcription System for Khmer Language, 6th edition of the Language Resources and Evaluation Conference (LREC), Marrakech, Morocco.

[7] Le, V., Tran, D., Castelli, E., Besacier, L., Serignat, J-F. (2004). Spoken and written language resources for Vietnamese. In *Proceedings of* LREC 2004, p. 599-602.

[8] Schultz, T., Waibel, A. (2001). *Language-independent and language-adaptive acoustic modeling for speech recognition*, Speech Communication, 35, vol. 1, pp. 31--52.

[9] Boersma, P. (2001). *Praat, a system for doing phonetics by computer,* Glot international, 5, vol. 9/10, pp. 341--345.

[10] Caelen-Haumont, G., Sam, S., Castelli, E. (2011). Automatic Labeling and Phonetic Assessment of the Mo Piu Under-Resourced Asian Language: early results. In *Proceedings of the International Conference on Asian Language Processing (IALP 2011),* Penang, Malaysia, 160-3D.

[11] Matisoff, J. A. (1978). *Variational semantics in Tibeto-burman: the 'organic' approach to linguistic comparison,* Philadelphia: ISHI Publications.

[12] Swadesh, M. (1971). *The origin and diversification of Language,* Sherzer Joel (ed). Chicago : Adline.

[13] Caelen-Haumont, G., Vittrant, A. (2012, to be published). Transcrire, écrire et formaliser en analyse phonétique, mélodique et tonale: l'exemple d'une langue d'Asie du Sud-est tonale (mo piu) et du français. In *Actes du 25^{e} Colloque du Cercle Linguistique du Centre et de l'Ouest (CERLICO),* Presses Universitaires de Rennes, France.

[14] Caelen-Haumont, G., (2012, to be published). Towards the Mo Piu Tonal System: First Results on an Undocumented South-Asian Language. In *Proceedings of Speech Prosody,* Shangaï, China.

[15] Caelen-Haumont, G. (2012, to be published). Towards the tonal system of an unknown language from south-east Asia: a deeper insight. In *Proceedings of the Tonal Aspects of Languages,* TAL 2012, Nanquing, China.

[16] Hirst, D. (2011). Online: http://uk.groups.yahoo.com /group/praat-users/files/Daniel_Hirst/.

[17] Weber, B., Caelen-Haumont G., Pham B. H., Tran D.-D. (2012 to be published). MISTRAL+: A dedicated tool for under-resourced languages study/. In *Proceedings of SLTU'12,* Cape Town, South Africa.

[18] Weber, B., Caelen-Haumont, G., Pham, B. H., Tra,n D.-D. (2012 to be published). MISTRAL+: A *M*elody, *I*ntonation *S*peaker *T*onal *R*ange semi-automatic *A*nalysis using variable Levels. In *Proceedings of LREC 2012,* Istambul, Turkey.