

MISTRAL+: DEDICATED TOOL FOR UNDER-RESOURCED LANGUAGES ANALYSIS

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

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, do-dat.tran, hai-binh.pham}@mica.edu.vn

ABSTRACT

This paper presents MISTRAL+, a dedicated tool for the study of under-resourced languages. MISTRAL+ is the upgraded version of an automatic tool created in 2004 called MELISM. The entire process has been modified in order to simplify and enhance the study of under-resourced languages. MISTRAL+ is composed of two separated modules: MISTRAL_Praat a plugin integrated to the tool PRAAT, and MISTRAL_xls a VBA module.

MISTRAL_Praat enables the creation of an approximation of the signal that is studied, it performs an automatic tonal annotation and exports all data in a xls standard file. Using MISTRAL_xls, the user is able to easily and quickly extract from the data generated by MISTRAL_praat the information he needs for his study.

In the first part, MISTRAL+ and its main functionalities will be presented. In the second part, a closer look will be put on MISTRAL_praat. The third part will describe the second module MISTRAL_xls. The last part will present the study done using MISTRAL+ on the study of the Mio Pu, a vietnamese under-resourced language.

Index Terms— F0 range, Phonetic IPA labeling, automatic annotation and segmentation, tonal languages, expressive 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, using continuous F0 parameters, MOMEL-INTSINT [4] calculates F0 targets and supplies a semi-automatic annotation grounded on the speaker register (minimum / medium / maximum) and on relative values

(higher / lower) in comparison with the previous target, and it also provides information about F0 variations.

MISTRAL+, standing for Melody Intonation Speaker Tonal Range Analysis using variable Levels is the upgrade of the MELISM tool [5]. As its previous version, MISTRAL+ is dedicated to the analysis of the expressive speech in any languages, enabling to study the weak significant F0 modulations and durations at the phonetic, melodic and lexical levels. Many works relating to the different perspectives of such a study have been presented at the international level, for instance [6, 7, 8, 9, 10].

In fact expressive speech and tonal speech are grounded on the same need: the study of weak F0 modulations. Thus the new version not only enables the study of the expressive speech of any languages (accented, stressed, tonal or not) at the lexical/intonation levels, but also at the acoustic level, the specific melody of tonal languages. MISTRAL+ is a tool enabling on one hand to take into account the weak F0 modulations and on the other hand to proceed stylization and simplification of the F0 curve.

2. MISTRAL+

MISTRAL+ is a combinaison of two modules: a plugin praat called MISTRAL_praat, and a macro excel named MISTRAL_xls. For a specific corpus, it performs phonological annotation enabling to:

- give automatically quantitative information (F0 value) about the speaker's targets issued from the segmentation levels or from a change in the F0 variation 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, tonal patterns, 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.

3. MISTRAL_PRAAT

The first module of MISTRAL+ is a plugin integrated to the tool PRAAT [1]. All the functions performed by its previous versions remain but have been updated and enhanced. Some new functionalities have also been added. MISTRAL_Praat is designed in order to facilitate the linguistic tasks of studying under-resourced languages. 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 tonal annotation based on the segmentation done by the linguist. To each of these functionalities exists a specific script praat.

MISTRAL_Praat is 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 and from the speech file. The user proceeds to the segmentation and annotation. He segments the file into words and/or phonetic items and annotates it in a textgrid. MISTRAL_Praat supports the International Phonetic Alphabet. The user may have added other tiers in order to help him in his study of the under-resourced language. For instance a specific tier for the words’ translation, a tier where he segments the speech into phonetic units, or also a tier for comments.

3.1. Extraction and Approximation

Once the annotation has been completed, the user launches MISTRAL+. The first functionality of MISTRAL_Praat enables the user 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 the user is studying. The extraction is processed every

10ms on the speech signal using an accurate autocorrelation method supplied by PRAAT. The user defines the floor and ceiling values based on the data he has already collected about his speaker or he can use default values: 75Hz and 600Hz. From the values extracted and the user parameters, the MOMEL algorithm [4] approximates the F0 variation of the signal. The aim is to keep most of segmental characteristics of the signal. Several ‘target points’ are created and represent specific changes in the F0 variation. The resulting curve is a sequence of entirely sonorant segments and constitutes the macro-prosodic component. From these calculations a manipulation file is created. It can be easily used, modified and studied in PRAAT.

The approximation done by the system aims to fit as good as possible the original signal. But at this point a manual step is needed before the study can go further. The manipulation 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 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 will need to study more efficiently the language. An example is given in figure 1 below:

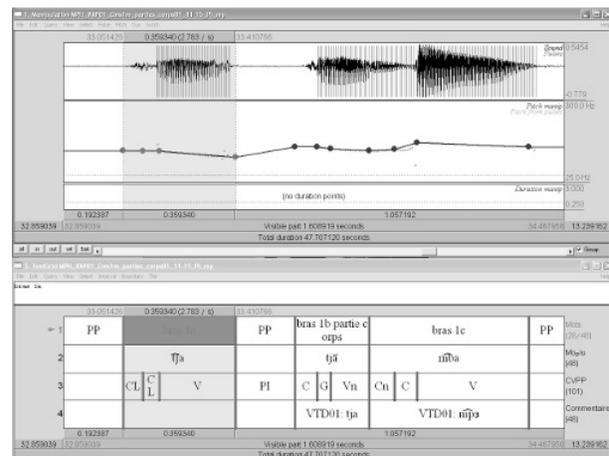


Figure 1. the word “arm” in Mo Piu in a Manipulation and its textgrid after segmentation and correction

3.2. Synchronisation

Once this annotation is completed, and once the manipulation file has been modified in order to be as

close as possible from the original sound, the user may want to use the second functionality of *Mistral_praat*.

As the annotation and modification of manipulation file are done manually, the user may have created unwanted issues that will impact the tonal annotation (see 3.3). While the user is modifying the frequency height of the wrong target point, he may add accidentally a time delay between the target and the corresponding boundary. The user has thus the possibility via the second step 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 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.

3.3. Tonal 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.

This functionality computes several information enabling the linguist to study the language according to the segmentation he has done: such tonal information in hertz and semi-tones, duration, and derived.

According to the French tradition, since Delattre 1966 [11], 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 different 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: Fmin = 140, Fmean = 200, Fmax = 300.

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 hertz and semi-tones, and the frequency values (minimum, mean, maximum). This file enables the linguist to quickly find the information he needs to interpret the data delivered in the text grid and in the excel file.

The *textgrid* is based on the *textgrid* 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. To give a more accurate description, MELISM then *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 ‘Derived’ 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 2.

Time	Segment	Hertz	SemiTones	Duration	Derived
0.000	bras 1a	132	-8.8	0.5085	0.83
0.130	Pause				
0.130	bras 1b partie corps	109	-12.1	0.191	25.08
0.321	End				

Figure 2. The word “arm” in Mo Piu and its tonal information in a textgrid

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).

4. MISTRAL_XLS

MISTRAL_xls is a excel VBA module. It enables to create easily macros on the data contained in the xls file exported from the module MISTRAL_praat.

Thanks to MISTRAL_xls, the user saves 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.

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. 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. This kind of table is shown in figure 3 below:

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	V	L	C	D	E	F	G	H	I	J	K	L	M
2	MPI	VAPU	Mos	mo pi	classes	Syll mel	Devee	Duree	Pt	Niveau	Debut	Fin	Commentaires
3	MPI	VAPU	bras 1a	tja	CL	33	-3.19	63	132	3	33.85	33.11	
4	MPI	VAPU	bras 1a	tja	CL	33	0	53	130	3	33.11	33.17	
5	MPI	VAPU	bras 1a	tja	V	32	-12.31	244	130	3	33.17	33.41	
6	MPI	VAPU	bras 1b partie corps	tja	PI	24	25.88	191	109	2	33.41	33.60	
7	MPI	VAPU	bras 1b partie corps	tja	C	44	0	72	145	4	33.60	33.67	VTD01: tja
8	MPI	VAPU	bras 1b partie corps	tja	G	44	-18.85	42	145	4	33.67	33.72	VTD01: tja
9	MPI	VAPU	bras 1b partie corps	tja	Va	43	-6.44	124	138	4	33.72	33.84	VTD01: tja
10	MPI	VAPU	bras 1c	mha	Ca	34	7.3	82	132	3	33.84	33.92	VTD01: mha

Figure 3. Report generated by MISTRAL_xsl according to the macro the user has created

- as Praat saves Unicode in Big Endean order which cannot be read properly in Excel, if needed, the procedure may convert the exported result from UTF-16BE to UTF-16LE.

5. ONE APPLICATION AMONG OTHERS

Until now different applications have been presented at the international level in the various fields of linguistic research that this tool enables to explore: for instance in the expressive speech and prominent words domain (French and a Bantu tonal language [6, 7], Bàsàa [12]), in the description of an unknown language, the Mo Piu one [8, 9, 10], both in the domain of phonetic and tonal investigation, or of the automatic phonetic annotation procedures for under-resourced languages. A paper [8] 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 5 below presents the final textgrid of the same speech extract than above (1st repetition of the word), by a Mo Piu male speaker with a pause between the 2 parts of the noun (/bras 1a / Pause / bras 1b partie corps/). The semi-automatic tonal annotation allows describing the vowel tones on a 5 levels scale as such: 1° for the vowel of 1st part of the noun (/bras 1a/ as /11- 2° for the 2nd part of the noun (bras 1b partie corps/) as /11-/ once more. 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 sent to automatically fill up an xls file for statistical use.

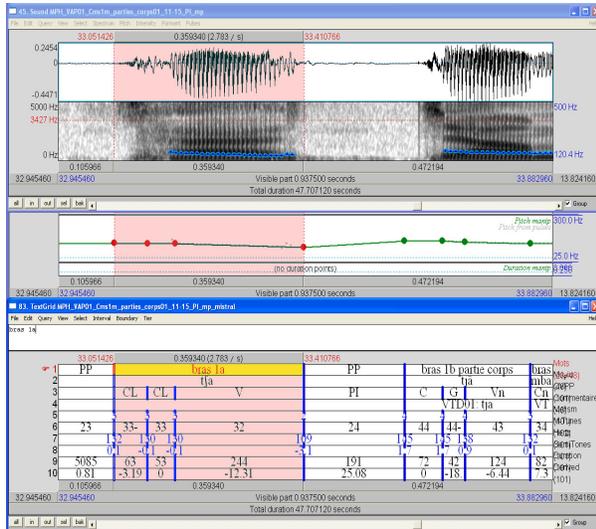


Figure 5. Three Praat windows: 1° above, it represents the speech signal and spectrogram showing the formants and the F0 contour 2° in the middle, the ‘Manipulation’ one with the stylized F0 line and targets points, 3° below from top, the manual annotation tiers (a- words translated in French, b- the transcription of phonetic Mo Piu items using IPA, c- vowel / consonants segmentation), then the automatic tiers (a- the F0 semi-tones values of the targets points, b- the dynamic inter-targets space or melodic syllable, c- the conversion of the target value in Hz, d- their conversion in semi-tones, e- the duration, f- the derived of the melodic syllables).

5. CONCLUSION

This paper presents MISTRAL+, a tool enabling the automatically annotation of melodic targets in compliance with the linguist 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, an endangered minority of North Vietnam, still unknown and without writing.

6. REFERENCES

- [1] Boersma P. and 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, 224-227.
- [3] Mertens P., and d'Alessandro Ch., 1995, "Pitch contour stylization using a tonal perception model". *Proc. Int. Congr. Phonetic Sciences 13*, 4, 228-231 (Stockholm 1995)
- [4] Hirst D.J. and Di Cristo A., 1998, "A survey of intonation systems", in Hirst and Di Cristo (Eds), *Intonation Systems: a Survey of Twenty Languages*. Cambridge: Cambridge University Press, 1-44.
- [5] Caelen-Haumont G., and Auran C., 2004, *The phonology of melodic prominence: the structure of Melisms*. In *Proceedings of Speech Prosody 2004* (Nara, Japon, March 23 – 26, 2004), 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, 95-138.
- [7] Caelen-Haumont G., Cortial B., Culas C., Hong T. D., LêT. 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, SLTU'10*, (Penang, Malaysia, May 3 – 5, 2010), 42-50.
- [8] Caelen-Haumont G. and Sam S., 2012, Confrontation of two models of language for the automatic phonetic labeling of an unknown ethnic language of the South-Asia: the case of Mo Piu, *Proceedings of LREC 2012*, (to be published).
- [9] Caelen-Haumont, G., Vittrant, A., 2011 : Transcrire, écrire et formaliser en analyse phonétique et tonale : l'exemple d'une langue d'Asie du Sud-est non documentée (mo piu), *Actes du 25^e Colloque du Cercle Linguistique du Centre et de l'Ouest (CERLICO)*, Presses Universitaires de Rennes (to be published).
- [10] Caelen-Haumont, G., Sam, S., Castelli, E., 2011 : Automatic Labeling and Phonetic Assessment of the « Mo Piu » Under-Resourced Asian Language:

early results, Proceedings of the International Conference on Asian Language Processing (IALP 2011), Penang, Malaysia, IEEE 260-263.

- [11] Delattre P., 1966, Les Dix Intonations de base du français. *The French Review*, Vol. 40, No. 1, 1-14.
- [12] Makasso E.-M., 2008, « *Intonation et mélismes dans le discours oral spontané en Bàsàa* », Thèse de doctorat, Université de Provence, Aix-Marseille I.