When diachrony is helping for a synchronic study: the case of the Mo Piu tones from the Hmong-Mien family in Northern Vietnam

CAELEN-HAUMONT, Geneviève,

MICA, Hanoi, Vietnam

BARTKOVA, Katarina,

ATIFL, Nancy, France

International Research Institute MICA

Multimedia, Information, Communication & Applications

UMI 2954

Hanoi University of Science and Technology

1 Dai Co Viet - Hanoi - Vietnam

The Mo Piu Project

- **MICA « Au Co » project, and 3 CNRS-ANR projects about Mo Piu language (Appsy, Langues Pi, PEPS April)**
- **The Mo Piu ethny is an *endangered minority of the Northern Vietnam* close to the Chinese border**
  - 2012 : 238 people
  - before our study
    - a branch still unknown of the ***Hmong-Miên family***
    - ***without writing, uncharted, undocumented***
- **Finding out the *Mo Piu phonology and tonology***

Aims of this study

- **to contribute to the knowledge of the Hmong-Mien family languages**
- **to better detect the number and patterns of the Mo Piu tonal system**
  - Focussing on a deeper insight on the *mo piu tonal system*
- **Using and checking 2 dedicated tools: MISTRAL+ and PROSOTRAN**
- **Pilot study: only data recorded by one (LAK01) of our speakers are studied**
  - LAK01 chosen because
    - he knows perfectly his language
    - he has linguistic skills (primary school teacher)

Some linguistic cues to link the Mo Piu language to the Hmong-Miên Family

- For instance, comparison of the Mo Piu numerals with the Proto-Hmong-Mien / Hmongic ones
  - **Ratliff 2010**
  - ***etlingweb.ewa.mpg.de***

Mo Piu

Hmong Nzhuab, China		Dananshan Miao (Hmong Njua), China		Mo Piu, Vietnam *(temporary transcription)*		*Reconstruction Proto-Hmong-Miên*
1. ʔi^{54}	6. ʈou^{44}	1. ʔi^{43}	6. ʈəu^{44}	1. ɛ [æ]¤	6. ʈɔ¤	1. *ʔɨ
2. ʔau^{43}	7. ɕaŋ^{44}	2. ʔau^{43}	7. ɕaŋ^{44}	2. wa¤	7.dʑãŋ¤	2. *ʔu̯i
3. pei^{54}	8. ʑi^{22}	3. pe^{43}	8. ʑi^{24}	3. pa¤	8. ɟi¤	3. *pjɔu
4. plou^{54}	9. tɕuɑ^{42}	4. pləu^{43}	9. tɕua^{31}	4. plɔ¤	9. tɕo¤	4. *plei
5. tʃɻ^{54}	10. kou^{22}	5. tʂi^{43}	10. kəu^{24}	5. p(h)i¤	10. khɔ [ə]	5. *pra
						10. *gju̯ɛp

List of Prototones (Ratliff 2010)

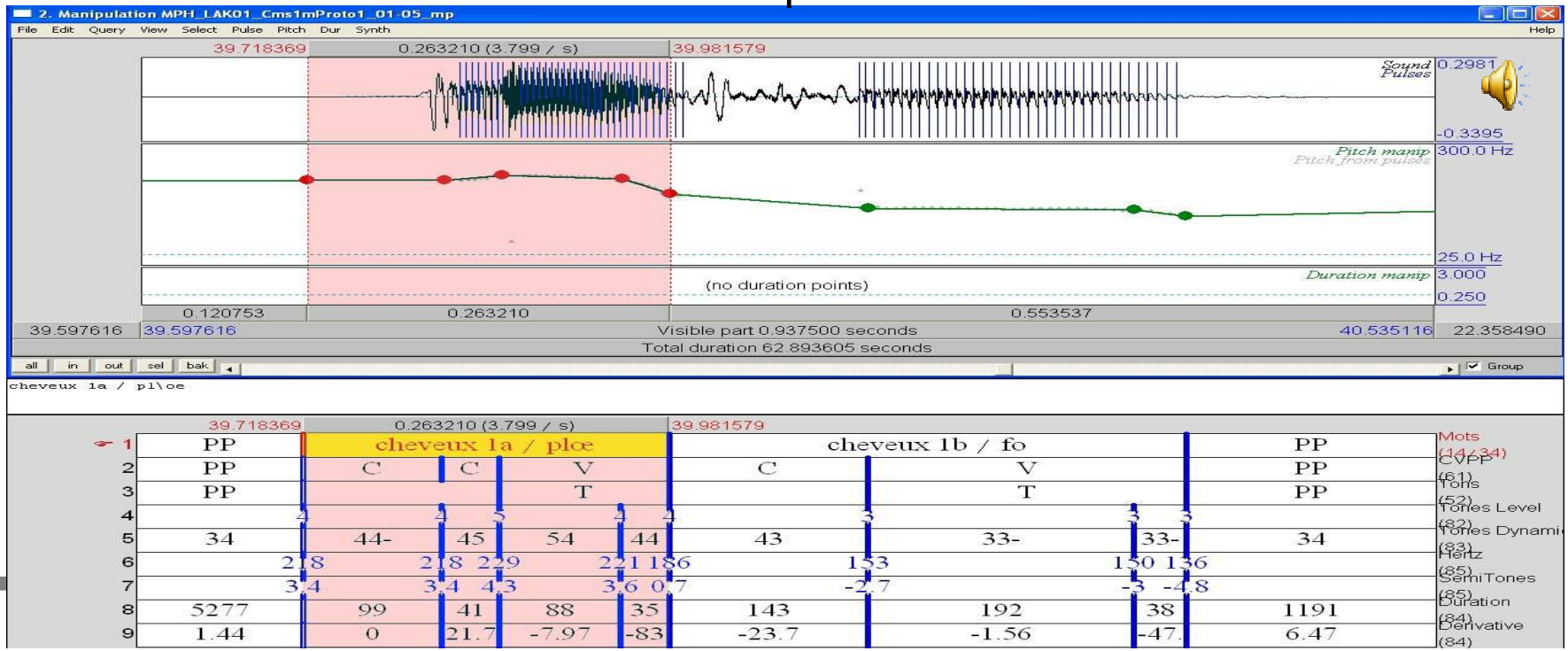
- **In this context of linguistic reconstruction, a *list of 8 prototones* was reconstructed**
  - Originally based on the Calmsea list
  - Use of 2 versions of the prototones list
    - ***Extended version***
      - 206 words x 3 repetitions x 2 speakers = 1236 items
      - Composed of words from the prototones list + speaker's addings (words semantically linked) during recording sessions
      - 1752 tonal sequences from 808 vowels
      - Here only one speaker's (LAK01) data are analysed
    - ***Restricted version (Martha Ratliff's one)***
      - 175 words x 3 repetitions x 2 speakers = 1050 items
      - But the 175 French words give rise to Mo Piu single or compound words → selection of the true item / true tone
      - the mo piu compound words
        - Periphrases (word-to-word translation): most of the cases is a loan, therefore divergent from the prototypical pattern
        - compound word with classifier or not
  - ***2 versions of the prototones lists → 2 studies***

About the history of the 8 prototones

- **Historically, the 8 prototones issued from 4 tonal patterns (Ratliff, 2010): A, B, C, D**
  - Difference on the first C: – voiced / + voiced
    - Pattern A: **C**V(N)
      - - voiced ➔ Prototone 1
      - + voiced ➔ Prototone 2
    - Pattern B: **C**V?
      - - voiced ➔ Prototone 3
      - + voiced ➔ Prototone 4
    - Pattern C: **C**VH
      - - voiced ➔ Prototone 5
      - + voiced ➔ Prototone 6
    - Pattern D: **C**V(C)
      - - voiced ➔ Prototone 7
      - + voiced ➔ Prototone 8
  - Some other cues on the 1st consonant when voiced
    - Breathy: Prototones 2, 4, 6
    - Creaky: Prototone 8

Example of compound word: *hair ~ cheveux*

- **An example of the mo piu compound word /hair /**
  - /plœ/ = /hair/ standing for /poils ≠ cheveux in French /
  - /fo/ = /head/
  - /plœ/ is the selected word though it is generic (as in English) because…
  - … the tone is in concordance with the 1 (or 2) representative(s) of the tones in the same class of prototones



Study 1: Prototone list extended Comparison of MISTRAL+ and PROSOTRAN

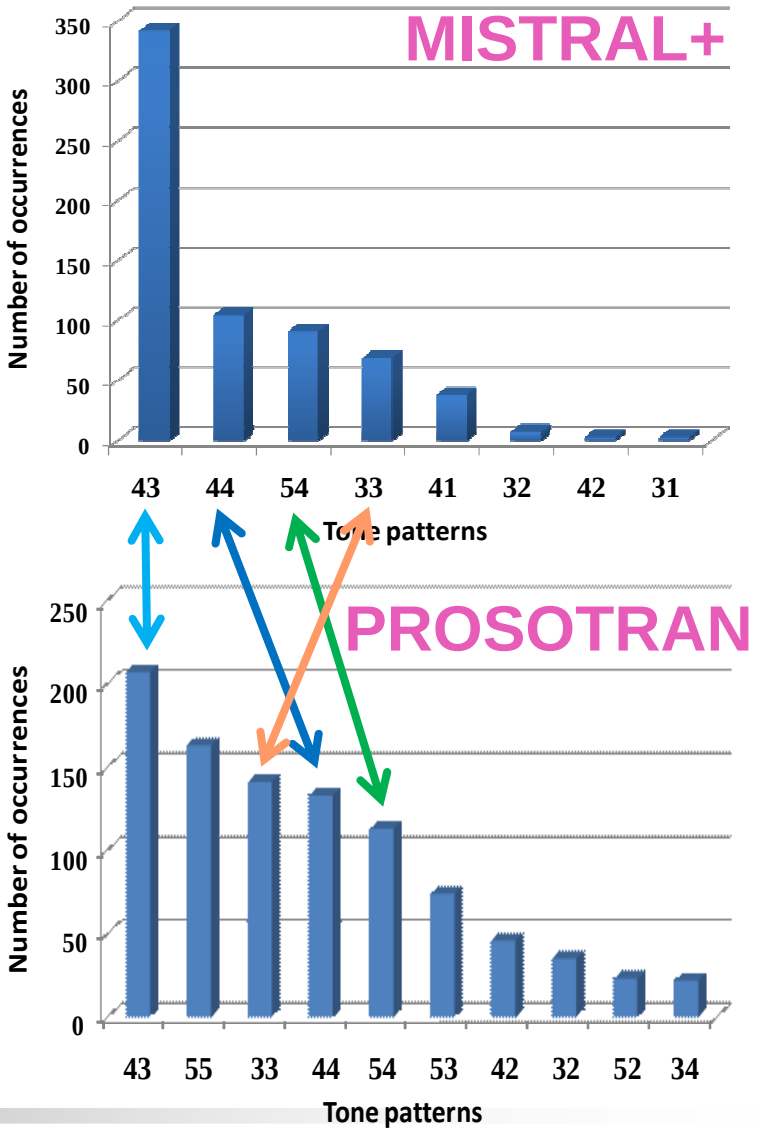
- ***MISTRAL+: a semi-automatic approach*** **under Praat (Weber and all, 2012)**
  - Computing :
    - the speaker tonal range (Hz, semi-tones every 10 ms)
      - choice of the number of levels: here 5 levels
    - the tone values at phonetic and word levels: boundaries and segment (dynamic perspective)
    - the duration of the different items, derivative…
  - Pre-alignment of the melodic curve but manual adjustement is needed, using vision and listening
  - Time synchronization between
    - boundaries of the phonetic items
    - and the rebuilt simplified melodic curve (under Praat Manipulations)
  - Creating an automatic xls file with all the manual and computed data
    - allowing IPA codes

Study 1: Prototone list extended Comparison of MISTRAL+ and PROSOTRAN

- ***PROSOTRAN*** **(Bartkova and all, 2012)**
  - Based on phoneme segmentation (manual for both tools)
  - Annotating automatically the tone patterns in semi-tones (10 ms)
  - Tone levels calculated from histograms of F0 distribution
    - discarding the extreme values
    - using 5 levels
  - Calculating
    - The tone level on 3 distinct points per vowel
      - at the beginning of the vowel
      - at the turning point between initial and final part,
      - at the end of the vowel
    - The steepness of the slope (from very steep to flat)…
    - …which enables to adjust the tone level detection in a post-processing step

Study 1: Results (1)
Comparison of MISTRAL+ and PROSOTRAN

- **Comparison of results (Prototone list extended) using two tone boundaries**
  - at the beginning
  - and the end
- **Occurrences of tone patterns annotated by each tool :**
  - Agreement between the results
  - As it is based also on the user's knowledge (threshold effect…), MISTRAL+ supplies less tone patterns



Study 1: Results (2)
Comparison of MISTRAL+ and PROSOTRAN

- **In continuation of previous studies, for both tools**
  - ***rising slopes are extremely seldom***
  - the tone ***/43/*** is the most often detected
  - Rate of agreement between the tools
    - when 3 points (beginning, pivot, end) taken into account : 60 %
    - when ***2 points (beginning, end) : 70 %***
      - among the remaining 30%
        - 26 % share one value
        - only 4% are mismatching for the 2 values
  - MISTRAL+
    - Falling slopes: 30 %
    - Plateaux: 68%
  - PROSOTRAN
    - Falling slopes: 43 %
    - Plateaux: 55%

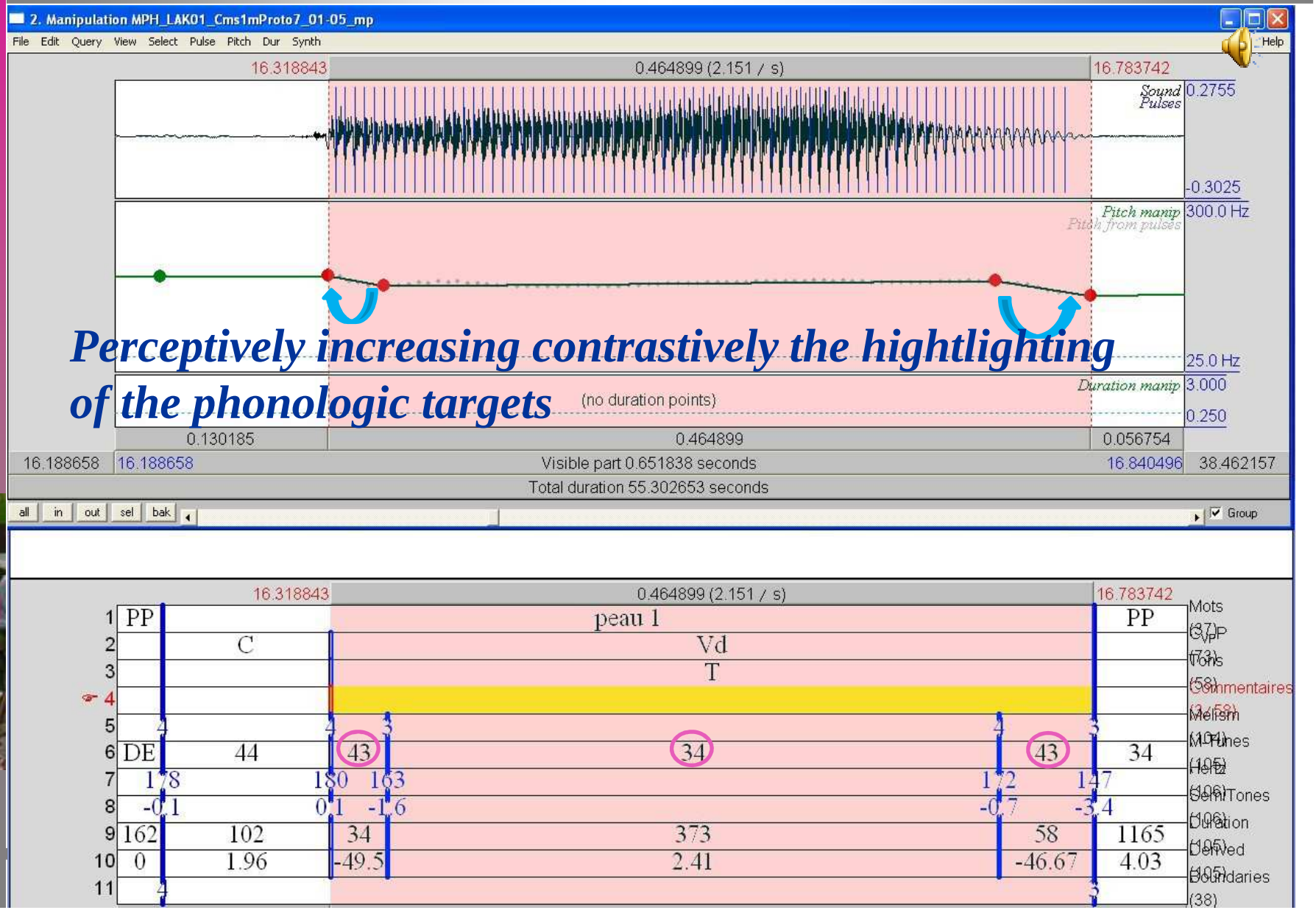
Study 2:
Restricted Prototone list and MISTRAL+

- ***660 words/vowels/tones***
- **175 lexical items x 3 repetitions x 1 speaker (LAK01)**
- ***525 mo piu lexical items*** **after selection of one unit from the compound words**
- ***Strong 'tonetic' manual stylization under Praat***
  - very close to the F0 slope in a first step (this study)
  - second step: generalization and phonologic perspective (later, so out of scope of this study)
- ***Corrections of the threshold effect*** **(± 5 Hz around its value) when the repetitions of the same item disagree**

Study 2: Speaker effect

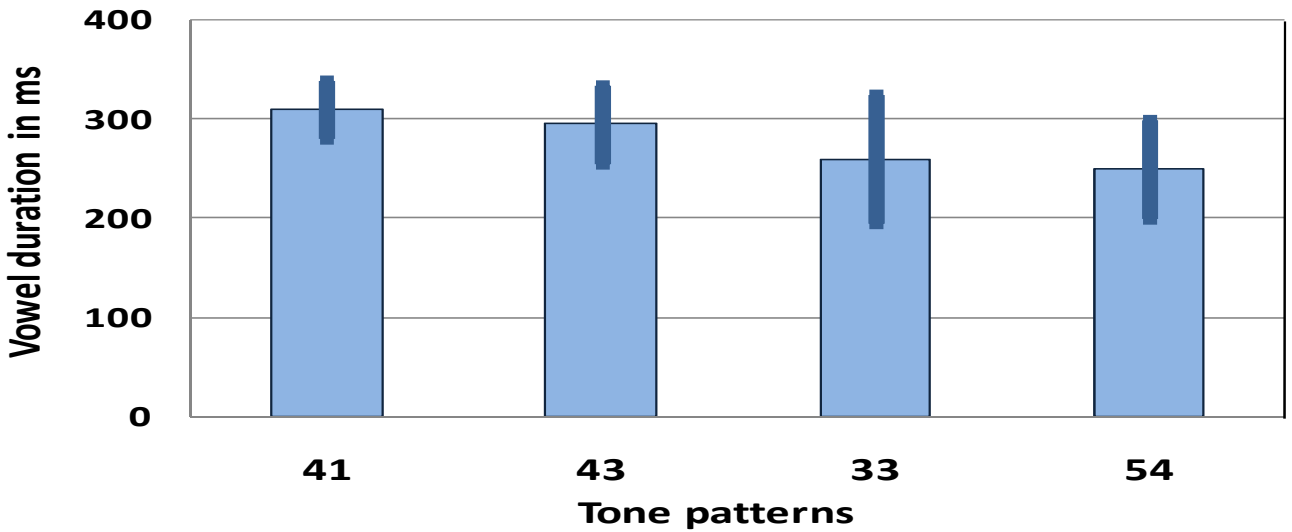
- **Each tone (mainly either simple or double slope) composed respectively of 2 or 3 fundamental points**
  - first boundary
  - turning point (pivot)
  - final boundary
- ***Between these tonologic points, relative tonal freedom*** **(F0 modulations)**
  - Most of the cases: the slope is continuously decreasing
  - However ***some overshooting / undershooting occurs***
    - the tone /44/ is sometimes '*tonetically*' structured as such /4544 = 45-54-44/
    - /54/ → /54554 = 54-45-55-54/
    - /43/ → /4343 = 43-34-43/
    - The seldom rising tones are due to the speaker effect and do not carry a discriminative function
  - An interesting issue seems that the ***final part*** of the modulated tone is ***duplicating the tonologic values of the targets***
    - /43***43***, /545***54***/, /43***43***/
    - Tonologic constraint on the tonetic layer ? Perceptive constraint ? Physiologic constraint ? A mix of constraints ?

Example: no tonologic F0 modulations - tone /43/ -



Perceptively increasing contrastively the hightlighting of the phonologic targets

Results 1: duration of the main tonal patterns



- **Mean duration of the 525 vowels/tones**
  - 293 ms (standard deviation, 72 ms): long duration
  - in total concordance with the previous studies
- **No clear tendency to make oppositions between**
  - 1 or 2 slope tones
  - plateaux and falling slopes

Results 2: distribution of the tones patterns across the prototones lists

Nb of lexical items / Vowels / Tones

Tone patterns found in the study (MISTRAL+)

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

8 Prototones lists

- **lexical items are unevenly distributed across the different prototone lists (6, 7 and 8 supply only a few examples)**
- **Tone pattern /43/ very frequent**
  - in each prototone list
  - ranked as 1st in 2, 3, 5, (6), 7, 8 prototones lists
    - ⋆ So no clear cues to 1-2, 3-4, 5-6, 7-8 tonal convergences

Discussion 2: the tones patterns distribution

- **Limited data restricted to only one speaker and containing an unbalanced number of prototones**
- **Only some hypotheses possible**
  - the prototone 1 seems to have converged towards the tone /54/
  - the prototones 2, 3, 5, 7, 8 seem to have merged into the tone /43/
  - the prototone 6 is distributed equally into tones /43, 41/ but its data are not enough numerous
  - though 1/3 of the tones belongs to tone /43/, the prototone 4 seems to mainly converge towards the plateau /33/
  - the tones /44/ scattered, and /32/ very seldom, seem to be overshot or undershot targets of the other tones
    - it may be possible also that the tone patterns /44, 32/ exist in Mo Piu words not included in the Prototones list
  - other tones may exist in the Mo Piu language than those presented in the prototone list
- ***Of course, the prototone list is not a Mo Piu tone list***
- **New data and several speakers are needed**

Results 3: Tonal patterns

- **Computings restricted to the tones**
  - …presenting the biggest population (/43, 54, 41, 33/) among the prototones
  - …which can be considered as Mo Piu tonal candidates
- **But a problem to solve**
  - Extraction of the right pattern: for example /433 = 43-33/
    - is it a **speaker modulation** of /43/ ???
    - ***or*** a ***tonologic*** twofold tone 43-33 ???
- **2 efficient embedded cues to answer this question:**
  - ***the difference of population*** between tones occurences
    - with 1, 2 or more than 2 slope orientations
    - either falling slopes or plateaux
  - ***the most frequently observed pattern***

Discussion 3: extraction of the right tonal pattern

Tones	1 direction	2 directions	> 2 directions	Ranked 1st
/54/	4%	*75%*	21%	54-44
/43/	15%	*53%*	32%	43-33
/41/	*50%*	28%	22%	41
/33/	*100%*	0%	0%	33

- **/54/ and /43/: biggest population for *2 slope directions* (respectively 75%, 53%)**
  - /544/ (/54-44/), and /433/ (/43-33/): ranked 1st for 2 directions
- **/41/, /33/: biggest population for *1 slope direction* (respectively 50%, 100%)**
  - /41/ and /33/ ranked 1st for 1 direction
- **One hypothesis about the remaining %**
  - simple variants of these 2 main patterns

Conclusion 1: Tonal patterns in Mo Piu langage

- **The goals of our study**
  - to give a description of Mo Piu tone patterns
  - to evaluate the detection and the annotation of the tone patterns via 2 tools
    - an automatic one: PROSOTRAN
    - a semi-automatic one, allowing manual adjustments: MISTRAL+
  - in the same conditions
    - manual segmentation of the speech signal made by the same expert phonetician
    - data calculated in semi-tones accross 5 levels
  - the results issued from the 2 tools are converging
    - overall agreement between the 2 tools concerning the slope direction: 70%
    - the pattern /43/ is the most frequent for both of them

Conclusion 2: comparing the present findings to the previous ones

- **the main findings concerning tone levels confirm these ones of the previous studies [Caelen-Haumont and al., 2010, Caelen-Haumont, 2012]: *same tonal patterns***
- **This present study allows**
  - ***to reduce the number of different tone patterns previously found***
  - to give ***more precision*** about them
    - bidirectional pattern for falling tones /54/ (/54-44/) and /43/ (/43-33/),
    - one direction pattern for /41/ and for the plateau /33/
- **On the concordance of these findings, our results can be considered as:**
  - a step towards tone phonology
  - supplying a pre-phonological scope about the Mo Piu tonal system

Conclusion 3: about the 8 prototones lists

- **In the restricted number of the 175 words of the Prototones list, …**
- **…the main issue is that the 8 prototones do not lead to 4 Mo Piu tones**
- **Under the pressure of history, contact with other languages, loan words, and internal organization, *a tone restructuration is made in Mo Piu language***
- **Nowadays *in Mo Piu, the main tone patterns are falling slopes and plateaux***
- **Some prototones (2, 3, 5, 8) seem to have merged into the same tone /43/…**
- **…while the prototone 1 seems to mostly correspond to the tone /54/**
- **other data are needed to confirm**
  - whether 6 is split up in the two tones /43, 41/
  - and whether the prototone 4 leads to the plateau /33/

Conclusion of all the conclusions

- I had the pleasure to reach Bangkok in order to join *my last international meeting point*
- I have attended a lot of international congress places since 1976…
- *Bangkok* is thus the last one and it will stay unforgotten in my heart and memory,
- *SEALS 23* is the *last conference* of my CNRS researcher career
- … and this *presentation is probably the last one*…
- So *very warm thanks* to all my research friends all over the world, my colleagues, my students, Mo Piu people and administrative staff, and all of you, for having given me a nice collaboration, help and fascinating intellectual discussions
- The *team* I built on Mo Piu minority language in Hanoi (MICA) and France is now strong enough to carry on alone the work. Warm thanks particularly to Alice Vittrant, Do-Dat Tran, Brigitte Cortial, Jean-Cyrille Ly Van Tu, Katarina Bartkova, Alexis Michaud, Jean-Pierre Salmon, Benoît Weber, Dang Khoa Mac, without forgotting the other contributors,
- *My research work was a very great happiness* and so will be the future of my life… keeping an eye on Mo Piu language research !