

MISTRAL+: A *Melody Intonation Speaker Tonal Range* semi-automatic Analysis using variable Levels

Benoît Weber, Geneviève Caelen-Haumont, Binh Hai Pham, Do-Dat Tran

MICA Institute, HUST - CNRS/UMI 2954 - Grenoble INP

Hanoi University of Science and Technology, 1 Dai Co Viet St., Hanoi, Vietnam

(benoit.weber, genevieve.caelen, hai-binh.pham, do-dat.Tran)@mica.edu.vn

Abstract

This paper presents MISTRAL+, the upgraded version of an automatic tool created in 2004 named INTSMEL then MELISM. Since MELISM, the entire process has been modified in order to simplify and enhance the study of languages. MISTRAL+ is a combinaison of two modules: a Praat plugin MISTRAL_Praat, and MISTRAL_xls. For specific corpora, it performs phonological annotation based on the F0 variation in prominent words, but also in any chunk of speech, prominent or not. So this tool while being specialized can also be used as a generic one. Now among others, new functionalities allow to use API symbols while labeling, and to provide a semi-automatic melodic annotation in the frame of tonal languages. The program contains several functions which compute target points (or significant points) to model F0 contour, perform automatic annotation of different shapes and export all data in an xls file.

In a first part of this paper, the MISTRAL+ functionalities will be described, and in a second part, an example of application will be presented about a study of the Mo Piu endangered language in the frame of the MICA Au Co Project.

Keywords

Automatic annotation and segmentation, tonal languages, expressive and emotional speech.

1. Introduction

Today, many efficient tools exist to perform speech analysis, each with its own distinctiveness. Some more generic are dedicated to annotation, labeling, and the study of prosody parameters: PRAAT [1], WINPITCH [2]. Some based on perception propose a transcription of melody, stylized, and related to tiers of speech descriptions: PROSOGRAM [3]. Also based on perception, MOMEL-INTSINT [4], using continuous F0 parameters, calculates F0 targets, then supplies an automatic annotation grounded on the speaker register (minimum / medium / maximum), on relative values (higher / lower) too in comparison with the previous target, and thirdly provides information about weaker F0 variations.

The tool we propose is called MISTRAL+, standing for *Melody Intonation Speaker Tonal Range Automatic analysis using variable Levels*. As its previous versions INTSMEL then MELISM [5], MISTRAL+ being their upgraded version, is dedicated on the analysis of the expressive speech in any language, enabling to study the weak and significant F0 modulations and durations at the lexical level. There are many works relating to the different perspectives of such a study, for instance [6, 7]. The new version not only enables the study of the expressive speech of any language (accented, stressed, tonal or not) at the lexical/intonation levels, but also at the acoustic level, the specific melody of tonal languages. For specific corpora, MISTRAL+ performs phonological annotation enabling to:

- give automatically quantitative information (F0 value) about the speaker's targets:

1° in several units (semi-tones, Hz),

2° in several scales (2 to 9 levels), and

3° in several graphic transcriptions (letters or numbers),

- propose on these grounds an automatic melodic annotation system and stylization, independent from the phrase, sentence, corpus, speaker, gender or age,
- supply F0 absolute levels according to the speaker's melodic range,
- phonologically describe in detail the internal structure of prominent words, or any chunk of speech segmented as user wishes it,
- compare their different patterns in function of their syntactic, semantic, pragmatic status, and phonologic, 'tonologic' or prosodic values,
- get a semi-automatic tonal annotation (tonal languages),
- automatically fill up an excel file corresponding to the user specifications,
- compare the different melodic systems existing in different languages,
- objectively test new theoretical hypotheses about the experimental procedures.

2. Presentation of MISTRAL+

MISTRAL+ is composed of two different modules: MISTRAL_PRAAT and MISTRAL_xls.

2.1. MISTRAL+ _Praat

The first module of MISTRAL+ is a plugin integrated to the tool PRAAT [1]. MISTRAL_Praat has been designed in order to facilitate the linguistic tasks of studying any language. All the efforts have been done in order to save

time: 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, there exists a dedicate script Praat.

The functionalities MISTRAL_Praat are included in a process which involves manual interventions of the linguist, such as annotation, or modification of manipulation files. These manual steps are part of the study and justify the split of MISTRAL_Praat into these specific parts.

The first step of the language study is done manually. From the speech file, the user segments it into words and/or phonetic items 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 automatic labeling and segmentation, tier for comments...).

2.1.1. Extraction and Approximation

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

The Praat script extracts the F0 values of the signal every 10ms 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 his speaker or otherwise he can use default values: 75Hz and 600Hz as a first step. From the values extracted and the user parameters, the MOMEL algorithm [4] approximates the F0 variation of the signal. Several 'target points' are created and represents specific changings in the F0 variation. The resulting curve is a sequence of entirely sonorant segments and it constitutes the macro-prosodic component. From these calculations results a manipulation file, which can be used, modified and studied in Praat.

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. In the Manipulation file, the user indeed modifies and adapts the target points of the F0 approximation in time and frequency till he perceives it in the same way than the original sound file.

To each boundary he had in his textgrid, the user may associate a new target point in the F0 approximation of the manipulation file 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 more effeciently the language.

2.1.2. Synchronization

As the annotation and modification of manipulation file are done manually, the user may have created unwanted issues that will impact the melodic annotation (see 2.1.3). While the user is modifying the frequency height of the wrong target point, he may add accidently a time delay between the target and the corresponding boundary. In fact this problem occurs very frequently in the Praat TextGrids. The user has thus the possibility via the second functionality of 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: 5ms, 10ms, or 20 ms. The Praat script will then modify all of the target points which are in the range chosen by the user. Only the time position of the target point is modified, its frequency level is kept in order to not modify the perception.

2.1.3. Melodic automatic annotation

At this point, the speech signal has been approximated as a manipulation file and has been annotated in a textgrid. The third functionality of MISTRAL_Praat bases its calculations on these two files.

It computes melodic information in Hertz and semi-tones, duration, and derivative enabling the linguist to study more precisely the language according to the segmentation he has done.

Following the French tradition, since Delattre 1966 [8], the speaker's range is traditionally divided into 4 levels. On the other hand, the studies about tonal languages are most of the time using 5 levels with a numeric scale. 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 differents levels according to the language and the kind of speech he is studying. 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.

The parameterization of the number of levels leads in an improvement of the precision in the description of the F0 great or more subtle modulations: it enables a more accurate study of shapes of prominent words or tones for instance in the case of tonal languages. The user is free to adapt the number of levels according to his particular task.

The user can also define the frequency range (Fmin, Fmax) and the mean frequency, if he has already collected data on his speaker. If it is his first study, he may choose default values.

Once he has defined the number of F0 levels, and the frequency range values(minimum, mean, maximum) 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 differents 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 hertz and semi-tones, and the frequency range values. This file enables the linguist to quickly find necessary information he needs to interpret the data delivered in the text grid and in the excel file.

The textgrid is based on the file made by the user during the annotation and segmentation step. Six tiers may be added (depending on the choice of the user). The tier ‘MESLIM’ gives the level value of each target points contained in the manipulation file. The M-Tunes tier describes the variations of levels between two target points. Depending to the kind of segments (and/or kind of language, and/or study) used, this information may concern either with the melodic modulation on the whole words, or with the voiced segment tones. To give a more accurate description, MISTRAL+ tools make a distinction between intra-level variations, and plateaux. By opposition to the plateaux which by definition are not oriented, intra-level variations are annotated with the symbols +/- (e.g.: mm+; ss-, etc.), which codes the existence of a weak rising vs. falling variation of the slope.

	A	S	H	e	m	c	b	i	g
A	AA	AS	AH	ae	am	ac	ab	ai	ag
S	SA	SS	SH	se	sm	sc	sb	si	sg
H	HA	HS	HH	he	hm	hc	hb	hi	hg
e	eA	eS	eH	<i>ee</i>	<i>em</i>	<i>ec</i>	<i>eb</i>	<i>ei</i>	<i>eg</i>
m	mA	mS	mH	<i>me</i>	<i>mm</i>	<i>mc</i>	<i>mb</i>	<i>mi</i>	<i>mg</i>
c	cA	cS	cH	<i>ce</i>	<i>cm</i>	<i>cc</i>	<i>cb</i>	<i>ci</i>	<i>cg</i>
b	bA	bS	bH	<i>be</i>	<i>bm</i>	<i>bc</i>	<i>bb</i>	<i>bi</i>	<i>bg</i>
i	iA	iS	iH	<i>ie</i>	<i>im</i>	<i>ic</i>	<i>ib</i>	<i>ii</i>	<i>ig</i>
g	gA	gS	gH	<i>ge</i>	<i>gm</i>	<i>gc</i>	<i>gb</i>	<i>gi</i>	<i>gg</i>

Table 1: The matrix of the melodic syllables used for the description of lexical F0 configurations, and especially for prominent high values (bold characters). The duplicated targets (ex: /HH/) are corresponding to a plateau.

The combination of these 9 levels results in 81 melodic syllables, which takes into account the dynamic dimension from a previous F0 target to the following one i.e. the melodic space between two successive F0 level (see Table 1 above).

For each boundary of the textgrid made beforehand by the user, the tier ‘Hertz’ gives the value in hertz of F0 and the tier ‘Semi-Tones’ gives its value in semi-tones (ST0, STmin, STmax) using Fmid as shown in Equation 1 below:

$$ST_i = 12 \times \log_2 \frac{F_i}{F_{mid}}$$

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

An example of the textgrid once it has been completed by MISTRAL+ is given in the Figure 1 below.

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 (see 4. MISTRAL_xls).

3. Mistral_xls

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. Once the macro is done, he can then apply it quickly and massively on several files.

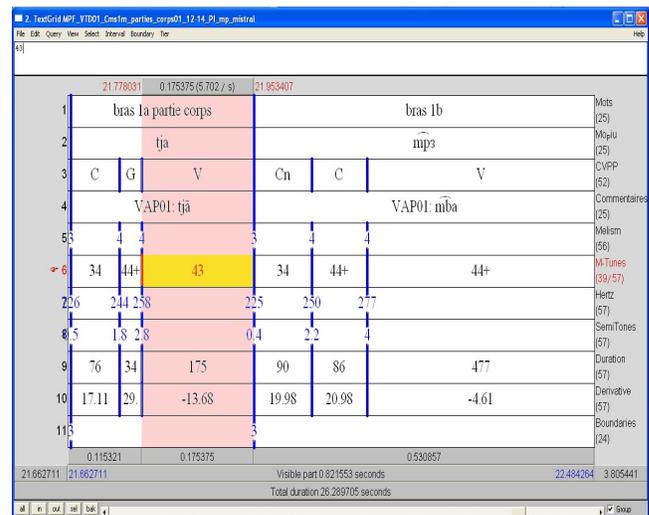


Figure 1: The word “arm” in Mo Piu and its tonal information in a textgrid

First, the user defines the format and order of the results he wants to obtain in excel sheet(s). Most of the input parameters are based on the files and tier names exported from MISTRAL_Praat. Once the sheet structures of the report are defined, the user can run the VBA module by clicking on the MICA logo (see Figure 2 below). The user defines which files he wants to use and which reports he wants to obtain from the selected files. After supplying these information, the user can generate the reports in an excel file using the definition table. Some different studies have been undertaken based on this tool for instance in the domain of the Mo Piu language [9, 11, 12, 13].

As Praat saves Unicode in Big Endian order which cannot be read properly in Excel, if needed, the procedure converts the exported result from UTF-16BE to UTF-16LE.

4. One application among others

Until now different applications have been presented at the international level in the various fields of linguistic

	A	B	C	D	E	F	G	H
	Click on the MICA's logo to run →							
								
1								
52	*M-Tunes*	2	1				Syll mel	6
53	*Hertz*	2	1				Hz	9
54	*M-Tunes*	2	2				Début	11
55	*M-Tunes*	2	3				Fin	12
56	*Derived*	2	1				Derivee	7
57	*Duration*	2	1				Duree	8
58	Sheet Name	Phonetic_5	Base Table	*Evaluation*	Base column Index	2	Additional Function	Correct_F0
59	Tier Name	Base Column Index	Select Column Index	Search Text	Replacement	Use Last Value	Column Name	Column Index
60	[Filename]		1			YES	N° fichier	1
61	[Filename]		2			YES	Locuteur	2
62	*Mots*	2	1				Mots	3
63	*Mo_Piu*	2	1			YES	Mo Piu	4
64	*CVPP*	2	1			YES	Classes	5
65	*Correction*	2	1				Correction	6
66	*FR-MA_aut*	2	1				Ph autom	7
67	*Evaluation*	2	1				Evaluation	8
68	*Evaluation*	2	2				Début	9
69	*Evaluation*	2	3				Fin	10
70	*Commentaires*	2	1				Commentaires	11
71	*Melism*	2	1				Tons frontières	12
72	*M-Tunes*	2	1				Tons	13
73	*Hertz*	2	1				Hertz	14
74	*SemiTones*	2	1				Demi-tons	15
75	*Duration*	2	1				Durée	16
76	*Derived*	2	1				Dérivée	17
77	*Boundaries*	2	1				Frontières	18

Figure 2: The VBA module

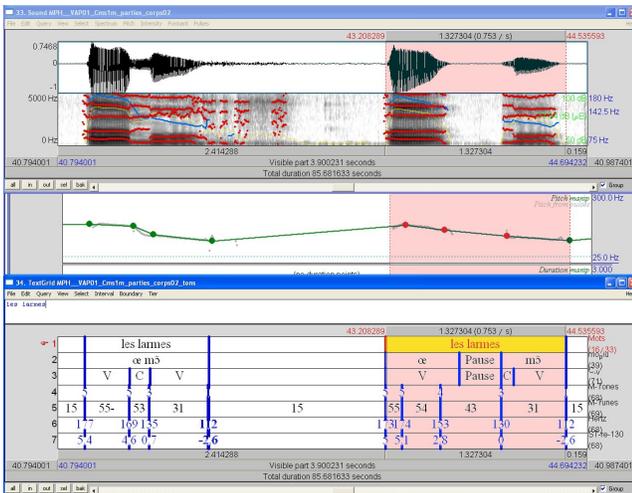


Figure 3: Three Praat windows : 1° above, the speech signal, spectrogram with formants and F0 contour 2° in the middle, the Manipulation file with the stylized F0 line and targets points, 3° below the manual and automatic tiers.

research that this tool enables to explore: for instance in the expressive speech and prominent words domain (French [6], and a Bantu tonal language, the Bàsàà [7]), in the description of the unknown language Mo Piu without writing (the Mo Piu people is an endangered minority, living in the mountains of North Vietnam). These studies are concerning first the domains of phonetic and tonal investigation, and secondly the automatic

phonetic annotation procedures for under-resourced languages. These studies are developed in the frame of a MICA Project (Au Co) which aims to study the south-east Asian ethnic languages with these two purposes. A paper in the same proceedings [13] is developing this last field of research showing a comparison between an automatic phonetic annotation using several models of language, and its assessment by a phonetician. The Figure 3 above presents an example of the MISTRAL+ procedure concerning its help in the tonal study. The Praat Textgrid is composed from bottom to top, of 4 automatic tiers:

- a- the F0 semi-tones values of the targets points,
- b- their conversion in Hz,
- c- the numeric melodic range of the targets ('melodic syllables'),
- d- the melodic level of the target,

and in this example, of 3 manual annotation tiers:

- a- the words in French,
- b- the transcription of phonetic Mo Piu items using IPA symbols,
- c- the vowel / consonants segmentation in phonetic classes.

of phonetic Mo Piu items using IPA, c- vowel / consonants segmentation).

The Figure 3 above presents the repetition of the same words (*les larmes / the tears*) by the same male speaker with and without a pause between the article and the noun. The semi-automatic tonal annotation allows describing the vowel tones on a 5 levels scale as such: 1° for the article as /55-/ or /54/ 2° for the noun as /31/. This procedure allows

getting the tone values in a quasi-real time. It spares a manual annotation taking, as everyone knows it, a lot of time, and avoids annotation mistakes.

Then as explained above in the paragraph 2, all the manual and automatic annotations and labels of the tiers are exported to an xls file for statistical use (Figure 4 below).

Items number	Models of languages	Speaker	Word number	French	Mo Piu	Phonetic Classes	Manual correction	Automatic labeling	Manual assessment	Item beginning	Item end	Item Duration	Tone at the left boundary	Tones in 5 levels	Tones in Hertz	Semi-tones	Comment
6784	FR-VN	VAP01	41	PP			PP %	PP	B	50.38	51.03		3	33-	120		
6785	FR-VN	VAP01	41	PP			PP %	l	F	51.03	51.53		3	33-	120		
6786	FR-VN	VAP01	41	souffle 2a	la	C	l %*%-%	l	B	51.53	51.55		3	34	125		
6787	FR-VN	VAP01	41	souffle 2a	la	C	l %*%-%	ø	F	51.55	51.68		3	34	125		
6788	FR-VN	VAP01	41	souffle 2a	la	C	l %*%-%	ɔ	F	51.68	51.72		3	34	125		
6789	FR-VN	VAP01	41	souffle 2a	la	V	ɑ %*-%	ɔ	V	51.72	52.07		4	43	145		
6790	FR-VN	VAP01	41	souffle 2a	la	V	ɑ %*-%	PP	F	52.07	52.09		4	43	145		
6791	FR-VN	VAP01	41	PP		PI	PP %*-%	PP	B	52.09	52.72		3	34	119		
6792	FR-VN	VAP01	41	PP		PI	PP %*-%	ɔ	F	52.72	52.74		3	34	119		
6793	FR-VN	VAP01	41	souffle 2b	pā	C	p %	ɔ	F	52.74	52.77		4	44-	141		a ou ā ? , V
6794	FR-VN	VAP01	41	souffle 2b	pā	Vn	ā %*-%	ɔ	V	52.77	53.08		4	43	139		a ou ā ? , V
6795	FR-VN	VAP01	41	souffle 2b	pā	Vn	ā %*-%	f	F	53.08	53.12		4	43	139		a ou ā ? , V
6796	FR-VN	VAP01	41	PP			*% -PP %*-%	f	A	53.12	53.21		3	33-	118		
6797	FR-VN	VAP01	41	PP			*% -PP %*-%	PP	B	53.21	53.90		3	33-	118		
6798	FR-VN	VAP01	41	PP			*% -PP %*-%	n	F	53.90	53.93		3	33-	118		
6799	FR-VN	VAP01	41	souffle 3a	la	C	l %*-%	n	V	53.93	54.05		3	35	113		la ou la ?
6800	FR-VN	VAP01	41	souffle 3a	la	C	l %*-%	a	F	54.05	54.08		3	35	113		la ou la ?
6801	FR-VN	VAP01	41	souffle 3a	la	V	*% -ɑ %*-%	a	V	54.08	54.14		5	53	173		la ou la ?
6802	FR-VN	VAP01	41	souffle 3a	la	V	*% -ɑ %*-%	ɔ	V	54.14	54.35		5	53	173		la ou la ?
6803	FR-VN	VAP01	41	souffle 3a	la	V	*% -ɑ %*-%	f	F	54.35	54.39		5	53	173		la ou la ?
6804	FR-VN	VAP01	41	PP		PI	PP %*-%	f	A	54.39	54.86		3	34	117		
6805	FR-VN	VAP01	41	PP		PI	PP %*-%	PP	B	54.86	55.14		3	34	117		
6806	FR-VN	VAP01	41	PP		PI	PP %*-%	h	A	55.14	55.17		3	34	117		
6807	FR-VN	VAP01	41	souffle 3b	pā	C	p %*-%	h	F	55.17	55.20		4	44-	166		a ou ā ? , V
6808	FR-VN	VAP01	41	souffle 3b	pā	C	p %*-%	r	#	55.20	55.20		4	44-	166		a ou ā ? , V
6809	FR-VN	VAP01	41	souffle 3b	pā	Vn	ā %*-%	r	F	55.20	55.43		4	43	166		a ou ā ? , V
6810	FR-VN	VAP01	41	souffle 3b	pā	Vn	ā %*-%	r	F	55.43	55.47		4	43	166		a ou ā ? , V
6811	FR-VN	VAP01	41	PP		PP	*% -PP	r	F	55.47	55.51		3	FIN	118		
6812	FR-VN	VAP01	41	PP		PP	*% -PP	PP	B	55.51	56.80		3	FIN	118		

Figure 4: The report of the PRAAT TextGrid tiers data, shaped as the user wishes, generated by MISTRAL via the VBA module

5. CONCLUSION

This paper presents MISTRAL+, a speech tool enabling to automatically annotate the melodic targets in compliance with the user needs. Being independent from any linguistic theory, it can be applied to any language, accented, stressed, tonal or not, or any kind of research. It concerns the study as well of the prominent words giving useful information about the F0 micro and macro-modulations in such words, as of the ordinary, emotional, or attitudinal speech. In the phonetic and/or the tonal domains, it enables to semi-automatically annotate the tones levels, *even though and especially* the language is still unknown, saving a great deal of time.

Its applications developed in the frame of the Au Co project and the Mo Piu language encompasses fundamental linguistic research as well as the automatic annotation procedures performed on the ‘Mo Piu’ language.

6. Acknowledgements

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

7. References

- [1] Boersma, P. Weenink, D. (2011). Praat : doing phonetics by computer, Version 5.2.22, downloaded 15/04/2011 from <http://www.Praat.org/>.
- [2] Martin Ph. (1996). WinPitch: un logiciel d'analyse temps réel de la fréquence fondamentale fonctionnant sous Windows, *Actes des XXIV Journées d'Etude sur la Parole*, Avignon, pp. 224--227.
- [3] Mertens, P., d'Alessandro, Ch. (1995). Pitch contour stylization using a tonal perception model. In *Proc. Int. Congr. Phonetic Sciences 13*, 4, Stockholm, Sweden, pp. 228—231.
- [4] Hirst D.J., Di Cristo, A. (1998). A survey of intonation systems”, in Hirstand Cristo (Eds), *Intonation Systems: a Survey of Twenty Languages*. Cambridge: Cambridge University Press, pp. 1--44.
- [5] Caelen-Haumont, G., Auran, C. (2004). *The phonology of melodic prominence: the structure of Melisms*. In *Proceedings of Speech Prosody 2004*, Nara, Japon, pp. 143--146.
- [6] Caelen-Haumont G. (2007). *Emotion, emotions and prosodic structure : an analysis of the melisms patterns and statistical results in the spontaneous discourse of 4 female speakers from four generations*,

- book chapter, in Sylvie Hancil ed., *The Role of Prosody in Affective Speech*, Peter Lang, li97, pp. 95--138.
- [7] Makasso, E.-M. (2008). Intonation et mélismes dans le discours oral spontané en Bàsàa. Thèse de doctorat, Université de Provence, Aix-Marseille I.
- [8] Delattre, P. (1966). Les Dix Intonations de base du français. *The French Review*, Vol. 40, No. 1, pp. 1--14.
- [9] Caelen-Haumont G., Cortial B., Culas C., Hong T. D., Lê T. X., Nguyen T. N., Pannier E., 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 Technologies for under-resourced Languages*. In *Proceedings of SLTU'10*, Penang, Malaysia, pp. 42--50.
- [10] 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.
- [11] 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*, Shanghai, China.
- [12] 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.
- [13] Caelen-Haumont, G., Sam, S. (2012 to be published), Comparison between two models of language for the automatic phonetic labeling of an undocumented language of the South-Asia: the case of Mo Piu. In *Proceedings of LREC 2012*.