



**HAL**  
open science

# Glottalization, tonal contrasts and intonation: an experimental study of the Kim Thuong dialect of Muong

Minh-Châu Nguyễn

## ► To cite this version:

Minh-Châu Nguyễn. Glottalization, tonal contrasts and intonation: an experimental study of the Kim Thuong dialect of Muong. Linguistics. Université de la Sorbonne nouvelle - Paris III, 2021. English. NNT : 2021PA030119 . tel-03652510

**HAL Id: tel-03652510**

**<https://theses.hal.science/tel-03652510>**

Submitted on 19 Dec 2022

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial - NoDerivatives 4.0 International License

# UNIVERSITÉ SORBONNE NOUVELLE

École doctorale n°622 : *Sciences du langage*

Langues et civilisations à tradition orale (LACITO, UMR 7107)

PH. D. THESIS

presented by

Minh-Châu NGUYỄN

defense: December 14th, 2021

to obtain the degree of doctor in Phonetics, phonology and speech sciences

## Glottalization, tonal contrasts and intonation: an experimental study of the Kim Thuong dialect of Muong

*Thesis supervised by:*

M. Alexis Michaud                      Researcher, LACITO/ CNRS UMR 7107  
Mme Lise Crevier-Buchman          Researcher, LPP/ CNRS UMR 7018

---

*Rapporteurs :*

M. Marc Brunelle                      Professor, University of Ottawa  
Mr. James Kirby                        Professor, Ludwig-Maximilians-Universität München

---

*Membres du jury :*

M. Marc Brunelle                      Professor, University of Ottawa  
Mr. James Kirby                        Professor, Ludwig-Maximilians-Universität München  
Ms. Thị-Ngọc-Yến Phạm              Professor, Hanoi University of Science and Technology (HUST)  
Mme Solange Rossato                 Maître de conférences, Université Grenoble Alpes



# Abstract

## Glottalization, tonal contrasts and intonation: an experimental study of the Kim Thuong dialect of Muong

All languages in the Vietic subbranch of Austroasiatic have at least one glottalized tone. This thesis zooms in on one of these languages: Muong (in Vietnamese orthography: *Mường*, endonym: /**mon**<sup>3</sup>/), spoken in Kim Thuong (Phu Tho, Vietnam). Twenty speakers recorded twelve tonal minimal sets of the five tones of smooth syllables, plus three tonal minimal pairs of the two tones of checked syllables, under two conditions: in isolation and in a carrier sentence. Acoustic and electroglottographic recordings allow for estimating fundamental frequency, glottal open quotient and duration. These parameters are compared across tones, experimental conditions and speakers, in order to contribute to a better understanding of glottalization as a feature of linguistic tones. First, allotones of the phonologically glottalized tone in Muong (Tone 4) are classified on a phonetic basis, confirming the consistent presence of *creak*. It is tempting to contrast it with the *glottally constricted* tones of Northern Vietnamese (with which Muong is in sustained language contact). However, the phonological discussion emphasizes that analysis of Tone 4 as a prototypical “creaky tone” would be a pitfall. Tone 4 behaves in key respects like the other tones in the system: it is not defined solely by phonation type. Moreover, the range of phonetic (allotonic) variation of Tone 4 includes cases of glottal constriction. Use of a phonetic nomenclature for types of glottalization serves as a basis for describing the interaction of glottalization with intonation.

**Keywords:** glottalization, creaky voice, phonation types, tone systems, experimental phonology, phonetic fieldwork, electroglottography, Vietic languages, Muong language



# Résumé

## Glottalisation, oppositions tonales et intonation : étude expérimentale du dialecte muong de Kim Thuong

Toutes les langues de la branche viétique de la famille austroasiatique possèdent au moins un ton glottalisé. La présente thèse se concentre sur l'une de ces langues : le muong (en orthographe vietnamienne : *Mường*, endonyme : /**mon**<sup>3</sup>/), parlé à Kim Thuong (Phu Tho, Vietnam). Vingt locuteurs ont enregistré douze ensembles minimaux des cinq tons des syllabes sans occlusives finales, et trois paires minimales des deux tons des syllabes à occlusives finales, dans deux conditions : à l'isolée et dans une phrase-cadre. Les signaux acoustiques et électroglottographiques recueillis permettent d'estimer fréquence fondamentale, quotient ouvert et durée. Ces paramètres sont comparés entre tons, entre conditions expérimentales et entre locuteurs, afin de parvenir à une meilleure compréhension de la glottalisation en tant que caractéristique d'un ton lexical. Tout d'abord, les allotones du ton phonologiquement glottalisé en muong (le ton 4) sont classés sur des bases phonétiques. Il est tentant d'opposer ce ton, caractérisé par la présence régulière d'une voix craquée, avec les tons B2 et C2 du vietnamien du nord (avec lequel le muong est en contact linguistique soutenu), caractérisés par une constriction glottale. Cependant, une analyse phonologique du ton 4 comme prototype de « ton en voix craquée » masquerait la complexité des faits : le ton 4 fait partie d'un système au sein duquel il n'est pas défini exclusivement par un type de phonation. En outre, la plage de variation allotonique du ton 4 comprend des cas de constriction glottale. Une nomenclature phonétique des types de glottalisation sert de base à la description du ton 4 et de son interaction avec l'intonation.

**Mots-clés** : langues viétiques, langue muong, systèmes de tons, types de voix, glottalisation, voix craquée, phonétique expérimentale, électroglottographie



# Contents

**Abstract** i

**Résumé** iii

**Abbreviations** xvii

**Acknowledgements** xix

<b>1</b>	<b>Introduction</b>	<b>1</b>
1.1	Linguistic uses of phonation types	1
1.2	Phonation types, tones and intonation	2
1.3	Language documentation and conservation in present-day Southeast Asia	3
1.4	Language description	5
1.5	Areal studies and typology in an Open Science perspective	5
1.6	Acoustics, production and perception	6
1.7	Motivation and research questions	6
<b>2</b>	<b>Background</b>	<b>9</b>
2.1	Prosody, intonation and tone: a framework	9
2.1.1	Phonetics, phonology and intonation	9
2.1.2	Tones and intonation	13
2.2	Linguistic uses of phonation types	15
2.2.1	Phonation-type registers	15
2.2.2	Phonation types and tones: typological observations	16
2.2.3	Phonation types and tones: Hanoi Vietnamese as a textbook example	17
2.3	Definitions: glottalization and creaky voice	22
2.3.1	Literature review of creaky voice	24
2.3.2	Creaky voice and laryngealization	32
2.3.3	Glottal constriction and glottal stop	35
2.3.4	Glottalization as a component of lexical tone: an enduring challenge for phonology	37
2.4	Electroglottography: principles and analysis methods	37
2.4.1	Principles	37
2.4.2	Methods to analyze the electroglottographic signal	38
2.4.3	Criticism of estimation of the glottal open quotient by electroglottography	40

## Contents

2.5	The Muong people and the Muong language: a general view	42
2.5.1	The Muong as an officially recognized ethnic group	42
2.5.2	The Muong language in its Vietic context	45
2.5.3	Literature review of Muong studies	47
2.6	The dialect under study: Kim Thuong Muong	48
2.7	Phonemic inventory of Kim Thuong Muong	52
2.7.1	Consonants	53
2.7.2	Vowels	54
2.7.3	Tones	54
<b>3</b>	<b>Method</b>	<b>61</b>
3.1	Speech materials	61
3.1.1	List of the (near-) minimal sets	62
3.1.2	Carrier sentence	75
3.1.3	Narratives	77
3.2	Experiment setup	79
3.2.1	Participants	79
3.2.2	Equipment	86
3.2.3	Recording procedures: training and performance	86
3.3	Data processing	99
3.3.1	Data segmentation and annotation with SOUND FORGE	100
3.3.2	Analyzing the electroglottographic signal: how the PEAKDET script was applied	107
3.3.3	Step-by-step description of the process of analyzing the electroglottographic signal with PEAKDET	110
3.4	Plotting and data visualization	143
3.4.1	Visualization of results	143
3.5	Data archiving and publishing	145
3.6	A bird's eye view of the full data set for the main experiment	146
<b>4</b>	<b>Results</b>	<b>151</b>
4.1	The Kim Thuong Muong tone system: observations based on fundamental frequency and open quotient curves averaged by speaker	151
4.1.1	Fundamental frequency	152
4.1.2	Open quotient	159
4.1.3	Duration	159
4.1.4	Cross-speaker similarities and differences in tonal spaces	161
4.2	An overview of the full tone system from the results of normalizing 20 speakers	164
4.2.1	Smooth tones	164
4.2.2	Checked tones	168
4.3	Glottalization as a component of lexical tone: A closer look at the glottalized tone across 20 speakers	169
4.3.1	A general look at twenty speakers	169

4.3.2	A comparative look at gender differences	171
<b>5</b>	<b>Discussion</b>	<b>185</b>
5.1	About the tone system of Kim Thuong Muong	185
5.2	A characterization and classification of sub-types of creaky voice based on audio and electroglottographic signals	190
5.3	Detection of creaky voice from the electroglottographic signal	208
5.3.1	An easy problem, but paradoxically without an off-the-shelf solution	211
5.3.2	Step 1: list of specifications	211
5.3.3	Step 2: writing and testing the Creakdet script	212
5.4	Glottalization in Kim Thuong Muong: phonetics and phonology	235
5.4.1	Vietnamese dialectology, a source of inspiration for Muong studies	237
5.4.2	Some free associations around the glottalized tone (Tone 4)	240
5.5	Intonation in Kim Thuong Muong: the view from sentence-final particles	243
5.5.1	Final particles and intonation in Southeast Asian languages, and the case of Vietnamese and Muong	243
5.5.2	Tone and intonation on sentence-final particles in Muong	246
5.5.3	Boundary tones or tone coalescence?	246
5.5.4	An intonational tone in Muong?	251
<b>6</b>	<b>Conclusion and perspectives</b>	<b>255</b>
	<b>Appendices</b>	<b>257</b>
<b>A</b>	<b>Challenges and strategies of studying an unwritten language in the context of a bilingual community</b>	<b>259</b>
A.1	Data collection software vs. hands-on monitoring of recording sessions by the investigator	259
A.1.1	Learning from an earlier pilot study: difficulties of operating at a phonemic level	260
A.2	Strategy for finding (near-) minimal sets in an unwritten language	266
A.3	Using illustrative photos to stimulate the target word: an appropriate method in unwritten language	268
<b>B</b>	<b>Some practical details about the recording settings</b>	<b>275</b>
B.1	Equipment and experimental setup in the field	275
B.1.1	Recorders and microphones	275
B.1.2	Electroglottographic device	276
B.1.3	Recording environment	282
B.1.4	Recording set-up	282
B.2	List of files: main experiment and additional materials	284
<b>C</b>	<b>Data analysis and visualization</b>	<b>293</b>
C.1	Labeling the tone system: numeric labels and association with colors	293

*Contents*

<b>D</b>	<b>Additional graphs produced as a result of this study</b>	<b>299</b>
D.1	The Kim Thuong Muong tone system by fundamental frequency in semitone	299
D.2	The glottalized tone in Kim Thuong Muong: the distribution of fundamental frequency and open quotient values are presented in box plot	299
D.3	The correlation between fundamental frequency and open quotient by a scatter plot	299
	<b>Bibliography</b>	<b>309</b>

## List of Figures

2.1	A highly schematic representation of the components of prosody. Reproduced from Michaud (2017), with a minor change: replacing the original term “Lexically distinctive properties” by “Lexical prosodic properties”.	13
2.2	Vietnamese tone A <sub>1</sub> ( <i>ngang</i> ), as realized in isolation over the syllable /kɿ/ by a speaker from Hanoi. Year of recording: 1900.	18
2.3	Vietnamese tone A <sub>2</sub> ( <i>huyền</i> ), as realized in isolation over the syllable /kɿ/ by a speaker from Hanoi. Year of recording: 1900.	19
2.4	Vietnamese tone B <sub>1</sub> ( <i>sắc</i> ), as realized in isolation over the syllable /kɿ/ by a speaker from Hanoi. Year of recording: 1900.	19
2.5	Vietnamese tone B <sub>2</sub> ( <i>nặng</i> ), as realized in isolation over the syllable /kɿ/ by a speaker from Hanoi. Year of recording: 1900.	20
2.6	Vietnamese tone C <sub>1</sub> ( <i>hỏi</i> ), as realized in isolation over the syllable /kɿ/ by a speaker from Hanoi. Year of recording: 1900.	21
2.7	Vietnamese tone C <sub>2</sub> ( <i>ngã</i> ), as realized in isolation over the syllable /kɿ/ by a speaker from Hanoi. Year of recording: 1900.	21
2.8	Vietnamese tone C <sub>2</sub> ( <i>ngã</i> ), as realized in isolation over the syllable /bɑ/ by a speaker from Hanoi. Year of recording: 1900.	23
2.9	Acoustic patterns of four aperiodic glottal vibration. Reproduced from Hedelin and Huber (1990, p. 361).	29
2.10	Acoustic patterns of six laryngealized types. Reproduced from Batliner et al. (1993).	30
2.11	Acoustic patterns of four glottalized tokens illustrating different sub-types. Angled brackets above each waveform indicate the region where each sub-type appears. Reproduced from Redi and Shattuck-Hufnagel (2001, pp. 415–416).	31
2.12	The sub-classification of creaky voice by Keating, Garellek, and Kreiman (2015). Reproduced from Keating, Garellek, and Kreiman (2015, pp. 1–3).	33
2.13	Nomenclature systems of creak sub-classification provided by previous studies.	34
2.14	Example of EGG and dEGG signals with indication of glottis closure and opening. Reproduced with permission from the author, Alexis Michaud.	39
2.15	Continuum of phonation types (Reproduced from Gordon and P. Ladefoged, 2001, p. 384).	40
2.16	Map of Vietic languages by Ferlus (1998).	43
2.17	Map of Vietic dialects by V.-T. Nguyen (2005).	44
2.18	Map of Kim Thuong Muong location in Phu Tho province, the dialect in red ( <a href="http://bando.tnmtphutho.gov.vn/map.phtml">http://bando.tnmtphutho.gov.vn/map.phtml</a> ).	49

List of Figures

2.19	Tone box of Kim Thuong Muong . . . . .	59
3.1	General outline of the $f_0$ curve of an affirmative statement (Vaissière, 1983). Reproduced from Vaissière and Michaud (2006), with permission. . . . .	69
3.2	An effort to illustrate abstract nouns . . . . .	72
3.3	Informed consent form (Page 1) . . . . .	88
3.4	Attestation of payment . . . . .	90
3.5	A demo slideshow showing illustrative photos for the elicitation of the first minimal sets. This is accompanied by an explanation in Table 3.4. . . . .	92
3.6	Basic procedure of data processing. . . . .	100
3.7	Segmentation principle: examples of different treatments for voiced and unvoiced initial consonants. . . . .	102
3.8	Segmentation principle: minimum span of three cycles in token 6512 (syllable / <b>ja</b> <sup>2</sup> /), data from speaker F3. . . . .	103
3.9	Segmentation principle: treatment of missing syllables/rhymes . . . . .	103
3.10	Segmentation principle: treatment of the initial voiceless consonant which becomes a voiced consonant under coarticulation and hypoarticulation. . . . .	105
3.11	Mono electroglottographic .wav file: the second input of PEAKDET . . . . .	107
3.12	Regions List: the second input of PEAKDET. . . . .	108
3.13	A schematic representation of data processing with PEAKDET. . . . .	111
3.14	Two ways to open and run PEAKDET on MATLAB: (i) (in red): PEAKDET is opened in CURRENT FOLDER WINDOW and executed by typing “peakdet_inter” in the COMMAND WINDOW; (ii) (in magenta): PEAKDET is opened in the EDITOR WINDOW and executed by selecting Editor > Run from the toolbar. . . . .	112
3.15	The two dialog boxes that appear after calling PEAKDET, to request the location of the two input files. . . . .	113
3.16	The difference between two versions (2016 and 2019) of the initial settings in the COMMAND WINDOW after executing PEAKDET. . . . .	115
3.17	A demo of four initial plots after automatic processing of a token by PEAKDET. The example is taken from the data of speaker M1, first token 0101, target syllable / <b>paj</b> <sup>5</sup> / in isolation. . . . .	117
3.18	Together with four plots as in Figure 3.17, the display of the current processing token on COMMAND WINDOW with identification information, $f_{0 \text{ dEGG}}$ information and $f_{0 \text{ dEGG}}$ processing options. The example is taken from the data of speaker M1, first token 0101, target syllable / <b>paj</b> <sup>5</sup> / in isolation. . . . .	119
3.19	Two operations for $f_{0 \text{ dEGG}}$ verification. The example is taken from the data of speaker M1, first token 0101, target syllable / <b>paj</b> <sup>5</sup> / in isolation. . . . .	121
3.20	Example of a token where the first option (type o) of the $f_{0 \text{ dEGG}}$ modification should be applied. Data from speaker M1, token: 03, UID: 0111, syllable / <b>ja</b> <sup>2</sup> / (first frame word). . . . .	123

3.21	Example of a token where the second option ( <i>enter 1</i> ) of the $f_{0 \text{ dEGG}}$ modification should be applied. Data from speaker F21, token UID: 0501, syllable /paj <sup>4</sup> /. . . . .	124
3.22	Example of a token where the third option ( <i>enter 2</i> ) of the $f_{0 \text{ dEGG}}$ modification should be applied. Data from speaker M14, token UID: 2001, target syllable /laj <sup>4</sup> / in isolation, first performance. . . . .	127
3.23	Example of a token where the fourth option ( <i>enter 3</i> ) of the $f_{0 \text{ dEGG}}$ modification should be applied. Data from speaker M14, token UID: 3502, syllable /kieŋ <sup>4</sup> / in isolation, second performance. . . . .	130
3.24	Example of a token where the sixth option ( <i>enter 5</i> ) of the $f_{0 \text{ dEGG}}$ modification should be applied. Data from speaker M14, token: 10, UID: 0142, syllable /tǎŋ <sup>3</sup> / the fourth word of carrier sentence, second performance. . . . .	133
3.25	Former display (in the 2016 version) of $O_{q \text{ dEGG}}$ calculated in four different ways: (i) maxima on unsmoothed dEGG signal (in green), (ii) maxima on smoothed dEGG signal (in blue), (iii) barycentre of peak on unsmoothed dEGG signal (in red), (iv) barycentre of peak on smoothed dEGG signal (in black). . . . .	137
3.26	An example of $O_{q \text{ dEGG}}$ curves with large differences across methods as a result of imprecise opening peaks. Reproduced from M.-C. Nguyễn (2016): data of speaker M1, experiment 2, item 1331. . . . .	138
3.27	An important tip when processing with PEAKDET: use two screens for a convenient view and time-saving operation. . . . .	138
3.28	An example of item with imprecise opening peaks. Reproduction of (M.-C. Nguyễn, 2016): item carrying UID 1331, speaker M1, experiment 2. Abscissa: in samples (1 sample = 1/44,100 second). . . . .	139
3.29	Some specific cases where selecting the method of the barycenter of the peaks on the smoothed signal (typing 3) is highly recommended. . . . .	141
3.30	Actions in the COMMAND WINDOW for the step of $O_{q \text{ dEGG}}$ verification	143
3.31	Calculation of the total corpus . . . . .	147
3.32	A brief summary view of the corpus . . . . .	147
4.1	The tone system of Kim Thuong Muong: speaker by speaker. $f_{0 \text{ dEGG}}$ on the left and $O_{q \text{ dEGG}}$ on the right. . . . .	154
4.2	The tone system of Kim Thuong Muong: normalization across speakers. $f_{0 \text{ dEGG}}$ on the left and $O_{q \text{ dEGG}}$ on the right. . . . .	165
4.3	The Kim Thuong Muong's tone system: smooth tones (left) and checked tones (right), normalized on 20 speakers. . . . .	166
4.4	The glottalized tone of Kim Thuong Muong: speaker by speaker. $f_{0 \text{ dEGG}}$ on the left and $O_{q \text{ dEGG}}$ on the right. . . . .	172

List of Figures

4.5	An example illustrates the unusual beginning of speaker M12’s <b>Tone 4</b> with unreasonably high $f_{0 \text{ dEGG}}$ values caused by micro-closing peaks. Data from speaker M12, token UID: 2002, target syllable / <b>laj</b> <sup>4</sup> / in isolation, second performance. . . . .	184
5.1	Example of single-pulsed creak: extreme case. Data from speaker M11, token: 41, UID: 0501, over syllable / <b>paj</b> <sup>4</sup> / in isolation. DOI: <a href="https://doi.org/10.24397/pangloss-0006782#W9">https://doi.org/10.24397/pangloss-0006782#W9</a> . . . . .	192
5.1	Example of single-pulsed creak: common case. Data from speaker M11, token: 47, UID: 0531, over syllable / <b>paj</b> <sup>4</sup> / in carrier sentence. DOI: <a href="https://doi.org/10.24397/pangloss-0006782#W10">https://doi.org/10.24397/pangloss-0006782#W10</a> . . . . .	193
5.1	Example of multiply-pulsed creak. Data from speaker F10, token: 551, UID: 6002, over syllable / <b>ku</b> <sup>4</sup> / in isolation. DOI: <a href="https://doi.org/10.24397/pangloss-0006784#W224">https://doi.org/10.24397/pangloss-0006784#W224</a> . . . . .	196
5.1	Example of jitter in harshness, illustrating a technical difficulty in automatic detection of multiply-pulsed creak. Data from speaker F12, token: 178, UID: 1832, over syllable / <b>laj</b> <sup>2</sup> / in carrier sentence. DOI: <a href="https://doi.org/10.24397/pangloss-0006790#W170">https://doi.org/10.24397/pangloss-0006790#W170</a> . . . . .	198
5.1	Example of aperiodic creak. Data from speaker F12, token: 541, UID: 5501, over syllable / <b>kaj</b> <sup>4</sup> / in isolation DOI: <a href="https://doi.org/10.24397/pangloss-0006790#W109">https://doi.org/10.24397/pangloss-0006790#W109</a> . . . . .	201
5.1	Example of a maximum pressed voice. Data from speaker F21, token: 41, UID: 0501, over syllable / <b>paj</b> <sup>4</sup> / in isolation. DOI: <a href="https://doi.org/10.24397/pangloss-0006812#W9">https://doi.org/10.24397/pangloss-0006812#W9</a> . . . . .	204
5.2	Example of a minimally pressed voice. Data from speaker F21, token: 42, UID: 0502, over syllable / <b>paj</b> <sup>4</sup> / in isolation. DOI: <a href="https://doi.org/10.24397/pangloss-0006812#W141">https://doi.org/10.24397/pangloss-0006812#W141</a> . . . . .	207
5.3	First example of pressed voice reproduced from the Github gallery of glottalized signals. . . . .	209
5.4	Second example of pressed voice reproduced from the Github gallery of glottalized signals. . . . .	210
5.5	Algorithm of CreakDet version 1 . . . . .	213
5.6	Comparison of the $\text{smoo\_delta}f_{0 \text{ dEGG}}$ : an illustration for the consideration of the threshold established in condition 1 and 2. . . . .	215
5.7	Illustration of how the “CRPdet.m” function works, through examples of (a) double-pulsed creak, (b) jitter, and (c) aperiodic creak. In each sub-figure, from top to bottom: (i) two variables: $\Delta f_0$ (in magenta and green) and $\text{smoo\_delta} f_0$ (in blue and red) with EGG signal; (ii) $f_{0 \text{ dEGG}}$ ; (iii) acoustic signal; and (iv) dEGG. . . . .	228
5.8	Overall results of CreakDet version 1. . . . .	232
5.9	Figure of $f_0$ curves of Hanoi Vietnamese tones, reproduced from Kirby, 2011, p. 386. . . . .	238

5.10	Acoustic signal, spectrogramme and fundamental frequency plot of the examples 2 and 3. . . . .	249
5.11	The final particle /hə/, taken from a dialogue prepared before recording. . . . .	252
A.1	Screenshot of the SpeechRecorder software from Vera Scholvin’s doctoral study on “Prosody in French-Vietnamese Language Contact”, with permission. . . . .	260
A.2	Experiment 2: Cards for target words. Yellow cards for smooth syllables and pink cards for stopped syllables. . . . .	263
A.3	Experiment 2: Cards used in an attempt to cue tones. . . . .	264
A.4	Four examples of illustrative photos that were taken in the field in 2018 to serve the experiment: the best way to pick photos to illustrate target words, especially local concepts. You can find the translations of these words in the Table 3.1. . . . .	270
A.5	Two examples of illustrative photos that were picked by the method of minimalist sketch in order to avoid distracting scenes as much as possible. . . . .	271
A.6	The illustrative photos for four adjective . . . . .	272
B.1	Roland 4-channel recorder and its accompanying devices. . . . .	277
B.2	The Glottal enterprises EG2-PCX and its accompanying components. . . . .	279
B.3	Some features to note regarding the use of EGG equipment Glottal enterprises EG2-PCX. Reproduced from Hao (2015) (here). . . . .	280
B.4	The room on the ground floor of my Muong teacher’s house was used as a “field recording studio” during all my field trips from 2016 to 2019. . . . .	283
B.5	The position of equipments on the recording (an illustration from the recording of speaker F5 in August 2015). . . . .	285
C.1	Color perception of French vowels. Reproduced (with permission) from the spectrogram reading course of Pr. Jacqueline Vaissière in 2018. . . . .	295
C.2	The distribution of representative colors of 5 tones on the color wheel. . . . .	297
D.1	Kim Thuong Muong’s tone system presented in semitones: speaker by speaker. Same data as in Figure 4.1. . . . .	300
D.2	The glottalized tone in Kim Thuong Muong: the distribution of $f_{0 \text{ dEGG}}$ and $O_{q \text{ dEGG}}$ values are presented in box plot. . . . .	304
D.3	The correlation between $f_{0 \text{ dEGG}}$ and $O_{q \text{ dEGG}}$ . . . . .	305



## List of Tables

2.1	Correspondences between Vietnamese and Muong for twelve body part names, adapted from André-Georges Haudricourt (1953). . . . .	46
2.2	Inventory of initial consonants . . . . .	53
2.3	Final consonants inventory . . . . .	54
2.4	Palatal final consonants <b>c</b> and <b>ɲ</b> in inherited vocabulary and in loan words	55
2.5	Vowel inventory . . . . .	55
2.6	A brief overview of the tone system of Kim Thuong Muong. St.V. = Standard (Hanoi) Vietnamese. . . . .	59
2.7	Examples of a merger of etymological categories A2 and C2 in Kim Thuong Muong . . . . .	60
2.8	Correspondences between Kim Thuong Muong – Vietnamese – Khmu, illustrating the presence of a final glottal stop in Palaung-Wa corresponding to tone B (B1 ‘sắc’, B2 ‘nặng’ in Vietnamese; and B1 – glottalized tone in Kim Thuong Muong.) . . . . .	60
3.1	Speech materials: eight minimal sets and four near-minimal sets that contrast for five tones in smooth syllables. . . . .	64
3.2	Speech materials: three minimal pairs that contrast the two tones of checked syllables. . . . .	67
3.3	A complete list of speakers / consultants who participated in the recordings from 2016-2019. . . . .	83
3.4	An explanation of how the slideshow of illustrative photos is employed at training and when performing the experiment. . . . .	96
3.5	Annotation scheme: four digits. Structure: AABC, detailed below. . . . .	106
3.6	PEAKDET output information: .mat file containing a 10×100 matrix . . . . .	144
3.7	Current status of corpus: 20/28 data files have been annotated with SOUND FORGE and processed with MATLAB. . . . .	147
4.1	Some general information about the speakers’ data: The values of the mean and standard variation of $f_{o\ dEGG}$ and $O_{q\ dEGG}$ and the ratio of the excluded $O_{q\ dEGG}$ values obtained by the quantitative analysis of the electroglottographic signal on twenty speakers. . . . .	153
5.1	Ten final particles involved in interrogative statements in Muong. . . . .	247
C.1	List of representative colors for tones accompanied by a brief explanation of reasons for the choice of colors . . . . .	296



# Abbreviations

- ddEGG** Second derivative of the electroglottographic signal
- DECPA** Derivative-Electroglottographic Closure Peak Amplitude
- dEGG** First derivative of the electroglottographic signal
- DOI** Digital Object Identifier: a persistent identifier used to identify digital objects uniquely. These identifiers, standardized by the International Organization for Standardization (ISO), are used for publications but also for data in certain repositories, including the Pangloss Collection.
- EGG** Electroglottography, *or* electroglottographic, as in “EGG signal”
- $f_0$**  Fundamental frequency of speech
- $f_{0 \text{ dEGG}}$**  Fundamental frequency as estimated from the derivative of the electroglottographic signal
- ID** Identifier (for speakers): F for Female and M for Male, followed by a number. Thus, F<sub>1</sub> is the first female speaker.
- IPA** International Phonetic Alphabet
- O<sub>q</sub>** Glottal open quotient
- O<sub>q dEGG</sub>** Glottal open quotient as estimated from the derivative of the electroglottographic signal
- ptcl** Short gloss for *discourse particle*
- UID** Unique Identifier



# Acknowledgements

The research included in this dissertation could not have been performed without the assistance, patience, and support of many individuals.

I would like to extend my gratitude first and foremost to my thesis supervisors Alexis Michaud and Lise Buchman, as well as to the two members of the *Comité de suivi*: Didier Demolin and Nicolas Audibert. I thank them for their confidence in me.

This research would not have been possible without the participation of the native speakers in Kim Thượng (Tân Sơn, Phú Thọ) who went gracefully through long and demanding elicitation sessions, not minding the heat and discomfort because they knew they were participating in building a precious corpus which would be available in open access in the [Pangloss](#) Collection. In particular I would like to thank my Muong teacher, Sa Thị Đinh, for her warm support during all my field trips from 2014 to 2019. I would additionally like to thank Đinh Thị Hằng for her assistance during the experiment that I conducted in my 2018 field trip and her collaboration in preparing the data set later down the line for its digital publication.

I am also grateful to colleagues (junior and senior) at LACITO as well as LPP (Laboratoire de Phonétique et Phonologie), two research centres that provide a most congenial environment. I would like to thank Annie Riolland, Jacqueline Vaissière, Rachid Ridouane, Séverine Guillaume, Balthazar Do Nascimento, Zlatka Guentchéva-Desclés, Benjamin Galliot, Isabelle Bril, Anne Armand, and fellow students Vera Scholvin, Albert Badosa Roldós, Valentina Alfarano, Camille Simon, Fatima Zahra Issaiene, Mezane Konuk, Maxime Fily, Neige Rochant, Yann Le Moullec, Eréndira Calderon, Alexandra Vydrina, Cécile Macaire and many more, as well as the Vietnamese phonology group around Marc Brunelle (Phạm Thị Thu Hà and Tạ Thành Tấn).

Compagnons courtois, / *Compaignons gentilz*  
Calmes et adroits, / *Serains & subtilz*  
Sans méchanceté, / *Hors de vilité,*  
La civilité / *De civilité*  
Est votre vraie loi, / *Cy sont les houstilz*  
Compagnons courtois. / *Compaignons gentilz.*  
(Rabelais, *Gargantua*, trad. Ludovic Debeurme)

Many thanks to Thầy Ferlus (Michel Ferlus), Thầy Dõi (Trần Trí Dõi), Cô Xuyên (Lê Thị Vũ-Xuân Xuyên), Mark Alves, Guillaume Jacques, and Elisabeth Delais-Roussarie for support and encouragement.

I am grateful to Thomas Pellard for making available under a permissive license (Creative Commons – Attribution: CC BY) the  $\LaTeX$  template used for this document,

## *Acknowledgements*

as well as an introductory course: these tools constitute by themselves a much appreciated encouragement along the path towards more rigorous typography.<sup>1</sup>

Many thanks to jury members, who played a role similar (I feel) to a North American-style dissertation committee, providing guidance and support along the way.

Needless to say, all shortcomings are my own responsibility.

Financial support from the “Empirical Foundations of Linguistics” Labex project (ANR-10-LABX-0083) and the “Computational Language Documentation by 2025” ANR project (ANR-19-CE38-0015-04) is gratefully acknowledged, as well as financial support for fieldwork from LACITO and *École Doctorale 622*.

Finally I would like to extend my deepest gratitude to my family: to my beloved mother who did not give up having me after seven miscarriages, to my lenient father who rarely talked to me but I know his silence meant “let’s do whatever you want as long as you happy with it”; to my little sister who cooked more than 3,000 meals so I can keep studying even when back home on vacations. Although it is technically incorrect to include him in this paragraph, as he doesn’t share the same bloodline as me, I am inclined to extend family-like thanks to Alexis Michaud for being available all along providing instructions, motivation, and encouragement in time, meticulously taking care of my studies. For me he is not only an ideal supervisor but also a great “superfather”.

---

<sup>1</sup>Template available from: <https://fr.overleaf.com/latex/templates/these-inalco/kpghvjrpwjr>.  
L<sup>A</sup>T<sub>E</sub>X course available from: <https://cel.archives-ouvertes.fr/cel-01527916>.

# Chapter 1

---

## Introduction

The study of glottalization, tonal contrasts and intonation in the Muong language (Kim Thuong Muong<sup>1</sup>) stands at the intersection of several fields, each of which has motivations, objectives, methods and requirements of its own: experimental phonetics/phonology (specifically, the study of tones and phonation types), language documentation and conservation, language description, and areal studies (Southeast Asian linguistics). These main strands are presented separately below. They set the stage for the research question addressed in the present thesis, which is presented in 1.7.

### 1.1 Linguistic uses of phonation types

The study of phonation types has consistently been a topical issue in phonetics and phonology over past decades, due in no small part to their relevance to expressive speech and hence to high-quality speech synthesis (Erickson, 2005). Specifically, glottalization serves pragmatic, stylistic and attitudinal functions, to different extents in different languages: thus, creaky voice is less prevalent in present-day French than in present-day English, with American English having more creak than British English (for a review, and some fresh evidence, see Pillot-Loiseau et al., 2019).

But the attention of phoneticians/phonologists had been drawn to phonation types at least since the middle of the 20th century, when their role in some phonological systems was brought out. Eugénie Henderson brought out the key role of phonation-type registers in the evolution of Cambodian (Henderson, 1952), an insight which proved fruitful for a wide range of other languages of Southeast Asia, in the Austroasiatic family (for an overview: Ferlus 1979) and beyond (on the case of Chru, an Austroasiatic language: Brunelle, Thành Tấn Tạ, et al. 2020). The present study is intended as a contribution to the latter strand of research: investigating a language where phonation (specifically: the use of creaky voice) plays a phonologically distinctive role as a part of the tone system.

The presence of one form or other of glottalization is a common characteristic of languages of the Vietic group<sup>2</sup> (Ferlus, 1998), so, as seen from Southeast Asia, it really

---

<sup>1</sup>Place names in Vietnam are provided in Vietnamese orthography, at least at first occurrence, for clarity of reference. Vietnamese author names are reproduced as they stand in the original.

<sup>2</sup>The Vietic languages are a branch of the Austroasiatic language family. The branch was once referred to by the terms Viet–Muong, Annamese–Muong, and Vietnamuong. The term ‘Vietic’ was proposed by Hayes Hayes, 1992, who proposed to redefine Viet–Muong as referring to a sub-branch of Vietic

does not appear as an out-of-the-way situation. The realization that glottalization plays a role in Vietnamese tone dates back at least to de Rhodes's 17<sup>th</sup> century Dictionary (Rhodes, 1651). This property, which has been lost in the Southern dialect of Vietnamese, is still present in the Northern dialect (Brunelle, 2009a).

The present study of Kim Thuong Muong aims to bring out how the linguistic categories match the phonetic diversity of observed phonation types. In the case of the five lexical tones of Kim Thuong Muong, an initial expectation is that creak will be mostly found on **Tone 4** syllables, but also bracing ourselves for the possible encounter of complex combinations of factors, including interaction between tones (some of which incorporate phonation-type specifications) and intonation.

## 1.2 Phonation types, tones and intonation

It was mentioned in 2.2 that phonation types have been a topical issue in phonetic research in recent years. The interaction of lexical tones and intonation likewise appears as a topic of enduring interest. This interest may be due in part to the simple reason that linguists who do not speak a tonal language natively encounter such languages in their studies and readings, and naturally come to realize that “[a]s tone and intonation, which have different functions, are materialized simultaneously by the use of pitch variation, interaction between the two is expected to occur” (W.-S. Lee and Zee, 2017, p. 345).

The same lexical tones are expected to be realized somewhat differently in different intonational contexts. In language groups where tone can be described in terms of pitch alone, the interplay of tone and intonation is reported to be language- and dialect-specific. The well-studied area of Sinitic languages (Chinese dialects) provides solid evidence on this topic: the phrasing adopted by W.-S. Lee and E. Zee, “the effect of intonation on tone differs in different Chinese dialects” (W.-S. Lee and Zee, 2017, p. 349), does not appear controversial. Precious insights are to be found from case studies of this vast language group: for instance, the tone contours of individual syllables are better preserved across sentence modes (declarative vs. interrogative) in Beijing Mandarin than in the (otherwise very similar) Mandarin dialect spoken in Chengdu, in Southwest China (*ibid.*).

The presence of phonation types as part of the phonological specifications of tones (as in Muong) adds another twist to the fascinating issue of the interaction between tone and intonation. How is the use of creaky voice modulated intonationally in a language where it plays a role in the tone system? This topic links up with so many dimensions of intonation (phrasing and prominence, but also sentence mode, attitudes, emotions and rhythm) that answers provided here are bound to be incomplete, if only because a considerable range of different experimental setups would be required to shed light on all these different dimensions. However, the topic is so pervasive in communication in the Muong language that literally any set of materials in Muong

---

containing only Vietnamese and Muong. This usage has become most widespread, and is followed here.

sheds some light on the issue. Even elicited materials such as a word list allow for insights, because even a word in isolation constitutes an utterance, and its realization is thereby complete with word-level and utterance-level intonational trappings that reflect the attitude with which it is uttered (more or less assertive, more or less cautious, and so on), and also phonetic correlates of sentence-level phenomena (junctures), in addition to its lexical tone. Seen in this light, the topic of the role played by creaky voice in communication in the Muong language does not appear as hopelessly complex: its complexity is real, but it is made up of smaller complexities, and the tried-and-tested scientific approach that consists of examining different issues separately constitutes a good guide to progress in this field. In this perspective, the obvious limitations of the materials used in this study (mostly tonal minimal sets in carrier sentences) also entail advantages in terms of ease of analysis. It does not appear unreasonable to start out from such materials, rather than confront headlong all topics at once by focusing primarily on samples of dialogues, as one might be tempted to do in view of the creativity found in lively dialogues.

It nonetheless appeared advisable to cast the net wide at the stage of data collection, and to record materials that are not limited to one context of elicitation: an interesting research question is to what extent spontaneous speech differs from materials that are elicited sentence by sentence with a strong efforts towards symmetry.

Mention of the balance (or lack of such) between different types of materials recorded for this study provides a handy transition to the topic of the contribution that the present study intends to make to language documentation and conservation, two dimensions of the activity of linguists which matter more than ever in a time of accelerating worldwide attrition of the diversity of languages and culture.

### 1.3 Language documentation and conservation in present-day Southeast Asia

The study of *minority ethnic languages*, as they are known in Vietnam (and neighbouring countries), is itself a ‘minority’ discipline within linguistics, with few people in the field, especially given the large number of languages that are currently on the wane in Southeast Asia (Enfield and Comrie, 2015, p. 12). Of course, what would really need to be done is to safeguard language diversity, for reasons that are obvious to scientists reflecting on the issue.

The knowledge systems and practices of Indigenous Peoples and local communities play critical roles in safeguarding the biological and cultural diversity of our planet. (...) Our warning raises the alarm about the pervasive and ubiquitous erosion of knowledge and practice and the social and ecological consequences of this erosion. (...) We appeal for urgent action to support the efforts of Indigenous Peoples and local communities around the world to maintain their knowledge systems, languages, stewardship rights,

ties to lands and waters, and the biocultural integrity of their territories – on which we all depend. (Fernández-Llamazares et al., 2021)

While the overall picture leaves little ground for optimism concerning the preservation of language diversity, linguists are in a position to create at least a decent archival record of languages before they go out of human memory. “Documentary linguistics is concerned with the creation of a longlasting, multipurpose record of the language use of speakers/signers” (Seyfeddinipur and Rau, 2020, p. 503). The work reported in the present dissertation is admittedly by no means a typical contribution to language documentation, as its focus is especially narrow (a phonetic/phonological study of the tone system), but it is nonetheless designed as a contribution to language documentation. Phonological elicitation can arguably be integrated within language documentation, thereby improving the record.

Several open-access digital libraries were created with the purpose of storing research publications and related information about languages in the world. Specifically for Southeast Asian languages, the SEAlang projects,<sup>3</sup> for “Southeast Asian Languages”, was established in 2005. Its library provides language reference materials for Southeast Asia. Through 2009, it focused on the non-roman script languages used throughout the mainland, and in 2010-2013 it concentrated on the many languages of insular Southeast Asia, including Malaysia, Indonesia, and the Philippines. It is to be hoped that this ambitious project will continue over the years and decades. On a much more modest scale, interested scholars maintain a Zotero group (set up under the impetus of Mark Alves), which pools references on Vietic languages and linguistics: “Vietic Languages and Cultures”.<sup>4</sup>

Attention paid to the endangered languages of Southeast Asia naturally links up with a concern for the conservation of language data.

As part of a project funded by the *Comité pour la Science ouverte* (formerly *Bibliothèque scientifique Numérique*), DO-RE-MI-FA (for *Données des Recherches de Michel Ferlus en Asie du Sud-Est*),<sup>5</sup> digitization and dissemination of Michel Ferlus’s audio data from fieldwork was undertaken from 2014 to 2016. The project was conducted with the aim of bringing data collection of languages in Vietnam and neighboring countries to the research community. This was hailed as an important resource for research and cooperation among linguists, anthropologists and engineers. Due in no small part to the DO-RE-MI-FA project, there has been notable expansion of Southeast Asian language corpora in the Pangloss Collection (hosted by LACITO) over the past decade.<sup>6</sup> As a participant in the DO-RE-MI-FA project project, I had the honor of handling Ferlus’s data of Mường language with a total of about 30 recordings

---

<sup>3</sup>SEAlang’s website: <http://sealang.net/>.

<sup>4</sup>[https://www.zotero.org/groups/956729/vietic\\_languages\\_and\\_cultures](https://www.zotero.org/groups/956729/vietic_languages_and_cultures).

<sup>5</sup>The information of DO-RE-MI-FA project is available in <https://lacito.hypotheses.org/251>.

<sup>6</sup>The Pangloss Collection is a digital library whose objective is to store and facilitate access to multimedia recordings in (mostly) endangered languages of the world. Developed by the LACITO centre of CNRS in Paris, the collection provides free online access to documents of connected, spontaneous speech. The Collection’s website is: <https://pangloss.cnrs.fr/>.

from eight different dialects, which were recorded starting in 1983. When conducting my own study, whose result is this dissertation, I had a desire to enrich this precious resource. All the materials collected for this dissertation are already archived in the Pangloss Collection (available here: <https://pangloss.cnrs.fr/corpus/Muong>).

## 1.4 Language description

For obvious reasons, it was not feasible to provide a well-rounded description of Muong within the frame of this dissertation. Writing a reference grammar constitutes full-time work (as does the writing of a study in experimental phonetics/phonology, which is the goal of the present dissertation). Such work is nonetheless an important long-term goal. It sets an overarching general plan for language description, which lends meaning to focused (even piecemeal) descriptive work such as the research on phonation types which constitutes the core of the present dissertation.

Among the three pieces which together constitute a full-fledged language description – the Boasian Trilogy, as it has been called –, the preparation of a dictionary has been initiated. Even though it currently remains at the stage of a word list, it is based on a solid framework provided by the EFEO-SOAS-CNRS wordlist (Pain et al., 2014) and was devised with gradual expansion in view. The second part of the trilogy, a corpus of texts, is relatively modest, and still mostly untranscribed, but the collection process was set in train since the earliest stages of fieldwork. Each of the participants in the phonetic/phonological data collection campaign was also asked to provide narratives about life in the village or other topics familiar to them, and these materials provide an initial basis for well-rounded language documentation. As my command of the language progresses in future, I will increasingly be able to build on these materials, transcribe (with the help of consultants), translate and annotate, also recording new data, in a snowballing process that aims at serious coverage, following practices in descriptive and documentary linguistics.

## 1.5 Areal studies and typology in an Open Science perspective

For someone interested in living languages, working in a nationwide and worldwide context of language endangerment is a subject of sadness. Another reason for chagrin could be the necessity to choose among the many languages and research topics one is interested in. Devoting years of research to the tone system of just one dialect of one language within one branch of one language family can look desperately narrow in view of the wealth of possible directions for research and documentation.

Importantly, the perspective of Open Science gives a fundamental twist to this situation. As soon as publications (like the present Ph.D. dissertation) are not viewed as standalone deliverables by themselves, but as elements within an environment that also comprises data and tools (Roettger, Winter, and Baayen, 2019), connecting studies together into wider patterns appears as a natural development. Typology and areal studies do not appear as separate strands of research, but as a straightforward continuation of research

that is done well *on the ground*, at the level of individual languages and highly specific research topics. As soon as data are curated in such a way as to be shareable with other researchers, cumulative progress and facilitated scholarly dialogue appear as very real and very close prospects.

## 1.6 Acoustics, production and perception

Acoustics, production and perception should be studied jointly, as a matter of course. The present work contains no investigation of physiology, and no work on perception (apart from the investigator's constant use of aural impressions, of course); fundamental notions of acoustics are used, but this is not a work of fundamental research into acoustics or signal processing, unlike e.g. Henrich (2001). These limitations are hereby acknowledged as resulting, not from oversight, but from practical difficulties conducting fiberoptic video recordings of the vocal folds (preferably through high-frequency laryngoscopy) and other types of investigations. It is hoped that, by accepting the compromise of focusing solely on audio and electroglottographic evidence, the present study gained somewhat in reliability of the data that were recorded, even at the cost of a narrowed experimental scope.

## 1.7 Motivation and research questions

The six research strands mentioned above constitute the backdrop to the present study; contributing to these research strands is a strong motivation, from the short term to the long term. As for the specific research question addressed by the study, it is grounded in recent developments in the cross-linguistic study of linguistic uses of phonation types. Specifically, the research topic is formulated in view of recent proposals about a widely-studied tonal language: Mandarin Chinese. My initial intuition about the language studied here (Muong) is that its use of creak differs widely from what is observed in Mandarin, allowing for useful refinements to cross-linguistic (typological) models.

The use of phonation types has consistently been a topical concern in phonetics and phonology in recent decades. It is fascinating to observe that the same phonation type can play different functions in different languages. The object of the present study, glottalization in general and creaky voice in particular, is well-known for serving the function of indicating phrasing (cueing intonational junctures) in many languages, such as English (Dilley and Shattuck-Hufnagel, 1996), Cantonese (Yu and Lam, 2014), and Mandarin (Kuang, 2017). The prevalence of this finding would suggest that it could turn out to be a linguistic universal.

However, the typological picture becomes more complex as one examines cases of non-modal phonation with phonemic function. Such cases are clearly a minority among the world's languages, and moreover occur in lesser-known languages, mostly in "exotic" languages. When phonemic tone is also involved, several types of relationships between non-modal phonation and tone are attested. A highly influential theory is that elaborated

by Silverman et al. (1995) and Silverman (1997). They point out that phonation-type characteristics and tone are perceptually recognizable (“recoverable”) through a temporal arrangement (“phasing”) that avoids the sort of confusion that would be expected in simultaneous production. Arranging tonal events and phonation-type changes in sequence (temporal phasing) makes them “segmentable” by the ear, so that each element can be perceptually recovered. Jalapa Mazatec is a key example for this theory (indeed, the theory seems tailored for cases like Jalapa Mazatec). On the other hand, another combination is not discussed by Silverman, but is firmly attested in East/Southeast Asian languages such as North Vietnamese (Michaud, 2004b; Brunelle, 2009b), Green Mong (Andruski, 2006) and White Hmong (Garellek, Keating, et al., 2013), where pitch and phonation type are two dimensions of the same phonological units. These units are called tones by convention. But they are complex templates, and they feel like the backbone of syllables. This is unlike Mazatec, where tone is just a pitch level, and is only one phonological ingredient among others.

Crucially, the object of the present study (Kim Thuong Muong) appears to fall squarely among the phonetically complex tone systems (like various other tonal languages of mainland Southeast Asia), since one of its tones has glottalization (canonically realized as creak).

In this perspective, a question comes into sharp focus: what is the phonological status of creak in Muong? Muong only has one glottalized (creaky) tone. That tone is also the lowest tone within the system. It reminds one of Tone 3 in Mandarin, which is the lowest tone and reported to be creaky. Kuang (2017) concludes that “creaky voice in Tone 3 is phonetically just a by-product of laryngeal adjustment for producing phonological low pitch”. The role of creak is marginal in another well-studied Sinitic language: Cantonese (Yu and Lam, 2014), which is closer to Muong both geographically and in some phonological respects (such as the number of tones and the existence of stop-final syllables). From this perspective, it could seem perfectly reasonable to expect Muong to pattern in a similar way to these two Asian languages. Therefore, for the purpose of this study, we raise two main research questions, which are:

- What is the role of creaky voice in Kim Thuong Muong? How does it combine with pitch to compose the tone system?
- To what extent is there variation in the glottalized tone?

An attempt is made to prove the phonological status of creaky voice as part of the glottalized tone. Through experience with the dialect prior to the experimental study reported here, my initial hypothesis was that Muong should be classified in the same set as Vietnamese in terms of the use of glottalization as a dimension of the tone system alongside pitch. But Muong is saliently different from Vietnamese in using creaky voice as a distinctive phonation type: the phonetic realization of the creaky tone in Muong is thereby unlike that of the glottalized Vietnamese tones (etymological B<sub>2</sub> and C<sub>2</sub>), which have glottal constriction as their canonical realization.

Such is the research project carried out in the following chapters.



# Chapter 2

---

## Background

This chapter sets out background knowledge: information that needs to be referred to later on in the study. It includes a section setting out a theoretical framework for the study of prosody, intonation and tones, a review about linguistic uses of phonation types, and general information about the target language, Kim Thuong Muong.

### 2.1 Prosody, intonation and tone: a framework

Intonation studies currently find themselves in a paradoxical situation: the number of published works is growing steadily, but we remain far from a consensus even on the most elementary points of method, starting with the very definition of what constitutes intonation.<sup>1</sup> The lack of consensus is an issue for beginners entering this field of studies. It warrants taking the time to set out at some length the general framework adopted here for the study of prosody, intonation and tone.<sup>2</sup>

A guiding thread along this path consists in taking into account information concerning the intonation of languages which have lexical (sometimes also grammatical) tones: such is what is meant here by *tone languages*. Tone languages arguably offer a privileged angle of approach to establish certain fundamental conceptual distinctions, and to maintain them consistently.

#### 2.1.1 Phonetics, phonology and intonation

What is intonation? The field as a whole remains marked by a diversity of apparently irreconcilable points of view, from which there stem many misunderstandings. In his book entitled *Intonation*, Cruttenden starts out by delineating a *suprasegmental* or *prosodic* domain by contrasting it (quite classically) with the segmental domain: that of phonemes. Thus, the adjective *nice* has three phonemes: a nasal consonant /n/, a diphthong /aɪ/, and a fricative /s/. Such is its segmental composition: /nais/.

---

<sup>1</sup>This point was made eloquently by Amalia Arvaniti in her keynote address at the 2019 International Congress of the Phonetic Sciences in Melbourne (there is apparently no published text based on that talk).

<sup>2</sup>No claim for originality is made for this section, which is essentially translated from Michaud, M.-C. Nguyễn, and Scholvin (2021). The writing of the relevant section in the original article (§1) was done by Alexis Michaud.

But there are clearly other features involved in the way a word is said which are not indicated in a segmental transcription. The word *nice* might be said softly or loudly; it might be said with a pitch pattern which starts high and ends low, or with one which begins low and ends high; it might be said with a voice quality which is especially creaky or especially breathy. Such features generally extend over stretches of utterances longer than just one sound and are hence often referred to as suprasegmentals (...). Alternatively, the shorter term PROSODIC is sometimes used and I shall generally prefer this term in this book. (Cruttenden, 1986, p. 1)

Thus far, intonation is not mentioned. It appears in the next sentence, but fairly indirectly, as part of a passage tacked on inside brackets.

Prosodic features may extend over varying domains: sometimes over relatively short stretches of utterances, like one syllable or one morpheme or one word (...); sometimes over relatively longer stretches of utterances, like one phrase, or one clause, or one sentence (intonation is generally relatable to such longer domains). (*ibid.*)

It seemed necessary to quote these passages in full to bring out a paradoxical observation: intonation is not defined at the outset of the book (entitled *Intonation*), as if the meaning of the word were self-evident. Cruttenden contrasts the segmental dimension of an utterance (its consonants and vowels) with its suprasegmental dimension, which is understood to possess structure: use of the phrase “prosodic features” implies that prosody comprises an organization based on discrete units, namely features, a fundamental concept in classical phonology. Within the domain of prosodic features, one gathers that only some have intonational value. Readers of Cruttenden’s textbook are left to clarify for themselves, as best they can, what the term *intonation* covers. Are we to understand that intonation is an abstract linguistic structure, which concerns non-segmental units?

In an encyclopedia article by the same name (“intonation”), Francis Nolan provides a more explicit characterization.

The term intonation refers to a means for conveying information in speech which is independent of the words and their sounds. Central to intonation is the modulation of pitch, and intonation is often thought of as the use of pitch over the domain of the utterance. However, the patterning of pitch in speech is so closely bound to patterns of timing and loudness, and sometimes voice quality, that we cannot consider pitch in isolation from these other dimensions. (...) For those who prefer to reserve ‘intonation’ for pitch effects in speech, the word ‘prosody’ is convenient as a more general term to include patterns of pitch, timing, loudness, and (sometimes) voice quality. (Nolan, 2006, p. 433)

Nolan takes up the negative characterization of intonation (as being independent of phonemes) but enriches it by an important observation concerning the level of the *word*. To say that intonation is “independent of the words and their sounds” is to clarify that intonation is placed above the level of the word, and thus above lexical phenomena such as stress (in a language like English or German) and tones (in languages like Bambara, Mandarin, Vietnamese, Muong and many more). However, intonation is not explicitly characterized by F. Nolan as an abstract structure. The characterization that he provides suggests that intonation is essentially a matter of the use, in spoken communication, of three acoustic parameters: fundamental frequency ( $f_0$ , and its perceptual counterpart: pitch), intensity and duration. Feeling confident about the merits of a method that had been tried and tested successfully in the “segmental” domain, linguists have attempted to approach intonation as a topic of phonology. Their ambition is to uncover, below the endless variation of phonetic substance, a structure characterized by discrete (categorical) oppositions and by identifiable principles of organization. Readings about intonation typically deal with the *phonology of intonation*, or *intonational phonology* (Gussenhoven, 2002; Jun, 2005; Ladd, 2008). The phonetician-phonologist shoulders the full burden of clarifying “what intonation consists of, and how we can visualize it and analyze it phonologically” (Nolan, 2006, p. 434).

On the face of it, the approach which consists in progressing from experimental phonetics to phonological modeling conforms to the requirements of the scientific method. It is rooted in the empirical observation of phonetic phenomena (in particular the analysis of fundamental frequency curves), and guided by ambitious goals in terms of theoretical modeling. However, there are signs that suggest that intonation studies carried out according to phonetic-phonological methods are up against limitations, which the authors foresee to some extent. An example of this can arguably be drawn from a reflection by David Crystal according to which “intonation is not a single system of contours and levels, but the product of the interaction of features from different prosodic systems – tone, pitch-range, loudness, rhythmicity and tempo in particular” (Crystal, 1975, p. 11). Crystal’s reflection points to the need to look beyond a conception of intonation that would reduce it to issues of melody: contours (rises and falls) and levels (on a pitch scale). To evoke “different prosodic systems” is to open the door to the recognition of various dimensions of intonation, each of which could be approached with specific methods. But the opening thus achieved is immediately closed down again by offering a list that is limited to the familiar phonetic dimensions, commonly called “suprasegmental”, as they also appeared in Francis Nolan’s characterization of intonation quoted above:  $f_0$ , intensity and duration. These three parameters appear in a list that is admittedly intended to be open (“in particular...”), but which in practice amounts to foregrounding these three specific phonetic dimensions – as is natural enough for a phonetic-phonological study. After one has entertained high hopes of holding intonation under our gaze (to repeat F. Nolan’s formula quoted above: “how we can visualize it and analyze it phonologically”), it is difficult to give up hopes of near-immediate access and to enter a long detour through other dimensions of linguistic complexity: dimensions that are neither phonetic nor phonological. Such a research program is all the more

unattractive as the intonation of English – the most widely taught language worldwide, and consequently the most studied by linguists – can be adequately described (better than that of many other languages) in terms of patterns of fundamental frequency alone: the patterns that are referred to as “tones” in the British school of intonation studies (O’Connor and Arnold, 1973).

Seasoned practitioners of language description point out, however, that the task of studying intonation is not a straightforward one. Intonation “has considerable importance in oral communication, but has specificities that make it really troublesome to the linguist, since the methods that have been tried and tested in other areas do not seem truly adequate for the analysis of intonation” (Creissels, 1994, p. 173). Is the phonetician-phonologist well equipped to deal with intonation? The question may come as a surprise, since the association between intonation (whose links with oral speech are self-evident) and phonetics-phonology seems obvious. However, to venture a critical reflection about the field of phonetics/phonology at large, the emphasis on the sounds of speech tends to place the researcher in the position of an outside observer. It can have the effect of exempting him or her from confronting languages as instruments of communication. It is quite possible to carry out an experimental phonetic analysis of a given phonological opposition without speaking the language at issue. Among linguists, phoneticians/phonologists are not necessarily General Practitioners of linguistic analysis: they are not necessarily the best connoisseurs of the functioning of languages in the diversity of their dimensions. Seen in this light, entrusting phonologists with the task of clarifying the functioning of intonation looks like a paradox, as the study of this domain requires paying attention to communicative dimensions that are not exclusively (or even primarily) phonetic-phonological.

To progress in the understanding of intonation, it therefore appears promising to adopt a functional characterization, rather than a phonetic-phonological characterization. The definition of intonation adopted here is that elaborated by Mario Rossi, and adopted by Jacqueline Vaissière and Alexis Michaud (with minor adaptations, explained below). Rossi’s definition is not based on phonetic properties (physiological, articulatory, acoustic or perceptual).

Intonation, which has long been confused with one of its privileged parameters, melody, is a linguistic system designed to organize and prioritize the information that the speaker intends to communicate to the addressee(s) in his message, and to linearize the hierarchy of syntactic structures. (Rossi, 2001).

In the study of intonation, it appears as a good method to start from a characterization that is placed at this level of abstraction: that of the linguistic system.

Rossi distinguishes several strands in the understanding of the term *intonation*. Understanding intonation as a melody leads to a single-parameter approach (focusing on fundamental frequency), and to an interpretation in terms of phonological units. In the autosegmental-metric theory, these units are H (High) and L (Low) levels, which appear on a line (tier) distinct from the segmental line. Rossi’s approach stands in

strong contrast to the single-parameter approach, as it recognizes several components within intonation: “intonation is a part of prosody; prosody comprises accentuation, intonation and rhythm” (Rossi, 2001, p. 103).

*Accentuation* here refers to lexical stress. This characterization is sufficient in the field of Romance languages (the primary object of Mario Rossi’s study). On the other hand, from a typological perspective, it needs to be given a broader meaning to take into account other prosodic phenomena that affect the level of the lexical word, such as lexical tones. Insofar as they ensure lexical oppositions, these features meet the definition of the phoneme. Thus, some authors speak of *tonemes* to refer to lexical tones (Pike, 1948): the term has the advantage of linking up with the notion of allotones, variants of tone realization, parallel to allophones for segmental phonemes. Since the meaning of the concept of accentuation is thus broadened, it is appropriate to change the name as well. The term *accentuation* has a relationship with the notion of accent that no redefinition can effectively neutralize. For these lexically distinctive phonological entities, the name proposed is *lexical prosodic properties*. Also replacing *rhythm* with the more general term of *performance factors*, which includes rhythm, one arrives at the schema of prosodic components presented in Figure 2.1.

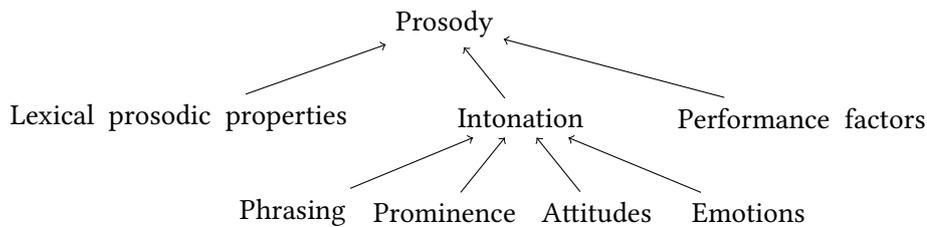


Figure 2.1: A highly schematic representation of the components of prosody. Reproduced from Michaud (2017), with a minor change: replacing the original term “Lexically distinctive properties” by “Lexical prosodic properties”.

With the above framework in mind, one is in a position to describe the relationships between tones and intonation without running into paradoxes.

### 2.1.2 Tones and intonation

Does the presence of lexical (or grammatical) tones imply that tonal languages are devoid of intonation? Clearly, the answer is no: tonal languages possess, in addition to tones, an intonation system of their own, so that it is not appropriate to distinguish “tone languages” and “intonation languages”. This necessary clarification has been reiterated many times through the history of linguistic theories (see in particular Hockett 1963; Zerbian 2010). In contrast, a more persistent misconception is that intonation necessarily plays a lesser role in tonal languages. Lexical tone and intonation share certain acoustic correlates (in particular fundamental frequency), and this use of the same communication channels (to put it in an information-theoretic perspective) leads to the belief that the

role of intonation in a tonal language is highly limited. This is the point of view expressed by Cruttenden, for example.

Tone and intonation are not completely mutually exclusive in languages. Languages with tonal contrasts may nevertheless make use of a limited amount of superimposed intonation. Such superimposed intonation may be manifested in four different ways: (i) the pitch level of the whole utterance may be raised or lowered; (ii) there will usually be downdrift in the absolute value of tones but downdrift may be suspended; (iii) the range of pitch used may be narrower or wider; (iv) the final tone of the utterance may be modified in various ways. (Cruttenden, 1986, p. 9)

Just as phonetically very similar phenomena can correspond, at the functional level, to a lexical accent in one language, and to intonational (pragmatic) emphasis in another, so intonational and tonal phenomena would be, if not in a mutually exclusive relationship, at least in a situation of competition, with intonation as the weaker competitor (emphasis added: “... *a limited amount of superimposed intonation*”). This point of view, which remains widespread, does not do justice to the richness of intonational phenomena observed in tonal languages. It seems more forward-looking to point out that “tone languages can have all features that characterize intonation-only languages, but not vice versa” (Steien and Yakpo, 2020, p. 5). The simultaneous presence of tonal and intonational phenomena is described in several works, including Downing and Rialland (2016) for various African languages, Vydrina (2017) for Kakabe (Mande group), and Michaud and Vaissière (2015) for some Asian languages. Without trying to summarize the literature, it seems worth highlighting the diversity of ways in which intonation and tone can coexist.

One such possible way is mutual avoidance. This is how Pittayaporn (2007) describes the situation of final particles in Thai, some of which have a lexical tone, others not. In his comparative study of the two categories of final particles, Pittayaporn suggests that intonation (described in terms of “boundary tones”, but functionally located on a different plane than lexical tones) is freely manifested on particles without tonal specification, whereas it is entirely neutralized in the case of particles with tonal specification. More complex modalities of coexistence were already identified by Luksaneeyanawin (1983), however. Various attitudes (assertion versus questioning, doubt, disbelief or surprise, for example) translate into global or semi-global patterns of fundamental frequency, on which lexical tones are superimposed. The phonetic literature on intonational cues conveyed by phonetic detail in the realization of vowels and consonants at a given point in the utterance (Fougeron, 1999; Kohler and Niebuhr, 2011; Niebuhr, 2013) highlights the diversity of prosodic cues, and hence the way in which tones and intonation pass simultaneously through channels that are certainly (in part) common, but diverse enough to circulate a wealth of information.

In light of the above clarifications, the topic of phonation types (specifically: glottalization) in relationship to tones and intonation can now be addressed.

## 2.2 Linguistic uses of phonation types

Below, a sample of publications about tone and phonation is reviewed in order to set the stage, without aiming at comprehensiveness: tones and phonation types would warrant a full Encyclopedia by themselves. The main goal here is functional: to shed light on the full range of terms used in the present study, so as to avoid (or at least limit) ambiguities and possible misunderstandings.

Phonetically, glottal stops occur in many languages, including English (Christophersen, 1952), French (Malécot, 1975) and German (Kohler, 1994). In British English, ‘glottaling’ of consonants (especially /t/) is common (Przedlacka, 2000). In many languages, glottal stop serves as an empty-onset filler: e.g. in French, empty-onset words that have little phonological material tend to be reinforced (set apart from the preceding words) by an initial glottal stop, as in *je m’appelle Yves*, which tends to be realized as /ʒmɛpɛlʔiv/ rather than /ʒmɛpɛliv/, or *je m’appelle Anne*, which tends to be realized as /ʒmɛpɛlʔan/ rather than /ʒmɛpɛlan/ (this observation was communicated by Michel Launey to Alexis Michaud).

Phonemic use of glottalization (glottal constriction or creaky voice) is relatively uncommon cross-linguistically. P. Ladefoged and Maddieson (1996, p. 48) mention phonation types as part of their discussion of stop consonants; the examples that they mention of languages making use of creaky voice in association to stops are Hausa and Mazatec. Another example is Hayu (Michailovsky (1988)). All three are likely to appear as highly “exotic” languages in the eyes of many phoneticians.

### 2.2.1 Phonation-type registers

Phonation-type registers are widespread in Southeast Asia, where they constitute a key topic in diachrony as well as in synchrony. However, they are reported to be fairly rare among the world’s languages at large. As a result, phonation-type registers may be unfamiliar to some readers, including persons with a thorough and well-rounded training in general linguistics. An introduction to this topic is therefore proposed below. It constitutes a digest of published materials; readers who are already familiar with the topic can safely skip this brief review, as well as the next, which goes into the topic of ways in which phonation types can be compounded with tones.

Register is a common phonological contrast in the Austroasiatic and Chamic languages of Mainland Southeast Asia (Henderson, 1952; Haudricourt, 1965; Gregerson, 1976; Huffman, 1976; Ferlus, 1979). It consists in a bundle of acoustic properties – the most important being  $f_0$ , phonation type, and vowel quality – that are realized on rhymes but originate from a voicing contrast in onsets (no voicing contrast in sonorants is reconstructed in Proto-Austroasiatic and Proto-Chamic). The low register (also second or lax register) derives from former voiced stops, while the high register (also first or tense register) stems from original voiceless stops. Low register syllables typically have a lower  $f_0$  and a laxer/breathier phonation and they

often have more close vowels (or vowels with close onglides). Note that register languages do not necessarily combine all of these properties, and that others, like VOT [Voice Onset Time] and vowel duration, are often associated with the contrast. This characterization of register, summarized in Table 1, is supported by acoustic evidence from Austroasiatic (Lee, 1983; L.Thongkum, 1987, 1989, 1991; Wayland, 1997; Watkins, 2002; Wayland & Jongman, 2003; Abramson, Luangthongkum, & Nye, 2004; Abramson, Nye, & Luangthongkum, 2007; DiCano, 2009; Abramson, Tiede, & Luangthongkum, 2015; Tạ, Brunelle, & Nguyễn, 2019) and Austronesian languages (Fagan, 1988; Edmondson & Gregerson, 1993; Hayward, 1995; Thurgood, 2004; Brunelle, 2005, 2009a, 2010; Matthews, 2015). (Brunelle, Thành Tấn Tạ, et al., 2020)

The literature review in DiCano's audio and electroglottographic study of Chong, a language of Cambodia (DiCano, 2009), offers a clear and handy introduction to the typological diversity of phonation-type registers.

Most register languages contrast only two phonation types, e.g. Middle-Khmer (Jacob 1968) and Wa (Watkins 2002). Those which contrast three are quite rare, but do exist, e.g. Jalapa Mazatec (Kirk, Ladefoged & Ladefoged 1993) and Bai (Edmondson & Esling 2006), and, of course, languages with a four-way phonation-type contrast are extremely rare. Only a few languages in the world have been found to use this number of contrasts: Chong (Thongkum 1991), !Xóõ (Traill 1985), Bai, and Bor Dinka (Edmondson & Esling 2006). Chong contains both dynamic (contour) and level registers. There is a modal register, a tense register, a breathy register, and a breathy-tense register.

Beyond the complexities of combinations among phonation-type registers, as explored by DiCano, a topic of special interest, which stands at the heart of the present work, is the association of phonation types with tones.

### 2.2.2 *Phonation types and tones: typological observations*

From a general linguistic point of view, there are interesting synchronic and diachronic ties between phonation types and tones. Phonation types arguably play a role in some tonogenetic processes. In the case of Vietic, the diachronic links between glottal stop (or glottal constriction) and tone are well-established (Ferlus, 2004). In languages of Europe, too, glottalization in a dialect can correspond regularly to tone in another.

Remarkably, the relation between tone contours in the north [of the Scottish Gaelic area] and glottalization in the south has an exact parallel in Scandinavian as well. Glottalization in Danish (the so-called *stød*) corresponds to word tone in Norwegian and Swedish in the following way: glottalization corresponds to tone 1, absence of glottalization corresponds to tone 2. (Ternes, 2006, p. 143)

Synchronically, even though each phonation type has natural association to a certain range of  $f_0$  (typically: low  $f_0$  and creaky voice), different languages clearly use different combinations of pitch and phonation types for tonal contrasts. An instance of synchronic combination between pitch and phonation type to produce tonal contrasts is found in Qingjiang Miao (Ch'ing Chiang Miao), a Black Miao dialect belonging to the Hmong-Mien (Miao-Yao) family. This language has five level tones, among a total of eight tones (the remaining consist of two rising and one falling). Phonation cues contribute to a comfortable auditory distance between the tones (Kuang, 2013b; Kuang, 2013a). In other words, the combination between pitch and phonation-type characteristics allows for well-dispersed tonal spaces (note that the term “tonal spaces” is used here in an abstract phonological sense, obviously not restricted to pitch).

Within the large and diverse landscape of phonation types and tones, there are several reasons that make a close look at Hanoi Vietnamese especially relevant as a preparation to the study of Muong.

### 2.2.3 Phonation types and tones: Hanoi Vietnamese as a textbook example

A look at Hanoi Vietnamese can serve as a handy introduction for readers with less familiarity with prosodic systems that use glottalization as part of lexical tones. There exists a relatively abundant literature covering the topics of the templates of the lexical tones as realized in isolation or in carrier sentences (T.-L. Nguyen et al., 2013), tonal coarticulation (Han and Kim, 1974), and topics of intonation (Brunelle, Hà, and Grice, 2012). Studies of Vietnamese tone set high standards for studies of tone in Muong, which should aim to vie with these achievements.

As an illustration of the acoustic outlook of glottalization, instead of taking up information from one of the above studies, let us examine data recorded in the year 1900 by a native speaker from the Hanoi area. The recording was done at the Paris *Exposition Universelle* by Léon Azoulay, of the Paris Society for Anthropology (*Société d'Anthropologie de Paris*).<sup>3</sup> Use of these historical materials adds a special twist to the inherent interest of this tone system, and arguably allows for relevant methodological insights concerning data collection as well as the estimation of fundamental frequency.<sup>4</sup> The recording includes the six tones of smooth syllables (syllables without a final stop consonant) as exemplified with the syllable *cơ* (IPA: /kɤ/).

Figure 2.2 shows tone *ngang* (etymological A1). Frequencies outside the range from 600 to 6,000 Hz are attenuated, presumably due to characteristics of the recording device, and do not show on the spectrogram at the chosen level of contrast (the dynamic range chosen for drawing the spectrogram is 40 dB). Also, there is noise in the range from 1,200 to 2,500 Hz, that likewise seems attributable to the recording device. It is

<sup>3</sup>The recordings can be listened to through the interface Telemeta: see <http://archives.crem-cnrs.fr/>. We are grateful to the curators of the archive for granting permission to access the master digital files for the purpose of the present research.

<sup>4</sup>This set of seven figures constitutes a slightly improved version of those already presented in my M.A. thesis (M.-C. Nguyễn, 2016). The discussion of these figures is taken up from that earlier thesis without any major modifications.

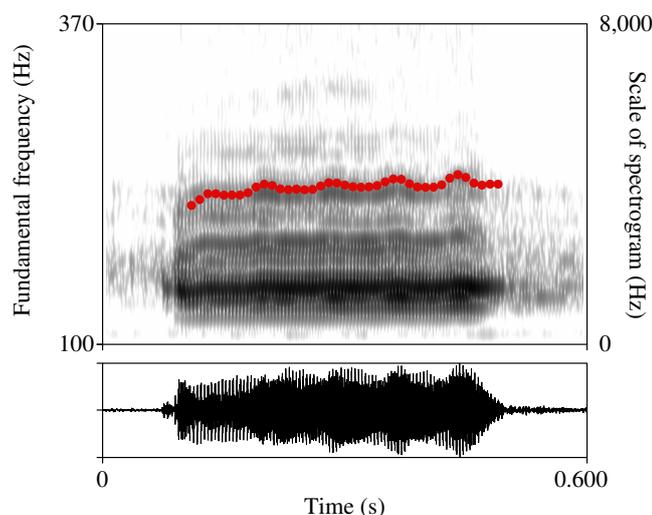


Figure 2.2: Vietnamese tone A1 (*ngang*), as realized in isolation over the syllable /kɤ/ by a speaker from Hanoi. Year of recording: 1900.

nonetheless possible to track fundamental frequency in these signals. Tone *ngang* (A1) looks phonetically level (relatively flat), in the upper part of the speaker's  $f_0$  range, but not at the top of the range. The noticeable jitter and shimmer are likely to be artefacts of the recording device. Shimmer and jitter apart, this tone does not seem noticeably different from present-day realizations, more than a century later. Phonologically, this tone is not among the glottalized tones of Vietnamese; phonetic observations about this token are in keeping with this phonological property: no glottalization is discernible either on the spectrogram, on the signal, or by ear.

Tone *huyền* (etymological A2), shown in Figure 2.3, is falling towards the end. Again, this is not a phonologically glottalized tone, and no glottalization is discernible in the phonetic realization of this token.

The items were originally recorded in the following order: *ngang huyền nặng sắc hỏi ngã* (etymological A1, A2, B2, B1, C1, C2). For the sake of clarity of exposure I shall nonetheless follow the etymological progression: A1, A2, B1, B2, C1, C2, and thus move on presently to tone B1 (Figure 2.4), tone *sắc*. It has a sustained phonetic rise, straightforwardly reflecting its characterization in the phonological literature as a rising tone.

Figure 2.5 shows tone B2, tone *nặng*, which is characterized by a strong final glottal constriction. This example introduces all at once several of the complexities associated with glottalization. The syllable shows strong changes in amplitude in the course of the vowel, pointing to glottalization in the first half of the syllable (contrary to the expectation that glottalization would be observed at the *end* of the syllable). Fundamental frequency then rises sharply, together with an increase in the signal's amplitude. Finally,

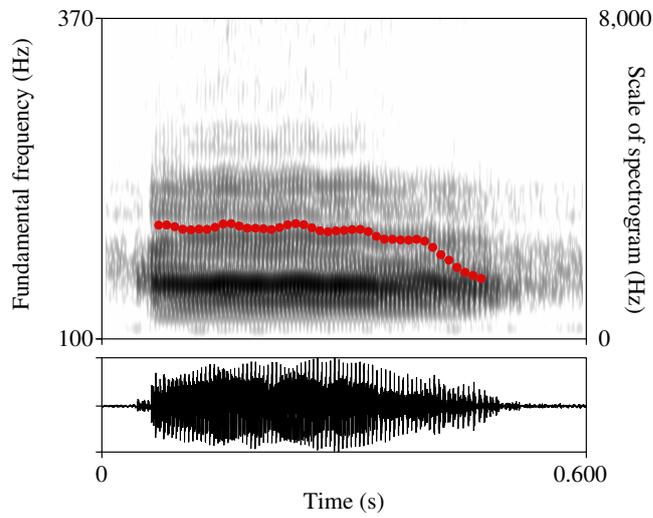


Figure 2.3: Vietnamese tone A2 (*huyền*), as realized in isolation over the syllable /*kɤ*/ by a speaker from Hanoi. Year of recording: 1900.

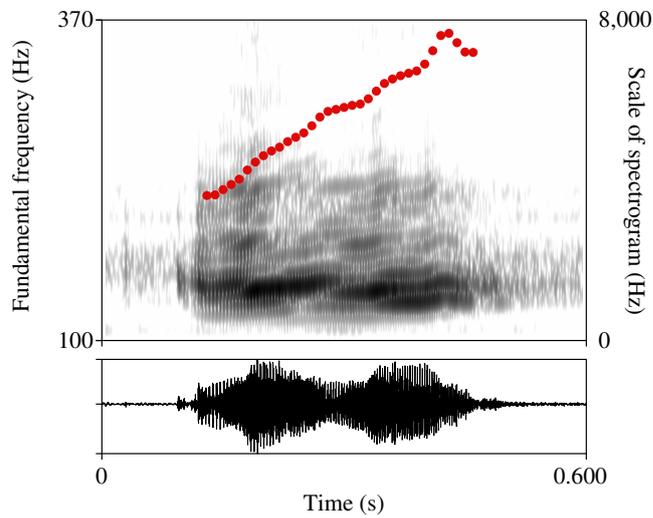


Figure 2.4: Vietnamese tone B1 (*sắc*), as realized in isolation over the syllable /*kɤ*/ by a speaker from Hanoi. Year of recording: 1900.

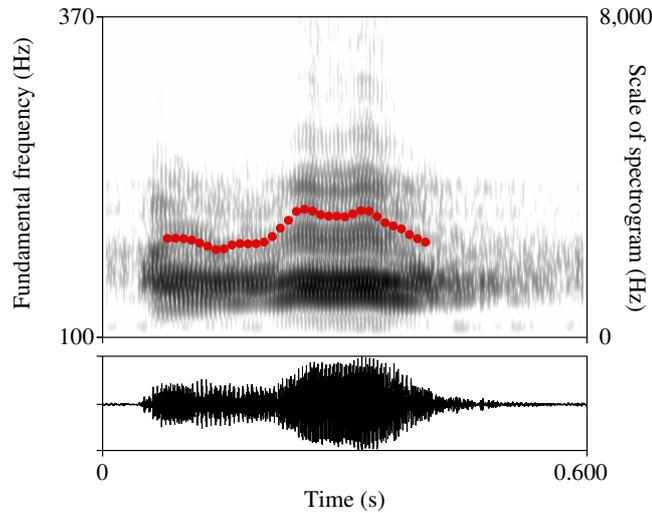


Figure 2.5: Vietnamese tone B2 (*nặng*), as realized in isolation over the syllable /*kɤ*/ by a speaker from Hanoi. Year of recording: 1900.

$f_0$  decreases, perhaps with some final glottalization. To my ear, this token of tone B2 sounds closer to *ngã* (C2): in my dialect (standard Northern Vietnamese), a syllable with glottalization followed by a modal interval with higher pitch is categorized as *ngã* (C2), whereas tone *nặng* (B2) is associated to syllable-final glottalization. In cases where the glottalization of B2 is followed by a modal-voiced interval, its pitch must be low, otherwise the syllable will be categorized as carrying tone *ngã* (C2).

If I allow myself to speculate as to why the speaker produced such a realization: it could be that this was the usual realization in his speech; but it seems more likely to me that during the rehearsal of this recording, the investigator felt that the standard realization of this tone would not be loud enough for a good recording: the syllable, cut short by glottal constriction, tends to sound as if it were strangled. It might therefore be that an artefact crept in: that the person in charge of the recording instructed the consultant to produce a stronger realization, resulting in a token which is even more hyper-articulated than the other tones: a longer syllable, with more voicing after the glottalization than would usually be observed for this tone. The diversity of intonational variability of the tones serves as a word to the wise: special care is necessary at elicitation of glottalized tones.

The last two tones, C1 and C2, are shown in Figures 2.6 and 2.7 respectively. When comparing the pair formed by this speaker's B2 and C2 tones, the difference becomes aurally clear, because his tone C2 has noticeably higher overall pitch, and ends on a much higher pitch than C2.

This observation illustrates (i) the importance of the phasing of glottalization, and (ii) the phonetic relativity of the cues to tone, including phonation types. It calls for

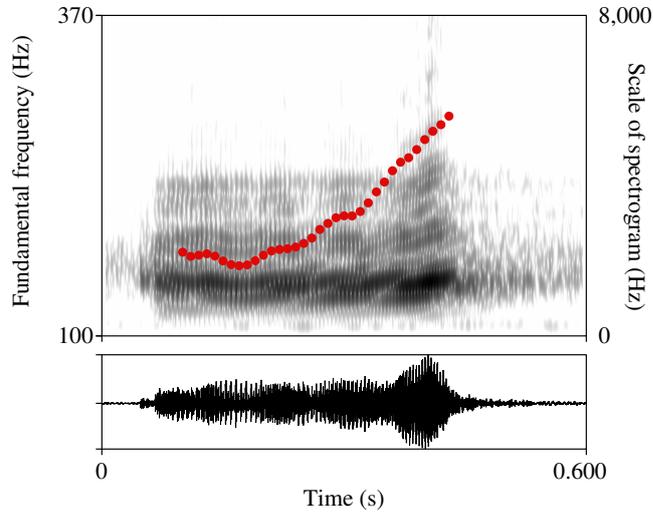


Figure 2.6: Vietnamese tone C1 (*hỏi*), as realized in isolation over the syllable /*kɤ*/ by a speaker from Hanoi. Year of recording: 1900.

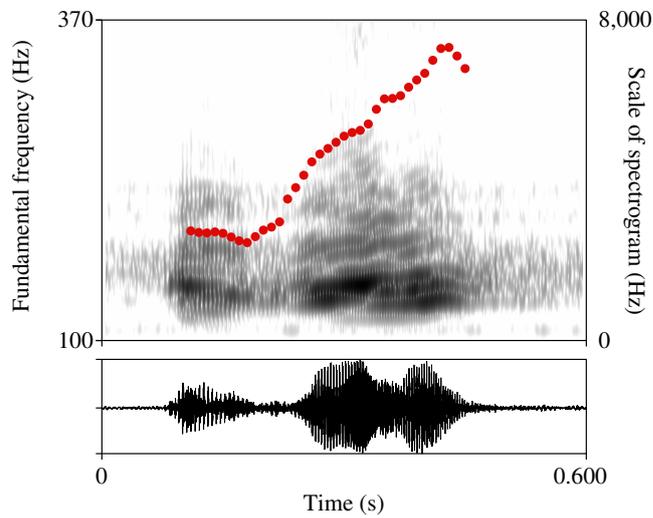


Figure 2.7: Vietnamese tone C2 (*ngã*), as realized in isolation over the syllable /*kɤ*/ by a speaker from Hanoi. Year of recording: 1900.

tools to describe these phenomena with high accuracy, to progress towards phonetic and phonological models that get as close as possible to these complex phenomena.

In this endeavour, an important aspect is the exploratory technique used. Fundamental frequency detection as implemented in phonetic software packages such as PRAAT can fail for glottalized portions, because the algorithm requires, if not quasi-periodicity in the signal, at least a degree of similarity between successive glottal cycles – a requirement which is often not met in audio signals corresponding to glottalized portions of speech, as exemplified by Figure 2.8a, also taken from the 1900 recording. The token is a realization of Vietnamese tone C2 (*ngã*) over the syllable /**ŋa**/.

If the “Advanced pitch setting” voicing threshold in PRAAT is set at 0.7, spurious periods of high  $f_0$  are detected (see Figure 2.8a). If this parameter is set at 0.8, these spurious periods disappear, but no  $f_0$  is detected on the initial portion of the syllable either: see Figure 2.8b. Over one hundred years after this speech signal was recorded, there exists no simple way out of this technical difficulty.<sup>5</sup>

This last observation, which highlights the usefulness of exploratory techniques to complement audio recordings, provides a handy transition to the topic of electroglottography.

### 2.3 Definitions: glottalization and creaky voice

Uses of glottalization and creak range from the phonemic to the ‘paralinguistic’.

... an American English speaker may have a very creaky voice quality similar to the one employed by speakers of Jalapa Mazatec to distinguish the word /**j̥á**/ meaning ‘he wears’ from the word /**já**/ meaning ‘tree’ (Kirk, J. Ladefoged, and P. Ladefoged, 1993). As was noted some time ago, one person’s voice disorder might be another person’s phoneme (P. Ladefoged, 1983).

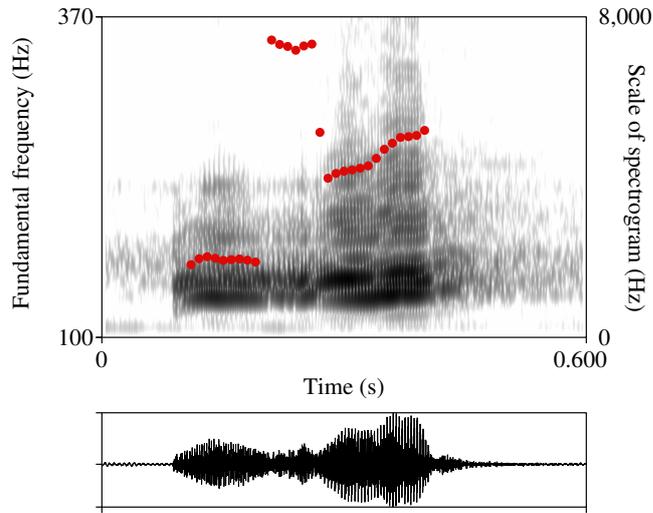
(Gordon and P. Ladefoged, 2001, p. 383)

The developments that precede – in particular the introductory note in 1.1, the review of linguistic uses of glottalization in §2.2, and the results in Chapter 4 – now place us in a position to put forward some definitions, and to progress towards a systematic characterization of glottalization in Kim Thuong Muong.

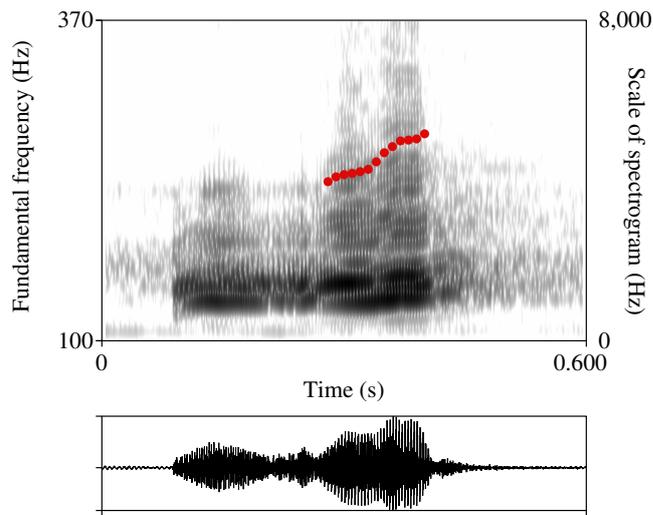
In this thesis, ‘glottalization’ is used as a cover term for glottal constriction, creaky voice, and all the phenomena that range in-between them. Inside the field of glottalization, a distinction is made between creak, on the one hand, and on the other hand glottal constriction and glottal stop – keeping in mind that there can be intermediate cases.

---

<sup>5</sup>The software package WinPitch (<http://www.winpitch.com/>) offers a range of algorithms for calculating speech fundamental frequency from audio signals. For this particular example, WinPitch’s implementation of the AMDF method (Average Magnitude Difference Function) yields good results. My point here is not that  $f_0$  can never be computed for glottalized syllables, but that its computation is by no means straightforward, and that this topic is well worth investigating further.



(a) Illustration of mistaken values at  $f_0$  extraction during glottalization. Voicing threshold set at 0.7 in Praat.



(b) Illustration of mistaken values at  $f_0$  extraction during glottalization. Voicing threshold set at 0.8 in Praat.

Figure 2.8: Vietnamese tone C2 (*ngã*), as realized in isolation over the syllable /**ŋa**/ by a speaker from Hanoi. Year of recording: 1900.

### 2.3.1 Literature review of creaky voice

#### 2.3.1.1 The process of becoming an independent object of study in the field of voice quality

Today, creaky voice is a mainstream topic in phonetics/phonology and beyond, witness e.g. a review article bringing to the attention of cognitive scientists “[t]he versatility of creaky phonation: Segmental, prosodic, and sociolinguistic uses in the world’s languages” (Davidson, 2021). That constitutes an ideal environment to carry out an in-depth monographic study like the present one. However, it may be worth highlighting that interest in creak as a topic of scholarly investigation is a relatively new development. It is natural enough that it had earlier (say, until the first half of the twentieth century) been considered mostly in relation to pathology, rather than to linguistic (phonemic and intonational) uses. Creak does not have phonemic status in the languages with which the *Association phonétique des professeurs de langues vivantes* was primarily concerned. Hence its absence from the table of phonetic symbols in the initial set, published in their journal, *Le Maître phonétique*, and which eventually expanded into the International Phonetic Alphabet as we know it, complete with a special symbol for creaky voice. In fact, creaky voice was at first not distinguished from harshness, considered as a voice disorder (Michel, 1964, p. 2) or a laryngeal pathology (Sherman and Linke, 1952). Paul Moore and Leden (1958) considered creak as “a peculiar voice quality” (p. 224) or even “a non-language sound” (p. 234). As noted by Peter Ladefoged in a famous quote:

What is phonemic in one language may be pathological in another. This is especially true with respect to voice quality. (P. Ladefoged, 1981)

It seems that the limitations of the vision of creak as a speech defect were early realized by some authors. There were efforts to clarify subsets within the phenomena pooled together under the label of ‘harshness’, adding other technical terms – although these efforts did not go without their own risks of bringing up more ambiguity. Anderson (1950) suggests several terms such as “breathiness, hoarseness, and huskiness, harshness and stridency, and throatiness” as labels for voice disorders. However, the boundaries among his categories are not crystal clear. He considers *harsh* as a partial description of huskiness and throatiness, while *breathiness* serves as one of the characteristics to describe hoarseness; I can see how these associations can match real examples, but would find it awkward to pin these together in hard-and-fast categories, as I would see no contradiction in different patterns, such as association of the notion of *harshness* with *hoarseness*. Van Riper and Irwin (1958) (cited in Paul Moore and Leden 1958, p. 255) use both terms “harsh voice” and “strident voice” for the same voice quality; they firstly recognize “glottal fry” as a distinct component within “harsh” voice quality, but later on in the same paper they mention “glottal fry” as an equivalent to “hoarseness”, which is treated apart from harshness and stridency. Paul Moore and Leden (1958) comments that there is “considerable ambiguity in the use of the terms. The semantic problems

arise from the fact that various writers have employed one term to designate different qualities, and conversely, a variety of names have been applied to a single quality.”

A clear stride towards clarity of phonemic nomenclature is achieved when the perspective on creaky voice (vocal fry) changes from “an abnormal voice quality relating to harshness” to a normal function. A landmark here is arguably the publication of two studies: Hollien (1963) and Hollien, P. Moore, et al. (1966, p. 246). They postulate that vocal fry exists as a “physiologically normal mode of laryngeal operation”. They also suggest that vocal fry can be “best described as a phonational register occurring at frequencies below those of the modal register”.

I hasten to add that the emphasis placed here on these two studies does not imply that there are no earlier published statements pointing in the same direction. Moser (1942) and Van Riper and Irwin (1958) “have alluded to the possibility that it may be a normal function” (Michel, 1964, p. 3). The idea here is that Hollien’s work contains a cogent, articulate picture.

Supporting Hollien’s proposing, John Frederick Michel distinguishes creaky voice (called *vocal fry* at that time) from harshness in particular and from voice disorder in general. In his PhD thesis, Michel provides the following clarification:

...the distinction must be made between clinical harshness and simulated harshness. The former may be characterized in terms of the Curtis definition and is typified by a functional or organic problem. Simulated harshness can be defined as a voice which is perceived to be rougher than normal but exhibits no pathology. Unfortunately, simulated harshness probably consists primarily of vocal fry. (Michel, 1964, p. 8)

Later, more and more authors and articles supported this viewpoint, based on different dimensions and approaches.

*Aerodynamics:* McGlone (1967), the first author who provided information about air flow during vocal fry phonation, agrees with Hollien, P. Moore, et al. (1966) that “this phonation involves a physiologically normal mode of laryngeal operation which results in a distinctive acoustic signal”.

*Production:* Hollien and Michel (1968) explores further the possibility that vocal fry is a separate phonational register from modal voice and falsetto. A broad data set is used: materials from 12 males and 11 females. “The results provide data that demonstrate vocal fry to be a register as specified” (Hollien and Michel, 1968, p. 600).

*Perception:* In the same year, Hollien collaborated with Wendahl to conduct a perceptual study of vocal fry. The result showed that “repetition rate can be successfully assigned to vocal fry. In addition, evidence was obtained indicating that vocal fry occurs within a low-frequency phonatory register, is perceived similarly in relation to single or double glottis pulses, and is regular, rather than aperiodic, in nature” (Hollien and Ronald W. Wendahl, 1968, p. 506). In this paper, in an attempt to separate the relationship between vocal fry and voice disorders, they make the following claim:

In the past, vocal fry has often been equated with rough or harsh voice quality. Recently, however, harshness and hoarseness have been associated

with irregular vibrations of the vocal folds, the presence of aperiodicity in the signal, or a range of frequencies falling within the modal register. Thus, the low frequency and the perceptual regularity of vocal fry found in this study provide evidence against classifying it among the voice pathologies. (Hollien and Ronald W. Wendahl, 1968, p. 509).

This long citation clarifies that they only recognize vocal fry with regular single or double pulses (the case of triple pulses is also reported but quite rare) as a normal mode of laryngeal production. Aperiodicity is still set on the side of voice disorder: this is used as a criterion for distinguishing *fry* from *harshness*. This criterion was subsequently challenged, however, as linguistic studies pointed to numerous cases where aperiodicity is among phonetic sub-types of creak (Redi and Shattuck-Hufnagel, 2001; Keating, Garellek, and Kreiman, 2015). In the same year, Michel explicitly proposed to base the fundamental distinction between vocal fry and harshness, not on the presence or absence of aperiodicity, but simply on fundamental frequency. He reported that: “A mean fundamental frequency of 36.4 Hz with a range of 30.9 to 43.7 Hz was found for vocal fry; the mean of the harsh voices was 122.1 Hz within a range of 103.7 to 180.0 Hz” (Michel, 1968, p. 590).

Of course, using  $f_0$  as a criterion requires that this parameter be measurable in the first place, and thus, that there be some kind of periodicity (cyclicity) in glottal behavior, thereby making it possible to exclude (or give separate treatment to) cases of wild aperiodicity. Not all instances of creaky voice are content with a neat phonetic corridor of low fundamental frequency such as reported in the above quotation, where the creaky voice is so well-behaved that the investigator ventures to describe it with a precision of less than 1 Hz (“30.9 to 43.7 Hz”). This quibble apart, the shift in emphasis away from periodicity as such appears well-warranted from a linguistic perspective. Laver (1980) agrees with Michel: “The low fundamental frequency of this creak type of phonation is one factor that distinguishes it from harsh voice, which is otherwise somewhat similar”. The present study is among those that gratefully take up the usage thus clarified.

*Physiology:* Using ultraslow motion pictures, Timcke R., H., and Moore P. (1959) are probably the first authors who explored the physiological mode of laryngeal operation that leads to perception of creaky voice. One of the most typical characteristics that must be met for its perception is a train of discrete excitations or pulses. This characteristic has been explained by a “nearly complete damping of the vocal tract between successive excitations”. The “double glottis pulse” has been described as a unique vibratory pattern in which “the vocal folds opened and closed twice (per cycle) in rapid succession and then remained closed for relatively long periods of time” (Hollien and Ronald W. Wendahl, 1968).

Nowadays, after more than 60 years of experimental studies of creaky voice, Hollien’s statement (already cited above) that this voice quality “is best described as a phonational register occurring at frequencies below those of the modal register” (Hollien, P. Moore, et al., 1966, p. 246) is uncontroversially acknowledged. However, there is still a margin for progress in terms of concepts and terminology, and a need for clarification concerning

the choices made in the present study among competing nomenclatures.

### 2.3.1.2 Creaky voice: nomenclature and sub-classification

The large terminological diversity for glottalization and creaky voice may be to some extent inherited from its relatively long history of nonstandardized nomenclature, prior to the clarifications recapitulated above. Creak has been identified in the literature under many different names, and this voice quality remains characterized by complex terminology. Before it came to be distinguished from harshness in the classifications of voice disorders, it was included under categories such as X-factor, Dysphonia ventricularis (cited in Paul Moore and Leden 1958, p. 235), Harshness or Harsh voice (Van Riper and Irwin, 1958; Bowler, 1964), Strident voice (Van Riper and Irwin, 1958), Dicrotic dysphonia, and Glottal Fry (Paul Moore and Leden, 1958). Commenting on this unclear nomenclature and the “considerable ambiguity in the use of the terms”, Moore and Leden suggest that: “The semantic problems arise from the fact that various writers have employed one term to designate different qualities, and conversely, a variety of names have been applied to a single quality” (Paul Moore and Leden, 1958, p. 225).

The specific voice quality has had limited identification in the literature under such names as glottal fry, x-factor, harshness (restricted definition), and dysphonia ventricularis. The specific quality is usually not recognized as a distinct entity in the literature and, therefore, is obscured in such term as harshness or hoarseness. The general result is a confused terminology. (Paul Moore and Leden, 1958, p. 235)

Pertaining to “Dicrotic dysphonia”: Some others believe this is an ideal term for this voice quality because it identifies the primary acoustic factors and associates them with the basic laryngeal physiology (...); “dicrotic dysphonia” refers to a faulty vocal sound produced by a vibratory pattern containing a double pulsation. (Paul Moore and Leden, 1958, p. 231)

Once it is recognized that the phenomenon at issue is a normal human phonational register, it is more consistent with names such as vocal fry (R. W. Wendahl, G. P. Moore, and Hollien, 1963; Coleman, 1963; Michel, 1964; McGlone, 1967) and creaky voice (Laver, 1980; P. Ladefoged and Maddieson, 1996; Keating, Garellek, and Kreiman, 2015; Kuang, 2017). No conceptual difference is made here between the labels ‘vocal fry’ and ‘creaky voice’.

These pairs of terms have been used somewhat confusingly, and it seems best not to try to attach specific meanings to each term. (P. Ladefoged and Maddieson, 1996, pp. 50–53)

A further clarification concerns the terms ‘laryngealized’ and ‘laryngealization’. Equivalence between ‘creaky’ and ‘laryngealized’ is widespread in linguistic usage. It is advocated by P. Ladefoged and Maddieson (1996), and taken up e.g. in D. Crystal’s *Dictionary of Linguistics and Phonetics*, where ‘creak’ is defined as follows:

A term used in the phonetic classification of voice quality, on the basis of articulatory and auditory phonetic criteria. It refers to a vocal effect produced by a very slow vibration of only one end of the vocal folds; also known as vocal fry. ... Creaky sounds are also called 'laryngealized'. (Crystal, 2011, pp. 121–122)

Some authors nonetheless reserve the terms 'laryngealized' and 'laryngealization' for a specific usage, e.g. as a cover term for creak and glottal constriction (as I did in my M.A. thesis, following Michaud 2004b). But in practice, the usefulness of such a cover terms appears somewhat limited, whereas the decision to redefine such a familiar term, whose diversity of usages remains strong, would place a burden on readers. The terms 'laryngealized' and 'laryngealization' are therefore avoided in this thesis, for the sake of legibility (the same reason why acronyms are avoided to the greatest possible extent).

Once the term 'creaky voice' is unambiguously adopted, there remains the topic of sub-classifications within creaky voice, as well as the topic of the overall nomenclature within which creaky voice is included, and which lends it its structural meaning.

It seems that, in the historical development of phonation type nomenclatures, there were initially no further distinctions: no taxonomy for sub-types within creaky voice, even though various authors recognized the existence of at least two kinds of creak (namely, single and double glottis pulses). Hollien described the two patterns as a unique vibratory mode. A perception test carried out on single and double pulses led to the conclusion that: "The number of pulses produced is immaterial to the perception of phonation as vocal fry" (Hollien and Ronald W. Wendahl, 1968, p. 509). This was clearly a strong reason not to enter into sub-classifications of creak at that time. Proposals for sub-classification of creaky voice apparently date back to the 1990s, with the studies of Hedelin and Huber (1990), Batliner et al. (1993), Gerratt and Kreiman (2001), Redi and Shattuck-Hufnagel (2001), and Keating, Garellek, and Kreiman (2015). Let us proceed to review them in some detail.

Hedelin and Huber (1990) use a study of Swedish intonation as a basis to suggest a distinction between four different patterns of aperiodic glottal excitation, observed to occur consistently at different kinds of text junctures. These are: glottalization, creak, creaky voice and diplophonic phonation. Their acoustic patterns are illustrated by Figure 2.9 (reproduced from Hedelin and Huber 1990, p. 361). "Laryngealization" is used as a cover term for the four types mentioned above. This is consistent with Batliner et al. (1993) but conflicts with P. Ladefoged and Maddieson (1996) who used the term laryngealization as interchangeable with creaky voice. In this classification, creak is even distinguished from creaky voice. The term creak refer to a pattern of low frequency (in the range between 20 Hz and 50 Hz) and strong damping of vocal tract between successive excitations. "Alternative names often used in the literature for this pattern of glottal excitation include vocal fry, pulse register and strohbass." Whereas creaky voice is characterized by "aperiodicity and fluctuations in intensity". The third type is diplophonic phonation which is characterized by "an alternation between strong and weak glottal excitations during phonation". Therefore, it correspond to the pattern of double glottis pulses as explored by previous studies.

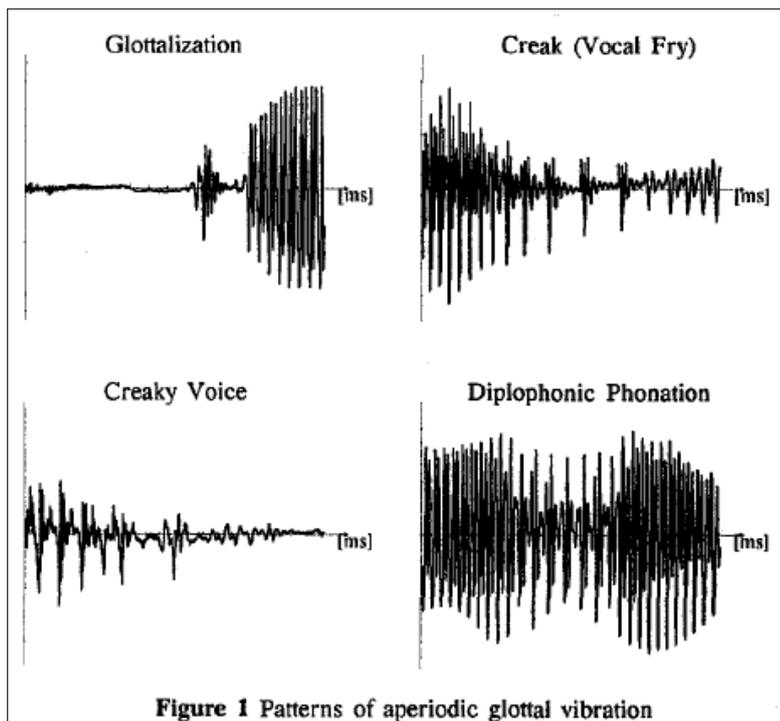


Figure 2.9: Acoustic patterns of four aperiodic glottal vibration. Reproduced from Hedelin and Huber (1990, p. 361).

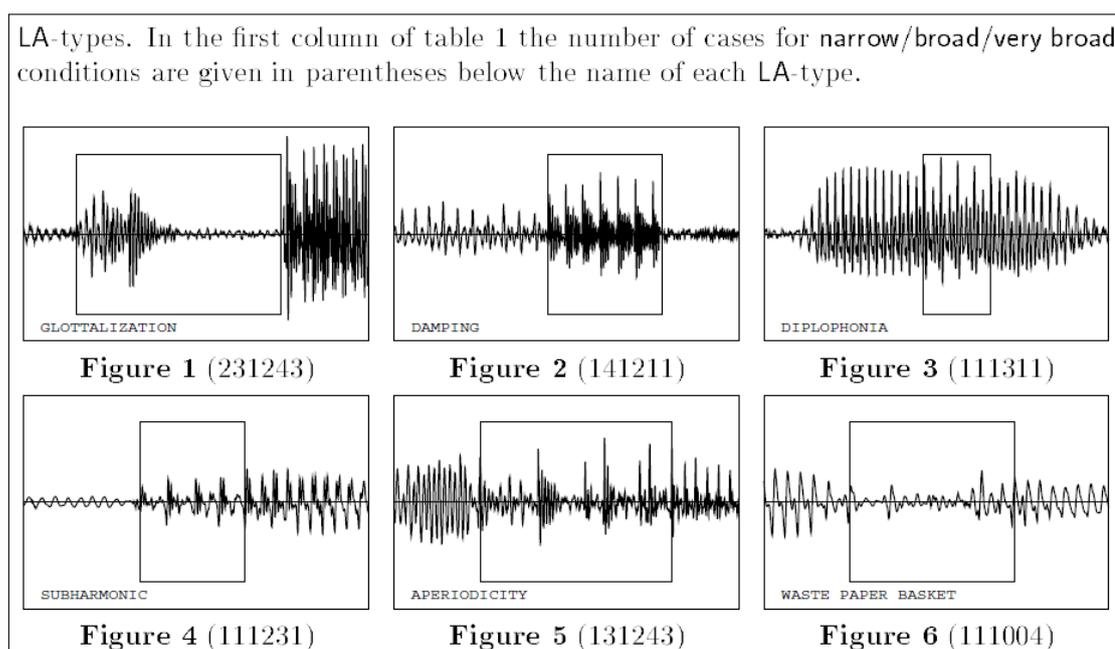


Figure 2.10: Acoustic patterns of six laryngealized types. Reproduced from Batliner et al. (1993).

As just mentioned, Batliner et al. (1993) agree with Hedelin and Huber (1990) in using “laryngealization” as a cover term for all phenomena of the voice quality under discussion. They use six acoustic properties to distinguish six different types of “laryngealization” as reproduced in Figure 2.10, but their brief report provides little discussion and in my experience it is exceptionally difficult to follow. The summary attempted in Table 2.13 is mainly based on the similarity of the acoustic signals with those in other studies.

There is a conflicting use of terms between Redi and Shattuck-Hufnagel (2001) and the two works just mentioned above. Whereas Hedelin and Huber (1990) and Batliner et al. (1993) classify glottalization as a sub-type within laryngealization, beside other sub-types of creakiness, Redi and Shattuck-Hufnagel (2001) use the term *glottalization* as a cover term for four sub-type of creak:

1. aperiodicity: irregularity in duration of glottal pulses from period to period;
2. creak: lowering of fundamental frequency with near-total damping;
3. diplophonia: regular alternation in the shape, amplitude, or duration of successive periods;
4. glottal squeak: a sudden shift to relatively high sustained  $f_0$ , with a signal that usually has very low amplitude.

A widely cited article about the classification of creak is Keating, Garellek, and Kreiman (2015). In this paper, they set out a “prototypical creaky voice” as a standard pattern with three key properties: (i) low  $f_0$ ; (ii) irregular  $f_0$ ; and (iii) constricted glottis:

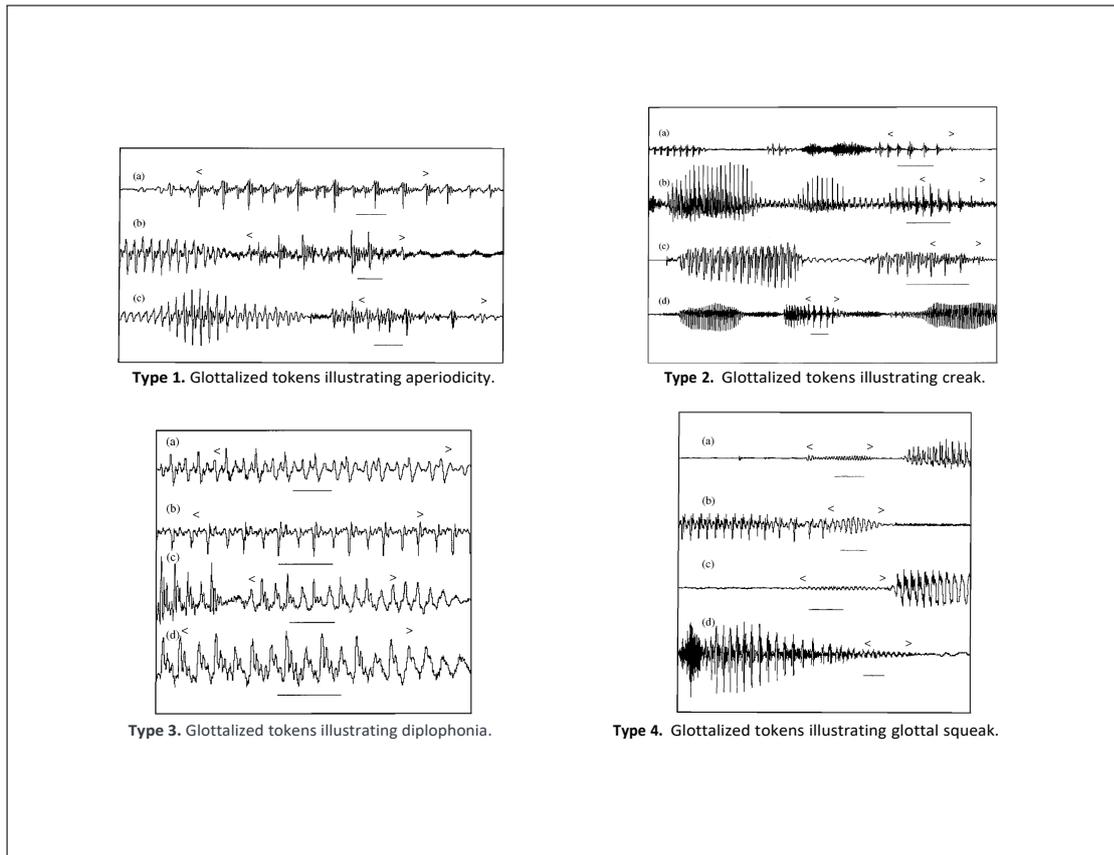


Figure 2.11: Acoustic patterns of four glottalized tokens illustrating different sub-types. Angled brackets above each waveform indicate the region where each sub-type appears. Reproduced from Redi and Shattuck-Hufnagel (2001, pp. 415–416).

a small peak glottal opening, long closed phase, and low glottal airflow. Then, “each of the three properties of prototypical creak can be lacking, yielding several further kinds of creak”. They provided a table to illustrate the five different kinds of creak with their own set of characteristics (as reproduced in Figure 2.12). Among them, tense or pressed voice is not considered as a type of creak but “can function phonologically as such in languages in which a creaky (or laryngealized) phonation can co-occur with high tone”.

Different classifications found in the literature are summarized in Table 2.13, where I attempt to connect terms that appear (at least roughly) equivalent from among terminologies used by different authors. This classification is approximative and relative, because not only do the labels differ across authors, but the conceptual boundaries that they draw inside the landscape of phonation types also differ. Different authors use the same term for what others consider different contents, and different terms for what others consider as the same contents. Therefore, it is often unclear to what extent the terms that seem equivalent actually coincide in their intended extension.

This topic will be taken up in the Discussion chapter (§5.2), where a working nomenclature will be proposed.

### 2.3.2 Creaky voice and laryngealization

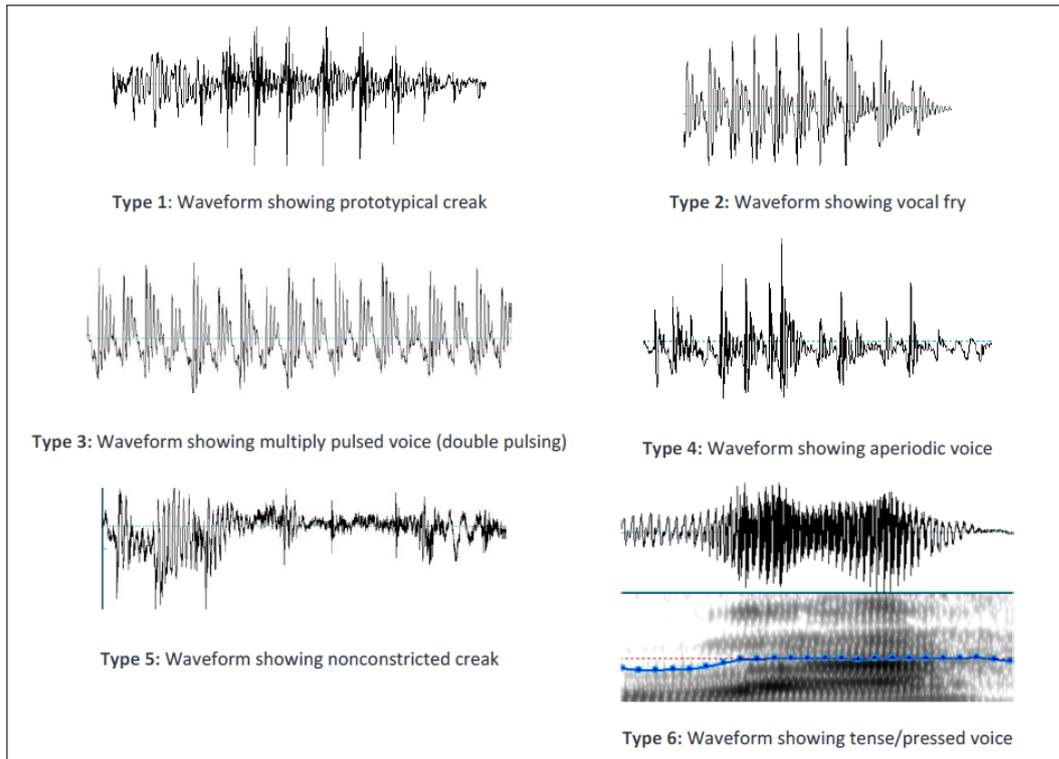
Auditory criteria indeed seem to be central to the notion of ‘creak’, since the term itself refers to an auditory impression. But paradoxically, it seems that phonetic studies of creak focus mostly on production and acoustics. This is a point that I would be interested to investigate further in future: the perceptual counterpart of different types of creak, and how perceptual data can help us find out more about which aspects of the acoustic signal are most relevant to the language users’ perception.

Now coming to more technical phonetic description, Patricia Keating, Marc Garellek and Jody Kreiman propose on the basis of a review of the literature that prototypical creaky voice has three key properties:

- (1) low rate of vocal fold vibration ( $f_0$ ), (2) irregular  $f_0$ , and (3) constricted glottis: a small peak glottal opening, long closed phase, and low glottal airflow. (Keating, Garellek, and Kreiman, 2015)

In detail, the articulatory and acoustic part of the story is not tidy, and departure from canonical realizations are numerous. The list of articulatory and acoustic correlates of creak put together in a recent review is impressive:

[Laver (1980) mentioned] low subglottal pressure and glottal flow, slack, thick, compressed vocal folds with a short vibrating length, ventricular contact with the folds, weak or damped pulses, low  $f_0$ , irregular  $f_0$ , period-doubled vibration. Later descriptions (e.g. Childers and C. Lee (1991), Gobl (2010), and D. Klatt and L. Klatt (1990)) added such properties as irregular amplitude, low Open Quotient, skewed glottal pulses, narrow formant bandwidths and sharp harmonics, abrupt closure of the folds, and low spectral tilt. (Keating, Garellek, and Kreiman, 2015)



(a) Waveform showing six kinds of creak.

**Table 1:** Properties characterizing different kinds of creak. Check mark means a property characterizes a type; NO means it does not; blank means variable or unknown.

Property	low F0	irreg F0	glottal constr	damped pulses	sub-harms
Main correlate	low F0	high noise	low H1-H2	low noise; narrow BWs	high SHR
Type					
prototypical	√	√	√		
vocal fry	√		√	√	
multiply pulsed		√	√		√
aperiodic	NO	√	√		
nonconstricted	√	√	NO		
tense	NO		√		

(b) Table summarizing the properties characterizing different kinds of creak. Check mark means a property characterizes a type; NO means it does not; blank means variable or unknown.

Figure 2.12: The sub-classification of creaky voice by Keating, Garellek, and Kreiman (2015). Reproduced from Keating, Garellek, and Kreiman (2015, pp. 1–3).

Authors	Sub-classification				
Hedelin & Huber (1990)	<b>Creak / vocal fry</b> - low f0 - strong damping	<b>Creaky voice</b> - aperiodic pulses - fluctuations in intensity	<b>Diplophonic phonation</b> - double glottis pulses		
Batliner et al. (1993)	<b>Damping / creak</b>	<b>Aperiodicity/ creak/creaky voice</b>	<b>Diplophonia</b>		
Redi & Shattuck-Hufnagel (2001)	<b>Creak</b> Lowering of fundamental frequency with near-total damping	<b>Aperiodicity</b> Irregularity in duration of glottal pulses from period to period.	<b>Diplophonia</b> Regular alternation in shape, duration, or amplitude of glottal period		
Keating, Garellek & Kreiman (2015)	<b>Prototypical creaky voice</b> - low rate of vocal fold vibration (F0), - irregular F0 - constricted glottis: a small peak glottal opening, long closed phase, and low glottal airflow  <b>vocal fry</b> - The glottis is constricted and F0 is low, but it is not necessarily irregular (quite periodic). - high damping of the pulses	<b>Aperiodic voice</b> - F0 irregularity: no periodicity in vocal fold vibration => no perceived pitch. - Lacks the prototypical property of low F0 (like multiply pulsed voice); instead, the property of irregular F0 is enhanced.	<b>Multiply pulsed voice</b> - A special kind of F0 irregularity: alternating longer and shorter pulses. - It can be double pulsing (or period doubling), or higher multiples are also possible. - Prototypical low-F0 is not necessarily present (indeterminate pitch)	<b>Nonconstricted creak</b> - F0 is low and irregular, as in prototypical creak; but the glottis is spreading, not constricted - Occur in utterance – finally	<b>Tense/pressed voice</b> - The glottis is constricted, but the F0 is neither low nor irregular - When a creaky (or laryngealized) phonation can co-occur with high tone

Figure 2.13: Nomenclature systems of creak sub-classification provided by previous studies.

Given this situation, it may appear necessary to phoneticians to confront the diversity headlong, and propose a phonetic categorization, together with recommendations on tools to tell apart the different categories. Building on earlier three- or four-way classifications of creaky voice into different types, Keating et al. propose a classification of the considerable diversity of phonetic realizations of creaky voice into five kinds. Not all of these five types have the characteristics of prototypical creak. This is a phonetic classification intended to help researchers choose acoustic measurement tools that are appropriate for the kind of creak that they aim to capture; this is not a classification primarily based on the linguistic functions of creak.

- **Vocal fry:** “Creak that is vocal fry with a regular  $f_0$  could instead show higher HNR [Harmonics-to-Noise Ratio] together with lower formant bandwidths.”
- **Multiply pulsed voice:** “Creak that is multiply pulsed can lack a clear  $f_0$  but instead show sub-harmonics (resulting in higher values of SHR).”
- **Aperiodic voice:** Creak that lacks the prototypical property of low  $f_0$ ; instead, its properties are low H1-H2 and irregular  $f_0$ .
- **Nonconstricted creak:** “Creak can instead show higher H1-H2, but still with a low and irregular  $f_0$ .”
- **Tense/pressed voice:** “Creak that is more like tense or pressed voice can have a mid or high, and regular,  $f_0$ .” (Keating, Garellek, and Kreiman, 2015)

The acoustic correlates of these five types are summarized in table form in Figure 2.12 of Chapter 2.

### 2.3.3 Glottal constriction and glottal stop

An overview of definitions is proposed by John Esling, placing the term ‘glottal stop’ in historical perspective. It seems well worth copying the entire passage in full, because it not only recapitulates a long historical line of proposed definitions, but highlights directions for fresh work taking into account the full physiological complexity of the larynx and more generally of what he names the “valves of the throat”.

In the phonetic literature, there is a long but not always consistent history of the definition of glottal stops. Holder (1669, pp. 60, 72) defined a glottal stop as “a stop made by closing the larynx”. Bell (1867, pp. 40, 60) defined a glottal stop as a glottal catch made with the glottis closed and a catch of the breath as in a cough, while specifying that the linguistic effect of a glottal stop is softer than in a cough. Sweet (1877, pp. 6–7) also called a glottal stop a glottal catch and defined it as a sudden opening or closing of the glottis. He also claimed that the most familiar example of a glottal catch is in an ordinary cough. Noël-Armfield (1931, p. 107), Heffner (1950, p. 125), and Jones (1956, p. 19) defined a glottal stop as a closure and opening of the glottis. Jones implied that the glottis must be tightly closed. P. Ladefoged (1975) and P. Ladefoged (1981, p. 50) states that a glottal stop is made by holding the vocal cords tightly together and also suggests

that glottal stops occur in coughs. Laver (1994, pp. 187–188, 206) defines a glottal stop as a maintained complete glottal closure.

In most modern phonetic literature, therefore, a glottal stop is defined simply as a closed glottis or tightly closed glottis without any reference to the ventricular folds, arytenoid cartilages, or supraglottal cavity activities. (Esling, Fraser, and Jimmy G Harris, 2005, pp. 385–386)

The two terms *glottal stop* and *glottal constriction* are considered to refer to two different phonetic phenomena by Sprigg (1966, p. 5), but according to Esling these belong to a continuum: “...glottal stops have been observed to occur on a continuum from a weakly constricted glottal stop to a strongly constricted glottal stop (...)” (Esling, Fraser, and Jimmy G Harris, 2005, p. 386).

A typical configuration observed in the production of a glottal stop “includes an adduction of the arytenoid cartilages, a complete adduction of the vocal folds, a partial adduction of the ventricular folds, and moderate narrowing of the laryngeal vestibule through its epilaryngeal sphincter mechanism (Esling, 1996, pp. 72–73, Esling, 1999, pp. 358–369; Jimmy G Harris, 1999; Jimmy G. Harris, 2001)” (Esling, Fraser, and Jimmy G Harris, 2005, p. 386).

The involvement of the ventricular folds is also noted in creaky voice by Esling, Moisik, et al. (2019).

Our research has confirmed that the ventricular folds can couple with the vocal folds during creaky phonation (Moisik, Esling, et al., 2015). This vocal-ventricular fold coupling has four effects: (1) it increases the overall effective vibrating mass, lowering frequency; (2) it adds damping, making vibration more likely to cease; (3) it adds increased degrees of freedom to the system, encouraging irregular vibration; and most importantly, (4) it perturbs the transmission of the mucosal wave by preventing its traversal across the surface of the vocal folds, causing the wave energy to be reflected back towards the midline earlier than that would occur for modal phonation (Moisik and Esling, 2014).

(Esling, Moisik, et al., 2019, p. 63)

The above review highlights the very high level of complexity of the “valves of the throat”, and thereby sheds light on why there is as yet no consensus on terminology. Similar gestures can have very different effects depending on their duration, their amplitude, and the overall configurations within which they take place. When examining electroglottographic signals, we should always remember that they constitute a projection onto a *linear* dimension of phenomena that are *non-linear*, and physiological interpretation needs to be put forward very carefully, as a matter of hypotheses, rather than reading gestures off the electroglottographic signal.

### 2.3.4 *Glottalization as a component of lexical tone: an enduring challenge for phonology*

The contribution that creaky voice makes to the tone system is apparent from the results set out in the previous chapter. Clearly, *tone* in Muong (as in Vietnamese) is to be understood in a broad sense that encompasses phonation-type characteristics associated to the phonologically distinctive suprasegmental units called tones. From the point of view of theoretical phonology, proposing models of phonation registers is no easy task.

...[P]honation register can usually be best viewed as a “package” comprising a variety of phonatory, pitch, and other properties, and it may sometimes be difficult to determine which of these, if any, is the most basic in a linguistic or perceptual sense. (Clements, Michaud, and Patin, 2011).

Elaborating models for phonological units that are conventionally referred to as tones, but which incorporate phonation-type characteristics, such as Muong **Tone 4**, is part of the same challenge. There have been proposals to integrate phonation into a system of tone features, as a ‘phonation register’ feature. However, a review concludes that “arguments for tone features typically suffer from difficulties which make arguments for a register feature less than fully convincing” (Clements, Michaud, and Patin, 2011).

In order to be in a better position to address this topic, which stands at the heart of the present research, it appears advisable to dwell further on the phonetics of the creaky tone in Muong, before proceeding to phonological generalizations. Accordingly, the following section (§5.2) enters into some detail on creak and its subtypes, and the next (§5.3) into the topic of detecting creak in recordings. The issue of glottalization in Kim Thuong Muong and its phonological modeling can then be returned to on a phonetically well-informed basis (§5.4). All along, a tenet of the present analyses is that **Tone 4** is best studied, not as a phenomenon of its own, but as a component of the tone system.

## 2.4 *Electroglottography: principles and analysis methods*

Since electroglottographic data play an important role in the present study, this exploratory technique warrants detailed presentation of its underlying principles and of the ways in which electroglottographic signals can be used in research.

### 2.4.1 *Principles*

The electroglottographic signal provides an estimate of variation in the contact area between the two vocal folds. Electroglottography (often abbreviated to EGG) was invented by Fabre in the mid-20th century. The initial report about the invention Fabre (1957) was followed by further studies by the same author over the following years (Fabre, 1958; Fabre, 1959; Fabre, 1961), initiating strands of research which continue to this day.

Electroglottography is a common, widespread technique that enables the investigation of vocal-fold contact area in phonation in an easy and noninvasive way. A high frequency modulated current ( $F = 1$  MHz) is sent through the neck of the subject. Between the electrodes, electrical admittance varies with the vibratory movements of the vocal folds, increasing as the vocal folds increase in contact. (Henrich et al., 2004, p. 1321)

The EGG signal is a continuous signal, like the audio. It can therefore be stored in the same format as the audio, and displayed with the same tools.

The following paragraphs address the topic of analysis of the EGG signal.

#### 2.4.2 *Methods to analyze the electroglottographic signal*

The method chosen here for analysis of the electroglottographic signal uses its derivative signal (dEGG). Glottis-closure instants are approximated through detection of positive peaks in the first derivative of the signal, and glottis-opening instants through detection of negative peaks in-between the positive peaks.

This method is set out in full by Henrich et al. (2004). Henrich's paper goes into technical detail concerning the four main phases of a glottal cycle: (i) closing phase, (ii) closed phase, (iii) opening phase, and (iv) open phase. Increase in vocal fold contact area is reflected by the closing phase (itself followed by the closed phase) in the electroglottographic signal, and the moment of fastest increase in vocal fold contact area corresponds to the glottis-closure instant. Decrease in vocal fold contact area begins during the closed phase, and continues into the opening phase; the moment of fastest decrease in vocal fold contact area is the glottis-opening instant.

But the correspondence between these four main phases, on the one hand, and detectable events on the electroglottographic signal, on the other hand, is not easy to establish. Instead, the electroglottographic signal corresponding to one glottal cycle can be divided into two portions only, as shown on Figure 2.14. These two portions are named *closed phase* and *open phase* of the vocal-fold vibratory cycle, and defined as follows: the closed phase extends from a glottis-closure instant to the next glottis-opening instant; and the rest of the cycle is the open phase. (Greater detail is provided in Henrich et al. 2004, pp. 1321–1322, as well as Childers and Krishnamurthy 1984, Colton and Conture 1990, Orlikoff 1998.)

The derivative of the electroglottographic signal (dEGG) signal typically has a positive peak at glottis closure and a negative peak at glottis opening. Figure 2.14 illustrates visually a synchronization of EGG and dEGG signals. In the case of clear signals, one closing peak is clearly visible for each cycle, corresponding to the peak increase in vocal fold contact area and considered as the beginning of the glottal closed phase, and one (less salient) opening peak, corresponding to a peak in the decrease in vocal fold contact area and considered as the beginning of the glottal open phase. These peaks in the dEGG signal serve as the basis for estimating glottal parameters: mainly fundamental frequency and open quotient, the two parameters most used in the present

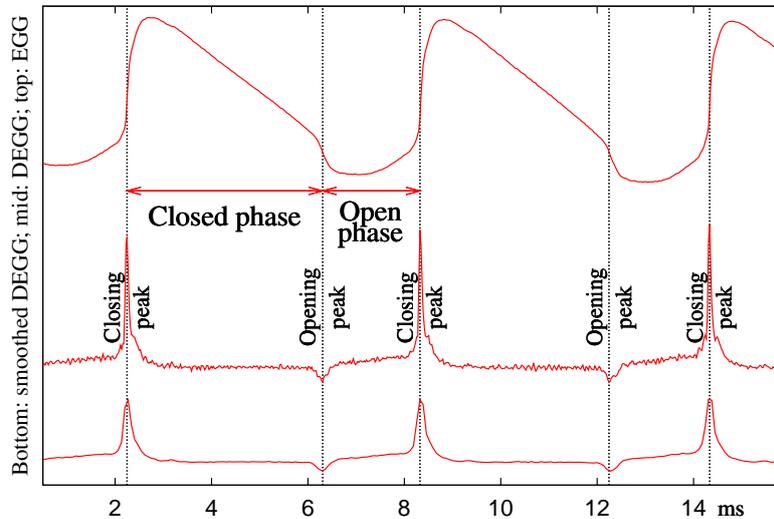


Figure 2.14: Example of EGG and dEGG signals with indication of glottis closure and opening. Reproduced with permission from the author, Alexis Michaud.

study, and also the height of the closing peak (Derivative-Electroglottographic Closure Peak Amplitude), about which more below.

Fundamental frequency ( $f_0$ ) (unit: Hz) is the inverse of the glottal period (i.e. the inverse of glottal cycle duration). Specifically,  $f_{0 \text{ dEGG}}$  is obtained by measuring the duration between two consecutive glottal closing instants, corresponding to a fundamental period. Its inverse gives the fundamental frequency of the voice (the formula is simple:  $F = 1 / T$ ). The values of  $f_0$  have *pitch* as their perceptual counterpart: low  $f_0$  is heard as low pitch, and high  $f_0$  as high pitch.

$O_q$  means glottal open quotient (unit: %). Measurement of  $O_{q \text{ dEGG}}$  requires measurement of the duration of the glottal cycle, plus detection of the glottal opening instant. This allows for computing the glottis-open interval; the open quotient is the ratio of the open-glottis interval to the entire cycle (the ratio between open time and fundamental period). This can be stated as the following equation:  $O_q = (\text{Open phase}) / (\text{Open phase} + \text{Closed phase})$ .  $O_q$  is a parameter that relates to phonation types: low  $O_q$  demonstrates pressed phonation; medium  $O_q$  reflects modal phonation; and high  $O_q$  reflects flow phonation (whispery voice, shading into breathy voice). This relates to the following observation:

There might be a continuum of phonation types, defined in terms of the aperture between the arytenoid cartilages, ranging from voiceless (furthest apart), through breathy voiced, to regular, modal voicing, and then on through creaky voice to glottal closure (closest together). This continuum is depicted schematically in Figure 2.15. (Gordon and P. Ladefoged, 2001, p. 384)

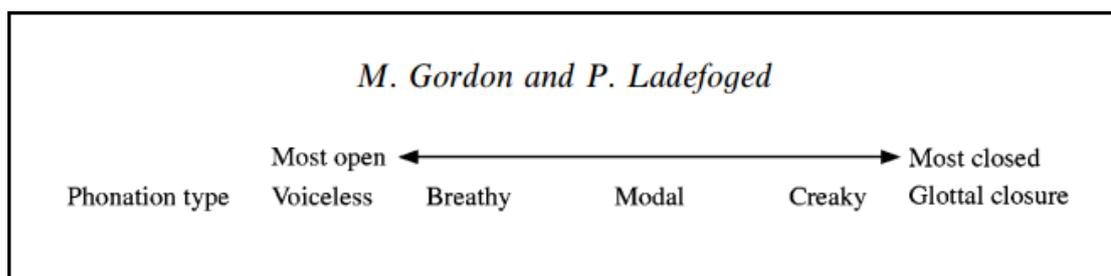


Figure 2.15: Continuum of phonation types (Reproduced from Gordon and P. Ladefoged, 2001, p. 384).

A strong motivation for adopting the dEGG method for estimating  $O_q$  in the present study of linguistic tone is comparability across studies. This method is fairly widely used in phonetic studies of phonation types published since Nathalie Henrich’s methodological article (Michaud, Tuân Vu-Ngoc, et al., 2006; Mazaudon and Michaud, 2008; Gao, 2016, e.g.), and also in various other phonetic studies (e.g. Recasens and Mira 2013). Use of similar algorithms facilitates comparison across studies, and hence across languages as well as across speakers and across datasets.

Moreover, the dEGG method is grounded in explicit assumptions that relate to physiological observations in a way which, although not simple and straightforward, is intuitively clear.

#### 2.4.3 Criticism of estimation of the glottal open quotient by electroglottography

Estimation of the glottal open quotient by electroglottography has come under criticism, which it appears useful to review here.

In a review article entitled “Electroglottography – an update”, Herbst (2020) recapitulates important caveats about the interpretation of the electroglottographic signal. Some of them are well-known: “Vocal fold vibration, a complex phenomenon taking place in three spatial dimensions, is mapped onto a single time-varying value” (Herbst, 2020, p. 4). It needs to be borne in mind that electroglottography provides a linear insight into phenomena that are not linear, and thus only offers glimpses into complex phenomena, which ideally need to be addressed through an array of exploratory techniques: a multisensor platform (Vaissière, Honda, et al., 2010).

But Herbst’s criticism cuts deeper. He questions the assumption that underpins the method employed here: that peaks on the derivative of the electroglottographic signal provide reliable estimates of the timing of glottis-closure instants. Reviewing recent studies, he considers that they “strongly suggest that positive and negative dEGG peaks do not necessarily precisely coincide with GCI and GOI, a notion that was already put forward by Childers and Lee, who maintained that the EGG signal *may not provide an exact indication for the instant of glottal closure*” (Herbst, 2020, p. 7). As emphasized by Hampala et al. (2016), “any quantitative and statistical data derived from EGG should be interpreted cautiously, allowing for potential deviations

from true VFCA [Vocal Fold Contact Area]”. But the criticism is then extended to the very notions of glottis-closure instant and glottis-opening instant: “vocal fold contacting and de-contacting (as measured by EGG) actually do not occur at infinitesimally small **instants** of time, but extend over a certain **interval**, particularly under the influence of anterior-posterior (...) and inferior-superior phase differences of vocal fold vibration” (Herbst 2020, p. 7; emphasis in original).

From a theoretical point of view, there may be a slight confusion here, as surely no one among users of the method of estimating  $f_0$  and  $O_q$  by means of peaks on the dEGG signal believes that glottal activity consists of instantaneous events of glottis closing and opening. The notions of glottis-closure instant and glottis-opening instant should, as a matter of course, be delivered complete with due precautions and careful hedging for their proper interpretation, but these precautions do not detract from the usefulness of these concepts. It should suffice to say once and for all that  $f_0$  and  $O_q$  as estimated through the dEGG method should not be confused with the physical parameter that they aim to capture. A good way to make this distinction consistently consists in embedding a reminder about the estimation method within the acronym used for the parameter. Therefore, the notations  $f_{0\text{ dEGG}}$  and  $O_{q\text{ dEGG}}$  are adopted throughout the present thesis to refer to the measured parameters, as distinct from  $f_0$  and  $O_q$ , the latter being understood either as abstract and ideal, or as generic label.

From a practical point of view, a key point here is what is meant by “precisely” when claiming that dEGG peaks do not coincide “precisely” with glottis-closure instants and glottis-opening instants. The weak claim that “perhaps the glottal area waveform, if available, would be a more suitable candidate” than the dEGG signal as a ground truth for glottal events is perfectly safe as a hypothesis, but hardly helpful for those to whom the glottal area waveform is simply not available. In practice, the difficulty of obtaining low-noise electroglottographic signals is a much more serious subject of concern to me than the fully accepted theoretical limitation whereby “the determination of contacting and de-contacting instants or events is an artificial concept” (Herbst, 2020, p. 10). The fact that glottis-opening instants as estimated from dEGG signals may be slightly earlier than those obtained by other methods does not detract from cross-token, cross-speaker and cross-language comparability, and common sense suggests that those are precious assets.

On topics of terminology, Herbst’s proposals are not particularly straightforward to implement. He uses ‘closed quotient’ ( $C_q$ ) rather than ‘open quotient’ ( $O_q$ ), which is not a real difference at all:

$$C_q = 1 - O_q$$

He argues that ‘closed quotient’ should be replaced by ‘contact quotient’:

Given that the underlying EGG signal measures relative vocal fold **contact** area and not glottal closure, the terminology for that parameter should be limited to “**contact quotient**” instead of “**closed quotient**”. Consequently, the term “open quotient” is also inappropriate, because EGG does not measure glottal opening. Instead, the term “quasi open quotient” (QOQ) might be used. (Herbst, 2020, p. 11)

I leave it to more established researchers to decide whether to take the turn towards use of the term “quasi open quotient” (QOQ). Trying to weigh the advantages, I find them very slight, compared to  $O_{q \text{ dEGG}}$ . The suggestion to prefix “quasi” to the term “open quotient” strikes me as standing in contradiction to the statement (made by the author earlier on in the same paragraph) that this parameter “is not an *ersatz* closed quotient” (Herbst, 2020, p. 11). Among prefixes, “quasi” sounds like a reasonable equivalent for description as “*ersatz*”: an inferior substitute or imitation, used to replace something that is unavailable and can only be approached, not equated.

Within studies related to electroglottography, “quasi” also brings to mind a proposal to build a “quasi-glottogram signal” from the electroglottographic signal (Kochanski and Shih, 2003). While I can make no claim to understanding the maths sustaining the attempt to build a “quasi-glottogram signal”, it is intuitively clear that the relationship between glottogram and electroglottogram in this proposal (published in the *Journal of the Acoustical Society of America*) was a much less straightforward one than that which links  $O_q$  with  $O_{q \text{ dEGG}}$ . It does not seem completely fair or productive to dismiss  $O_{q \text{ dEGG}}$  along with all other estimations of the glottal open quotient through electroglottography.

These considerations close the section about electroglottography, and all that remains in the present Background chapter concerns the Muong language and the people who speak it.

## 2.5 The Muong people and the Muong language: a general view

### 2.5.1 *The Muong as an officially recognized ethnic group*

As a brief introduction to the target language and its speakers, this section recapitulates census information based on administrative categories, and historical and geographical information about the Muong (in Vietnamese: *người Mường*).

The Muong are currently the second largest of Vietnam’s 53 officially recognized ethnic minority groups (after the Tày, speakers of a Tai-Kadai language), with an estimated population of roughly 1.4 million, based on [the 2019 census](#). Their present area of settlement spreads out over an area west, southwest, and south of the Red River (Lewis, Simons, and Fennig, 2009). They inhabit mountainous regions of northern Vietnam, with the greatest concentration in the provinces of Hoa Binh, Thanh Hoa (districts of Ngoc Lac, Thach Thanh, Cam Thuy, Ba Thuoc, Nhu Xuan, Lang Chanh), Phu Tho (districts of: Tan Son, where the fieldwork reported in this thesis was conducted; also: Thanh Sơn, Yen Lap, Ha Hoa), Son La (district of Phu Yen, Bac Yen, Moc Chau), and Nghe An. This distribution is reflected on the map of Vietic languages by Ferlus Ferlus (1998), reproduced here as Figure 2.16. In addition, Figure 2.17 by V.-T. Nguyen (2005) shows an overall picture of 91 Vietic dialects in northern Vietnam and neighboring areas, with special emphasis on Muong.

Muong communities are generally situated in low mountain valleys surrounded by peaks, i.e. in geographical zones contiguous with the Kinh (Vietnamese-speaking)

### Les langues du groupe viet-muong

1. maleng
2. arem
3. chut
4. aheu
5. hung
6. thô
7. muong, nguồn
8. vietnamien

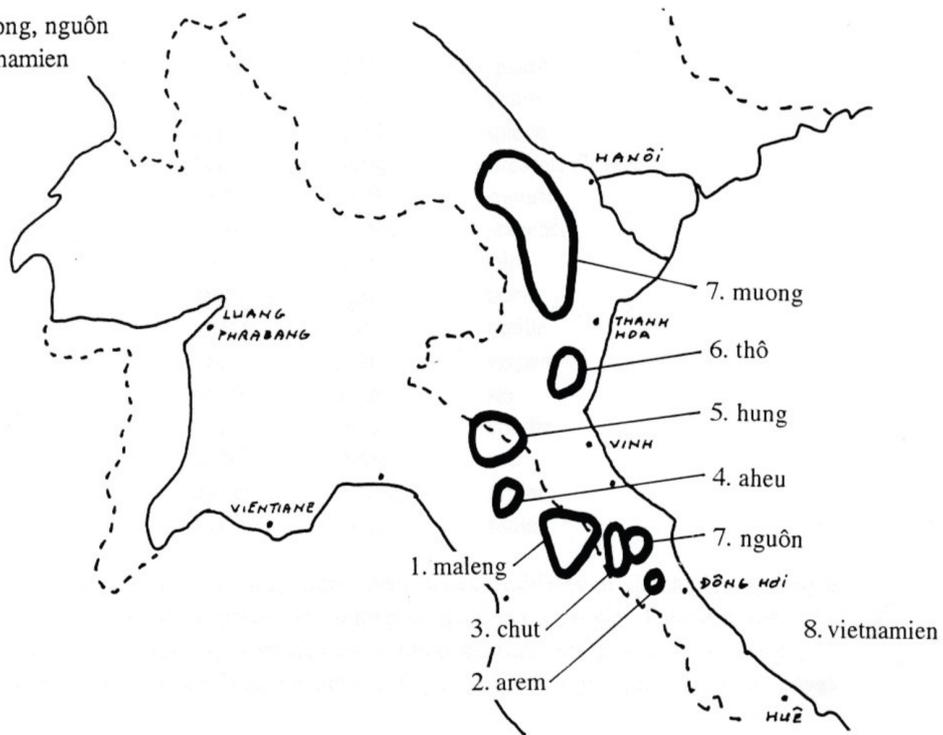


Figure 2.16: Map of Vietic languages by Ferlus (1998).

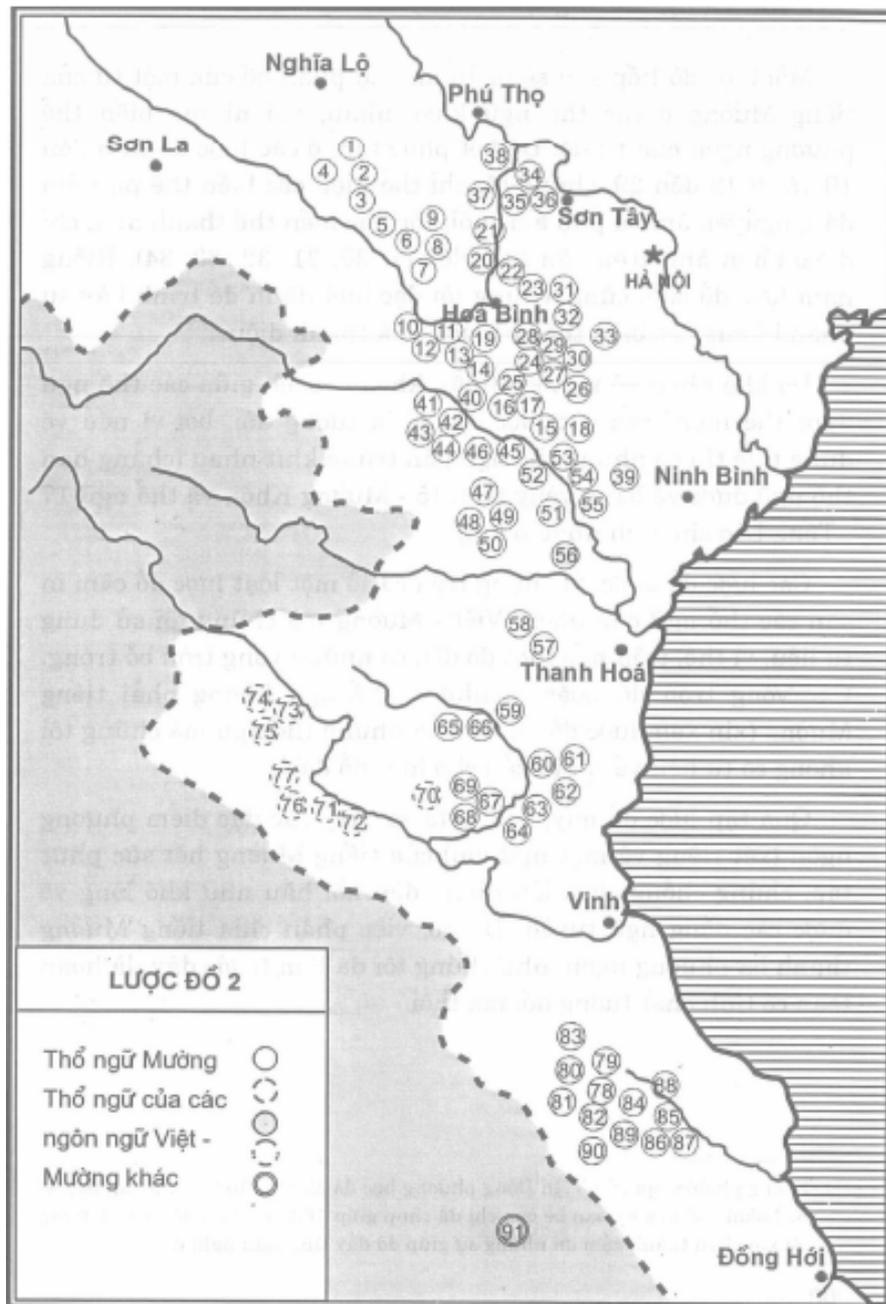


Figure 2.17: Map of Vietic dialects by V.-T. Nguyen (2005).

majority, as opposed to the higher elevations typically inhabited by the Hmong or Dao. The Muong are subsistence farmers who cultivate rice (and more recently corn) as staples, alongside a number of cash crops including tea (in Phú Thọ), sugarcane (in Thanh Hóa, Phú Thọ, and Hòa Bình), and recently, acacia lumber (in Phú Thọ and Hòa Bình) (Phan, 2012).

The name *Muong* derives from a Tai word for ‘principality’, found in place names such as Muong Thanh, considered as the first center of Tai people when they settled in present-day Vietnam in the seventh century (in the area currently named Điện Biên Phủ). According to Trần Tử (i.e. Nguyễn Đức Tử Chi) (Tử, 1988), *Muong* is a term used to describe an area of Muong residents that included many villages and that was ruled by an Assembly referred to in Vietnamese as “*Nhà Lang*” or “*Thổ Lang*”. The original report is by Jeanne Cuisinier, in *Les Muong: Géographie humaine et sociologie* (Cuisinier, 1948).<sup>6</sup> No details are provided about the historical time window when this system was in place; it may nonetheless be relevant to the linguistic history of the area that the administration tended to be entrusted by the imperial government to local rulers who apparently belonged to the local communities and can therefore be hypothesized to be speakers of the same language as the commoners.

The people officially classified as Muong call themselves /**mol**/, /**mwǎn**/ in Hòa Bình, or /**mon**/, /**mwǎl**/ in Thanh Hóa. In Phú Thọ, especially in the districts of Thanh Sơn and Tân Sơn, where the concentration of Muong people is greatest, and also in Yên Lập district and some communes of Thanh Thủy district, Muong people call themselves /**mol**/ or /**mon**/. These names are similar in terms of meaning: ‘person’; this has been analyzed as having the etymology of ‘burrower, digger’, referring to the agricultural activity of horticulturists Ferlus, 1996, p. 8. They were borrowed into Vietnamese as “Mọi” or “Mọn”, with a derogatory meaning carrying associations of backwardness. These labels have undergone a gradual demise since the introduction of “Muong” as an official name.

### 2.5.2 *The Muong language in its Vietic context*

Muong has unmistakable similarities with Vietnamese. This was pointed out by André-Georges Haudricourt in his article “The place of Vietnamese in Austroasiatic” (André-Georges Haudricourt, 1953). He discusses twelve words of basic vocabulary about names of body parts in Mon-Khmer languages, including: Vietnamese, Muong, Phong, Kuy, Mon, Bahnar, Mnong, and other languages. Among these, it can be seen from Table 2.1 that the correspondences between Vietnamese and Muong are really neat.

---

<sup>6</sup>*Nhà Lang* have been known as ruling class who have both rights and obligations to people in the place they control. In comparison with the Vietnamese (or Kinh) feudal landlord class, *Nhà Lang* can be assumed to have been more sensitive to the local realities. In Marxist perspective, *Nhà Lang* is praised as more “progressive”, and as representing not only itself as a class but also the whole Muong community. *Nhà Lang* had a responsibility to help Muong residents in special situations such as drought, crop failure and famine and the planning and realization of major work (infrastructure work). In addition, they were in charge for the higher-level ceremonies of this region (see more in Cuisinier 1948).

Table 2.1: Correspondences between Vietnamese and Muong for twelve body part names, adapted from André-Georges Haudricourt (1953).

numbering	English	Vietnamese (IPA)	Muong (IPA)
1	head	ɗ̥w <sup>A2</sup> , ʈok <sup>D1</sup>	ʈok
2	hair	tɔk <sup>D1</sup>	t <sup>h</sup> ak
3	eyes	măt <sup>D1</sup>	măt
4	ear	taj <sup>A1</sup>	t <sup>h</sup> ai
5	nose	muj <sup>B1</sup>	mui
6	mouth	miəŋ <sup>B2</sup>	mɛŋ
7	tooth	răŋ <sup>A1</sup>	<i>not mentioned</i>
8	tongue	luəj <sup>C2</sup>	lai
9	neck	ko <sup>C1</sup>	kel, kok
10	lip	moj <sup>A1</sup>	<i>not mentioned</i>
11	chin	kăm <sup>A2</sup>	kăŋ
12	arm (or hand)	tăj <sup>A1</sup>	t <sup>h</sup> ai

Muong actually provided decisive evidence for the Austroasiatic affiliation of the Vietnamese language, which had heretofore variously been considered as belonging to Tai or Chinese. There is thus no need to labor the point that studies of the Muong language have some relevance to a better understanding of the history of the Viet-Muong group: Muong dialects can yield critical evidence for the historical study of Vietnamese.

Taking a closer look at the relationship between Muong and Vietnamese – a central issue in Muong studies –, various views have been expressed. Since the Vietnamese (Kinh) and the Muong are officially two distinct ethnic groups, each tends to be considered as having its own language, hence a perception that there exists one Muong language and one Vietnamese language, each with various dialects and sub-dialects. The linguist's perspective, on the other hand, consists in taking stock of the Viet-Muong language sub-group's internal diversity, and exploring the diachronic processes of evolution that shaped this diversity, without assuming a neat two-way divide.

Within the broader Vietic sub-group, Viet-Muong in the narrow sense, also known as Northern Vietic (including Vietnamese, Muong and Nguon) is characterized by irregular tonal reflexes in some words of basic vocabulary: the loss of proto-Austroasiatic initial voicing in Northern Vietic languages resulted in high-register reflexes, as against the expected low-register reflexes in Southern Vietic. This is interpreted by Ferlus (1999) as a substratum effect dating back to the time when proto-Vietic spread northwards onto an Austroasiatic substratum of languages that lacked voiced stops (a set of languages of which the Khmuic language Ksing Mul arguably constitutes a remnant). At that time, Vietnamese, Muong and Nguon dialects were not yet distinct. The historical scenario is that their common ancestor (proto-Northern Vietic) then spread over large areas of the plains of present-day Northern Vietnam. Overall, the varieties spoken in the area of present-day Hanoi underwent more rapid evolutions, due to a variety of factors that

doubtlessly included contact between various dialects and languages. Among these, one variety became dominant, and gained the status of a regional, then a national standard: the Vietnamese language. By contrast with speakers of Northern Vietic in peripheral areas, those in the area of Thăng Long (present-day Hanoi) tended to be identified by their status as townspeople, hence the name *Kinh* (the Sino-Vietnamese word for ‘capital city’, 京: the same that constitutes the second syllable in ‘Beijing’ 北京; this is also the second character found in *Tokyo*, whose name in Chinese characters is identical to one of Hanoi’s former names: *Đông Kinh* 東京). This process is still active today: Frédéric Pain (p.c. 2014) observed that some proud city dwellers prefer to be identified as ‘Hanoians’ (*người Hà Nội*) in preference to the label ‘Kinh’, which is now used over the entire country as the label for the ‘majority’ ethnic group, and is hence less prestigious, having lost its specific association to citizenship of the capital.

Detailed ethnolinguistic study of the sense of ethnic belonging of the various communities of speakers of Northern Vietic falls outside the scope of the present study; the above remarks simply aimed to convey a feel for the gradual individuation of the Vietnamese language within Viet-Muong, a process which sheds light on the special synchronic closeness between Muong and Vietnamese.

The phonological evolution of the peripheral dialects now identified as “Muong” was less dramatic than that of Vietnamese, resulting in a more conservative phonological system, closer in some respects to proto-Northern Vietic. One of the telltale evolutions that single out Vietnamese within the Vietic sub-group is the spirantization of medial obstruents, which results in phonological correspondences such as Muong /ka/ :: Viet /ya/ ‘chicken’. In Muong, the proto-Vietic presyllable was lost without compensation; in Vietnamese, it caused the spirantization of /\*ʔ/ (Ferlus, 1982).

### 2.5.3 *Literature review of Muong studies*

A full review of linguistic studies of Muong will not be attempted here. Let us simply mention that Milton and Muriel Barker collected abundant lexical data, made available as a *Muong-Vietnamese-English Dictionary*: an extensive wordlist with Vietnamese and English glosses (M. E. Barker and Muriel A. Barker, 1976). In addition, Miriam Barker published a bibliography of Muong and other Vietic languages in 1993, with over 40 pages of references not only in the field of linguistic studies, but also in fields such as literature, geography and history (Miriam A. Barker, 1993). This was a key reference in gathering documentation for the present work.

Later, V.-T. Nguyen (2005) collected data from an impressive range of 30 dialects. This is a remarkable resource to get a feel for the diversity of dialects; on the other hand, the book suffers from fairly unreliable tonal transcriptions (Ferlus p.c. 2015, confirmed by my own observations). It is a general observation that native speakers of Vietnamese can find it difficult to cast off the perceptual filter of their mother tongue and attune their ear to different tone systems.

Within the limited scope of the present study, other references about the Muong language have not yet been taken into account.

## 2.6 The dialect under study: Kim Thuong Muong

The variety of Muong under study here is one of the many Muong dialects found in Phú Thọ province. The exact location of fieldwork is Kim Thuong commune, Tan Son district,<sup>7</sup> hence the label used here for this dialect: Kim Thuong Muong. Over 85% of the population in this district belongs to ethnic minorities. Among them, Muong make up more than 75%, and Dao account for about 6,4%, the remaining belongs to Hmông and Tày. Locally, the nation's "majority" ethnic group, the Kinh, are thus among the smallest minorities in terms of population. The Dao and Muong each form a community with festivals and habits of its own; members of other groups, such as Kinh, Tày and Hmông people, are new settlers or married into one of the two local communities.

This is a mountainous region that is difficult to access: a relatively locked area surrounded by mountains and rivers. This can be predicted to be conducive to the preservation of linguistic diversity. To the northwest is Xuân Sơn National Park; to the southwest is Đà River, which separates this area from Hoà Bình province, which contains areas of denser Muong population (see map: Figure 2.18). The East and the South border on other Muong dialects of Tan Son district. This configuration goes a long way towards explaining why Kim Thuong Muong is still spoken today, and has not yet been entirely replaced by Vietnamese. All the local people I encountered proudly referred to themselves as /**mɯəŋ**<sup>2</sup> **tǎn**<sup>2</sup>/ (in Vietnamese orthography, this can be rendered as *Mường Tản*), meaning 'the original Muong' (from my field notes, it would seem that /**tǎn**<sup>2</sup>/ means 'original, pristine').

To this day, Muong is essentially an unwritten language. Some romanized writing systems were devised, and used in sizeable publications of collections of oral literature (e.g. Bui 2010) but have not been widely adopted by the general public. Schooling is exclusively in Vietnamese, and is compulsory from age 6 to age 17.

While the target dialect of the present study (Kim Thuong Muong) has never studied before in its own right, it is geographically close to areas whose dialects have been studied, such as Kha Cuu Muong studied by Phan (2012), and Giap Lai Muong, one of the 30 dialects studied in Van-Tai Nguyen's book (V.-T. Nguyen, 2005). In regard to the current sociolinguistic situation, the Muong language in general and Kim Thuong Muong in particular are still vital and in everyday use. The risk of weakening and loss is not really obvious. However, it does not mean that the study of Muong is not urgent. From my point of view, research on Muong Kim Thuong at this time is reasonable because we can avail ourselves of the many advantages of studying a language that is still widely used: (i) ease for gathering information and data for study since Muong is still a living language; (ii) ease for verification Muong language's characteristics since the contact between it and other languages, especially Vietnamese, is not too complicated. Moreover, facing the powerful influence of Vietnamese, Muong language is clearly in danger of being weakened, and disappearing in the mid run.

Because of its relationship to Vietnamese and to the other languages of the Vietic

---

<sup>7</sup>Tân Sơn is a new district which was separated from Thanh Sơn district and established under the resolution 61/2007/NĐ-CP of the government from 09/04/2007.

2.6 The dialect under study: Kim Thuong Muong

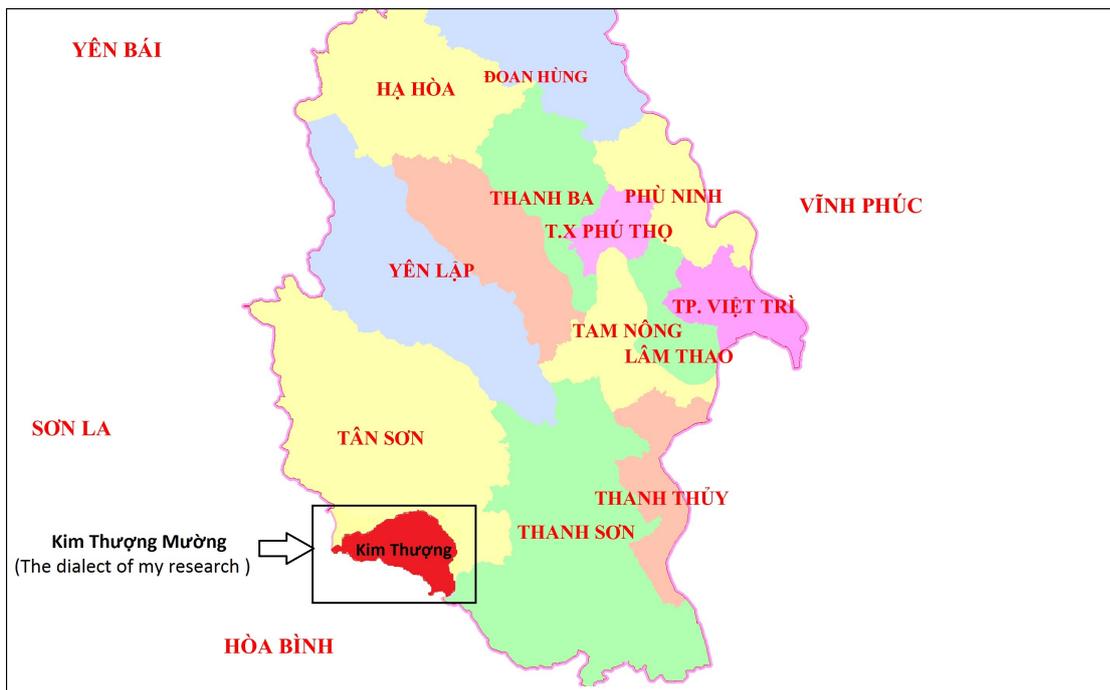


Figure 2.18: Map of Kim Thuong Muong location in Phu Tho province, the dialect in red (<http://bando.tnmtphutho.gov.vn/map.phtml>).

group (such as: Chút, Cuối, and Thà Vững), the growth of Muong studies since the early twentieth century not only helps us to have a better understanding of this individual language, but also is useful for comparative study and diachronic study.

Clearly, the amount of research on Muong is much more limited than on Vietnamese, in terms of both quantity and quality. Given the overall context of Muong studies, my dissertation was done with the desire to make a contribution (however small) towards a better understanding of this language, starting with a fully worked out phonemic inventory.

**Some notes about the social dynamics of Kim Thuong Muong.** Kim Thuong Muong is a non-written language variety inside a dialect continuum characterized by in-depth influence from the national standard – Hanoi Vietnamese. The Muong language remains locally vigorous; at the same time, Hanoi Vietnamese is clearly a prestige language, whose acquisition is key to success in school – opening prospects for a brighter future and better work opportunities outside Kim Thuong.

Kim Thuong Muong is defined as the dialect of the (few) villages of Kim Thuong commune. In terms of social status, this is the strongest group in this area with a population that makes up more than 75% of the total. The rest is divided among four groups: Dao, H'mông, Tày, Kinh, which differ sharply from one another. Locally, the nation's "majority" ethnic group, the Kinh, are thus among the smallest minorities in terms of population. The Dao and Muong each form a community with festivals and habits of its own; members of other groups, such as Kinh, Tày and Hmông people, are new settlers or married into one of the two local communities. Its peripheral location in geographic terms would seem to make Kim Thuong Muong more likely to endure than other varieties of Muong. Contact between the Muong and Dao is not strong: Dao group lives separately in areas that have more difficult terrain. Local people refer to the respective habitats of the Dao and Muong as "khu trên" and "khu dưới" (meaning upper and lower zone, respectively). This can be seen as the acknowledgment of a difference between the two groups' current settlement: the lowlands to the Muong, the highlands to the Dao. It could be that this description also reflects that the two should remain distinct: affirming their distinctness. Due to their current living habits and standards, the Dao people are not engaged in particularly active trade exchanges, nor are they greatly integrated to the network of health and education services. Therefore, despite geographic proximity, contact between the Dao and Muong in Kim Thuong is relatively loose. As to ethnic H'mông and Tày, they live interspersed within the Muong community but their number is very small and studying their interaction with their surroundings, and their potential influence on Kim Thuong Muong, would require specific case studies (a monograph about a household, for instance) which I can hardly hope to conduct, for want of having any knowledge of their mother language.

As for 'Kinh' newcomers to the area, their influence is linked to their status as representatives of the ethnic majority of Vietnam. From my field observation, local people in Kim Thuong have a positive attitude to their mother language. All the local people that I have met use their mother language for day-to-day communication. Individuals who belong to a different ethnic group but have married into the Kim

Thuong Muong community also tend to fall in line with local people by learning and using this language, yielding to a demand that seems overwhelming. Therefore, the Kim Thuong Muong dialect keeps the role of the strongest language in this community to this day. On the other hand, it seems clear that Kim Thuong Muong is influenced by the “Kinh” language (Hanoi Vietnamese) much more than by any other. There is strong contact in many aspects. Firstly, the development of commercial relations created an opportunity for Muong people in Kim Thuong to use Vietnamese to communicate and trade. Vietnamese now becomes the main language in the market, standardly used between people unacquainted with each other. Secondly, the role of Vietnamese is becoming increasingly important in the Kim Thuong Muong community, as the language of education, health services, and administrative affairs. Nowadays, children are educated in Vietnamese from the beginning of their schooling. Vietnamese is also the language used for television, radio, and other media mass. Because Kim Thuong Muong in particular and Muong language in general have no writing, they have to use Vietnamese for administrative documents to address the affairs of their community. Last but not least, local people recently tend to go out of their area, mainly going into “Kinh” communities, to seek employment when farm work ends. Those reasons explain why almost 100% Muong people in Kim Thuong are able to use Vietnamese (to different extents). The stage is thus set for strong contact between Kim Thuong Muong dialect and Vietnamese, leading to language variation and change in Kim Thuong Muong.

The predicament encountered when running an experiment using nonsense words (as part of my Master’s study: M.-C. Nguyễn 2016) is a case in point: the consultants tend to identify Muong tones with their phonetically closest equivalent in Standard Vietnamese, resulting in confusion of Tone 1 and Tone 2. The consultants’ great difficulties when required to steer clear of Standard Vietnamese tones and focus solely on the pronunciation of their mother tongue shows that they have much greater metalinguistic awareness about their second language than about their mother tongue. For them, metalinguistic notions of ‘tone’ are so strongly linked to Vietnamese (the language which they learned to analyze into phonemes in the process of learning Vietnamese writing) that special training is required for them to develop an awareness of their mother tongue’s categories. This could seem paradoxical, but is probably widespread in communities whose mother tongue is an unwritten language in a context of unequal bilingualism.

Although sociolinguistics and language contact were not intended as the main topic of the present research, it is clear that these are simply unavoidable. Any study of Kim Thuong Muong in particular and Vietic languages in general needs to include special precautions and verifications concerning possible interferences from Vietnamese. Since there is no way to deny or avoid this complexity, the way forward seems to be to embrace the complexity and give it a central place in the research. As a speaker of Vietnamese, and a ‘Kinh’ newcomer to the community (no matter how much effort is spent making my presence inconspicuous and unobtrusive), I am fortunate to be a witness to cases of code-switching, blending and other phenomena, which I intend to document through field notes, with the possible mid- and long-term prospect of a

sociolinguistic survey using the methods gradually developed in sociolinguistics from Weinreich (1957), Weinreich, Labov, and Herzog (1968), and Weinreich (2011) to Labov (1994), Labov (2001), and Labov (2010).

## 2.7 Phonemic inventory of Kim Thuong Muong

The present section was written in the spirit of “Illustrations of the IPA”: phoneme inventories accompanied by some observations about the phonological system.

In order to explore and propose a phonemic inventory of a language/dialect, a prerequisite is that the researcher knows [the International Phonetic Alphabet \(IPA\)](#)<sup>8</sup>. In order to phonemicize the data, one needs to conduct a distributional analysis along the basic principles of phonemic theory (Chao, 1934; Pike, 1947; André Martinet, 1956).

The inventories below are synthesized by vocabularies collected from Michel Ferlus’ word list, an expanded version of the EFEO-CNRS-SOAS word list for linguistic fieldwork in Southeast Asia, available online : <https://hal.archives-ouvertes.fr/halshs-01068533/>

Concerning the practical methodology in the field, the ideal way to collect the vocabularies is to have an assistant with you during the recording session who can read the words so that the speaker can pronounce the corresponding words in his/her language. At the same time, the researcher can transcribe them directly to a computer using the word list. Of course, this is only convenient in case there is an IPA keyboard available on your computer<sup>9</sup>. Otherwise, we can also print out the word list and transcribe them manually and later type them into the Excel file. This admittedly takes twice as long, but in my experience it allows for better contact with the consultants, focusing on a meaningful exchange. Data inputting can cause an interruption of sorts within the recording session: using a computer almost amounts to having a silent ‘third party’ present, out of touch with the consultant (unlike a research assistant: a human ‘third party’).

Kim Thuong Muong is the first language that I really described in the field. Beginners’ work tends to be slow: Michel Ferlus (personal communication 2015) reports that the first language for which he carried out the phonemic analysis task, namely Khmu, took him a full six month to work out. He reflected that a couple of decades later, rich with experience from working on dozens of languages, he could have done it in the space of three weeks. But as far as vowels and consonants are concerned, I did not have much difficulty in accessing the dialect being studied. For one thing, Muong is not a

---

<sup>8</sup>Beginners can learn about the IPA by using the following two related web pages: (i) Interactive IPA Chart (<https://www.ipachart.com/>), where one can hear recordings of the sounds that the symbols represent, (ii) the IPA charts: sample translation batch ([https://linguistics.ucla.edu/people/keating/IPA/IPA\\_charts\\_2019\\_trans.html](https://linguistics.ucla.edu/people/keating/IPA/IPA_charts_2019_trans.html)), a project to collect and post IPA charts whose metatext is in languages other than English – now including Vietnamese. There is also a free [iPA Phonetics](#) App for iPhone, iPod, and iPad from the Apple App store. This App provides an intuitive touch interface for exploring the International Phonetic Alphabet as well as numerous voice qualities and articulations.

<sup>9</sup>For beginner-level information on creating and customizing an IPA keyboard on Microsoft Windows, a tutorial is available here: <https://msklc-guide.github.io/>

difficult language in terms of its sound system (as noted by Michel Ferlus: again, this is a personal communication from 2015). Moreover, my mother tongue is Vietnamese, which is the language most closely related to Muong. Knowledge of the IPA and training in transcribing the Vietnamese in IPA (Kirby, 2011) are basic requirements in undergraduate linguistics courses, along with knowledge of the historical phonology of the Viet-Muong (Vietic) branch of Austroasiatic languages. In addition, there is a broad array of references about various other dialects of Muong, as well as on other languages of the Vietic group. Special mention needs to be made of the work of V.-T. Nguyen (2005) on the phonetics of no less than thirty Muong dialects, and of the successive publications by Michel Ferlus on several dialects of Vietic languages such as Lang Lo Tho (Ferlus, 2001), Arem (Ferlus, 2013), and heterodox dialects of Vietnamese such as Cao Lao Ha (Ferlus, 1995) and Phong Nha (a study in which I was fortunate to participate: (Michaud, Ferlus, and M.-C. Nguyễn, 2015)). Familiarity with these sources makes it possible to anticipate a set of expected phonemes, as well as phonotactic rules (concerning expected combinations of phonemes). Thus, I was able to notice in Kim Thuong Muong the contrastive palatal finals /c/, /ɲ/ because I was aware that these sounds, even though they do not exist anymore in Vietnamese as separate phonemes, used to be present at earlier stages of Vietic and are preserved in several varieties of Muong.

Thus, the phonemic system was collected, synthesized and systematized as follows.

### 2.7.1 Consonants

#### Initial consonants

An inventory of initial consonants is shown in Table 2.2. Among the phonemes in Table 2.2, /w/ is as a newcomer to the system, appearing on loanwords from Vietnamese, which have initial /v/ in Vietnamese.

Table 2.2: Inventory of initial consonants

	Labial	Dental	Alveolar	Retroflex	Palatal	Velar	Glottal
Plosive	p p <sup>h</sup>	t t <sup>h</sup>		ʈ		k k <sup>h</sup>	ʔ
Nasal	m		n		ɲ	ŋ	
Trill			r				
Fricative	β		s			ɣ	h
Approximant	w				j		
Lateral							
Approximant			l				
Implosive	ɓ		ɗ				

Table 2.3: Final consonants inventory

p	t	c	k
m	n	ɲ	ŋ
w		j	

### Final consonants

This system resemble the inventory of final consonants which is proposed by Ferlus for Làng Lỗ Thổ (Ferlus, 2001), a dialect considered as part of “Central Muong”<sup>10</sup> (Maspero, 1912). The two palatal consonants /c/ and /ɲ/ in Làng Lỗ Thổ, according to Michel Ferlus, “are found (i) in borrowings of Vietnamese words with final /c ɲ/ (orthographic *ch* and *nh*), which originate in velars following high front vowels, and (ii) in a few inherited words, such as /kăc<sup>24</sup>/ (Viet. *cắt* [kătD1]) ‘to cut’ and /sɲ<sup>24</sup>/ (Viet. *rắn* [rănB1]) ‘snake’”. The origin of **c** and **ɲ** in Kim Thuong Muong seems to be the same; interestingly, the list of inherited words is longer: see Table 2.4. In addition, palatals originating in the fronting of velars are not restricted to the context of a following high front vowel (a structure illustrated by /kɲ<sup>1</sup>/ ‘Vietnamese’) but also appear after mid-low front vowels, e.g. in /seɲ<sup>1</sup>/ ‘green’. These observations are observed with a low degree of certainty, as I am still a beginner in Vietic historical phonology; for want of any certainty I have not systematically indicated in Table 2.4 if a word is inherited or borrowed.

### 2.7.2 Vowels

Basically, the vowel system of Kim Thuong Muong is close to Vietnamese<sup>11</sup> with nine monophthong /i e ɛ a u ɤ o ɔ/, three falling diphthongs /iə uə uə/, and two pairs of long and short vowels /ɤ-ɤ̃/ /a-ă/. As one goes into sub-phonemic detail, some interesting differences appear. For instance, Kim Thuong Muong has a falling diphthong [iɛ], in complementary distribution with [ɛ]. The former appears before all final consonants other than the two palatals, /ɲ/ and /c/. Examples include: /teɲ<sup>4</sup>/ ‘to fight’, /tec<sup>7</sup>/ ‘to put’, as opposed to /riep<sup>7</sup>/ ‘sandals’, /k<sup>h</sup>iet<sup>7</sup>/ ‘lightning’, /kiek<sup>7</sup>/ ‘armpit’, /kiem<sup>1</sup>/ ‘ice-cream’, /tien<sup>4</sup>/ ‘narrow’, and /tiej<sup>2</sup>/ ‘nail’.

### 2.7.3 Tones

For the study of tones, prior knowledge of closely related languages may paradoxically bring disadvantages. Speakers of Vietnamese as a first language realize, to their dismay and sometimes to their utter distress, that re-training their ear to a new tonal language, as is necessary in fieldwork (Rice, 2014), is especially difficult when one has Vietnamese tones in one’s mind since infancy. A piece of anecdotal evidence on this topic is

<sup>10</sup>Henri Maspero used the term “Muong” to refer to all languages of the Vietic group except “Annamese” (Vietnamese)

<sup>11</sup>Unless specified otherwise, “Vietnamese” refers to the Northern dialect, in its “standard” Hanoian variety.

Table 2.4: Palatal final consonants **c** and **ɲ** in inherited vocabulary and in loan words

Num.	Kim Thuong Muong	Vietnamese	English	Note
1	/mɛɲ <sup>1</sup> /	nhanh, mau	quick	
2	/tɛɲ <sup>1</sup> /	đan	to weave	
3	/nwɛɲ <sup>2</sup> /	nhỏ, bé	small	
4	/kɪɲ <sup>2</sup> /	gần (nơi)	near to	
5	/kɛɲ <sup>4</sup> /	cắn	to bite	
6	/rɛɲ <sup>4</sup> /	cái đó, lờ	fish basket	
7	/t <sup>h</sup> ɛɲ <sup>4</sup> /	rắn	snake (general)	
8	/nɻɲ <sup>4</sup> /	nặn, vắt	to squeeze	
9	/kac <sup>6</sup> /	cắt, chặt	to cut	
10	/k <sup>h</sup> ac <sup>6</sup> /	sắt	iron	
11	/kac <sup>7</sup> /	cát	sand	
12	/ke <sup>4</sup> lac <sup>7</sup> /	dây lạt	bamboo strips	
13	/tɛc <sup>7</sup> yuəŋ <sup>4</sup> /	đặt, để xuống	put	
14	/mic <sup>7</sup> /	mật	honey	
1	/sɛɲ <sup>1</sup> /	xanh	green	
2	/kɪɲ <sup>1</sup> /	Kinh	Vietnamese	synonym: /taw <sup>4</sup> /
3	/mɛɲ <sup>3</sup> /	manh	strong	
4	/tɛɲ <sup>4</sup> /	đánh (nhau)	to fight	

Table 2.5: Vowel inventory

i	u	u
e	ɣ/ɣ̃	o
ɛ	a/ă	ɔ
iɛ		
iə	uə	uə

that several Vietnamese linguists who heard about initial successes of Automatic Speech Recognition applied to language documentation (Adams, Cohn, et al., 2017) immediately asked whether the software could identify tones for them, solving what they felt to be an impossible task for their ear. Whether the language under study is from the Austroasiatic family (to which Vietnamese belongs) or to Tai-Kadai, Hmong-Mien or Sino-Tibetan, making out the tone system is perceived as excruciatingly difficult and uncertain by scholars with Vietnamese as their mother tongue (Alexis Michaud, personal communication 2018). To what extent this impression matches real performance remains to be investigated. It could be the case that adjusting one's ear to a new tone system is a challenge for anyone, and that some people (for reasons not directly related to the languages they speak) are less attracted to this challenge than others. There nonetheless seem to be some converging indications to the effect that speakers of a phonetically complex tone system such as Vietnamese experience special difficulty in casting aside the phonological filter of their native language and adopting that of another tonal language.

Be the statistic as it may, I for one can certainly report experiencing this difficulty. When I discovered the Kim Thuong Muong tone system, carry-over of tone identification patterns from Vietnamese was a major hurdle, and the tendency to fall back on Vietnamese categories caused me to make some worrying mistakes in perceiving and classifying Muong tones. I found myself constantly trying to figure out which tone in Kim Thuong Muong was similar to which tone in Vietnamese. As a consequence, I initially missed the (contrastive) distinction between the falling tone (Tone 2) and the low level tone (Tone 1). I perceived both as the same tone: a tone characterized by a falling modulation. I failed to notice the differences because I based myself on a comparison with the Vietnamese A2 tone (called *thanh huyền* in Vietnamese orthography). Therefore, back in 2015 I initially came up with an analysis of the tonal system of Kim Thuong Muong with only 4 tones on smooth syllables. This mistake was detected only when my attention was brought to the minimal pair /ka1/ “chicken” and /ka2/ “big”. It took me quite a while to progress from initial recognition of this distinction (thanks to the patient help of language consultants) to consistent and confident use.

If anything could make me believe in *hard-wired* linguistic competence, then that is Vietnamese tones, which I still feel ‘hard-wired’ to perceive, as opposed to the painstaking adoption of the tonological filter of Kim Thuong Muong. It felt like a process of unlearning one's prior tone system to learn the new one – even though that is not literally the case, since I have not un-learned Vietnamese tones. Maybe the challenge can be compared to that of speakers of Dutch learning German and vice versa: there is special difficulty in keeping separate two linguistic systems that have so many points of similarity, tantalizing close to identity. Studies of language acquisition may shed light on my feeling that the Vietnamese tone system plants deep roots in native speakers.

This learning experience sheds light on possible processes of language change in cases of language replacement: it could well be the case that a group of Vietnamese learners

who switched to Kim Thuong Muong (for reasons which in the present sociolinguistic scene would be truly implausible) would end up speaking a four-tone language, collapsing the falling tone (Tone 2) and the low level tone (Tone 1) together.

### Diachronic background

Tonogenesis in Vietnamese is among the success stories of 20th-century historical linguistics, and specifically, of Panchronic Phonology: an approach to diachronic phonology that searches for regularities independent of a specific language group and a specific point in time, i.e. true laws of sound change. The seminal findings of André-Georges Haudricourt (1954a) and André-Georges Haudricourt (1954b) are now part of the common background knowledge of specialists of Asian languages (for a summary, see e.g. Michaud 2012). A study of the historical development of tone in all the Vietic languages for which documentation was available at the time (1998) is proposed by Ferlus (1998). This constitutes fundamental background for the present study: the system for etymological notation of tones used here (from A<sub>1</sub> to C<sub>2</sub> for smooth syllables, those without final stops, and D<sub>1</sub> and D<sub>2</sub> for stopped syllables – also called checked syllables –, those with final stops) is taken up from the usual conventions set out in these diachronic studies, to which the reader is referred for further details.

### The tone categories

The nature of tone itself raises issues in choosing notations for some tones of Southeast Asian languages. It seems that the tones are defined through the set of contrasts into which they enter with one another, rather than by a neat phonological property (such as the H, M and L tones of the Naxi language, for instance: Michaud, Vaissière, and M.-C. Nguyễn 2015). Transcribing this type of tone in phonetic notation remains an unsolved issue. The 1949 edition of the *Principles of the International Phonetic Association* mentioned the case of Vietnamese (referred to as ‘Annamese’): “ $\overset{\sim}$  and  $\overset{\vee}$  (preceding the syllable) have been suggested for the two rising tones of Annamese,  $\overset{\sim}$  implying the use of breathy voice and  $\overset{\vee}$  for creaky voice” (p. 18). The intended tones are presumably B<sub>1</sub> and C<sub>2</sub>. To my knowledge, these symbols (the macron [ $\overset{\sim}$ ] and the down arrowhead [ $\overset{\vee}$ ]) have not been taken up in later studies. An issue is that use of the macron [ $\overset{\sim}$ ] for one of the tones threatens to conflict with its use to indicate vowel shortness: in Vietnamese orthographic representation,  $\overset{\sim}{a}$  is a short vowel, contrasting with  $a$ . As orthographic representation happens to coincide with IPA transcription for these two vowels, using the macron [ $\overset{\sim}$ ] to indicate tone would contradict a well-established convention.

One possibility is to assign numbers to tones, without committing oneself to an analysis. This is the choice made for Mandarin Chinese, whose four tones are referred to as ‘first tone’, ‘second tone’, ‘third tone’ and ‘fourth tone’ in the linguistic literature – a practice that speaks volumes about linguists’ puzzle: the use of numbers amounts to giving up the analysis, and recognizing that there is no easy way to pinpoint the exact nature of each of the tones. Use of numbers for tones is also widespread in the

field of Thai (Siamese) studies: numbering the tones of Standard Thai from one to five allows for a straightforward identification of the synchronic categories. Interestingly, the earlier categorization of Thai tones, dating back to a time when there were only three contrastive tones, was also by means of numbers: *Mai Ek* เสียงเอก means ‘tone one’ and *Mai Tho* เสียงโท means ‘tone two’. These contrasted with a third category, left unmarked: เสียงสามัญ. This system is adopted here: classical distributional analysis conducted during fieldwork brought out five distinctive tones for smooth syllables and two tones for stopped syllables, and these tones are numbered from 1 to 7 in Table 2.6.

Another possibility is to adopt Chao Yuen-ren’s system of five levels (Chao, 1930), to propose a stylized representation of the perceived pitch along a scale from 1 (lowest) to 5 (highest). A mid-rising tone may thus be represented as 24, 25, 34, 35, or 45. Parameters guiding the choice of one level along the five-point scale include phonetic precision and phonemic economy, which may occasionally be at odds with each other. In particular, it is recommended to use no more numbers than is phonemically necessary, i.e. use the numbers sparingly: for instance, if only three levels are necessary for the target language, these should be transcribed as 1, 3 and 5, and use of 2 and 4 would be avoided. It is for each researcher to strike a balance between economy and precision. A limitation of this system is that it does not transcribe phonation types. In Hanoi Vietnamese, laryngealization plays an important role for two of the tones, namely B<sub>2</sub> and C<sub>2</sub>; moreover, phonation-type characteristics may be found on other tones, to a greater or lesser extent (see Brunelle 2009b). Keeping this limitation in mind, Chao’s system is nonetheless useful, and is a *de facto* standard in the description of tone in East Asian languages; for these two reasons, stylizations of the tones of Kim Thuong Muong using Chao’s system are provided in Table 2.6.

The description of Vietnamese tones by Kirby (2011) served as a reference in the present description. Of course the tone system of Kim Thuong Muong is treated here as a system in its own right, not as a variant of the tone system of Vietnamese; on the other hand, Vietnamese offered a convenient point of comparison. For instance, the adoption of the label 33 in J. Kirby’s description of Hanoi Vietnamese tone A<sub>1</sub> (*ngang*) led me to adopt this label for Tone 33 of Kim Thuong Muong, which is somewhat similar though a bit lower than the Vietnamese tone A<sub>1</sub>.

### Etymological tone categories in Kim Thuong Muong

The tonogenesis of the Viet-Muong group, as in most East Asian languages, is the result of two distinct phenomena (André-Georges Haudricourt, 1954b; Ferlus, 1998):

- (i) a two-way splitting composed by the confusion of the two kinds of initial consonants: voiceless (made up the series of high register) and voice (made up the series of low register);
- (ii) a three way splitting caused by the loss of two kinds of final consonants: on the one hand the glottal the glottal stop /-ʔ/ or glottal constriction /-ʔ/, on the other hand the final spirant /-h/, which made up tonal categories B and C respectively.

Table 2.6: A brief overview of the tone system of Kim Thuong Muong. St.V. = Standard (Hanoi) Vietnamese.

tone	shorthand name	label	comparison with St.V.	etymology
<b>Tone 1</b>	Mid-level	[33]	lower than A1	C1
<b>Tone 2</b>	Falling	[53]	similar to A2	A1
<b>Tone 3</b>	Rising	[35]	similar to B1	B2
<b>Tone 4</b>	Glottalized	[3ʔ3]	similar to C1 with glottalization	B1
<b>Tone 5</b>	Top-high	[55]	higher than A1	A2 and C2
<b>Tone 6</b>	high-checked	[5]	similar to D1 but flat	D2
<b>Tone 7</b>	low-checked	[3]	similar to D2 but flat	D1

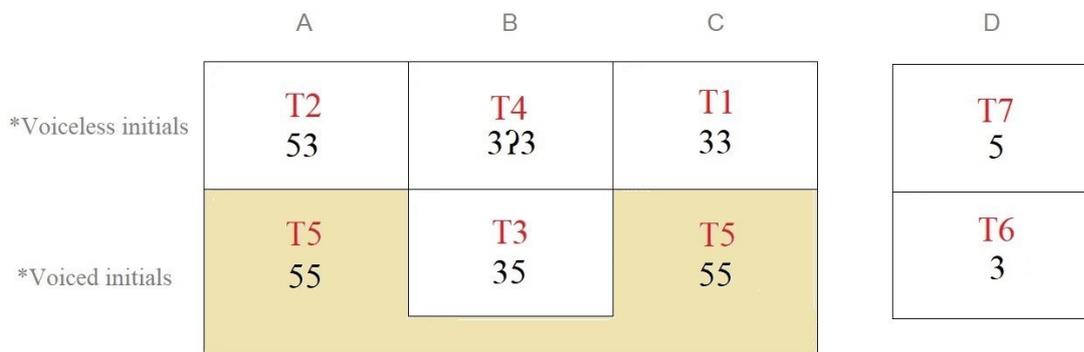


Figure 2.19: Tone box of Kim Thuong Muong

Basically, this process brought out a tonal frame for languages of Viet-Muong group with 6-tones (in smooth syllables). The basic tonal system and the diversity of realizations in the dialects Muong are the same as in Vietnamese. There are a few dialects of Muong that have 6 tones like northern Vietnamese, such as Muong Dam in Hoa Binh which reported by Ferlus (1998). Most other dialects have 5 tones with B2 - C2 confusion (V.-T. Nguyen, 2005). Figure 2.19 shows the tone box of Kim Thuong Muong. It indicates a merger of etymological categories A2 and C2. This is not an unprecedented phenomenon. In fact we already have encountered similar cases in a Nguon dialect in Quang Binh, which is supposedly Muong language in central Vietnam (Ferlus, 1998). Table 2.7 provides some examples of this confusion.

Going back to the topic of the glottalized tone, **Tone 4** of Kim Thuong Muong derived from an ancient final glottal constriction /-ʔ/, i.e., B position according to tone box's label. We can find numerous correspondences between this tone of Muong and two Vietnamese tones which belong to the same category: tone 'sắc' B1 in most cases and tone 'nặng' B2 in a few cases. Refer to the discussion of André-Georges Haudricourt (1961) on this topic, he provided four examples to demonstrate that: tones of category B (in Vietnamese) come from a final glottal stop which are still preserved

Table 2.7: Examples of a merger of etymological categories A2 and C2 in Kim Thuong Muong

Muong	Vietnamese	English	French
/pɤj <sup>5</sup> /	trời /tɤj <sup>A2</sup> /	sky	ciel
/ɲa <sup>5</sup> /	nhà /ɲa <sup>A2</sup> /	house	maison
/pun <sup>5</sup> /	bùn /bun <sup>A2</sup> /	muddy	boueuse
/waŋ <sup>5</sup> /	vàng /vaŋ <sup>A2</sup> /	gold	or
/ruŋ <sup>5</sup> /	rừng /ruŋ <sup>A2</sup> /	forest	forêt
/paw <sup>5</sup> /	bão /baw <sup>C2</sup> /	typhoon	typhon
/ŋɤŋ <sup>5</sup> /	ngỗng/ŋoŋ <sup>C2</sup> /	goose	oie
/muj <sup>5</sup> /	mũi /muj <sup>C2</sup> /	nose	nez
/tuə <sup>5</sup> /	đũa /đuə <sup>C2</sup> /	chopsticks	baguettes
/mɤ <sup>5</sup> /	mỡ /mɤ <sup>C2</sup> /	grease	graisse

in Palaung-Wa dialects and in Khmu. We reproduced his examples in Table 2.8 and added the correspondences in Kim Thuong Muong to confirm that Muong unmistakably follow this rule.

Table 2.8: Correspondences between Kim Thuong Muong – Vietnamese – Khmu, illustrating the presence of a final glottal stop in Palaung-Wa corresponding to tone B (B1 ‘sắc’, B2 ‘nặng’ in Vietnamese; and B1 – glottalized tone in Kim Thuong Muong.)

	Palaung – Wa	Vietnamese	Muong
Leaf	/laʔ/ (Khmu), /laʔ/ (Riang)	lá /la <sup>B1</sup> /	/la <sup>4</sup> /
Fish	/kaʔ/ (Khmu, Riang)	cá /ka <sup>B1</sup> /	/ka <sup>4</sup> /
Dog	/soʔ/ (Khmu)	chó /cɔ <sup>B1</sup> /	/cɔ <sup>4</sup> /
Rice	/rəŋkoʔ/ (Khmu), /koʔ/ (Riang)	gạo /ɣaw <sup>B2</sup> /	/kaw <sup>4</sup> /

Glottalization in category B also found in other Muong dialects according to Ferlus’ report (Ferlus, 1998, p. 8), such as Muong Dam (Hoa Binh) – tone [21’], Muong Tan Phong (Son La) – Tone [312’], and Nguon (Boc Tho) – Tone [313’] and [54’]. He use an apostrophe to mark the existence of glottalization. It seems that this prevalence of glottalization, especially of creaky voice in Muong dialects was a factor that led Diffloth (1989) to the reconstruction of a phonation type contrast between creaky voice and clear voice in Proto-Austroasiatic. Indeed, his proposal was based mainly on evidence from Katuic, Pearic and Vietic languages. However, this hypothesis was later re-examined and denied by Ferlus (2004).

# Chapter 3

---

## Method

Openness and clarity on topics of method are especially crucial to research which aims at the highest standards of Open Science. Hence, much detail will be provided in this chapter on how the present study was conducted. I will elaborate on each of the successive steps I have taken: selecting speech materials, designing the speech production experiment, and analyzing and visualizing the data.

In addition to the above topics, which are routinely covered in the Method section of phonetic studies, the present chapter will place emphasis on two aspects that have often received somewhat less attention in phonetic/phonological studies. One is the process of preparing data sets for archiving and online distribution, with a view to data reusability in future, and to hand-in-hand cumulative progress in documentation and research. Another aspect concerns the sociolinguistic context to the study: it appears relevant to start out the present chapter by explaining some of the specificities of *fieldwork experiments* to an audience of phoneticians/phonologists, without assuming any experience of immersion fieldwork on the part of readers.

To facilitate the reading of this chapter, all details related but not directly focused on the methodology of this thesis will be organized in the appendix with a mention in the main text.

### 3.1 Speech materials

The use of minimal sets is ideal for the study of tones, as it offers a way to compare tokens while keeping intrinsic properties of segmental phonemes constant: such as intrinsic pitch and co-articulatory influences from surrounding consonants (Hombert, 1978; Whalen and Levitt, 1995). Keeping in mind that “[s]tudies of non-citation forms of tones in different prosodic positions help elucidate listener behavior in tonal identification”,<sup>1</sup> some narratives were also elicited.

The target words of the experiment belong to 12 minimal sets that contrast for tone in smooth syllables (i.e., open syllables or syllables ending with a nasal coda) and 3 minimal pairs that contrast for tone in checked syllables (i.e., syllables ending with a stop coda).

---

<sup>1</sup>This observation is found in a LabPhon conference presentation by Christina Esposito and Marc Garellek, 2018. Available: <https://pdfs.semanticscholar.org/cc6e/3959782ff8158ba0315ba7ffec2aff71e42.pdf>.

These minimal sets and pairs were an extension of the list available with only 2 minimal sets in 2016. The method was based on 2961 vocabularies from the [EFEO-CNRS-SOAS word list](#). The details of strategies for searching minimal sets and pairs for the study of the tone system will be mentioned in the appendix A. This appendix is intended to clarify the entire procedure of this study and hopefully it will be a useful documentation and suggestion for people who want to start studying the tone system in particular and phoneme inventories in general, in undocumented and unwritten languages.

### 3.1.1 List of the (near-) minimal sets

Tables 3.1 and 3.2 provide full detail about the minimal sets and pairs. The tables include:

- **First column:** The numbering of minimal sets (from 1 to 12) and minimal pairs (from 1 to 3).
- **Second column:** The numbering of target syllables, labeled as “UID” (for “Unique Identifier”) because this number constitutes the unique identifier of target syllables. This number is used in the annotation of audio files, and in data processing down the line.
- **Third column:** The target syllables. These constitute the actual speech material of the recording session. In other words, the speakers were asked to pronounce these monosyllabic morphemes (roots).
- **Fourth column:** The full form of the target words from which monosyllables were extracted, in cases where the usual form of the word at issue is disyllabic. This point will be elaborated on below.
- **Fifth-sixth-seventh columns:** The translations in English, French and Vietnamese, respectively.

Initially, I was able to put together a total of 13 minimal sets and 6 minimal pairs from my vocabulary lists. However, one set and three pairs were removed from the list when it turned out, in the course of recordings, that they were not exactly minimal contrasts: there were mistakes in my initial transcriptions of the rhymes of some of the words. It needs to be kept in mind that there is as yet no dictionary or reference grammar of Kim Thuong Muong, and the analyses still constitute work in progress, even for some aspects of the language’s phonological system. The lines in gray text at bottom of tables reflect the ‘bootstrapping’ nature of the work: they contain the spurious set and pairs, which were included in the recording sessions from several speakers (before the issue was noticed). Even though these tokens were not processed at the stages of data annotation and parameter extraction, they are present in the recordings. The files are archived *as is* for the sake of preserving the integrity of the original recordings; for the sake of consistency between the primary data and the analyses based on it, it seemed advisable to include the stray syllables in Tables 3.1 and 3.2.

Overall, true-and-tested minimal sets are not easy to come by. Near-minimal sets, on the other hand, are easier to find.

Pairs that show segments in nearly identical environments, such as azure/assure or author/either, are called near-minimal pairs. They help to establish contrasts where no minimal pairs can be found. (Dobrovolsky and Katamba, 1996).

As only eight proper minimal sets (distinguished only by tone) were found, four near-minimal sets were added. The syllables in the latter sets differ not only by tone but also by initial consonant.

Even though the constraint on segmental identity was relaxed in order to obtain these additional sets, it was not lifted altogether: the alternative consonants are immediately adjacent to the target phonemes in the International Phonetic Alphabet chart. In other words, they have the same manner of articulation as well as a nearby point of articulation (i.e. a strong similarity in terms of place of articulation).<sup>2</sup> Use of near-minimal pairs is widespread in phonetic studies to circumvent the commonly encountered difficulty of finding minimally contrasting environments for phonemes (see e.g. Barlow and Gierut, 2002; De Boer, 2011; Garellek, 2015).

---

<sup>2</sup>I am grateful to James Kirby for issuing this recommendation for the present study.

Table 3.1: Speech materials: eight minimal sets and four near-minimal sets that contrast for five tones in smooth syllables.

N.	UID	Target syllable	Complete word	English	French	Vietnamese
1	1	pa <sup>5</sup>	pa <sup>5</sup> t <sup>h</sup> aj <sup>1</sup>	arm span	empan (de bras)	sài tay
	2	pa <sup>3</sup>	ke <sup>4</sup> pa <sup>3</sup>	a cylindrical jar to ferment vegetables	un pot cylindrique pour la fermentation des légumes	vại
	3	pa <sup>2</sup>	t <sup>h</sup> p <sup>6</sup> pa <sup>2</sup>	barrage	barrage	đập tràn
	4	pa <sup>1</sup>	pa <sup>1</sup>	cloth	tissu	vải
	5	pa <sup>4</sup>	pa <sup>4</sup>	fruit	fruit	quả
2	6	rɔ <sup>5</sup>	kɔn <sup>2</sup> rɔ <sup>5</sup>	tortoise	tortue	rùa
	7	rɔ <sup>3</sup>	rɔ <sup>3</sup> kuə <sup>2</sup>	to find crab (by hand) in rice field	attraper des crabes (à la main) dans une rizière	mò
	8	rɔ <sup>2</sup>	rɔ <sup>2</sup>	to be sated	être rassasié	no
	9	rɔ <sup>1</sup>	t <sup>h</sup> r <sup>5</sup> rɔ <sup>1</sup>	idle	désœuvré	rảnh rỗi
	10	rɔ <sup>4</sup>	pa <sup>4</sup> rɔ <sup>4</sup>	banana flower	fleur de bananier	hoa chuối
3	11	pa <sup>5</sup>	pa <sup>5</sup>	grand-mother	grand-mère	bà
	12	pa <sup>3</sup>	pa <sup>3</sup>	to touch on one's shoulder	se toucher l'épaule	bầu vai
	13	pa <sup>2</sup>	pa <sup>2</sup>	three	trois	ba
	14	pa <sup>1</sup>	pa <sup>1</sup> t <sup>h</sup> ien <sup>5</sup>	to pay	payer	trả tiền
	15	pa <sup>4</sup>	pa <sup>4</sup>	to patch	rapiécer	vá (xăm)
4	16	la <sup>5</sup>	la <sup>5</sup>	tongue	langue	lưỡi
	17	la <sup>3</sup>	p <sup>x</sup> 2 la <sup>3</sup>	to return	revenir	trở lại
	18	la <sup>2</sup>	la <sup>2</sup>	carry stuff or people on motorcycle	transporter des objets ou des personnes à moto	lái
	19	la <sup>1</sup>	la <sup>1</sup> t <sup>h</sup> ym <sup>5</sup>	a bamboo fence to keep fish in the lake	une barrière de bambou pour garder les poissons dans le lac	cái rào ao
	20	la <sup>4</sup>	la <sup>4</sup>	to drive	conduire	lái

5	21	taj <sup>5</sup>	taj <sup>5</sup> kaw <sup>4</sup>	to wash (rice)	laver (riz)	đai gạo
	22	taj <sup>3</sup>	taj <sup>3</sup> tɔŋ <sup>4</sup>	to pull bamboo by hand or by motorbike	tirer le bambou à la main ou à moto	lôi búông
	23	taj <sup>2</sup>	taj <sup>2</sup> nan <sup>3</sup>	accident	accident	tai nạn
	24	taj <sup>1</sup>	taj <sup>1</sup>	cascade	cascade	thác nước
	25	taj <sup>4</sup>	taj <sup>4</sup>	urinate	uriner	đái
6	26	kɔ <sup>5</sup>	kɔn <sup>2</sup> kɔ <sup>5</sup>	heron	héron	con cò
	27	kɔ <sup>3</sup>	kɔ <sup>3</sup>	to speak	parler	nói
	28	kɔ <sup>2</sup>	kiəw <sup>4</sup> kɔ <sup>2</sup>	tug of war	lutte acharnée	kéo co
	29	kɔ <sup>1</sup>	kɔ <sup>1</sup>	grass	herbe	cỏ
	30	kɔ <sup>4</sup>	kɔ <sup>4</sup>	to have	avoir	có
7	31	kieŋ <sup>5</sup>	kieŋ <sup>5</sup>	a earthenware jar to store liquids	une jarre en faïence pour conserver des liquides	bình sứ
	32	kieŋ <sup>3</sup>	ʔy <sup>2</sup> kieŋ <sup>3</sup>	beside	à côté	ở cạnh
	33	kieŋ <sup>2</sup>	kieŋ <sup>2</sup>	soup	soupe	canh
	34	kieŋ <sup>1</sup>	kieŋ <sup>1</sup>	gong	gong	kiêng
	35	kieŋ <sup>4</sup>	kieŋ <sup>4</sup>	wing	aile	cánh
8	36	ma <sup>5</sup>	ma <sup>5</sup> luəŋ <sup>2</sup>	ell's cave hole	trou de l'anguille	lỗ lươn
	37	ma <sup>3</sup>	ma <sup>3</sup>	rice seedings	plants de repiquage	mạ
	38	ma <sup>2</sup>	ma <sup>2</sup>	ghost	fantôme	ma
	39	ma <sup>1</sup>	ma <sup>1</sup>	tomb	tombeau	mả
	40	ma <sup>4</sup>	ma <sup>4</sup>	cheek	joue	má
9	41	ŋa <sup>5</sup>	ŋa <sup>5</sup>	to fall	tomber	ngã
	42	ŋa <sup>3</sup>	ŋa <sup>3</sup>	itch	prurit	ngứa
	43	ŋa <sup>2</sup>	ŋa <sup>2</sup> mət <sup>6</sup>	dazzle	éblouissement	chói
	44	ŋa <sup>1</sup>	ŋa <sup>1</sup> luŋ <sup>2</sup>	to recline	s'incliner	ngả lưng
	45	na <sup>4</sup>	ke <sup>4</sup> na <sup>4</sup>	archery	archerie	cung tên

10	46	ka <sup>5</sup>	paj <sup>4</sup> ka <sup>5</sup>	eggplant	aubergine	cà
	47	ta <sup>3</sup>	ta <sup>3</sup>	dumbbell	haltère	quả tạ
	48	ka <sup>2</sup>	kɔn <sup>2</sup> ka <sup>2</sup>	chicken	poulet	gà
	49	ka <sup>1</sup>	ka <sup>1</sup>	big	grand	to
	50	ka <sup>4</sup>	ka <sup>4</sup>	fish	poisson	cá
11	51	kaj <sup>5</sup>	kaj <sup>5</sup>	to button	boutonner	cài (cúc)
	52	paj <sup>3</sup>	paj <sup>3</sup>	a cylindrical jar to ferment vegetables	un pot cylindrique pour la fermentation des légumes	vại
	53	kaj <sup>2</sup>	kaj <sup>2</sup>	thorn	épine	gai
	54	kaj <sup>1</sup>	ta <sup>6</sup> kaj <sup>1</sup>	cabbage	chou	(rau) cải
	55	kaj <sup>4</sup>	kɔn <sup>2</sup> kaj <sup>4</sup>	the female	la femelle	con cái
12	56	ku <sup>5</sup>	ku <sup>5</sup>	old	ancien	cũ
	57	tu <sup>3</sup>	tu <sup>3</sup> mɿw <sup>4</sup>	hematoma	hématome	tụ máu
	58	ku <sup>2</sup>	ku <sup>2</sup>	buffalo	buffle	trâu
	59	ku <sup>1</sup>	ku <sup>1</sup>	tuber	tubercules	củ
	60	ku <sup>4</sup>	ku <sup>4</sup> miew <sup>5</sup>	owl	hibou	củ mèo
13	x	pɿj <sup>5</sup>	pɿj <sup>5</sup>	sky	ciel	trời
	x	pɿj <sup>3</sup>	pɿj <sup>3</sup>	to dig (by people)	creuser (par une personne)	(người) bới đất
	x	pɿj <sup>2</sup>	pɿj <sup>2</sup>	swim	nager	bới
	x	pɿ <sup>1</sup>	pɿ <sup>1</sup>	from	à partir de	từ
	x	pɿj <sup>4</sup>	pɿj <sup>4</sup>	to dig (by chicken)	creuser (par le poulet)	(gà) bới đất

Table 3.2: Speech materials: three minimal pairs that contrast the two tones of checked syllables.

N.	UID	Target syllable	Complete word	English	French	Vietnamese
1	61	pat <sup>6</sup>	pat <sup>6</sup> ja <sup>5</sup>	floor made by a kind of bamboo	plancher fait d'une sorte de bambou	sàn nhà sàn
	62	pat <sup>7</sup>	pat <sup>7</sup>	bowl	bol	bát
2	63	rwec <sup>6</sup>	rwec <sup>6</sup>	intestine	intestins	ruột
	64	rwec <sup>7</sup>	rwec <sup>7</sup>	to pour	verser	rót
3	65	lak <sup>6</sup>	lak <sup>6</sup>	peanut	cacahuète	lạc
	66	lak <sup>7</sup>	lak <sup>7</sup>	squint eye	strabisme	lác
4	x	kap <sup>6</sup>	kap <sup>6</sup>	quack (sound made by a duck)	coin-coin (son fait par un canard)	tiếng vịt kêu
	x	kap <sup>7</sup>	kap <sup>7</sup>	cable	câble	dây cáp
5	x	kup <sup>6</sup>	ke <sup>4</sup> kup <sup>6</sup>	trophy	trophée	cúp
	x	kup <sup>7</sup>	kup <sup>7</sup> taj <sup>2</sup>	lop-eared	aux oreilles tombantes	cup tai
6	x	kec <sup>6</sup>	kec <sup>6</sup>	cut	couper	cắt
	x	kac <sup>7</sup>	kac <sup>7</sup>	sand	sable	cát

The sets in Table 3.1 and the pairs in Table 3.2 are not ordered by a principled criterion such as the order of the alphabet: they are simply listed in the order in which I put them together. The first two sets were ready to use from my previous study (M.-C. Nguyễn, 2016, p. 20), and the rest were added since then based on the progress of the vocabulary collection task (which constitutes one of the language documentation “background tasks” to the present study). The strategy for finding the (near-) minimal sets is elaborated on in Appendix A.

The eight *bona fide* minimal sets are placed before the four near-minimal sets. The three minimal pairs are listed separately, in another table, to reflect the fact that from a structural perspective stop-final syllables belong to another tonal sub-system than the others (the smooth syllables).

The fact that Unique Identifiers (UIDs) were assigned early on (for the sake of data annotation) led me to retain the same ordering of sets later on, even though it is not grounded in phonological principles. A paramount concern is to ensure consistency between the corpus description and the actual data files. Updating a system of numeral unique identifiers is an error-prone process, so I chose to avoid changing the labeling of items, and thus I kept the original order. I wish to apologize to readers who have difficulties finding their way through these lists.

Another thing to note here is that the order of the tones in the minimal sets was also not placed in ascending or descending order according to tone label numbers but in the order: Tone 5 – Tone 3 – Tone 2 – Tone 1 – Tone 4. This is not a random order, we arrange them with the aim of maximizing the distinction between the tones. The labeling of tones in numbers is completely arbitrary, unrelated to any phonetic or phonological characteristic. During designing the experiment of minimal sets, we reorder them according to their relative descent in pitch level in order to optimize the tonal space. This follows the common tendency of  $f_0$  declination in paragraph prosody, which is similar to sentence but in the next level. The work of Vaissière and Michaud (2006, p. 5) affirms this tendency that:

The paragraph is the largest unit. The highest  $f_0$  value in each sentence tends to decline from the first to the last sentence in a paragraph, in French and in other languages (Lehiste, 1975). The end of the paragraph typically ends on an extra-low  $f_0$  (often leading to a change in voice quality) and intensity.

Figure 3.1, reproduced from Vaissière and Michaud (2006), shows an attempt to model this phenomenon.

Therefore, in the series, the first tone is Tone 5, the highest tone, and the last tone is the lowest tone, that is Tone 4. Tone 1 was easily placed in the penultimate position, before Tone 4, since it is consistently in the low-mid range, just above the creaky tone. The remaining two tones are an ascending tone (Tone 3) and a descending tone (Tone 2). Although normally the rise in Tone 3 is not as prominent as the fall in Tone 2, I placed Tone 2 after Tone 3 on purpose. Tone 2 is thereby placed just before Tone 1, highlighting the contrast between these two tones which I sometimes

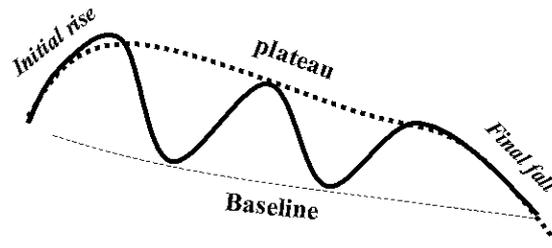


Figure 3.1: General outline of the  $f_0$  curve of an affirmative statement (Vaissière, 1983).  
Reproduced from Vaissière and Michaud (2006), with permission.

struggle to distinguish aurally from each other. By placing them close together, the speaker will be led to pronounce them with maximum distinction. That will help me to perceive the difference better, and thus to monitor recording sessions better, making sure that there is no confusion between the two tones during the experiment.

### *Some limitations of the materials*

Although efforts were made to improve over shortcomings noted in a previous study – reported in my 2016 M.A. thesis: M.-C. Nguyễn 2016, pp. 20–21–, the materials here still suffer from certain limitations. Some of these limitations may well be unavoidable, as they relate to important dimensions of linguistic structure.

First, there are a few cases of differences in the syllabic composition of the lexical items used as target words. Muong is a monosyllabic language, in the sense that “a syllable = a morpheme = a word” (Post, 2006, p. 43). Therefore, an ideal structure for the data would be for each minimal set to be composed of five monosyllabic target words of the same segmental composition, differing only in tone. However, this ideal is not always met. To deal with the difficulties in finding a complete set, in cases where it is not feasible to find the exact monosyllable needed for a complete set, two workarounds were used: (i) finding a disyllabic word which contains the needed target syllable, or (ii) finding a near-minimal set instead, relaxing the constraint on identity of initial consonants and widening the search to syllables whose initial consonant is similar to that of the other items in the set. The first solution, in fact, always came from the suggestion of the consultant(s). It is a very encouraging testimony to their engagement with the study that the native speakers understood the goal of the search and could offer such suggestions. Clearly, to the consultants who participated in the present study, this option is favoured over the option of building near-minimal sets, which was only used as a last resort. As a consequence, some target words of the minimal sets are less monosyllabic than the others from a lexical point of view. Outside context, they are not recognized as the intended morpheme, so that, strictly speaking, they have no meaning on their own, but they can nonetheless be pronounced stand-alone as a monosyllable, as part of a language game (no deep Wittgensteinian implications intended) between the investigator and the consultant.

Taking advantage of this characteristic, those “faux monosyllables” are gratefully and confidently employed for the current study. This is the case of /paj<sup>2</sup>/ in /tʰp<sup>6</sup> paj<sup>2</sup>/ ‘barrage’ in the first minimal set, /rɔ<sup>1</sup>/ in /ʔv<sup>5</sup> rɔ<sup>1</sup>/ ‘idle’ and /rɔ<sup>4</sup>/ in /paj<sup>4</sup> rɔ<sup>4</sup>/ ‘banana flower’ in the second minimal set, /laj<sup>1</sup>/ in /laj<sup>1</sup> tʰm<sup>5</sup>/ ‘a kind of bamboo fence’ in the fourth minimal set, and so forth.

Additionally, a few items for which a disyllabic form is provided in the relevant column of “complete word” in Tables 3.1 and 3.2 are in fact just fine as independent monosyllabic words. These words exemplify the tendency towards disyllabification (or disyllabication) whereby some words receive an increasingly frequent accompaniment (phonologically: an extra syllable) that helps disambiguate their meaning. Two cases can be distinguished here:

1. The target syllable needs a supplement (commonly a noun) to specify its exact meaning. The target item could be:
  - a verb, such as (minimal set N<sup>0</sup>5) /taj<sup>3</sup>/ in /taj<sup>3</sup> tɔng<sup>4</sup>/ ‘to pull kinds of bamboo by hand or by motorbike’ (whereas /paj<sup>2</sup>/ means ‘to pull’ in general, for various objects).
  - a noun, such as (minimal set N<sup>0</sup>8) /ma<sup>5</sup>/ in /ma<sup>5</sup> luəŋ<sup>2</sup>/ ‘ell’s cave hole’ (otherwise, the simple, monosyllabic word for ‘hole’ is /hɔŋ<sup>5</sup>/).
  - a partitive noun, such as (minimal set N<sup>0</sup>1) /paj<sup>5</sup>/ in /paj<sup>5</sup> tʰäj<sup>2</sup>/ ‘arm span’.
2. The target syllable carries the full meaning by itself but addition of a classifier allows for easier identification. This is the case of the classifier for animals, /kɔn/, in /kɔn<sup>2</sup> rɔ<sup>5</sup>/ ‘tortoise’ (minimal set N<sup>0</sup>2), in /kɔn<sup>2</sup> ka<sup>2</sup>/ ‘chicken’ (minimal set N<sup>0</sup>10), and in /kɔn<sup>2</sup> kaj<sup>4</sup>/ ‘the female’ (minimal set N<sup>0</sup>11); or of /ke<sup>4</sup>/ in /ke<sup>4</sup> paj<sup>3</sup>/ ‘a cylindrical jar to ferment vegetables’ and in /ke<sup>4</sup> na<sup>4</sup>/ ‘archery’.

An additional case seems worth mentioning here: the item /kieŋ<sup>3</sup>/ in /ʔv<sup>2</sup> kieŋ<sup>3</sup>/ “beside” in the minimal set N<sup>0</sup>7. The target syllable /kieŋ<sup>3</sup>/ can in fact be a monosyllabic word meaning “edge”. However, I decided to pick the compound word instead because I found that it is easier to find a picture to represent “beside” than “edge”. (The criteria to seek and select the photos will be explained in Appendix A.) In this case, the target syllable is a complement for a preposition.

To deal with this lack of homogeneity, in the experimental procedure, part of the training consisted in explaining to the speakers that the intended target items for the task are always monosyllabic, and clarifying to them exactly the intended syllable in cases where disyllabic words are involved.

It is paradoxical to note that the less the complementary syllables are obligatory, the easier it is for a speaker to forget the instruction to trim the word down to a monosyllable when speaking it out: pronouncing monosyllabic words only. This means that they are more likely to say a disyllabic word for ‘chicken’ (/kɔn<sup>2</sup> ka<sup>2</sup>/) and ‘archery’ (/ke<sup>4</sup> na<sup>4</sup>/), which would be just fine as monosyllabic /ka<sup>2</sup>/ and /na<sup>4</sup>/ respectively, than for /ma<sup>5</sup> luəŋ<sup>2</sup>/ ‘ell’s cave hole’ or in /paj<sup>4</sup> rɔ<sup>4</sup>/ ‘banana flower’, which they take care to trim down to a monosyllable as instructed. This observation provides a glimpse into the considerable importance of classifiers as a handy grammatical tool in Muong (as also in Vietnamese and a large number of languages of the area). Since

the recording sessions were monitored in-person throughout, speakers were reminded of the instruction immediately after each deviation, asking them to correct by saying the expected monosyllable during the recording. The original recordings keep track of ‘deviant’ realizations, so that someone who wished to verify their frequency and analyze their patterns could carry out a study (somewhat like the observations about tonal ‘slips of the tongue’ in Yongning Na: Michaud, 2017, pp. 180–181).

A second dimension of asymmetry in the word list, which directly affects the choice of photos that serve as stimuli, is that different parts of speech are used. As will be detailed in Appendix A, finding an optimal visual prompt for eliciting speech materials constitutes a difficult task (in retrospect, I find that it was the greatest difficulty in data collection on an unwritten language). The illustrative photos method turns out to be the most effective in this situation. To be able to show by photos, the target words must be visually clear (*visually tangible*, as it were). Hence, as a general rule, lexical words are to be preferred over function words. One of the exception is prepositions that relate to space: those can be conveniently illustrated. Thus, in the list of words for this study, a preposition have been included in the minimal set N<sup>0</sup>7. That is /*kien*<sup>3</sup>/ in /ʔ<sup>2</sup> *kien*<sup>3</sup>/ “beside”. As mentioned earlier, this target syllable has a homonymous word meaning “edge”, which is monosyllabic and a lexical word. However, we decided to pick the preposition instead because it was found easier to find an illustrative photo for “beside” than “edge”.

When using function words, it should be taken into account that they are known cross-linguistically to be realized with less articulatory energy – they are phonologically ‘weaker’, as it were. However, the difference between content and function words in Viet-Muong might be significantly less than in most other languages. Phonologically, the function words in Vietnamese and Muong have lexical tone, like content words. This is a typological difference from, for instance, Mandarin Chinese dialects, where some function words are toneless (this is sometimes called ‘light tone’ or ‘neutral tone’ in Sinological parlance: see W.-S. Lee 2003; Zhang and Hu 2020). Phonetically, experimental research about Southern Vietnamese (based on over 60,000 syllables of conversational materials) reveals differences in duration between function words and lexical words: “function words are independently shorter than lexical words”, but the difference is on an order of magnitude that is relatively limited (Brunelle, 2015). This suggests that the inclusion of function words in minimal tonal sets could be an acceptable compromise for Vietnamese (and Muong, which appears fairly similar in this respect).

On the other hand, within the range of lexical words (or content words), the possibilities of visualization are also different. A noun can usually be illustrated more easily and directly than a verb or adjective, but abstract nouns (such as happiness, patriotism, knowledge, commerce, etc.) are conspicuous counter-examples: they are almost impossible to show directly, and it is easy to get lost in the cultural meanders of allegorical representations. For instance, Figure 3.2 gives examples of efforts to illustrate the abstract nouns “patriotism” and “commerce”.

Considering the materials of this study, it can be seen generally that most of the target words picked up for the word list here are pretty common and lend themselves



(a) An illustrated photo for "patriotism"



(b) An illustrate photo for "commerce"

Figure 3.2: An effort to illustrate abstract nouns

to not-too-ambiguous visual representation. They are likely to be often used, as they belong to relatively basic vocabulary. Nouns constitute nearly 63% of the list, the rest is made up of about 24% for verbs and 12% for adjectives. There is only one preposition: /ʔ<sup>2</sup>kieŋ<sup>3</sup>/ “beside”, as mentioned above. However, it is still inevitable that some words are relatively abstract and difficult to be illustrated. The word that was the most challenging is the verb /kɔ<sup>4</sup>/ “to have” in the minimal set N<sup>o</sup>7. To illustrate this word, we made a detour through one of its uses: in Muong, the verb “to have” is fixed in the compound word /kɔ<sup>4</sup>chuuə<sup>1</sup>/ which means “to be pregnant”, and using the image of a pregnant belly proved a satisfactory way to elicit this verb. As explained above, elicitation sessions involved a training phase where the desired target items were explained and discussed.

A third issue which could be considered a flaw of the corpus is the diversity among initial consonants, which differ in terms of voicing feature as well as in terms of place and mode of articulation. Specifically, the voiced initial consonants in the data set are:

- The alveolar trill /r/ in syllable /rɔ/ (minimal set N<sup>o</sup>2), which is sometimes pronounced as a flap consonant /r/ (there is no phonemic distinction between an alveolar trill and an alveolar flap in Muong),
- The alveolar lateral approximant /l/ in syllable /laj/ (minimal set N<sup>o</sup>4),
- The bilabial nasal /m/ in syllable /ma/ (minimal set N<sup>o</sup>8),
- The velar nasal /ŋ/ in syllable /ŋa/ (minimal set N<sup>o</sup>9).

The use of such a wide range of initial consonants constitutes an important factor of complexity. Effects of different consonant types on the tone of the following vowel are a fairly well-documented and intensely researched topic. As a general rule, after voiceless obstruents, the fundamental frequency contour is relatively higher (in particular at the onset of voicing, towards the beginning of the vowel); whereas after voiced obstruents, the fundamental frequency contour is relatively lower (Gandour, 1974, p. 337). Onset pitch perturbations (sometimes called “pitch-skip effects”) thus influence the observable fundamental frequency curves of tones, in terms of overall register but also of modulation of  $f_0$ . The relationship between Voice Onset Time (VOT) and onset  $f_0$  perturbations is documented in a vast literature (for instance Kingston, 2009; Hanson, 2009; Kirby, 2018; Michaud and Sands, 2020).

In the segmentation process, all voiced initial consonants are excluded, as a matter of course, retaining only the rhymes for comparison with other syllables whose initial consonants are voiceless stops (detailed in Section 3.3.1). Arguments for the division of the syllable into an initial and a tone-bearing rhyme are recapitulated in Michaud and Kühnert (2006). This manipulation bears a relationship with the measured duration of the tones. The transition between voiced consonants and the following vowel is not always obvious. As a general principle, I usually segment at two or three periods later from the point in the transition where I believe that the vowel has started to be present.

Of course, excluding voiced initial consonants does not amount to factoring out their effects on the following tone (their co-intrinsic properties). The diversity of initial consonants is to be kept in mind when looking at averaged curves, as in 4.1.

A final caveat about the word-list experiment is the difficulty to control whether or not the speaker runs into code switching when the target syllable has a segmentally matching syllable in Vietnamese. This issue had been avoided in the previous study when we had only two minimal sets. The two initial consonants of the minimal sets in the previous study were /p/ and /r/, which exist in Muong but not in Northern Vietnamese (except a few cases of /p/ in borrowed words, such as /pin/ “battery” from French *pile*, and a learned pronunciation of orthographic *r* as /r/ instead of /z/). Therefore, the realization of /p/ as /b/ and /r/ as /z/ (their closest Vietnamese equivalent) would reveal that the speakers had somehow shifted to Vietnamese mode. In the current study, on the other hand, the vowels and consonants of eight out of a total of twelve syllables (“segmental skeletons”) are found in both languages. They are:

1. /laj/ in minimal set N<sup>o</sup>4,
2. /taj/ in minimal set N<sup>o</sup>5,
3. /kɔ/ in minimal set N<sup>o</sup>6,
4. /ma/ in minimal set N<sup>o</sup>8,
5. /ŋa/ in minimal set N<sup>o</sup>9,
6. /ka/ in minimal set N<sup>o</sup>10,
7. /kaj/ in minimal set N<sup>o</sup>11,
8. /ku/ in minimal set N<sup>o</sup>12.

In these cases, the only phonetic/phonological difference on which we can base our intuitions about possible code-switching is tone. However, in the case of rising tones (i.e. Tone B<sub>1</sub> in Vietnamese and Tone 3 in Muong), I do not feel at all confident distinguishing a target Tone B<sub>1</sub> (Vietnamese) and a target Tone 3 (Muong). These tones are both glottalized, and while the Vietnamese tone tends to have strong glottal constriction whereas the Muong tone has creaky voice, they share some allophones. Whether the syllables are truly undistinguishable, or (more plausibly) some statistical difference can be brought out, is not relevant to the experimental predicament encountered: it is clear that my aural impressions are not a reliable tool to distinguish the Muong rising tone and the Vietnamese rising tone, much less to ascertain possible intermediate forms (mild cases of interference between the two languages). In such cases, where the difference between the two tone systems is not salient, code switching between Muong and Vietnamese can easily escape detection. Fortunately, the method of illustrative photos instead of reading test somewhat appears as a good way to prevent this issue.

To conclude this list of limitations on a positive note: there is no way to overcome limitations completely, but this need not be a cause for chagrin. First, for the immediate purposes of the study, the slight asymmetries and imperfections noted above do not affect the experiment adversely to an extent that would jeopardize the whole enterprise. Secondly, at an important epistemological level, sifting through asymmetries in the corpus amounts to facing constraints that stem from important linguistic and cultural facts. We are thereby led to accept these facts and to understand them. The process serves as a healthy reminder that phonetics-phonology is not a world of its own, blissfully free from ties with the rest of the universe. Reading through books and articles about phonetics-phonology, I sometimes have a feeling (shared by some others

at ‘fieldwork-intensive’ research centres such as LACITO, LLACAN and SEDYL) that some phonologists could benefit from increased practice analyzing data pertaining to syntax, morphology and communication generally. These levels should be central in building phonological theories. As a result, I sometimes find myself uncertain about the purpose of phonologists’ enterprises. The fact that doing phonology and phonetics does not always involve thorough contact with the target language may result in a drift away from social and psychological phenomena and factors that one has to face in immersion fieldwork. Learning a language and being in touch with speakers on a day-to-day basis leads one to take heed at the evidence that the lexicon, syntax, morphology, pragmatics all constitute relevant levels at the heart of linguistic systems, not to mention ethnolinguistic and social factors. Factors from all levels should be embraced as objects of study partaking in spoken communication, rather than as embarrassing impediments. Asymmetries in the data, as also exceptions and counter-evidence to phonological generalizations, need not be seen as embarrassing failures of phonology as a discipline. The ultimate goal is not to arrive at a future superior phonological analysis to account for all glitches in the data, but to arrive at a balanced linguistic understanding of language.

### 3.1.2 *Carrier sentence*

The use of carrier sentences is widespread among phoneticians-phonologists in order to stabilize the phonetic realization of target words. The approved order of business consists in starting out by establishing the tones’ templates as realized in carrier sentences. Those can then be used as a reference for the study of coarticulation and, in due course, the full range of intonational phenomena.

When a word is spoken in isolation, there can be various effects due to the fact that the word is at the same time in utterance-initial and utterance-final position, since it constitutes an utterance by itself. Using the same carrier sentence for all target words allows for keeping the segmental and linguistic context constant. This method is commonly used in phonetic experiments, such as Michaud (2004b), Mazaudon and Michaud (2008), and Mac et al. (2015), among many others.

In this study, the corpus includes target syllables both in isolation (“citation form”) and in a carrier sentence. For each minimal set, after being collected in isolation, the target words were embedded one by one in pre-final position in the following carrier sentence:

- (1) /ja<sup>2</sup> măt<sup>6</sup> \_\_\_\_\_ tǎŋ<sup>3</sup>/  
 2SG to\_know target item INTERROG  
 ‘Do you know \_\_\_\_\_?’

These stimulus sentences were not presented in orthography, since Muong is an unwritten language. In order to demonstrate the carrier sentence to the speakers as an instruction on what they were expected to do in the experiment, I asked my Muong teacher to utter it. In cases where she could not be present at the recording session,

the method I resorted to consisted in doing my best to pronounce the carrier sentence myself, and then translating it into Vietnamese to double-check that the speakers were fully clear what it was.

A fictional indication of context was also provided. Each speaker can (and, in a sense, must) imagine a context of some sort when saying a sentence, and as a consequence, uncontrolled variability in the suprasegmental features of the carrier sentence can occur. Not providing any indication about context therefore introduces an uncontrolled factor of variability (as pointed out e.g. by Niebuhr and Michaud, 2015). For this reason, we provided a specific context for the sentence, so it would be easier for participants to have a reasonably clear and consistent image of a situation of communication. The speakers were instructed to imagine that they were teaching their language (i.e., Muong) to a non-native learner. Before teaching a new word, they would ask the learner whether or not he/she already knew it.

This carrier sentence had been devised and used in our previous study in 2016. We reused it for this study for at least three reasons: (i) it was safe to use a tied-and-tested carrier sentence whose advantages and disadvantages I already knew, (ii) it helps to sync the corpus of the two studies, which is handy for proper comparison, and last but not least, (iii) the carrier sentence was already familiar to those speakers who had participated in the 2016 experiment, and who also participated in the second study.

Among the advantages of this frame is the fact that the sentence, in its imaginary context, tied in neatly with the actual situation in which the investigator, who was not a native speaker, came to learn and study the dialect from local people. It helped participants remember that sentence easily and produce the speech as naturally as possible. A further advantage is that a four-syllable sentence (including the target word) proved a reasonable length to memorize and repeat. Brevity of the carrier sentence is particularly valuable when the number of target words is large. Yet another positive side of the chosen carrier sentence is that, within this frame, the target word plays the role of new, central information, which leads the word to be articulated with clarity (“hyper-articulated”, to use Lindblom’s term: Lindblom 1990). The target word is thereby protected against phonetic shortening and hypo-articulation – phenomena that are amply attested in phonetic realizations of the other syllables in the sentence: those that are constant in the frame, i.e. the first, second and fourth (as reported in Table 3.7).

On a less positive note, using a question as a carrier sentence, as done here, does not go without saying. It brings both advantages and disadvantages. A declarative sentence seems to be favored for this kind of phonetic experiment because it is the most neutral one among the different sentence modalities. Interrogation comes with some sort of intonational flourish in many languages. “It is almost invariably the case that high or rising pitch signals the former [a question] whereas low or falling pitch, the latter [a statement]” (Ohala, 1983, p. 1). For instance, in English, as well as in many other non-tonal languages, interrogative intonation, especially in yes/no questions, tends to have an ascending pitch contour, whereas declarative intonation has a descending pitch contour (Pierrehumbert, 1980; Ladd, 1981). In a tonal language like Chinese or Vietnamese, however, the difference between declarative and interrogative intonation

appears less straightforward because it involves interaction of tone and intonation, as well as the use of sentence-final particles that play an important role in conveying sentence mode and speaker attitude generally. The difference between declarative and interrogative intonation in Vietnamese has been widely studied, but no consensus has yet been reached on this issue. Some studies have attempted to show that there is a robust difference between declaratives and interrogatives in several acoustic correlates, such as overall  $f_0$ , duration and intensity (T. T. H. Nguyen and Boulakia, 1999; Brunelle, Hà, and Grice, 2012; Đao and A.-T. T. Nguyen, 2018). For example, in term of global  $f_0$ , while declaratives are found to have a slight overall declination in  $f_0$  (Do, Thien Hương Tran, and Boulakia, 1998; T. T. H. Nguyen and Boulakia, 1999), interrogatives are described as having a high overall  $f_0$  range (Hoang, 1985), or a high range plus a rise which begins much earlier than the final question marker of the sentence (Do, Thien Hương Tran, and Boulakia, 1998; T. T. H. Nguyen and Boulakia, 1999). On the other hand, others claim that the final particles in Vietnamese, which are frequent in spontaneous speech, carry the bulk of the functional load in expressing interrogation: once they are taken into account, the associated intonational patterns appear to be largely redundant (Seitz, 1986; Do, Thien Hương Tran, and Boulakia, 1998).

Our observations on Kim Thuong Muong's intonation suggest that the situation in this dialect is quite similar to Vietnamese as viewed through the prism of the argument mentioned above: intonational phenomena related to interrogation tend to cluster towards the very end of the utterance. Muong has a final rise in interrogatives, which is firmly grafted onto the last question markers or particles, which are really mandatory to express interrogation. (Details are to be found in a recent paper to which I contributed: Michaud, M.-C. Nguyễn, and Scholvin 2021, and also in section 5.5 of this thesis.) Specifically, the final question marker /cảj<sup>3</sup>/, employed in the carrier sentence, is the locus where salient intonational phenomena related to sentence mode appear. There is no ground for concern about a possible neutralization (or even reduction) of tonal contrasts on the syllable that precedes this particle.

### 3.1.3 *Narratives*

It is rather awkward to break the flow of the description of the highly controlled phonetic experiment conducted on minimal sets by introducing here a mention of spontaneous speech materials recorded as part of the same data collection campaign. It is all the more awkward as these materials are not analyzed systematically in the present study. However, they do have an important methodological goal, if only from a theoretical point of view: the ultimate goal is of course to understand the interplay of  $f_0$  and phonation type *in actual communication*, not only in one highly controlled context that might turn out to be somewhat marginal in terms of ordinary, real-life communication settings. Therefore, different types of speech data were collected during this study and could be used for current or future analyses.

In addition to the isolated words and words in the isolated carrier sentence that are obtained from the minimal set experiment, I also recorded many types of spontaneous speech, including monologues or dialogues telling narratives, or singing folk songs. A

complete full freedom was granted to speakers as to the topics for spontaneous speech. On the other hand, it has been anticipated that suddenly being asked to talk about a random topic without any preparation in advance would be quite awkward and difficult for them. In order to facilitate this task, some general topics has been prepared in advance to evoke speakers in case they are confused and do not know what to say. For instance, the following simple and safe topics was used as a stimulus for a start:

- The stories of the works throughout the year or year-round agricultural process
- The stories of the construction of the traditional houses (nhà sàn)
- Traditional cuisine
- Traditional medicine
- Traditional festivals and celebrations
- Childhood memories
- Folk games and songs
- etc.

The actual topics taken up were the following:

1. The stories of the works of agriculture and farming throughout the year
2. Daily activities
3. Isolated life and work on the family's farm
4. Occupational accidents
5. How to construct a traditional houses: in the past, the houses were built with wood and with a kind of bamboo, today they are built with bricks.
6. How to cook traditional sticky rice in bamboo tube (cơm lam)
7. The process of harvesting mosses from the stream and making them into food
8. The process of picking bamboo shoots and making dried bamboo shoots (măng khô)
9. The process of making the traditional rice cake (bánh chưng)
10. How to cook and drink alcohol locally
11. Types of local fish, the local techniques of fishing and cooking fish
12. Traditional wedding: past and present
13. Festivals and rituals during the year
14. The practice of money saving (góp hụi)
15. The traditional techniques of textile
16. Giving birth to and raising children
17. Uncle Tu's funeral (a local man (also the speaker M2) who died in the village in 2018)
18. Local remedies and medicinal plants
19. Some folk tales
20. Folk games
21. Lullabies and folk songs
22. Make a hammock to cradle children
23. The techniques of slaughtering cattle and distributing equally in the village
24. Local epidemic (HIV/AIDS)
25. Childhood in wartime

26. Family stories
27. Stories of going to see a fortune-telling
28. Tales of incantation
29. Tales of ghost
30. The custom of adoption in the village by name of new born baby
31. The custom of calling parents' names after their children

This list was created gradually during the field trips, hence there are some topics related to the specific characteristics or situation of the region.

Most of the Muong songs that have been recorded were composed recently, so the lyrics mix a lot of Muong language and Vietnamese language. The most valuable song is the "ant song" and a lullaby performed by speakers F6, M9, F13, F14, F16 and F17. All these recordings are mentioned and detailed in the metadata (Appendix B). They will be prepared and published on [Pangloss](#) soon.

As part of this recording, we also asked all the speakers to recount the process of making "banh chung" (the traditional square glutinous rice cake). Everyone knows this process, so it was feasible to collect it. Also, using a method similar to "[Pear story](#)" (Eliciting narratives through a film), I prepared a few slides in advance to remind them of all the details. This way, they were more likely to use the same vocabulary that frequently carries the glottalized tone. This benefits a further study of glottalization on spontaneous speech.

## 3.2 Experiment setup

This section includes information about the participants, a brief introduction to the equipment, and the procedure for the experiment. In the appendix B, details of the types of equipment and setup as well as several problems encountered with the equipment and devices will be provided as a reference for those interested in some practical problems of recording and conducting phonetic field experiments or looking for a detailed guide to starting phonetic fieldwork.

### 3.2.1 Participants

This study comprises data from twenty speakers (ten male, ten female) past the age of eighteen. None reported any particular health or language-related difficulties. For ease of reference, each speaker is assigned an alphanumeric label, with the initial letters M or F standing for male or female respectively. I hasten to add that the use of speaker codes is by no means intended to hide the speakers' identity: they proudly agreed to be mentioned by name as contributors to the study, and their name is indicated in the metadata of the online recordings in the [Pangloss](#) Collection.

The number following the letter marks the order in which each speaker contributed to my study in its broadest sense, i.e. a process of exploring, experimenting, and documenting this language from 2015 to 2019. As a consequence, the full list provided in Table 3.3 includes not only the information of the 20 speakers just mentioned

above, but also another 15 speakers. Among these fifteen speakers, some participated in recordings since my Master's study (in 2015-2016), whereas others only participated in the second data collection campaign (in 2018-2019); their data were not processed and analyzed in this thesis due to difficulties relating to the quality of the electroglottographic signal (mostly women). Additionally, the two speakers F6 and F15 did not participate in the experiment but were recorded for some folk stories and songs, which provide relevant evidence on Muong tones (and the Muong language generally) beyond the specific set of carefully controlled phonetic materials that sits at the heart of the present work. They were also assigned identifiers in order to facilitate the archiving of the whole corpus, as explained in Appendix A. Such are the reasons why there is no continuity in the identifiers of the 20 speakers in the current study's central experiment. The full metadata information on the contents collected from each participant will be listed in the Appendix B. In order to focus on the current study, we have grayed out in Table 3.3 the information of individuals whose data are not processed and analyzed here.

Recruiting native speakers for recording was an important part of the data collection process. Since the recordings were carried out in the field, with the help of my Muong teacher, who is also speaker F1, it was not difficult to find and approach local people native to the area. Most of the participants (33 out of the total 35: 20+15, as explained above) are from the three villages of Kim Thuong, and there are two speakers who come from Xuan Dai, the commune next door and speak the same dialect.

Before carrying out the experiment on a large scale, I ran the test with my Muong teacher. This had two purposes: (i) actual testing to find out problems that need to be overcome to improve the feasibility and success of the experiment; (ii) allowing my teacher to find out all about the experiment, which she would then explain to other community members when searching for participants to the experiment. Recording speech and collecting an electroglottographic signal are unfamiliar undertakings for participants; being told about the experiment by someone who had already taken it was a major asset for convincing native speakers to participate in the experiment. Such a data collection campaign is, to the best of my knowledge, simply unprecedented in the area, and confirmation from my teacher (as a local person) of the feasibility, ease of implementation, and complete absence of illegal contents did wonders to persuade native speakers to participate.

Another advantage of F1's help in recruiting speakers was that the chosen speakers' dialect backgrounds were fairly homogeneous. She knew them well as members of the same rural community, and she selected proficient speakers, rather than the occasional newcomer to the community. Accordingly, it did not come as a surprise that all participants distinguished the five tones on smooth syllables and the two tones on stopped syllables that had been identified during the vocabulary list elicitation earlier in the fieldwork. This is crucial because the experimental setup was designed for this tonal system, and cannot be applied if a consultant has a different phonological system.

A further advantage of being based in the village is that it was not difficult to find consultants for the recording. Instead of randomly selecting available subjects, we were

able to take the initiative in selecting people who would be suitable for this recording and available for the task, by communicating with them and understanding how they would relate to the task. F1 and myself took care to avoid involving of consultants who would presumably have difficulty understanding the instructions; of consultants who are so eager to give advice that they become ‘experiment hijackers’, discussing the experiment instead of launching into it; and of consultants who have pronunciation peculiarities that make them strongly atypical. Thanks to her diplomatic skills, the screening process prevented awkward situations and saved face for all involved (an important cultural parameter throughout Asia).

On the other hand, recording at field location also implied a certain number of difficulties. Among these, the greatest difficulty (in my experience) was to find male consultants for recordings: this proved much harder than finding female consultants. Because of agricultural activities, local residents are constantly busy, especially men. The majority of men in the village are only at home twice a year, for the rice planting and rice harvest seasons. The remaining period of the year, they leave home to earn a living in the cities or to look after their family’s farm in the forest.<sup>3</sup> Moreover, a few male speakers turned down the proposal to participate in linguistic fieldwork because they did not like the notion of being recorded: lack of confidence and fear of ridicule combined led them to turn down the offer to take part in the experiment.

Another concern is that in addition to acoustic recording, we are simultaneously recording an electroglottographic signal, which is key to the research reported here. Based on the experience from previous research in 2016, it was clear that obtaining a crisp and clear electroglottographic signal would be difficult, as it is by no means as simple to obtain as a good audio signal (especially for female speakers). Due to the specific characteristics of the throat, the Adam’s apples of women are generally less visible than those of men, which creates a difficulty in correctly placing the electrodes. Moreover, the size of vocal folds and consequently the amount of variation in vocal fold contact area (monitored by the electroglottograph) is smaller in women. As a consequence, the electroglottographic signals obtained from women are frequently much weaker, with a lower signal-to-noise ratio, which makes them more difficult to analyze, to the point where some signals appear simply too noisy for reliable analysis. In order not to waste anyone’s time and effort, before giving the invitation to participate in the experiment, we invited native speakers to come and do a pretest. This was possible because the experiment took place directly in the field. The speakers came on foot, or by a short bike or motorcycle ride, on the day of the pretest, at a time convenient to them. If the test was successful, we conducted the experiment right after or later on the same day, when the recording environment was sufficiently quiet. There were no problems with long drives, or taking speakers to unfamiliar surroundings, as can happen when recordings are made far from home. However, there was a number of cases where the test signal was all right but the actual recording was not as expected.

---

<sup>3</sup>In Vietnamese, these forest farms are called *trang trại*: in addition to the main settlement, close to the paddy fields, families usually have a smaller farm higher up on the mountain, where they raise poultry and grow additional crops.

Hence the existence of participants for whom recordings were done but not used later down the line, as mentioned above and recapitulated in Table 3.3. In order to sidestep these problems in future studies, in addition to collecting basic information, I made private notes about technical glitches, aiming to uncover the reasons why recording sessions could fail to yield the desired data despite a successful pretest. These logs containing information about recording sessions as they unfolded from beginning to end proved useful for later interpretation of the data.

All in all, thanks to the kind help of local contacts (in particular the speaker F<sub>1</sub>), the data collection campaign was carried to completion. Total: 28 participants (18 women and 10 men), 29 files (the speaker M<sub>12</sub> have recorded twice), thus 20/29 files have been processed. The detail information could be found in Tables 3.3 and 3.7.

Table 3-3: A complete list of speakers / consultants who participated in the recordings from 2016-2019.

N <sup>o</sup>	ID	Name	Date of birth	Address	Gender	Rec. years	Note
1	F1	Sa Thị Đình	08.02.1983	Chiềng I	female	2016 2018	2015: EGG signal was OK but 2018: getting crackling noise
2	F2	Sa Thị Bích	15.02.1975	Chiềng I	female	2016	Weak EGG signal
3	F3	Sa Thị Đang	13.07.1984	Chiềng I	female	2016 2018	Good EGG signal
4	F4	Sa Thị Linh	14.12.1987	Chiềng I	female	2016	2Crakling noise weak EGG signal
5	F5	Trần Thị Thắm	10.12.1961	Chiềng I	female	2016	Weak EGG signal
6	F6	Phùng Thị Thanh	10.09.1956	Chiềng I	female	2016 2018	Weak EGG signal
7	F7	Sa Thị Thảo	12.03.1988	Xóm Xuân 2	female	2018	Speech ready gesture Weak EGG signal.
8	F8	Sa Thị Hứu	06.08.1973	Chiềng I	female	2018	A loud machine noise because the volume has been turned up too high. Strange EGG signal like F16.
9	F9	Hà Hồng Khanh	25.01.1991	Chiềng I	female	2018	Weak signal , up-and-down signal, but creak is clear. She make creak consistency with T4 (in continuous speech)
10	F10	Hà Thị Hòa	15.04.1973	Chiềng II	female	2018	Good EGG signal
11	F11	Sa Thị Hoài	27.08.1982	Chiềng I	female	2018	Weak EGG signal
12	F12	Trần Thị Thảo	15.08.1972	Chiềng II	female	2018	Good EGG signal, good creak.
13	F13	Hà Thị Coi	16.05.1967	Chiềng I	female	2018	Very good EGG signal. A talkative speaker. Can sing.
14	F14	Hà Thị Xoan	20.08.1965	Chiềng I	female	2018	She can sing folksongs and tell folklore stories

15	F15	Sa Thị Tiến	10.03.1943	Chiềng 2	female	2018	She knows a lot of folk songs, lullabies, but the songs (hát ví, hát đối đáp) mix a lot of Vietnamese.
16	F16	Phùng Thị Hoan	15.01.1956	Chiềng I	female	2019	EGG Signal is ok but look a bit strange
17	F17	Hà Thị Thàn	01.05.1955	Chiềng I	female	2019	Good EGG signal, good creak (different sub-types of creak: single pulse, multi pulse, glottal constriction).
18	F18	Hà Thị Thùy	16.10.1988	Chiềng I	female	2019	Weak EGG signal
19	F19	Hà Thị Nha	05.05.1975	Xóm Dù, Xuân Đài	female	2019	Good EGG signal (8-10 dots)
20	F20	Phùng Thị Xuân	22.02.1989	Chiềng I	female	2019	EGG signal is OK (5-7 dots), slightly going up and down.
21	F21	Phùng Thị Thành	15.02.1988	Chiềng I	female	2019	EGG signal is ok. No creak
22	M1	Hà Văn Chí	05.12.1979	Chiềng I	male	2016 2018	
23	M2	Sa Mạnh Hùng	25.08.1963	Chiềng I	male	2016	He died in 2018. Some speakers have recounted his death.
24	M3	Hà Văn Quyết	20.11.1977	Chiềng II	male	2016	Good EGG signal
25	M4	Đình Văn Mới	04.01.1971	Chiềng II	male	2016	Bad EGG signal
26	M5	Sa Văn Dẫn	05.11.1955	Chiềng I	male	2016 2018	Speech ready gesture. Lots of jitter
27	M6	Sa Mạnh Hồng	21.10.1971	Chiềng I	male	2016	Good EGG signal
28	M7	Hà Ngọc Sùu	18.02.1984	Xóm Nâu, Xuân Đài	male	2018	Good EGG signal
29	M8	Sa Mạnh Hoàng	04.06.1975	Chiềng I	male	2018	Some crackling noise but ok for processing
30	M9	Hà Công Quán	19.05.1990	Chiềng I	male	2018	Good EGG signal
31	M10	Xa Đình Lập	29.05.1988	Chiềng I	male	2018	
32	M11	Trần Văn Hoàn	01.07.1987	Chiềng III	male	2018	Ideal low and constricted creak
33	s M12	Hà Vi Thiển	15.10.1937	Chiềng I	male	2018 2019	Lots of jitter
34	M13	Sa Văn Bất	12.02.1962	Chiềng I	male	2018	His voice gives an impression that he speaks with constant creakiness
35	M14	Sa Văn Sơn	02.12.1959	Chiềng Is	male	2018	EGG signal is OK

### 3.2 *Experiment setup*

### 3.2.2 Equipment

For the recording of the experiment, this study used the Roland 4-channel recorder which enables up to four recording channels, thus the electroglottographic signal could be recorded simultaneously with the acoustic signals. This is a high quality recording solution that linguists can bring to their fieldwork locations.

To obtain electroglottographic signal, the model used was the Glottal enterprises EG2-PCX electroglottograph. This device boast a high signal-to-noise ratio based on the use of two electrodes on each side of the throat (Rothenberg, 1992). This appeared as a major advantage, since the analysis method used here (set out in Section 3.3.2) relies on the detection of peaks on the derivative of the electroglottographic signal (dEGG), which can easily get drowned in background noise. The device used in the present study has the same design as those used in studies by Nathalie Henrich (Henrich, 2001), Marc Brunelle and collaborators (Brunelle, Thành Tấn Tạ, et al., 2020). This somewhat facilitates comparison across studies, by removing a factor of difference (electroglottographic hardware).

The detailed description of these equipments with the list of accompanied devices, how to setup and use them, as well as some notes on the recording environment will be provided in Appendix B.

### 3.2.3 Recording procedures: training and performance

This part is to elaborate on all the steps that were taken to accumulate the data for this study. In writing these details, I have thought about and put myself in the context of a student of linguistics who would like to explore and study new languages and dialects but does not know how to begin. Therefore, a description of the process before, during and after the experimental recording will be detailed here with some notes on the specific case of an unwritten language.

#### Informed consent and payment

After finding a potential local native speaker, together with my Muong teacher (who guarantees reliability), I briefly explained the purpose for which I want to make the recording, and then invited them to my teacher's house. The devices were set up and tested carefully beforehand, otherwise they would be invited in half an hour after the invitation. Once they arrived, before going into the details of the experiment, I asked them to do a brief check with the EGG equipment because this signal is not always obtained as easily as the audio signal. Especially in the case of women when their adam's apple is small and not obvious like men. Learning from a few failed recordings with the EGG signal and in order to avoid wasting everyone's time and effort, checking the EGG signal was implemented as a first step.

After confirming the possibility of recording the EGG signal, the native speakers were officially engaged in the experiment. At this point, they were presented with two forms: (i) the informed consent form as in 3.3 and (ii) the attestation of payment as

in 3.4. Both forms are written in Vietnamese, so the speakers can read and verify what I present about my study and my purpose of the experiment and recording. They were informed that what they recorded that day was for my doctoral research on the Muong tone system and would be archived in the open library [Pangloss](#). The content of the recordings contains absolutely no sensitive information. They have full discretion to decide whether or not to allow publication of the contents of their recordings and photos, or only to allow it to be studied. All recruited speakers agreed and signed the informed consent to allow me to use their contribution for my study and also for publication on [Pangloss](#).

The certificate of payment was then shown to inform them that they will be paid for their participation. For each speaker participating in a 30-45 minute recording session (including training and performance), a remuneration of 100,000 VND (i.e. about 3.5 euros at the current exchange rate at the time), was paid. This amount is reasonable because it is higher than the average hourly income of a person in this region who is mainly engaged in agricultural work. They have been informed in advance of the amount they will receive after the recording session. This somehow encourages them to be enthusiastic and focused on the experiment and increases their tolerance to certain discomforts during the recording, such as having to wear the dual electrodes on the neck for about 20-30 minutes in the local heat of 30-35 degrees Celsius, 85-90% humidity. As soon as they have finished the recording session and signed the attestation of payment, they are paid in cash.

### Training

The objective of training was to enable the consultants to perform the experiment smoothly. The purpose of the experiment was not hidden from the consultants: I explained from the beginning that my study focuses mainly on the tones of Kim Thuong Muong, and that the most important thing I expect from them is an accurate pronunciation of the tones, so that the study can bring out the contrasts between the tones, and the specific phonetic and phonological characteristics of the Kim Thuong Muong tone system. The consultants who participated in this study did not previously have a metalinguistic awareness of Kim Thuong Muong tones: for instance, they did not know how many tones there were. On the other hand, they have an awareness of differences between Vietnamese and Muong tones, reflected in statements such as that Kim Thuong Muong “does not have tones” (“*không có dấu*”), an observation by speaker M6 which I see as suggesting that the tone systems of the two languages are to a large extent kept distinct.

Elicitation by means of photos appears to be the best choice for a non-written language like Muong, as it avoids the use of another language (which could introduce a bias due to interference between languages) and also avoids the use of writing, which also introduces biases of its own. Using photos requires some preparation. A total of 66 photos were carefully prepared in advance and purposely arranged in a digital slideshow to facilitate the training and performance of the experiment, as can be seen in

**GIẤY CHO PHÉP SỬ DỤNG VÀ CÔNG BỐ  
CƠ SỞ DỮ LIỆU TIẾNG NÓI VÀ HÌNH ẢNH**

Họ và tên người tham gia xây dựng cơ sở dữ liệu:  
.....

Địa chỉ: .....  
.....

Ngày sinh: ngày ..... tháng ..... năm .....

Tài liệu bao gồm:  tiếng nói  hình ảnh  loại khác: .....

với các nội dung như sau:

<input type="checkbox"/> thí nghiệm ngữ âm thực nghiệm	<input type="checkbox"/> chuyện kể
<input type="checkbox"/> thí nghiệm nội soi thanh quản	<input type="checkbox"/> bài hát
<input type="checkbox"/> hội thoại	<input type="checkbox"/> loại khác:.....
<input type="checkbox"/> vốn từ	<input type="checkbox"/> loại khác:.....

Trong khuôn khổ đề tài nghiên cứu tiến sỹ của NCS. NGUYỄN Thị Minh Châu, hiện đang học tập tại Đại học Sorbonne Nouvelle Paris 3 và Viện Nghiên cứu Lacito – UMR 7107, trực thuộc Trung tâm Nghiên cứu Khoa học Quốc gia Pháp (CNRS), tôi đã được mời tham gia để xây dựng cơ sở dữ liệu (bao gồm tiếng nói và hình ảnh) vào thời gian ngày ..... tháng ..... năm .....

Tôi đã được NCS. NGUYỄN Thị Minh Châu giới thiệu, giải thích rõ về mục đích và nhiệm vụ tham gia trong đề tài. Tôi hoàn toàn đồng ý cho NCS. NGUYỄN Thị Minh Châu sử dụng và công bố cơ sở dữ liệu tiếng nói và hình ảnh của cá nhân tôi, theo giấy phép Creative Commons BY-NC-SA 3.0<sup>1</sup>. Tôi cũng hoàn toàn đồng ý cho Viện Lacito – UMR7107/ CNRS công bố các tài liệu này ở thư viện Pangloss Collection<sup>2</sup>, kho lưu trữ (truy cập mở trực tuyến) tài liệu về các ngôn ngữ có nguy cơ tuyệt chủng và chưa được bảo tồn trên thế giới.

---

<sup>1</sup> Chi tiết về nội dung giấy phép Creative Commons Ghi nhận công của tác giả 3.0 Việt Nam tìm hiểu tại “[creativecommons.org](http://creativecommons.org)”

<sup>2</sup> Tìm hiểu về thư viện Pangloss Collection tại [http://lacito.vjf.cnrs.fr/pangloss/index\\_en.html](http://lacito.vjf.cnrs.fr/pangloss/index_en.html)

1

Figure 3.3: Informed consent form (Page 1)

Các thông tin cá nhân của người tham gia hoàn toàn được tôn trọng. Tại đây, người tham gia toàn quyền quyết định về việc có tiết lộ danh tính của mình cùng với cơ sở dữ liệu trong các công trình khoa học của NCS. NGUYỄN Thị Minh Châu và trên website của thư viện mở Pangloss Collection.

Cho phép tiết lộ danh tính

Không cho phép tiết lộ danh tính

....., ngày ..... tháng .....  
năm .....

Người tham gia ký xác nhận  
(Ký, ghi rõ họ tên)

2

Figure 3.3: Informed consent form (Page 2)

**GIẤY NHẬN TIỀN**  
**Attestation de paiement**

Họ và tên/ Nom et prénom / Family name and given name:  
.....

Địa chỉ / Adresse / Address: .....

.....

Đã nhận số tiền là / a reçu la somme de / received the sum of:  
.....

Bằng chữ / en toutes lettres / in full words:  
.....  
.....

Từ nghiên cứu sinh NGUYỄN Minh Châu (viện nghiên cứu Lacito)/ du doctorate  
Minh-Chau NGUYEN (Laboratoire Lacito)/ from PhD Minh-Chau NGUYEN  
(Laboratory Lacito), với lý do/ pour/ object:.....  
.....  
.....

Xác nhận của nghiên cứu sinh ..... , ngày..... tháng ..... năm 20...  
Confirmation du doctorant Fait à ..... le ..... / ...../20...  
Confirmation of PhD Chữ ký người nhận / Signature

Figure 3.4: Attestation of payment

a demo of the first minimal set in Figure 3.5<sup>4</sup> which is accompanied by an explanation in Table 3.4 to show how they were employed during the training and recording.

The first step in the training session is to guide the speaker to identify and become familiar with the illustrative photos that evoke the target words. Before the recording, consultants were trained. I explained photo by photo and made consultants remember which photo corresponds to which word, reminding them to say monosyllables (not disyllables), and bringing them into the imagined context that goes with the carrier sentence. Consultants were required to practise one or several times before recording: until they were able to go through the task with fluency.

It would be simplistic to expect that we would simply show the photos and the consultant would say exactly the desired word. However, as mentioned in Appendix A, not all the target words are easy to illustrate visually. There remain words that are difficult to show directly through photos, such as those in Figure A.6. In such cases, I clarified my purpose in choosing each ambiguous photo, making consultants understand and remember the indirect meaning of them. General slides as the one in Figure 3.5b were employed for this step. Since the photos of all five target syllables appear together, it helps speakers recognize the intended concepts more quickly, as they can realize that it is the same syllable but the tone is different.

A drawback of this method is that the consultants may focus on remembering (memory task). This can compete with the demands of good phonetic realization: it is hoped that the consultants can concentrate on clear phonetic realization of tone; if some items are more difficult to remember than others, a sudden remembrance after a moment of hesitation (which can be awkward or stressful for the consultant) may be accompanied by different phonetic realization. This is of course not to say that the task is beyond the ability of the speakers. After being guided through the list once or twice, item by item, they would be able to quickly remember the 66 target syllables. This experiment, in fact, did not cause great difficulty to consultants in understanding my explanation and performing as requested: the association between photos and desired target words, once explained to them, proved easy to remember. In addition, they were reassured that it is not an issue if they hesitate or make mistakes: that they would then be asked to repeat immediately, to get better results. Therefore, after training and rehearsal they were (by self-report) comfortable and (in my view) consistent in their tone realization during the actual performance.

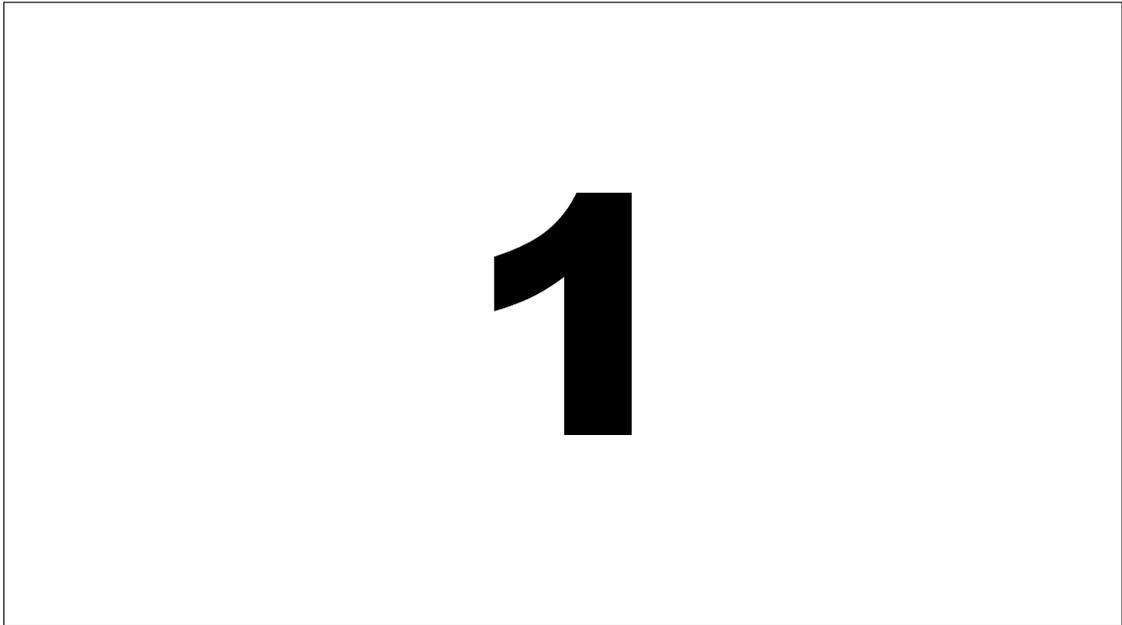
In general, it took about twenty minutes to explain the purpose and process of the experiment, and to train and rehearse.

### Performance

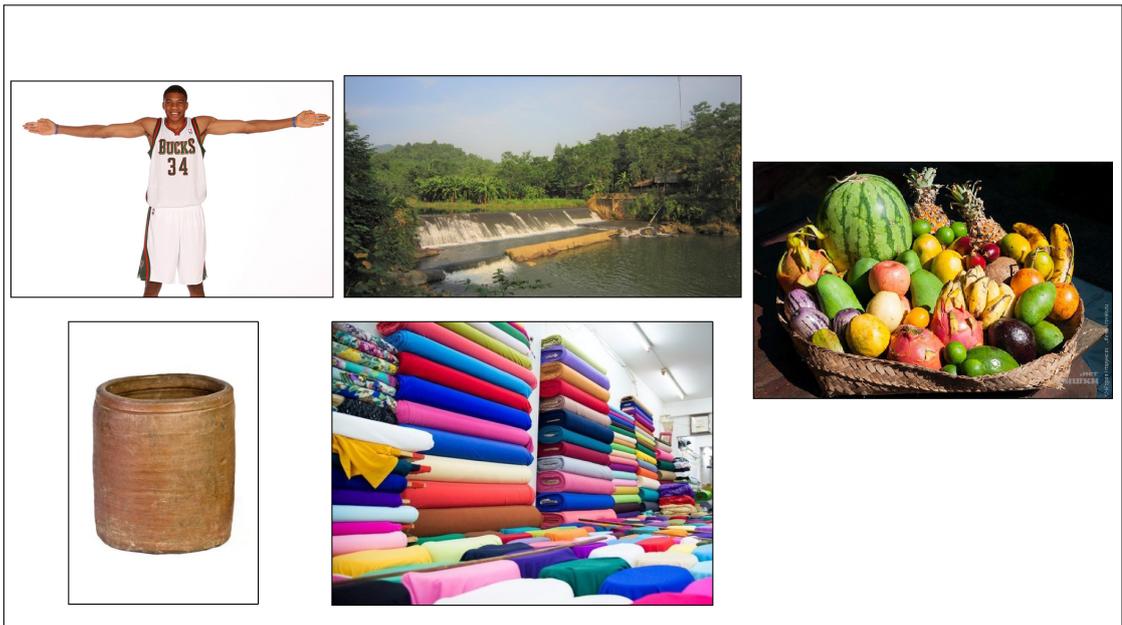
The recording took place right after training and rehearsal, so that the speaker's memory would be really fresh for the best performance. Recording proper usually took twenty five to thirty minutes: fifteen minutes for the minimal sets experiment (two times) and

---

<sup>4</sup>The entire slideshow could not be shown here due to the copyright of the photos on the internet. If anyone is interested, please contact me for a booklet.



(a) Slide 1: The start of the minimal set is displayed by a blank page with the digit of its order in the list, i.e. number 1 for the first minimal set.



(b) Slide 2: Full set of 5 photos in one slide.

Figure 3.5: A demo slideshow showing illustrative photos for the elicitation of the first minimal sets. This is accompanied by an explanation in Table 3.4.

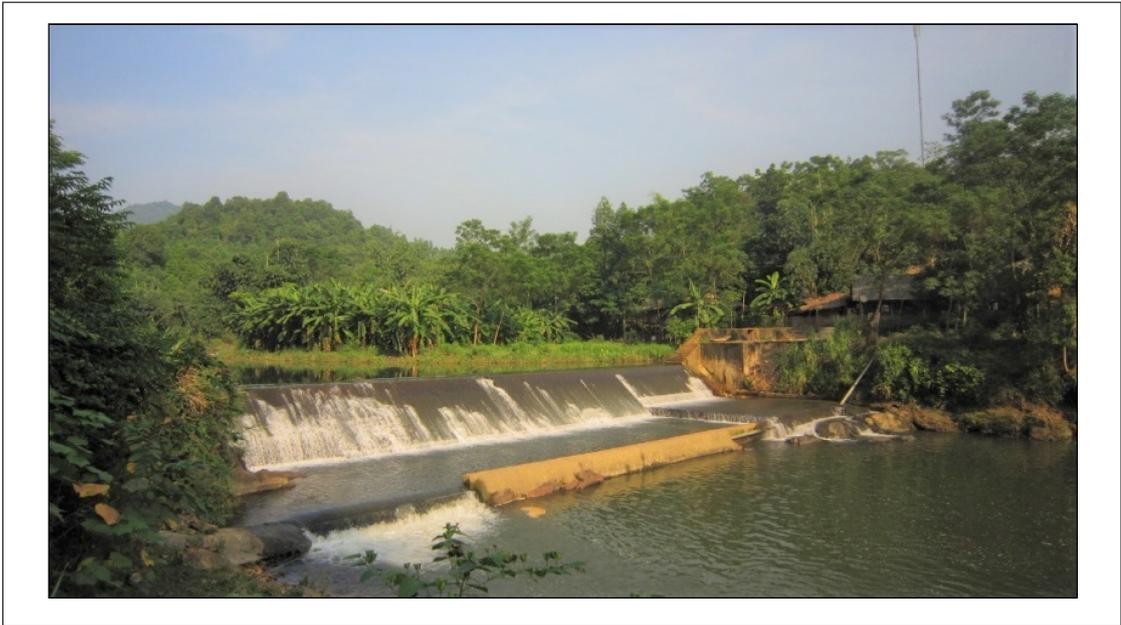


(c) Slide 3: illustrative photo of the first target syllable /paj<sup>5</sup>/ ‘arm span’

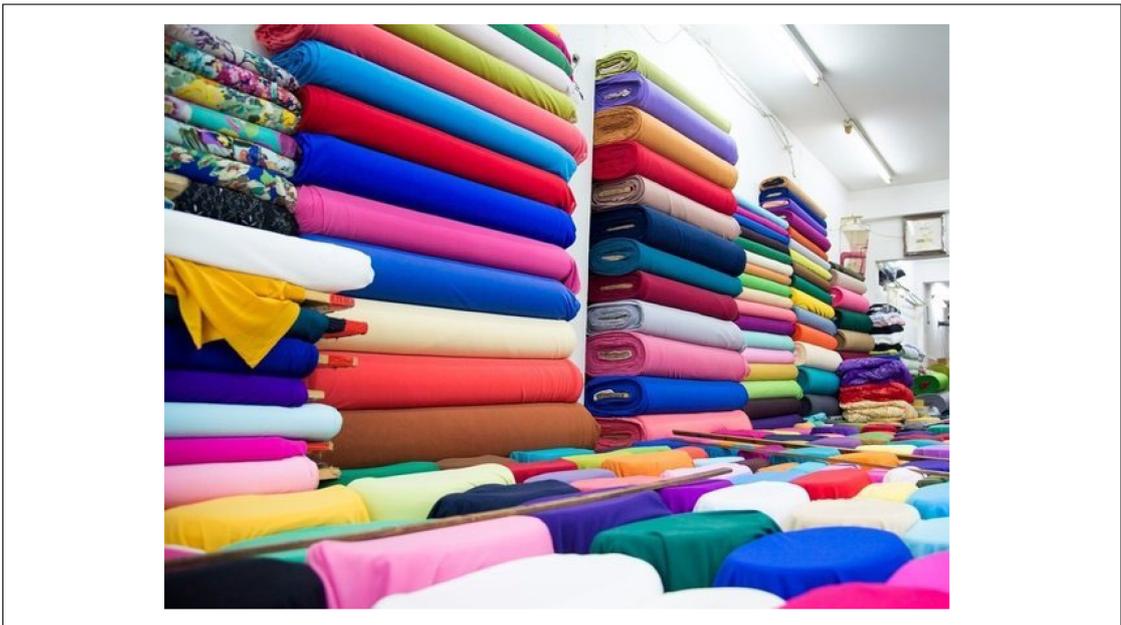


(d) Slide 4: illustrative photo of the second target syllable /paj<sup>3</sup>/ ‘cylindrical jar’

Figure 3.5: A demo slideshow showing illustrative photos for the elicitation of the first minimal set. This is accompanied by an explanation in Table 3.4.



(e) Slide 5: illustrative photo of the third target syllable /paj<sup>2</sup>/ ‘barrage’

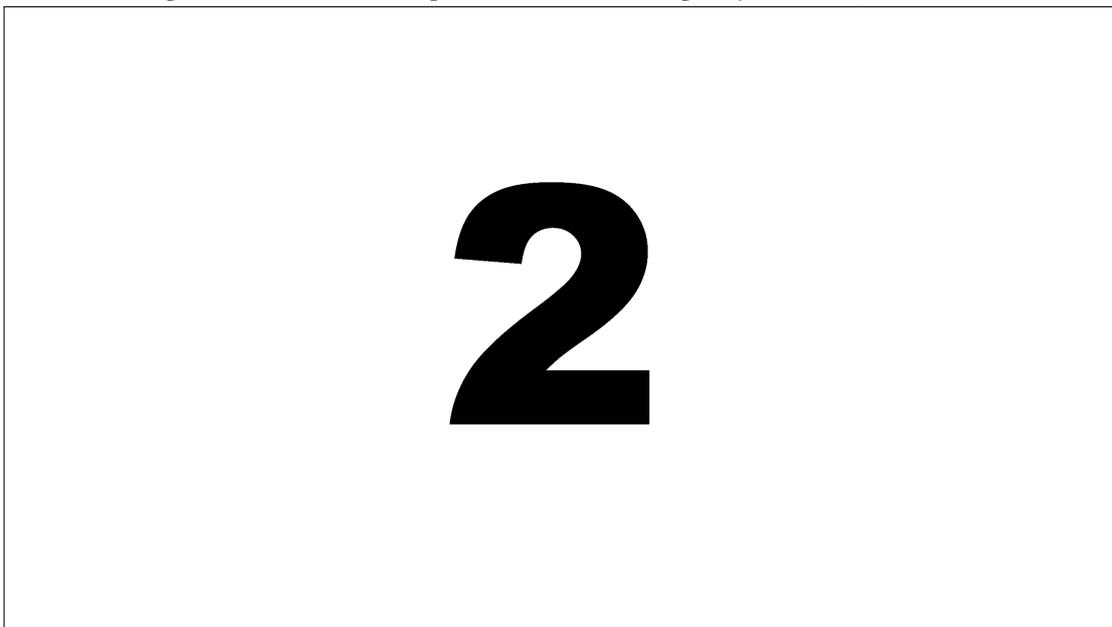


(f) Slide 6: illustrative photo of the fourth target syllable /paj<sup>1</sup>/ ‘cloth’

Figure 3.5: A demo slideshow showing illustrative photos for the elicitation of the first minimal sets. This is accompanied by an explanation in Table 3.4.



(g) Slide 7: illustrative photo of the fifth target syllable /paj<sup>4</sup>/ ‘fruit’



(h) Slide 8: a break page showing the digit 2 to mark the end of the first minimal set and the beginning of the second one.

Figure 3.5: A demo slideshow showing illustrative photos for the elicitation of the first minimal sets. This is accompanied by an explanation in Table 3.4.

Table 3.4: An explanation of how the slideshow of illustrative photos is employed at training and when performing the experiment.

Slide - Display	Purpose	At training	At performance
<p><b>Slides 1:</b> break slide (3.5a)</p> <p>A blank slide showing the order number of current minimal set</p>	<ul style="list-style-type: none"> <li>- Help speakers to be aware of the beginning of a minimal set.</li> <li>- A sign for separation of minimal sets in the audio files</li> </ul>	They are informed of the purpose of this page and are instructed to read the number displayed on the page during the rehearsal	The speakers read the order number when the break page is displayed
<p><b>Slide 2:</b> general slide (3.5b)</p> <p>Five photos are placed together</p>	<ul style="list-style-type: none"> <li>- Help speakers recognize the minimal set with five identical syllables but only different in tone.</li> <li>- In experiment, used for recording the minimal set in isolation</li> </ul>	Walk the speaker through each photo and explain the target word expected	Speakers say the expected words in isolation, following the researcher's prompt: pointing to each photo in turn
<p><b>Slides 3 - 7:</b> particular slides  (3.5c, 3.5d, 3.5e, 3.5f, 3.5g) Each slide presents a particular photo</p>	<ul style="list-style-type: none"> <li>- Draw the speaker's attention to each word and ask them to say the word in the carrier sentence.</li> <li>- In experiment, used for recording the minimal set in carrier sentence</li> </ul>	Let speakers remember the expected word in each photo. Instruct them to practise saying the word in the carrier sentence.	Speakers say the expected words in the carrier sentence
<p><b>Slide 8:</b> break slide (3.5h)</p> <p>A blank slide showing the order number of next minimal set</p>	<ul style="list-style-type: none"> <li>- Help speakers be aware of the switch to a new minimal set.</li> <li>- A sign for separation of minimal sets in the audio files</li> </ul>	They are informed of the purpose of this page and are instructed to read the number displayed on the page during the rehearsal	The speakers read the order number when the break page is displayed

five to ten minutes for recording some narratives. There were also two short breaks, one after each chunk of experiment performance. This way, the speaker could drink water, get a short rest after an intense moment, and I could check the condition of the electrodes, making sure they were not too wet from the speaker's sweat: drying them and reapplying the electrogel.

Thus, in general, the training and recording put together took about forty five minutes in total.

Due to the complexity of setting up the equipment, the training and recording always took place on the same day. In 2018, with the assistance of Thi-Hang Dinh (Đinh Thị Hằng), a colleague at the Institute of Linguistics in Hanoi, we efficiently conducted the experiment in one week with three to four native speakers per day.

The equipment was set up in the morning. Potential speakers were invited from the day before, or right in that morning. While I was checking the EGG signal obtained from the speaker(s), Dinh could help me train them. During the recording, I could focus on controlling the recorder and the EGG equipment, checking the sound during the session while Dinh showed the slides and pointed at the illustrative photos for the speakers to recognize the desired word as learned at training. Based on that lead, the speakers were required to speak aloud from minimal set N<sup>o</sup>1 to N<sup>o</sup>12 and from minimal pair N<sup>o</sup>1 to N<sup>o</sup>3. For each minimal set, they first spoke the five target syllables in isolation before repeating them once more in the carrier sentence. Exactly the same routine was performed with 3 minimal pairs. The slideshow of photos (as explained in Table 3.4) are organized for this purpose. The instruction with photos during experiment allows for (approximative) control on tempo, avoiding large changes in speech rate (speeding up or slowing down) that often happen in this type of tasks (for a review, see Niebuhr and Michaud, 2015).

A check-list was prepared so as to follow the same procedure consistently with all speakers. It is admittedly pedestrian, but it arguably saves a lot of trouble: having the sequence of basic gestures neatly arranged with no real risk of oversight frees the investigator's attention to focus on meaningful interaction with the consultants, without sudden bursts of stress in the midst of a recording session.

At the start of the session:

- Turn off mobile 'phones
- Unplug the electroglottograph from the power grid
- Put electrogel on the two electrodes of the electroglottograph
- Place the electrode collar on the participant's neck
- Take a picture of the speaker with full gear
- Turn on the electroglottograph
- Check the signals, using earphones
- Start recording (Turn on the recorder)

At the end of the session:

- Turn off the electroglottograph
- Stop the recording (Turn off the recorder)
- Take off the electrode collar and head-worn microphone

- Take a picture of the speaker’s throat: the slight marks made by the electrodes’ edges indicate their exact location
- Wipe the electrodes
- Turn on mobile ‘phones again

#### After a recording session

Once the recordings were completed, the speakers were asked to sign the payment attestation and were paid immediately thereafter.

The last step of data collection is to transfer the audio files from the recorder, saving them on several backups (one on my computer, one on a USB stick and one on a hard drive) to keep them in complete safety in the short run. In the field, backups were created on hard drives: there was not enough bandwidth over the mobile ‘phone network to sync gigabytes of data on remote servers. Back at LACITO, the data were stored on servers until the process of preparing annotations came to fruition, at which point the audio files were made public (in 2021). The electroglottographic files were not made public immediately, so as to preserve a window in time when I can continue to prepare publications on this data set. Interested colleagues are welcome to get in touch to set up collaborations on this data set. On the other hand, if the electroglottographic files were simply made available under a Creative Commons license, it would be perfectly lawful for colleagues to use the data without prior notice, only acknowledging the source of data, but it would be awkward for me. The electroglottographic files are already in the archive and making them public is now a matter of a few clicks for the archive managers. I am committed to taking this step without unreasonable delays.

To facilitate data management, a copy of the original files was created and renamed with the following file naming conventions: “crdo-MTQ\_KTM\_[SpeakerUID]\_[Content]”. All file names are in capital letters and the underscore is allowed as a separator. The **crdo-MTQ\_KTM** part is fixed. These identifiers follow conventions used in the [Pangloss Collection](#) (Michailovsky et al., 2014), an archive of language recordings in which the data is hosted. The identifier for each document begins with <crdo-> (the acronym of the former name of the Cocoon archive, which hosts the [Pangloss Collection](#): *Centre de Ressources pour la Documentation de l’Oral*) followed by the ISO code of the language in the Ethnologue catalogue of languages (the code for Muong is MTQ). KTM is the abbreviation for the target language, Kim Thuong Muong. The next two pieces of information are the speaker ID and the file content. The speaker ID can be found in the second column of Table 3.3. The file content is marked as “MINIMALSET” if it is the experiment recording, whereas if it is a narrative, a short keyword is used to reflect its broad topic. For instance, the data of speaker M1 was renamed from the original file as:

- *crdo-MTQ\_KTM\_M1\_MINIMALSET.WAV* for the file of experiment of minimal sets
- *crdo-MTQ\_KTM\_M1\_FOLK-GAMES* for the file of narratives about folk games
- *crdo-MTQ\_KTM\_M1M7\_CUSTOMS* for the file of a dialogue between speaker M1 and M7 talking about some traditional Muong customs

These renamed files are the files that are used further down the line for all processing

and analysis operations. Needless to say, the original (master) files were retained for reference, without any processing after recording.

Then, the information of all the files were stored in a metadata file. This key step allows for relating relevant pieces of information together: the file name (modified name and original name), speaker information, recording time, duration, storage location, file quality, as well as various notes.

As emphasized relentlessly by archive curators and information management specialists, rich metadata are crucial to keep the data well-organized. The more information we can store immediately after recording, the easier it is for us to organize and work with the data later, because it is much easier to do this in the field than to bring data back to the lab and try to retrieve metadata from memory. Accordingly, the perspective of data preservation and dissemination was taken into account from the earliest stages. Metadata were gathered at the stage of data collection: writing down information that goes without saying at the time, but which may become difficult to access (or lost altogether) if not put to writing in good order. Data management was handled in an Open Science spirit, keeping in mind the goal of practical usefulness of the data to other users in future.

### 3.3 Data processing

Figure 3.6 shows the basic procedure that was applied to the data set of twenty speakers to obtain quantified results: values of  $f_0$  and  $O_q$ , as plotted in numerous graphs in Result chapter.

For the most part, the initial input materials for this process are audio files obtained from the minimal set experiment. They are first segmented and annotated (using the `SOUND FORGE` software) to obtain the annotated electroglottographic file in mono-channel format and the Regions List indicating the time codes for each token, together with its unique identifier (UID). These are the two inputs required by `PEAKDET`, a semi-automatic tool for estimating  $f_0$  `dEGG` and  $O_q$  `dEGG` from the electroglottographic signals. After a meticulous (and time-consuming) verification process, summarized in the algorithm shown as Figure 3.13, the data matrix from `PEAKDET` is used to plot figures (by means of scripts) to visualize the result. This process will be elaborated step by step in the following sections.

In my experience, during the long process of extracting parameters from the electroglottographic files, it is most practical for all the files to be put in one place, with an online file storing service, to ensure that files are kept safe while being processed and can be reverted back to the previous version at any time in case of an incidental error. This prevents unfortunate computer crashes or artifact errors during the analysis process. Otherwise, such glitches can be very costly in time and effort.

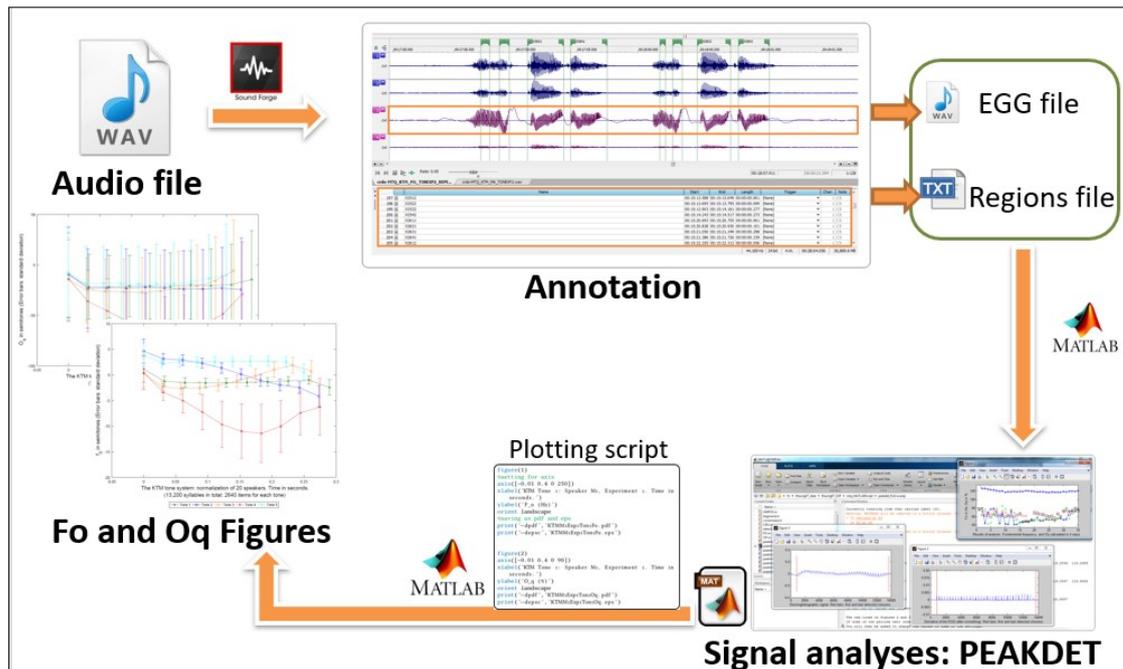


Figure 3.6: Basic procedure of data processing.

### 3.3.1 Data segmentation and annotation with Sound Forge

The first step in the process is to take the data (with the modified name) from each speaker to segment each word (both the target word and the frame word) into tokens and annotate them with a unique identifier (or UID), so that they can be analyzed in the next step, item by item. This step is carried out on SOUND FORGE, a professional digital audio editing software. The version used is SOUND FORGE10.

The bit-depth of the recordings is 24-bit, which allows for recording at a safe recording level, leaving headroom for sudden increases in voice intensity that may happen at any point during speech. Digital volume amplification was then conducted: the volume of the audio channel was normalized only once for each session, to preserve comparability across items within the same session. (The magnitude of amplification was on the order of 6 to 18 dB depending on sessions.) Normalization was conducted using SOUND FORGE, before data annotation. The audio was not subjected to any other processing, such as filtering.

**The general principle of segmentation: based on electroglottographic signal and audio signal**

Using a 4-channel recorder yields files that are not in the stereo format, which (at the time of writing) is most widespread in digital audio. Each audio files has four channels: (i) acoustic signal of native speaker from headset microphone, (ii) acoustic

signal of native speaker from regular microphone (referred to colloquially as a “table mic” because it can be placed on a table), (iii) electroglottographic signal of native speaker, (iv) acoustic signal of interviewer (instructor or cooperator). The four-channel signal can be handled without difficulty with `SOUND FORGE`.

The audio signal is not the subject of data analysis and parameter calculations in this study, but it was used as the main cue for segmentation since the transitions in the electroglottographic signal are not as clear-cut as those in the audio signal: as a sweeping statement, it can be said that segmental events do not appear as such in the electroglottographic signal, which is recorded at the voice source (the larynx) prior to filtering by the vocal tract. For segmentation, the audio signal offers key evidence to recognize the boundary and transition between the initial consonant and the rhyme, especially based on the shape change between periods of the audio signal.

The three most fundamental principles of segmentation are:

1. (Figure 3.7): Segment the whole syllable if the initial consonant is voiceless and take only the rhyme if the initial consonant is voiced (Michaud and Kühnert, 2006)
2. (Figure 3.8): The segmentation of a token must have a minimum of three cycles, otherwise the token is considered to be absent as a distinct interval.
3. (Figure 3.9): In cases where the rhyme is completely missing, it cannot be segmented. It is then marked (as a time marker on `SOUND FORGE`, rather than as an interval; it appears as a single orange vertical line) at the beginning of the missing syllable. It is then stored in the Regions List, but not processed by `PEAKDET`.
4. (Figure 3.10): Phonologically voiceless consonants that are phonetically voiced (sonorized) are treated as voiced consonants, i.e. the voiced portion corresponding to the initial is excluded at the stage of annotation, so that only the rhyme part is taken.

In detail, the segmentation and annotation task consisted in indicating the temporal beginning and ending point of each syllable rhyme, for the target syllables and also for each syllable of the carrier sentence. Tone is non-segmental and it normally stretches over all of the syllable’s rhyme part. Therefore, to measure glottal parameters relative to tone, one needs to annotate the rhyme part, excluding the initial consonant if it is voiced. One can take in the whole syllable if the initial consonant is (phonetically) voiceless. Examples are shown in Figure 3.7.

In the case of a voiced consonant, at the boundary between the consonant and the vowel, it is recommended to exclude the first two or three cycles of the vowel to be sure to take only the rhyme. This is to say that the indication of the beginning was placed a little later than where the transition can be recognized. My experience confirms that it is more difficult to segment the rhyme of a voiced fricative consonant (e.g. /**ja**2/), and a trilled consonant (e.g. /**r**ɔ/), than a voiced nasal consonant (e.g. /**ma**/).

For tone analysis, a rhyme part should have at least three glottal cycles (as Figure 3.8). In fact, one cycle (i.e. the interval between two closing peaks on the derivative of

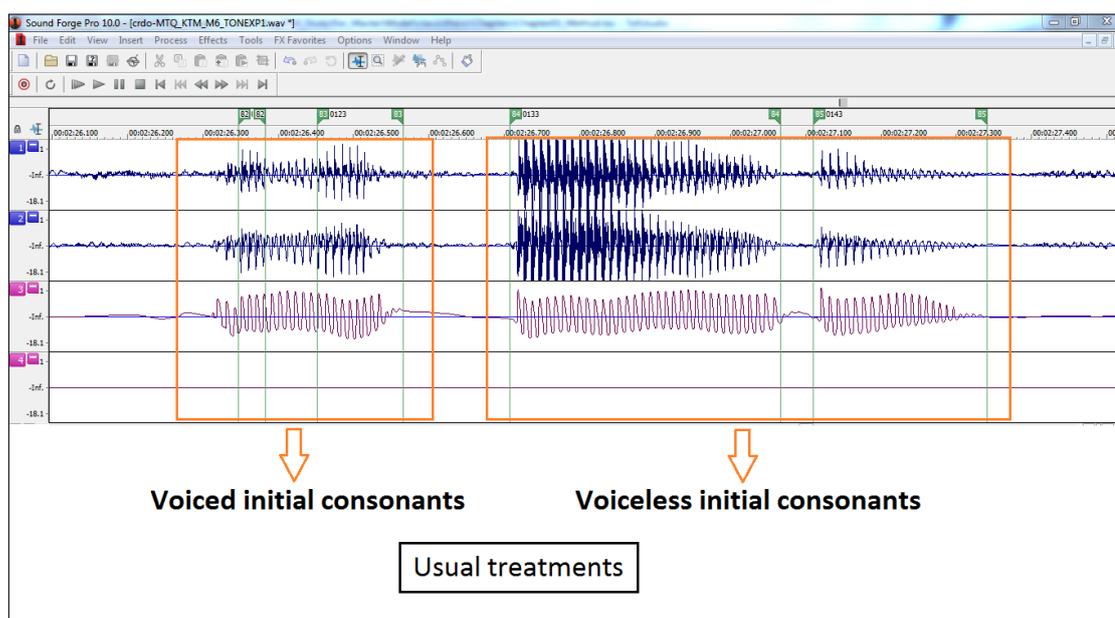


Figure 3.7: Segmentation principle: examples of different treatments for voiced and unvoiced initial consonants.

the electroglottographic signal, as explained in section 2.4) is sufficient for calculating one  $f_{0 \text{ dEGG}}$  value, making for well-formed results. However, the decision to choose a minimum of three cycles appeared safer because data processing involves averaging across tokens. Averaging tokens with just one or two cycles with others that have more than ten times more appeared as a possible cause of interpretive conundrums down the line. Since the emphasis of the present study is on the modulation of the  $f_{0 \text{ dEGG}}$  (and other parameters) in the course of tones, it appeared safer to reserve the study of borderline cases of elided or quasi-elided syllables for later.

During annotation, there were two special treatments for some uncommon cases, as illustrated on Figures 3.9 and 3.10.

(i) In a few cases the rhyme at issue is not present at all as a distinct temporal interval, so that it is not possible to assign it a beginning and endpoint – there is no identifiable portion of signal (not even one cycle) corresponding to the rhyme at issue. For this case, instead of a *Region*, I just place a *Marker* in SOUND FORGE, with the mention “*missing syllable*”, as in Figure 3.9. The item is then deleted (manually) when exporting the Regions List from SOUND FORGE for processing of the electroglottographic signal. (More details on the Regions List and how it is used further down the processing line will be presented in section 3.3.2.)

(ii) Because of coarticulation and hypoarticulation, there are a few cases in which a voiceless initial consonant has its manner of articulation modified: loss of closure for stops, realized phonetically as weak voiced fricatives (spirants in the sense of André Martinet, 1981). This phonetic phenomenon is called spirantization, a tendency that had

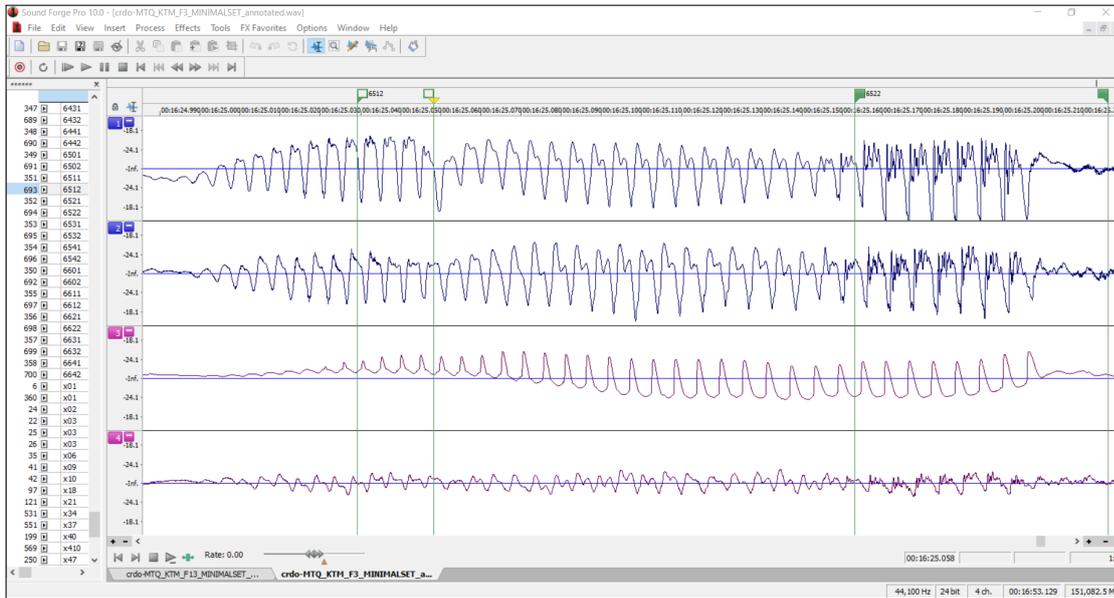


Figure 3.8: Segmentation principle: minimum span of three cycles in token 6512 (syllable /ja<sup>2</sup>/), data from speaker F3.

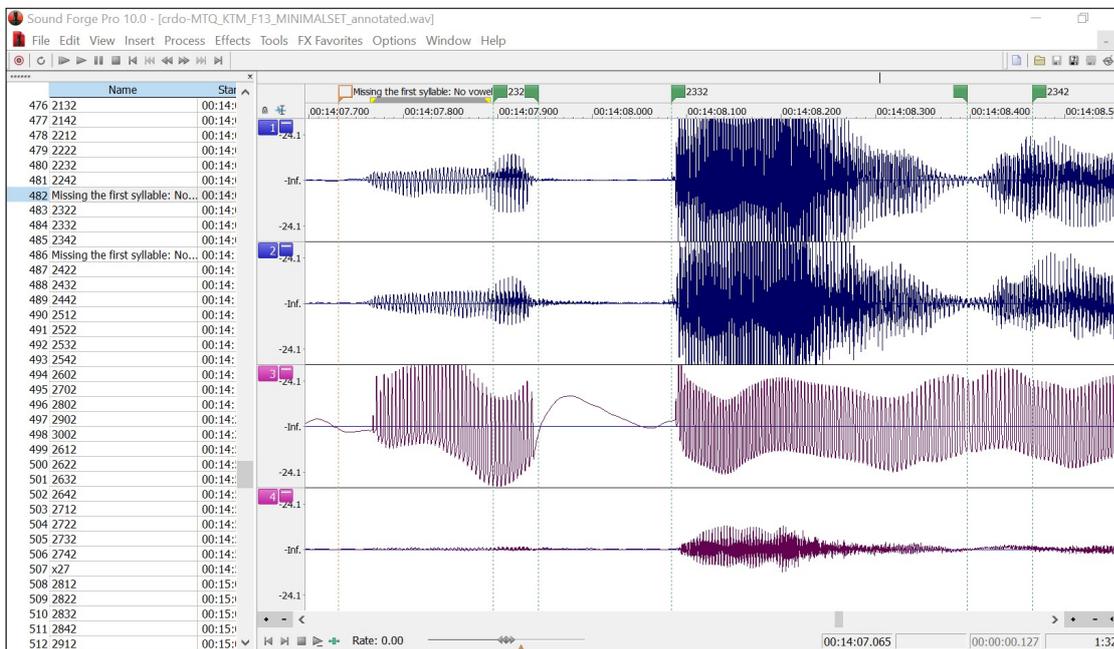


Figure 3.9: Segmentation principle: treatment of missing syllables/rhymes

a strong influence in the diachronic development of the Vietnamese consonant system Ferlus, 1982. This is not an issue for annotation: in that case, I just handle the rhyme the usual way, excluding the initial consonant, as in Figure 3.10b.

These phenomena do not have a direct bearing on the target items: they only occur on frame syllables, not on the target syllable, which is generally hyperarticulated rather than hypoarticulated. For instance, issue (i) only occurred on the first syllable /ja2/ when the speaker tends to shorten the carrier sentence during a series of repetitions; and issue (ii) only occurred to the last syllable /tǎŋ/ because it is the end of sentence, so these syllables tend to be pronounced voiced.

At the same time as the beginning and endpoint of each rhyme was indicated, I also labeled every token. Labels are strings of numbers that consist in the concatenation of three part, referred to as AA,<sup>5</sup> B and C.

Position AA is the UID of the target syllable as provided in second column of Tabs. 3.2 and 3.1. It is, in fact, the order in which the target syllables appear in the list of materials. Therefore, position AA runs (i) from 01 to 40 for target from eight minimal sets, (ii) from 41 to 60 for four near – minimal sets, and (iii) from 61 to 66 for three checked minimal pairs.

Position B relates to the position of the syllable rhyme within the sentence. Each sentence has four syllables, hence: value 1 means first token /ja<sup>2</sup>/ (2SG), value 2 means second token /mǎt<sup>6</sup>/ ‘to know’, value 3 means third token “target syllable”, and value 4 means fourth token /tǎŋ<sup>3</sup>/ (INTERROG). Besides, the value 0 is used to label the target syllable in isolation.

The last value in the scheme, position C, indicates the repetition of the token. As this experiment is performed twice, the value 1 indicates the first time, and the value 2 is for the second time.

A summary of the scheme is provided in Table 3.5. The precision of the annotation is key to precision in the quantitative values obtained at the next step: electroglottographic analysis, to which we now proceed.

In addition, there are two additional conventions:

1. xAA: good token but out of place since there was more repetition than necessary.
2. z: problematical token with disfluency, hesitation, background noise, overlapping speakers, etc.

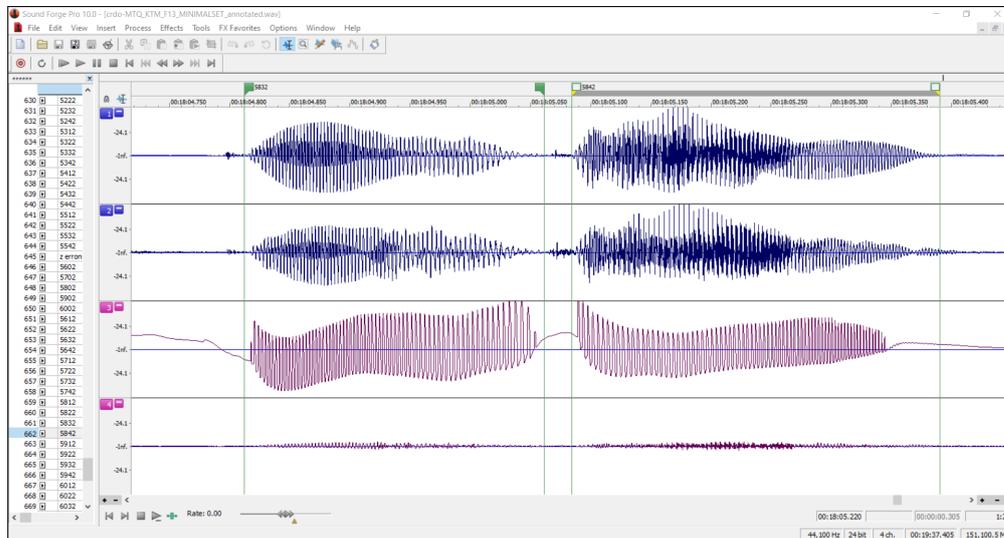
These annotations are not used for the next step of data processing with PEAKDET, thus they are deleted from the Regions List like the case of missing word. Later, in the data archiving work, they are all stored in the XML file in order to give as detailed as possible of the actual state of the corpus.

By understanding these labeling conventions, we can easily retrieve the content and information of any UID token in the data. For example:

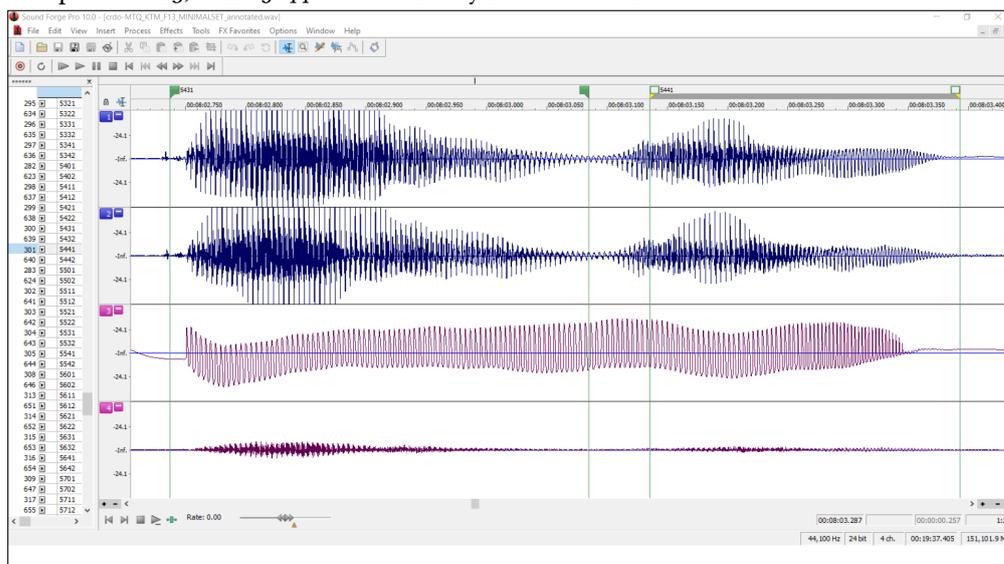
- Item with UID 3632 means: the target syllable /ma5/ in carrier sentence, second time.

---

<sup>5</sup>AA means that this part always contains two digits. The digits from 1 to 9 are padded with a zero for a systematic annotation scheme that always contains 4 digits.



(a) Usual treatment on the voiceless initial consonant on the syllable /tǎŋ<sup>3</sup>/, data from speaker F13, UID 5844: the entire syllable is taken.



(b) When the voiceless consonant becomes voiced on the same syllable /tǎŋ<sup>3</sup>/ (the same speaker, F13, but another token, UID 5441): only the rhyme is taken.

Figure 3.10: Segmentation principle: treatment of the initial voiceless consonant which becomes a voiced consonant under coarticulation and hyparticulation.

Table 3.5: Annotation scheme: four digits. Structure: AABC, detailed below.

Pos.	Meaning	Number code
AA	The UID of the target syllable, in two digits. Numbers from 1 to 9 are padded up with a zero.	01–40: words of 8 minimal sets 41–60: words of 4 near-minimal sets 61–66: words of 3 checked minimal pairs
B	The nature of the annotated rhyme: B = 0: target syllable said in isolation; 1 ≤ B ≤ 4: token from carrier sentence. B indicates the order of the token inside carrier sentence, 3 means target syllable.	0: target syllable (in isolation) 1: 1st syllable of frame - /ja2/ 2: 2nd syllable of frame - /mät6/ 3: 3rd syllable of frame - target syllable (in carried sentence) 4: 4th syllable of frame - /täŋ3/
C	Repetition	1: first time, 2: second time

- Item with UID 3611 means: the first syllable /ja2/ of carrier sentence in case /ma5/ is target syllable, first time.
- Item with UID 6801 means: the check syllable /kap7/ in isolation, first time.
- etc.

We can also know the tone of the target syllable based on the number in position AA :

- **Tone 1:** 04 09 14 19 24 29 34 39 44 49 54 59;
- **Tone 2:** 03 08 13 18 23 28 33 38 43 48 53 58;
- **Tone 3:** 02 07 12 17 22 27 32 37 42 47 52 57;
- **Tone 4:** 05 10 15 20 25 30 35 40 45 50 55 60;
- **Tone 5:** 01 06 11 16 21 26 31 36 41 46 51 56;
- **Tone 6:** 61 63 65;
- **Tone 7:** 62 64 66;

For each speaker's data, a total of 660 tokens (full size, without missing syllables) are segmented and labeled. The exact status is shown in Table 3.7. The last step that needs to be carried out on SOUND FORGE is the extraction of two files after the annotation is accomplished: (i) the electroglottographic signal (on the third channel) as a mono .wav file, (ii) the Regions List as a .txt file (as Figure 3.12).

The electroglottographic signal can be extracted easily on SOUND FORGE by simply selecting the desired channel, pressing the key combination **ctrl+C** to copy it and then **ctrl+E** to paste it into a new file, and finally saving it (**ctrl+S**) as a .wav file in the input folder for the next processing with **PEAKDET**.

In the mono annotated electroglottographic file just obtained (or it is the same in the 4-channel annotated file), I continue extract the Regions List by selecting in the SOUND FORGE toolbar (as displayed in Figure 3.12): **Edit > Regions List > Copy onto Clipboard**. Then, using Notepad++ to paste it and save the file as a .txt file in

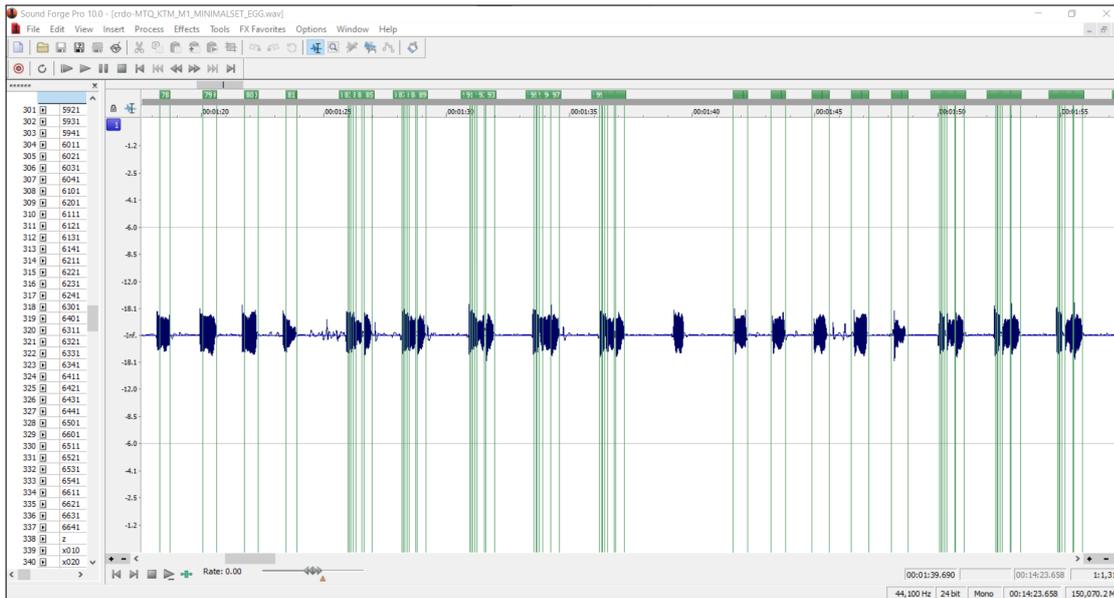


Figure 3.11: Mono electroglottographic .wav file: the second input of **PEAKDET**

the same folder where the electroglottographic file has been stored . This is also the time to remove from the Regions List all annotations for missing words, out-of-place words and problematical words that were mentioned earlier. They are always preserved in the original annotated .wav files for later archiving.

### 3.3.2 Analyzing the electroglottographic signal: how the **PeakDet** script was applied

After annotating all 20 speakers' data, the next step is to analyze the electroglottographic signal using the **PEAKDET** tool in the MATLAB computing environment to estimate  $f_{0 \text{ dEGG}}$ ,  $O_{q \text{ dEGG}}$  (by four methods), and several other parameters as detailed in Table 3.6.

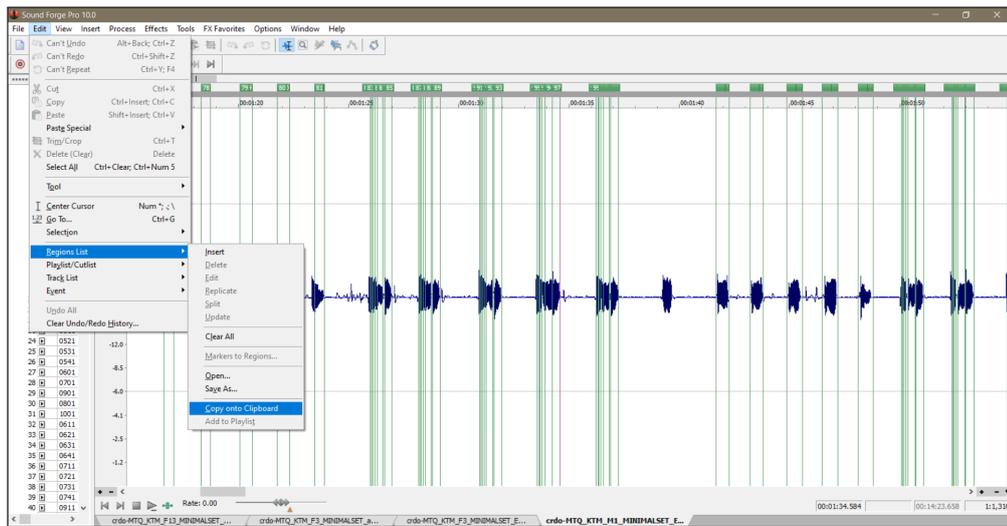
**PeakDet**: an implementation of an analysis method based on peak detection

There exist several scripts for analysis of the electroglottographic signal. That created by Nathalie Henrich (Henrich et al., 2004) is designed for the singing voice. It is based on autocorrelation, and is not designed to handle portions of voicing that are close to the onset or offset of voicing. In speech, however, voicing is often interrupted and resumed. This difference in the input signals led linguists to write algorithms that are suitable for tracking glottal parameters in spoken language.

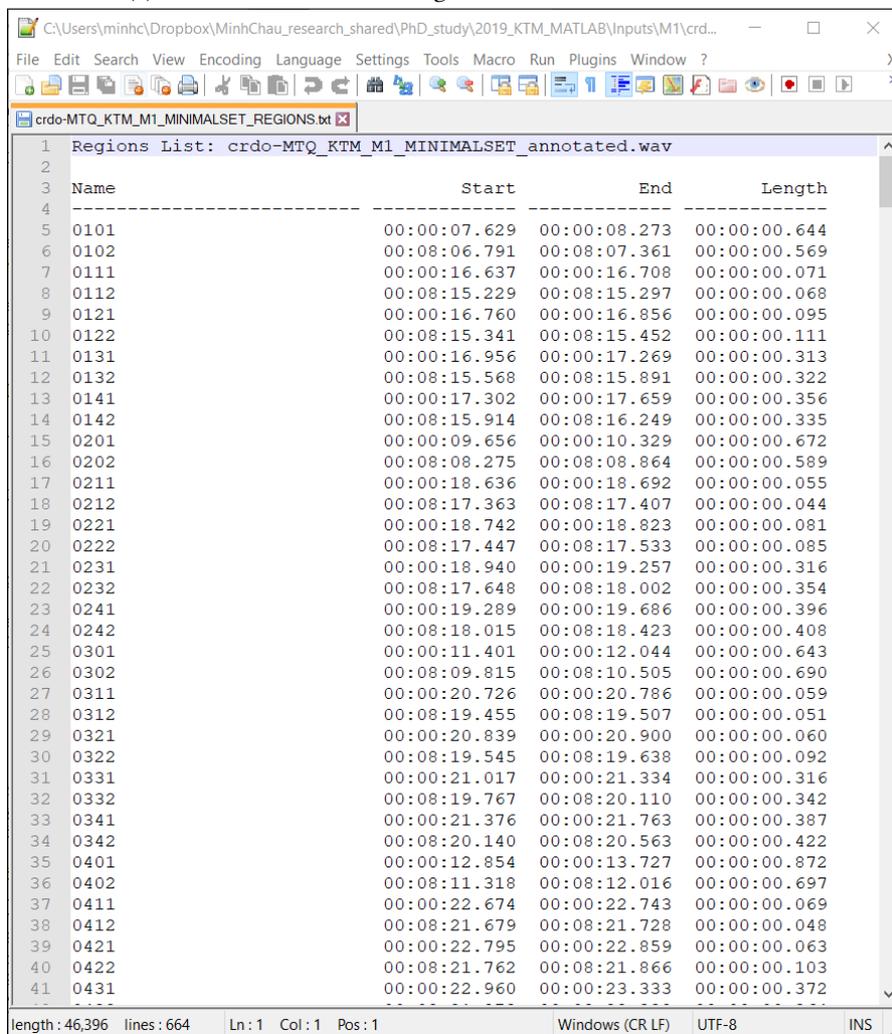
Peak Detection (henceforth **PEAKDET**) is a MATLAB<sup>6</sup> script for calculating  $f_{0 \text{ dEGG}}$

<sup>6</sup>MATLAB is a technical computing language and interactive environment for algorithm development,

Chapter 3 Method



(a) How to extract the Regions List on SOUND FORGE



(b) The Regions List is copied and save using Notepad++

and  $O_q$  from electroglottographic signals. It is available online from the website of *Tools for Electroglottographic Analysis: Software, Documentation and Databases* (link: <http://voiceresearch.free.fr/egg/index.html>). It is also available from COVAREP, a Cooperative Voice Analysis Repository for Speech Technologies (link: <http://covarep.github.io/covarep/>). COVAREP is an open-source repository of advanced speech processing algorithms hosted on GitHub: researchers in speech processing can store original implementations of published algorithms on COVAREP Degottex et al., 2014. The latest version is provided in a Github repository (link: [https://github.com/MinhChauNGUYEN/egg/tree/master/peakdet\\_inter](https://github.com/MinhChauNGUYEN/egg/tree/master/peakdet_inter)). The advantage of this repository on Github is that one can easily download the script, use it under the LGPL-3.0 license. Then, during the process of applying the tool, if there are difficulties or suggestions for improvement, it is possible to “open an Issue” on Github to communicate directly with the author. For instance, during this study, I created a bunch of Issues to suggest improvements to **PEAKDET** (as can be seen by following this link: <https://github.com/alexis-michaud/egg/issues>). Having privileged access to the code maintainer (being my supervisor) was very helpful in solving some of the practical issues that arose when applying this tool to my data. Specifically, the addition of a display of glottal cycle numbers in the dEGG signal figure was a great improvement in facilitating the verification of closing and opening glottal cycles. This improvement saved me a lot of time in processing the data with **PEAKDET**. More details on this topic will be provided later in this section.

**PEAKDET** is designed for semi-automatic measurement: the results for each token are verified visually, and some parameters can be modified to adjust to the input signal.

**PEAKDET** shows the position of the first and last detected glottis-closure-instants on the electroglottographic signal and on its derivative, so that the user can appraise visually whether the full interval of voicing has been taken into account or not. This is useful in cases where the amplitude of the signal varies considerably within the portion of signal under analysis. (Excerpt from the **PEAKDET** user guide.)

This current study used the **PEAKDET** tool for analyzing the electroglottographic signal in order to study Kim Thuong Muong’s tone system. The fact that the estimation of glottal parameters is not 100% automatic makes it all the more important to be as explicit as possible about decisions that I made in the process of semi-automatic processing. Hence, the paragraphs that follow provide a step-by-step description of how I applied the **PEAKDET** script for my data, what improvements were made to **PEAKDET** as compared to the version that I used for my Master’s work, as well as how I dealt with some particular, unusual issues that arose when dealing with certain

---

data visualization, data analysis, and numerical integration (Website: <http://www.mathworks.com/products/matlab/>). It comes complete with all the bells and whistles that one would expect from a full-fledged computing environment, in particular a comprehensive and convenient documentation manual. The **PEAKDET** script worked well with MATLAB 2014a, which I used in my study. Other versions, both earlier and later, have not yet been verified.

recordings. Hopefully, this piece will be useful to beginners to get some information on the experience of applying `PEAKDET` to a given data set.

### 3.3.3 Step-by-step description of the process of analyzing the electroglottographic signal with `PeakDet`

Use of `PEAKDET` assumes minimum knowledge of MATLAB. The script does not have a graphic user interface: it requires some interaction using the MATLAB command line, such as opening and running a script (.m file), setting the path to the right folders, and saving the work in progress. Elementary knowledge of this type is sufficient for a beginner to be able to apply `PEAKDET` without going into how the script works. This is the stage I was at for my Master's study, and it proved manageable. In the course of the present thesis work, I started to be able to understand the script: to follow how `PEAKDET` actually operates, step by step. It is much better thus, not only because I can now debug some simple problems by myself, but also because I now understand clearly how `PEAKDET` encounters some problems that are related to the characteristics of the electroglottographic signal. Then I can describe the issue on Github and ask for a solution. Therefore, it is highly recommended that the user has a basic knowledge of the MATLAB language.

The diagram in Figure 3.13 summarizes the steps of the implementation of `PEAKDET`, dividing them into four basic steps: (i) indicating the path to the two input files, namely the electroglottographic signal and the time codes in a text file (a Regions List) exported from `SOUND FORGE`; (ii) conducting a verification of fundamental frequency values,  $f_{o\_dEGG}$ ; (iii) conducting a verification of open quotient values,  $O_{q\_dEGG}$ ; (iv) saving the output. The first and fourth actions are executed only once, respectively at the beginning and at the end of the process of analyzing one electroglottographic recording. The second and third actions are repeated in a loop: once for each token in the file. Such are the main operations in `PEAKDET`, which have not changed across versions. These steps are described one by one in the paragraphs that follow.

As a preparation, a prerequisite is to prepare the two input files and to download `PEAKDET` on Github, i.e. `Peakdet_inter` (from here: <https://github.com/alexis-michaud/egg>).

#### Importing the data

The first thing to do is, of course, to open MATLAB, then run `PeakDet_inter`. There are two ways to run `PeakDet_inter`.

1. In the `CURRENT FOLDER WINDOW`, click on the folder icon with a green downward arrow like the red rectangle numbered 1 in Figure 3.14. A new window requesting "Select a new folder" will appear so that the user can navigate to the location of the `PeakDet_inter` folder. Once the `PeakDet_inter` folder is selected, all the files in that folder, including the file `peakdet_inter.m`, will appear in the `CURRENT FOLDER WINDOW` (as the red rectangle numbered 2) and the address leading to that folder will be updated in the directory path (as the red

**Algorithms of peakdet process:**

Set up the values for multiple closing peaks, threshold, and DEGG smoothing

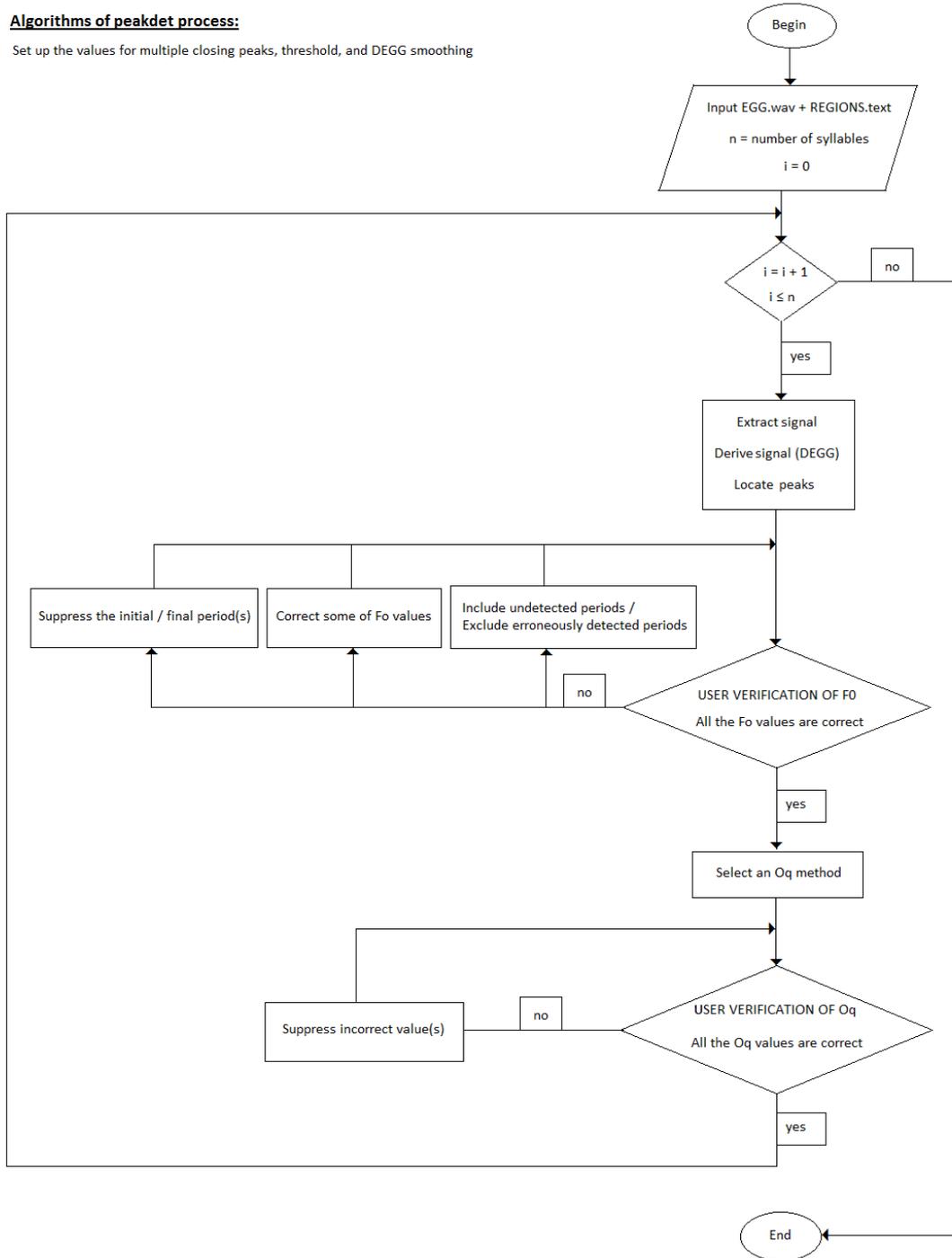


Figure 3.13: A schematic representation of data processing with **PEAKDET**.

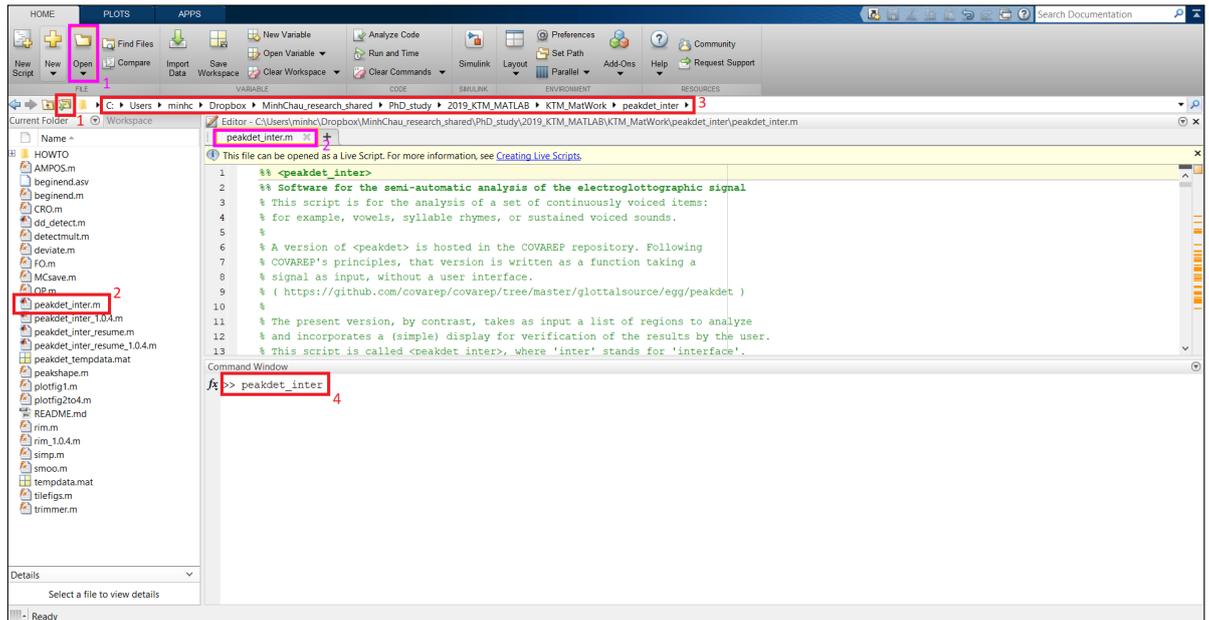


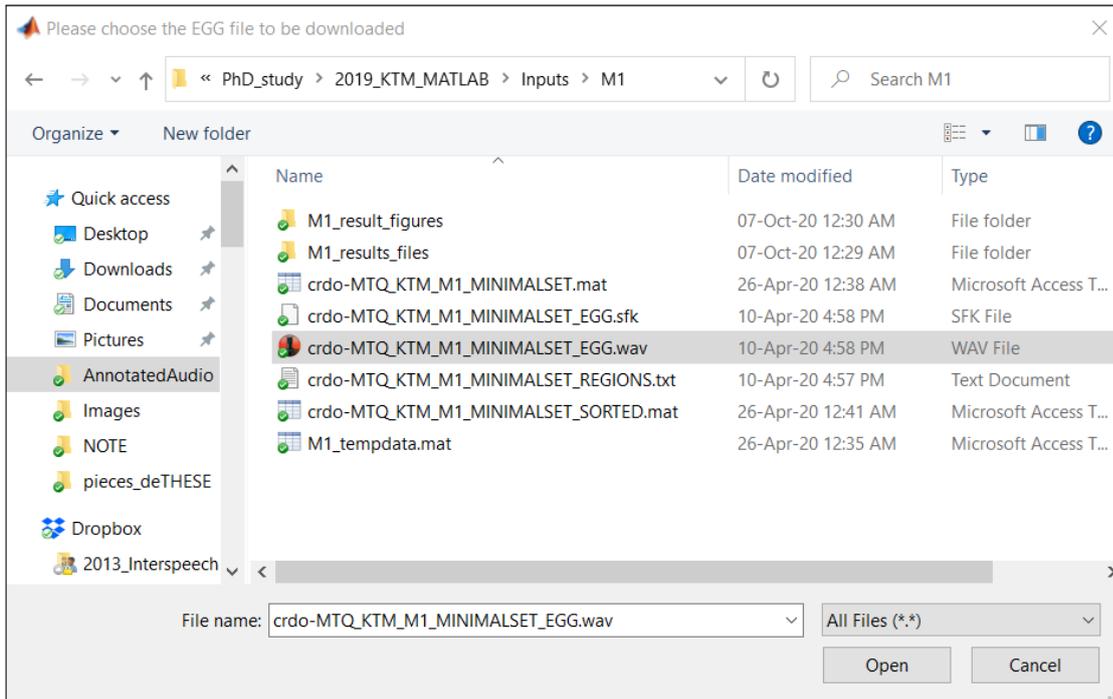
Figure 3.14: Two ways to open and run **PEAKDET** on MATLAB: (i) (in red): **PEAKDET** is opened in **CURRENT FOLDER WINDOW** and executed by typing “peakdet\_inter” in the **COMMAND WINDOW**; (ii) (in magenta): **PEAKDET** is opened in the **EDITOR WINDOW** and executed by selecting **Editor > Run** from the toolbar.

rectangle numbered 3). To run it, either select the peakdet\_inter.m file, then drag and drop it into the **COMMAND WINDOW**, or type “peakdet\_inter” (without capital letters) into the **COMMAND WINDOW** (as the rectangle numbered 4).

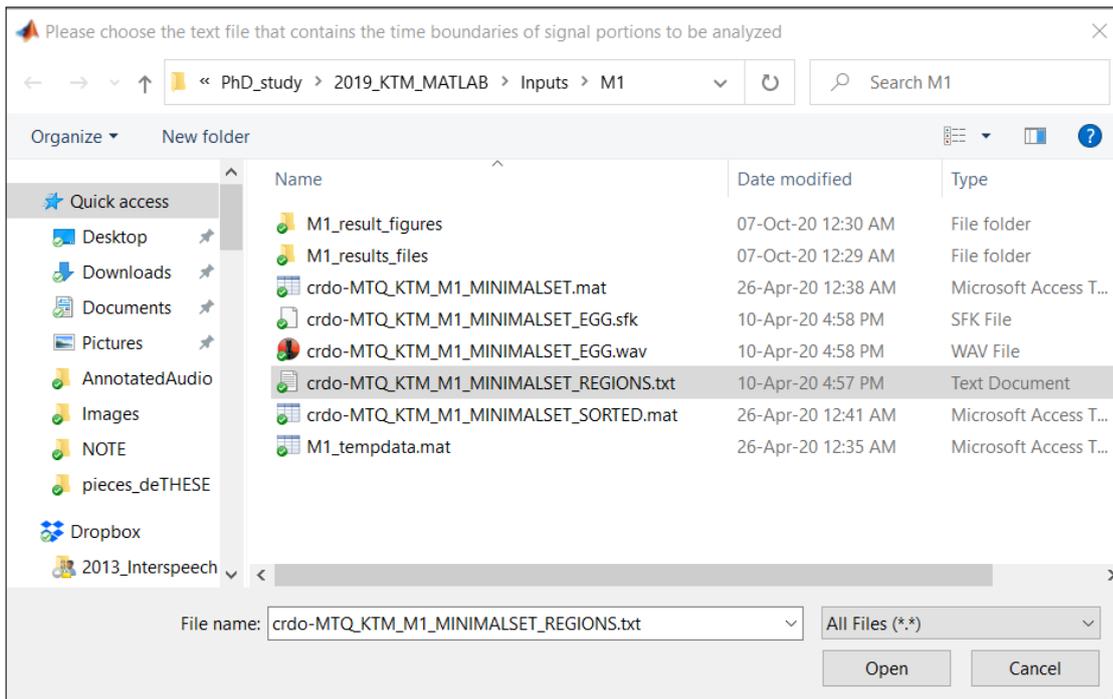
2. In the toolbar of MATLAB, select **Home > Open**, navigate to the location of the peakdet\_inter.m file, and select to open it in the **Editor window**. To run it, go back to the toolbar of MATLAB, select **Editor > Run**.

Note that if PeakDet\_inter is run in the second way, it is also necessary that the PeakDet\_inter folder is open in the **CURRENT FOLDER WINDOW**, otherwise MATLAB will not be able to find the path and run the command. Therefore, the second method is mentioned here only to provide all the possibilities to execute **PEAKDET**, but it is more practical and faster to execute it with the first method.

When **PEAKDET** has thus been called, two dialog boxes appear one after the other to prompt the user for the location of the two inputs: (i) the EGG file and (ii) the **Regions List**, as shown in Figure 3.15. The user’s work is to direct **PEAKDET** to the folder containing these files and to select them by double-clicking.



(a) The first dialog box asking for the indication of EGG file (.wav)



(b) The second dialog box asking for the indication of Regions List file (.txt)

Figure 3.15: The two dialog boxes that appear after calling `PEAKDET`, to request the location of the two input files.

### Default analysis parameters

At the same time as the input requests, some default settings are displayed in the `COMMAND WINDOW` as shown in Figure 3.16b. These parameters can also be adjusted later on, in case it is found that the values initially chosen are not suitable in view of the characteristics of the data under analysis. For example, some men's data have a much lower  $f_{0 \text{ dEGG}}$  range than women's, so the upper threshold for  $f_{0 \text{ dEGG}}$ , which is set to 500 Hz by default, can be changed to 300 Hz or 350 Hz. Changing this setting is among the options proposed at the stage of  $f_{0 \text{ dEGG}}$  verification. It is possible to make an adjustment for each token.

This is a change from the previous (2016) version, in which these settings were set as shown in Figure 3.16a.

In particular, for the three parameters in Figure 3.16a, the three frequent choices for my 2016 data were:

- Parameter #1: selecting 3 to use a value calculated by the *barycentre* method for peak detection in case of multiple closing peaks.
- Parameter #2: setting 300 Hz for male voices and 500 Hz for female voices as a  $f_{0 \text{ dEGG}}$  ceiling in the detection of double peaks.
- Parameter #3: selecting one among the values from 1 to 3 for the number of points for dEGG smoothing, depending on the level of noise in the signal.

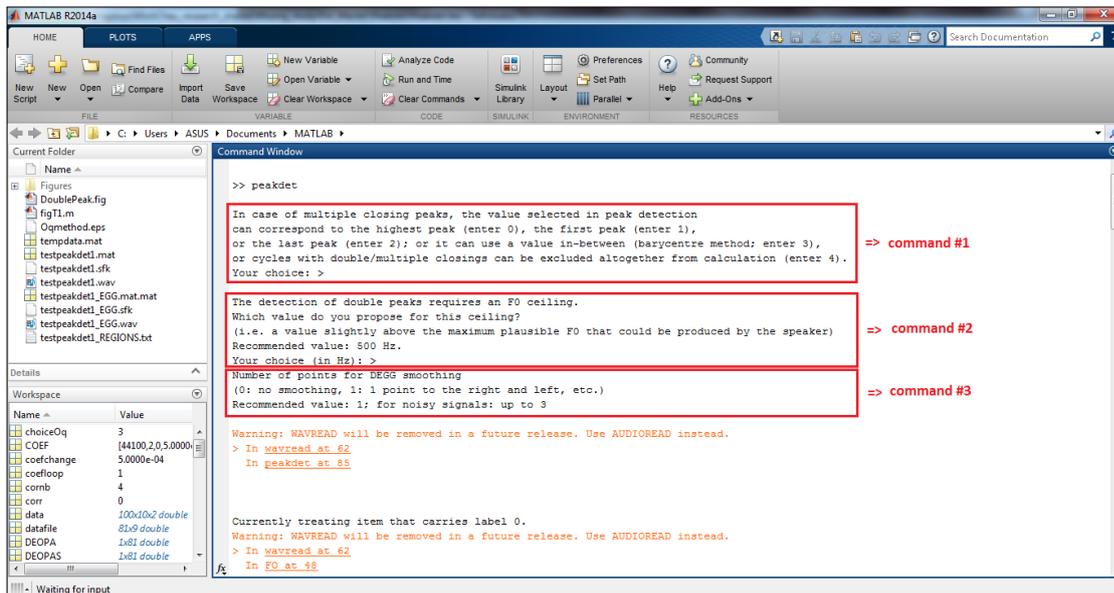
In the version of `PEAKDET` used here (version 1.0.5, released in 2019), these three variables are set by default to the following values:

- Parameter #1: as previously, use the *barycentre* method for peak detection in case of multiple closing peaks.
- Parameter #2: The  $f_{0 \text{ dEGG}}$  ceiling is set at 500 Hz as a threshold for the detection of double closing peaks.
- Parameter #3: the smoothing step for the derivative of the electroglottographic signal is set at 3, i.e. a window of a total length of seven data points (three before and three after the data point itself).

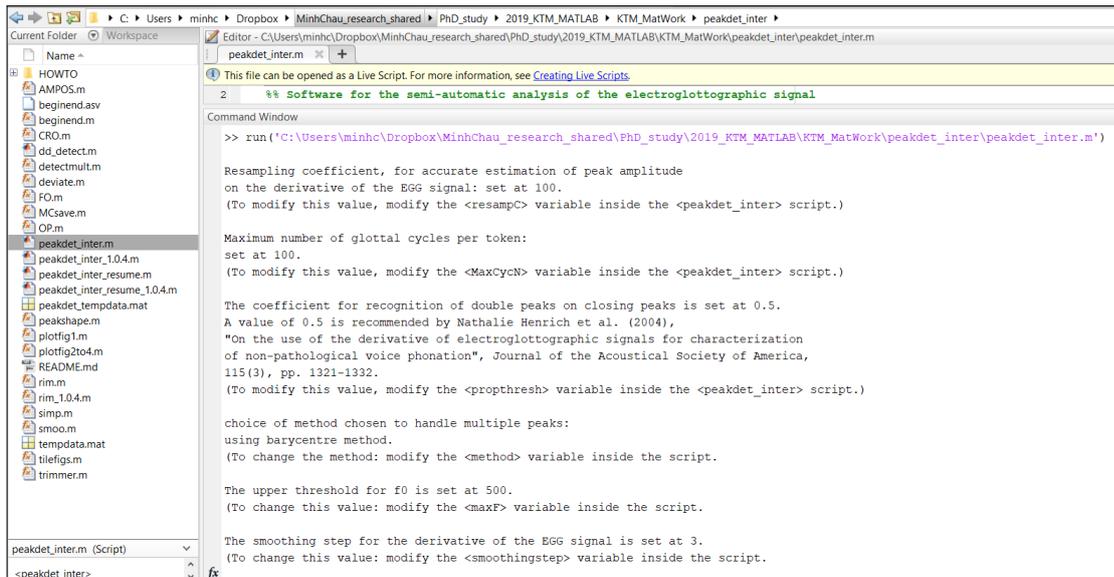
This modification is suitable for beginners who do not really go into the way in which `PEAKDET` operates: the above default settings are safe and can be applied to most data, including cases where the quality of the data is not really good (with noise) or if there are many instances of multiple closing peaks. Each default setting comes with an instruction that tells the user where to change the setting inside the script: see Figure 3.16b. Of course, users can only avail themselves of these functionalities if they understand the problem that causes `PEAKDET` to not work on their data or to give a misleading output. The user also has to understand the function and options of each parameter to modify it to suit his/her data.

### Graphic displays

When `PEAKDET` has received the path to the EGG file and its corresponding Regions List file (containing the time codes of each portion of signal to be analyzed), it proceeds to the next step. The tokens are processed in turn in a loop to detect open and



(a) The 2016 version: The display requesting selections for using three parameters. Reproduced from M.-C. Nguyễn (2016).



(b) The 2019 version (version 1.0.5): The parameters are set as default and displayed for the user to notice.

Figure 3.16: The difference between two versions (2016 and 2019) of the initial settings in the COMMAND WINDOW after executing `PEAKDET`.

closed peaks in the dEGG signal, from which the parameters ( $f_{o\_dEGG}$  and  $O_{q\_dEGG}$ ) are calculated.

With each token, the operations performed automatically by `PEAKDET` include: (i) smoothing and derivating the signal, (ii) detecting closing and opening peaks, (iii) calculating parameters, and (iv) storing values in a 10-column<sup>7</sup> × 100-lines matrix. The visible part, to which the user has access for each token after the automatic processing, consists of four plots:

- Figure 1 (exemplified in Figure 3.17a): the curves of two parameters  $f_{o\_dEGG}$  and  $O_{q\_dEGG}$  automatically calculated by `PEAKDET`;
- Figure 2 (exemplified in Figure 3.17b): the EGG signal;
- Figure 3 (exemplified in Figure 3.17c): the first derivative of the EGG signal, also called dEGG signal;
- Figure 4 (exemplified in Figure 3.17d): the second derivative of EGG signal or also called ddEGG signal.

This is also a change from the 2016 version as in the previous version there were only the first three plots. The ddEGG signal was added to address one of the issues raised regarding erroneous detection due to the occasional presence of a ‘hill-shaped’ pattern at the opening phase in glottalized pulses. The second derivative of the EGG signal offers a way to address this problem. The details of this issue will be provided at a more appropriate place in this thesis in Chapter 5 where we discuss some specific cases of glottal opening peaks.

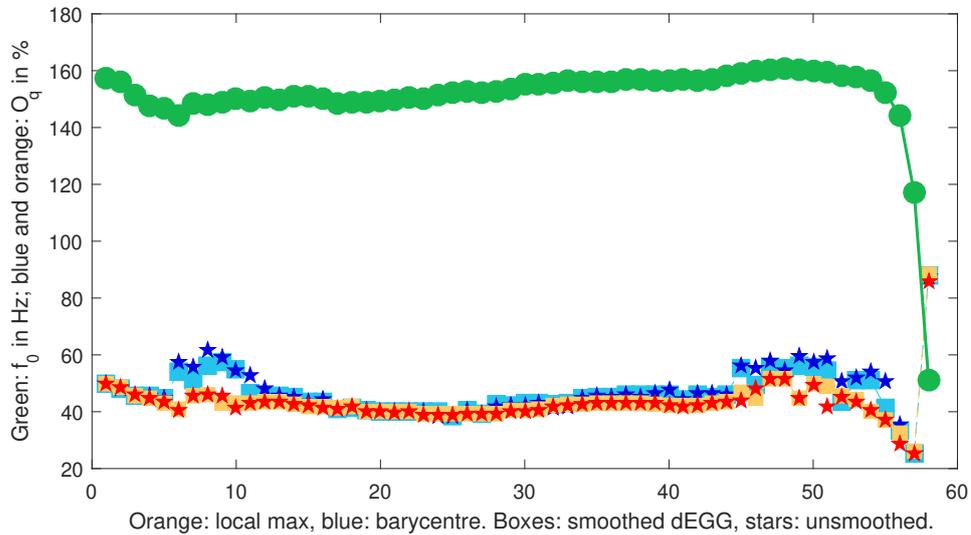
At the same time that those four plots are provided at once in four new `FIGURE WINDOWS`, `MATLAB` will provide some information about the current processing token in the `COMMAND WINDOW` (as in Figure 3.18), namely:

- UID of the token. In the example in Figure 3.18, the UID of the token is 101. According to the conventions used for the data set, this UID indicates that this is the target syllable /paj<sup>5</sup>/ in isolation at first performance.
- The status of the EGG signal: (i) first smoothed and derived as plotted in Figure 3 (e.g., Figure 3.17c), (ii) second smoothed and derived as plotted in Figure 4 (e.g., Figure 3.17d).
- The values of  $f_{o\_dEGG}$ . Thanks to this information, we can know the number of automatically detected glottal cycles and the specific value of each detected glottal cycle. For example, in the example being used here, the number of cycles is 50 and the specific  $f_{o\_dEGG}$  values of each cycle can be seen in Figure 3.18 which is also plotted in the green line in Figure 3.17a.

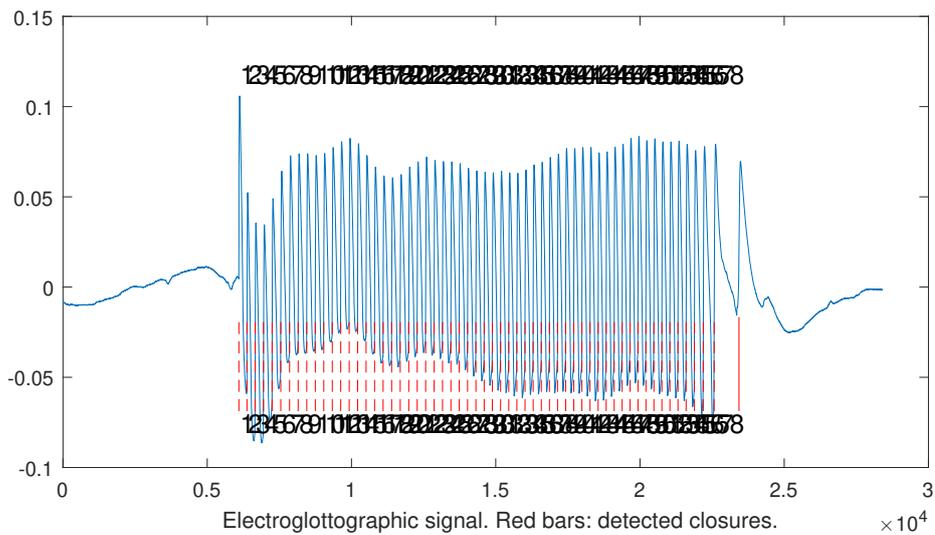
### User verification of $f_{o\_dEGG}$

In the `COMMAND WINDOW`, the next part following the information on the values of  $f_{o\_dEGG}$  are instructions on the options of  $f_{o\_dEGG}$  treatment (as can be seen in the bottom part of Figure 3.18).

<sup>7</sup>Note that the 10th column is set at 0 at this step and will be filled after user’s  $O_{q\_dEGG}$  verification step.

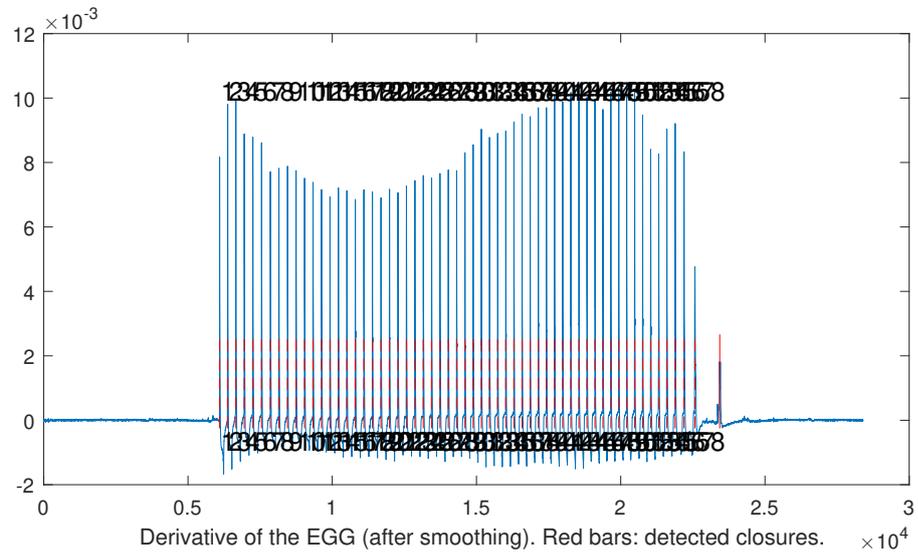


(a) Figure 1 of *PEAKDET* display: Two parameters automatically calculated by *PEAKDET*:  $f_0$   $_{dEGG}$  (in green) and  $Q_q$   $_{dEGG}$  (in blue and orange).

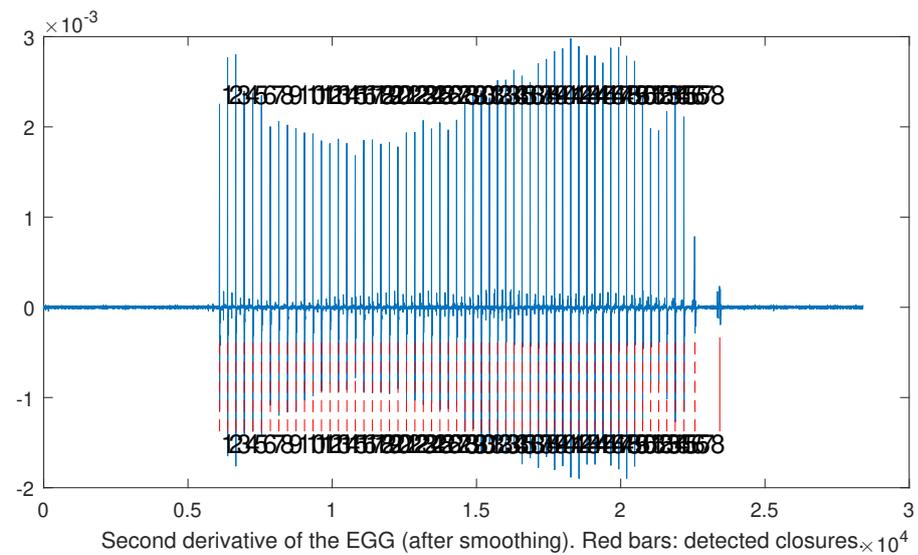


(b) Figure 2 of *PEAKDET* display: EGG signal

Figure 3.17: A demo of four initial plots after automatic processing of a token by *PEAKDET*. The example is taken from the data of speaker M1, first token 0101, target syllable /paj<sup>5</sup>/ in isolation.



(c) Two closing peaks are not detected at the 18<sup>th</sup> cycle by the automatic Peakdet.



(d) Two parameters after resetting the threshold at 0.001 to catch two undetected closing peaks.

Figure 3.17: A demo of four initial plots after automatic processing of a token by `PEAKDET`. The example is taken from the data of speaker M1, first token 0101, target syllable /paj<sup>5</sup>/ in isolation.

```

Currently treating item number 1.
This item carries label 101 in the input file.
Smoothing the first derivative of the electroglottographic signal...
Smoothed.
Smoothing the second derivative of the electroglottographic signal...
Smoothed.
Setting threshold for detection of closing peaks based on highest opening peak.

Fundamental frequency values (inverse of cycle durations):
Columns 1 through 16
157.3763 155.9130 151.5776 147.7738 146.7110 144.4150 148.5699 147.9072 148.7873 150.0951 149.5169 150.6662 149.9031 150.8621 150.8008 150.0418
Columns 17 through 32
148.6134 149.0923 148.8507 149.2083 149.9337 150.6353 150.3426 151.4059 152.3790 152.6744 152.4633 152.7590 153.4340 155.0851 155.1178 155.8469
Columns 33 through 48
156.4940 156.0730 156.1449 156.2113 156.9618 156.8725 156.5996 156.3442 157.0345 156.3996 156.8000 158.0815 158.9533 160.1017 160.4278 160.8550
Columns 49 through 58
160.1365 159.7768 159.5012 158.3099 157.8752 156.6324 152.3632 144.0706 117.0057 51.2056

If all the f0 values are correct, type 0 (zero).

The red lines on figures 2 and 3 indicate the first and last detected glottal cycles.
If some of the glottal cycles went undetected, or extra cycles were erroneously detected:
- enter 1 (one) to change the settings for automatic detection, or
- enter 2 to split one of the automatically detected cycles into two
  (by visual detection of a cycle not detected by the script)
- enter 3 to merge two automatically detected cycles
  (if visual detection reveals a spurious cycle).

If you wish to correct some of the f0 values manually, enter 4.
If the coefficient is correct but the initial/final cycle(s) must be suppressed, enter 5. >

```

Figure 3.18: Together with four plots as in Figure 3.17, the display of the current processing token on COMMAND WINDOW with identification information,  $f_0$  dEGG information and  $f_0$  dEGG processing options. The example is taken from the data of speaker M1, first token 0101, target syllable /paj<sup>5</sup>/ in isolation.

$f_0$  dEGG, as explained in section 2.4, is computed from closing peaks because “*The fundamental period is given by the position of the first maximum, which corresponds to the time between two consecutive closing peaks*” Henrich et al., 2004, p. 1328. A closing peak is identified as a good peak when it is well-defined with a single peak, as shown in figure 3.20.

At this stage, the user’s task consists in looking at Figures and  $f_0$  dEGG values, following the instructions of PEAKDET to make choices as appropriate in view of the results. If all the values are correct, type 0 (zero) to retain the result and move to the next step. If some of the glottal cycles went undetected, or extra cycles were erroneously detected, there are five options:

1. enter 1 (one) to change the settings for automatic detection;
2. enter 2 to split one of the automatically detected cycles into two (by visual detection of a cycle not detected by the script);
3. enter 3 to merge two automatically detected cycles;
4. If you wish to correct some of the fo values manually, enter 4;
5. If the coefficient is correct but the initial/final cycle(s) must be suppressed, enter 5.

In order to know which option to choose and use, the user has to look at the dEGG signal (i.e., Figure 3) to verify if all the glottal cycles are detected correctly or not. To facilitate this task, it is necessary to use three functions provided in the toolbar of the Figure 3 Window (as shown in Figure 3.19b), including: (i) zooming in (magnifying

glass icon with plus sign), (ii) zooming out (magnifying glass icon with minus sign), and (iii) moving the position (hand icon), in order to check every detected cycle. This operation is also carried out in exactly the same way when checking the opening peaks.

As shown in Figure 3.19b, when the signal is zoomed in, we can see the detected glottal cycles more clearly and easily based on two information:

- The vertical dashed red lines indicate precisely where the closing peaks are detected and the two solid red lines indicate the first and last detected glottal cycles. This means that the interval between two lines is counted as one glottal period. Inverting the time of each interval gives the  $f_{o\text{ dEGG}}$  value of a period.
- The numbers at the top and bottom of each interval indicate the order of each glottal cycle. Based on this, it is easy and rapid to find out which cycle(s) is (are) detected or undetected erroneously, and then to enter this order number in the Command Window for a modification. Note that after each modification, the glottal cycles will be recounted and the order number of the cycles will be changed.

These features were not offered in the `PEAKDET` 2016 version. This is really an important improvement as it speeds up the analysis process and improves accuracy. Before these improvements, I had to manually count each cycle to know which period needed to be adjusted. It was taking me over a year to process data from ten speakers. And after this upgrade in the current `PEAKDET`, it only took me about three more months to complete the remaining ten.

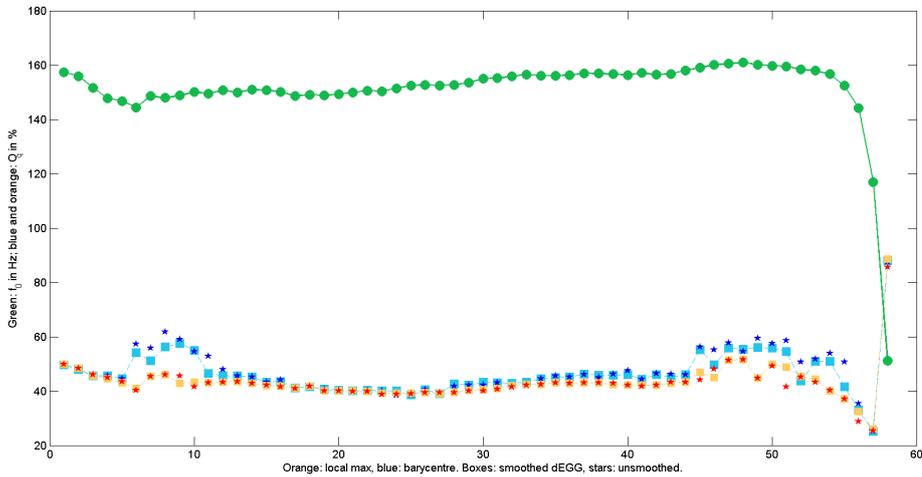
In my experience, if all the closing peaks are precise and are well detected, the  $f_{o\text{ dEGG}}$  curve (green) in Figure 4 will be continuous without sudden breaks. To save time, before verifying the closing peaks in dEGG signal, I first look at  $f_{o\text{ dEGG}}$  curve in Figure 4. If detecting unusual value(s) that cause the curve to break abruptly, then I will refer to dEGG signal in Figure 3 to verify how the `PEAKDET` detected at that cycle or cycles.

In terms of options for modifying the  $f_{o\text{ dEGG}}$  outcome, the following will explain and give an example to demonstrate when each of the above options should be chosen and what should be done for each option.

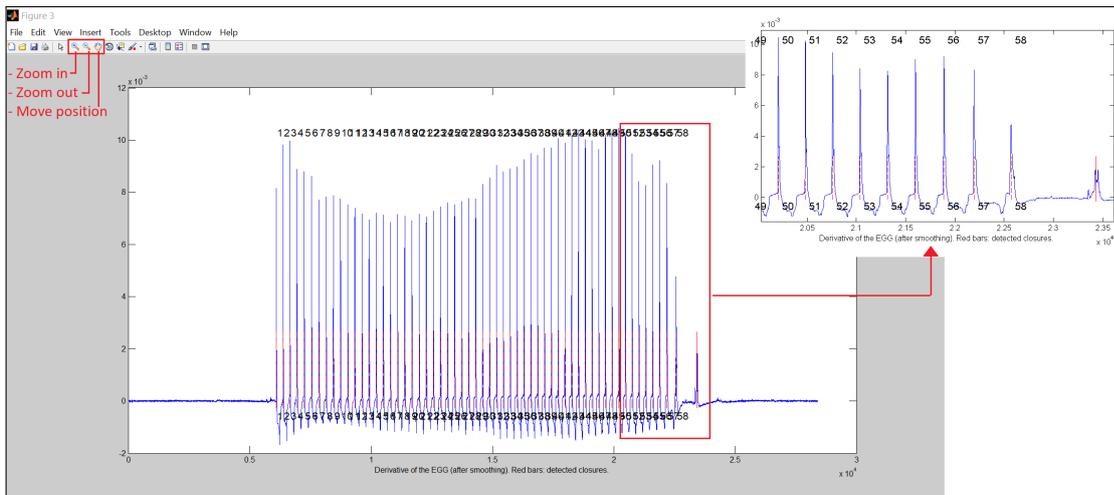
Thanks to analyzing a large amount of data, I had the opportunity to go through and use all the options offered in `PEAKDET`. All the examples here were chosen are the most clear and typical example for each option should be used. In reality, in some complicated signals we have to decide which options are the most optimal and maybe even combine several options to deal with one problem. Gradually, such experiences arise after dealing with more specific cases. The more I use `PEAKDET`, the more I understand how it works to use it more effectively and get more accurate results.

The first option is selected when all the closing peaks are detected correctly and no modification is required. The other options (second through sixth) are used when some of the glottal cycles went undetected, or extra cycles were erroneously detected.

Note that the options from the second to the sixth (options to modify) can be selected many times and do not have to be in a certain order. Only when option 0 is selected does this process end. One then moves on to the next step: the verification of  $O_{q\text{ dEGG}}$ .



(a) First action: look at  $f_0$  dEGG curve (green) in Figure 4 to check if there are any unusual values.



(b) Second action: zoom in on the dEGG signal in order to verify closing peaks.

Figure 3.19: Two operations for  $f_0$  dEGG verification. The example is taken from the data of speaker M1, first token 0101, target syllable /paj<sup>5</sup>/ in isolation.

**The first option:** “If all the  $f_0$  values are correct, type 0 (zero)”. Example taken from data of speaker M1, token: 03, UID: 0111, syllable /ja<sup>2</sup>/ (first frame word).

In my data, closing peaks are mostly well-defined with a single peak. Therefore, in most cases, I can quickly select the first choice by typing “0” (zero) in the COMMAND WINDOW) to indicate that no correction is necessary, and go on to the next step.

Figure 3.20 shows an ideal example, where all the closing peaks are clear, resulting in good detection. The corresponding  $f_0$  dEGG curve is continuous, without breaks: see Figure 3.20b.

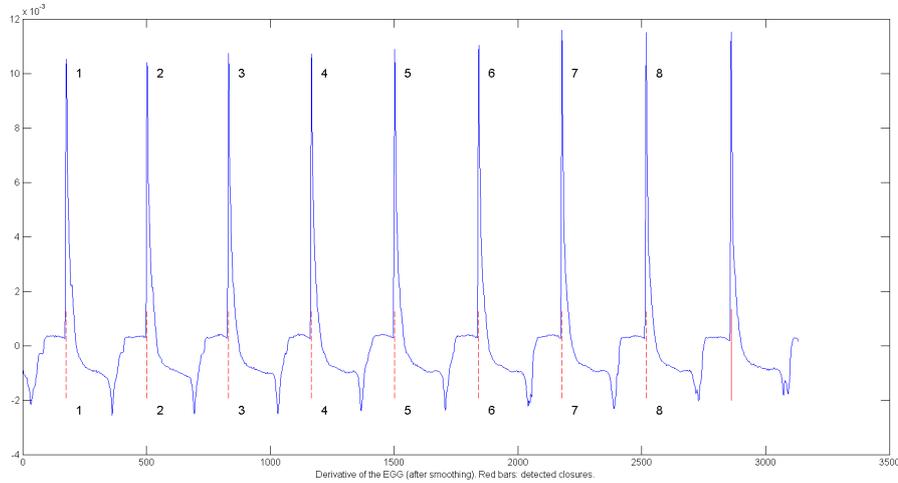
On the other hand, let us return to the example of Figure 3.19: token 0101, target syllable /paj<sup>5</sup>/ in isolation, data from speaker M1. It can be seen that the last three  $f_0$  dEGG values decrease dramatically. However, this is not a detection error of PEAKDET. Referring to the dEGG signal of these cycles, shown in Figure 3.19b, we can observe that all cycles from the 56<sup>th</sup> to the 58<sup>th</sup> have precise closing peaks and are correctly detected. In fact, this is a frequent phenomenon in this language. When the word is spoken in isolation, it tends to be realized with a tense/hard offset, involving a final glottalization. The three long glottal cycles (especially the last one) at the end of the syllable exemplify this phenomenon. So the detection results can safely be validated.

**The second option:** “enter 1 (one) to change the settings for automatic detection” This is useful in cases where some the glottal cycles went undetected. An example is found in data of speaker F21, token with UID 0501: syllable /paj<sup>4</sup>/ in isolation, first performance.

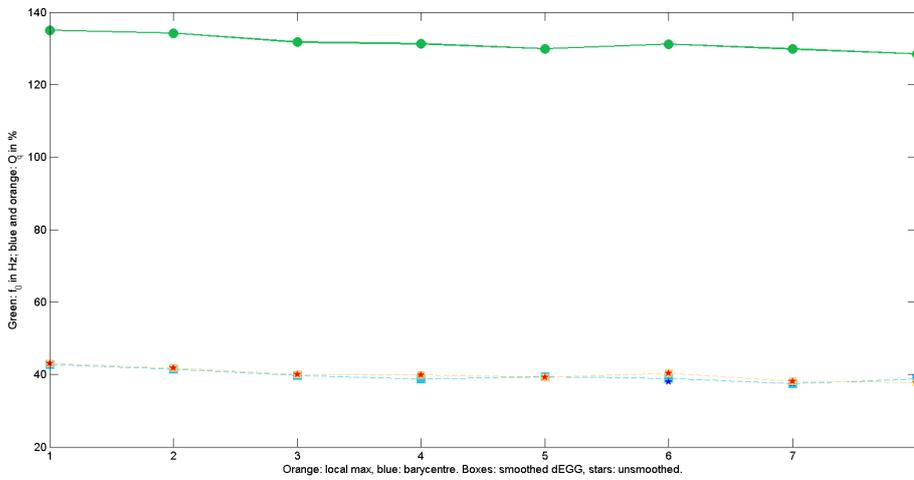
In this case, it is necessary to adjust the value used as a threshold for the detection of closing peaks, so that PEAKDET can detect all periods (or leave out the spurious peaks, if extra periods were detected erroneously). To illustrate the case where a change in the threshold value is necessary because of undetected peaks, Figure 3.21a shows the location of undetected peaks in the dEGG signal, and Figures 3.21b and 3.21c shows the difference in the resulting  $f_0$  dEGG curve before and after threshold correction.

**The third option:** “enter 2 to split one of the automatically detected cycles into two (by visual detection of a cycle not detected by the script)”. Example taken from data of speaker M14, token UID: 2001, target syllable /laj<sup>4</sup>/ in isolation, first performance.

Since this example is a token carrying the glottalized tone, it is difficult to say to what extent these six small symmetrical oscillations in the middle of the syllable reflect a technical problem (a difficulty for the electroglottographic device to monitor vocal fold contact area for some technical reason) and to what extent they reflect a phenomenon related to glottalization. On the one hand, these cycles occur in the middle of the rhyme, a position where glottalization is typically (and phonologically) present in Tone 4 in Muong. On the other hand, it is not a typical glottalized signal, but rather looks similar to a soft offset of voicing. Considering the audio (DOI: <https://doi.org/10.24397/pangloss-0006798#W39>, we can perceive a constriction in the middle: there is no relaxing at all. These small, symmetrical oscillations might reveal an interesting case in which, I suppose, the vocal folds have been completely adducted. during this period, but the surface of the vocal folds continues to vibrate with

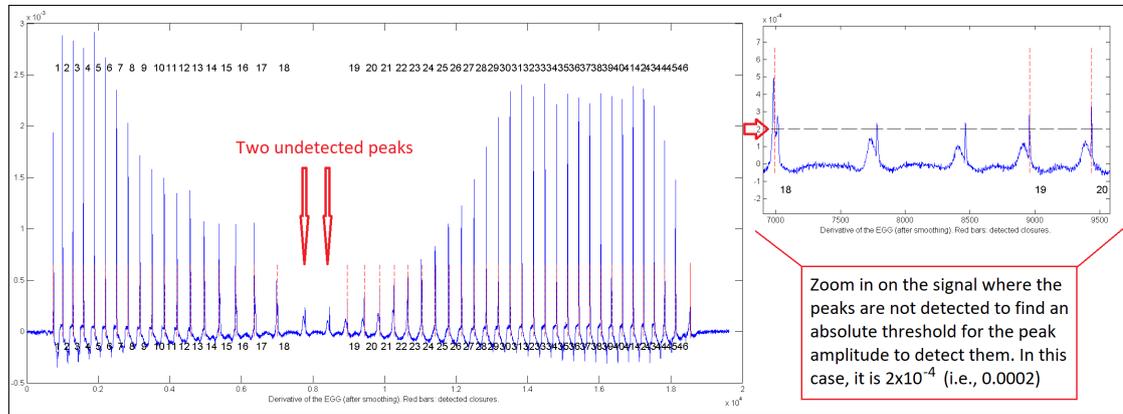


(a) dEGG signals. An example of good and clear signal with all well-defined closing and open peaks.

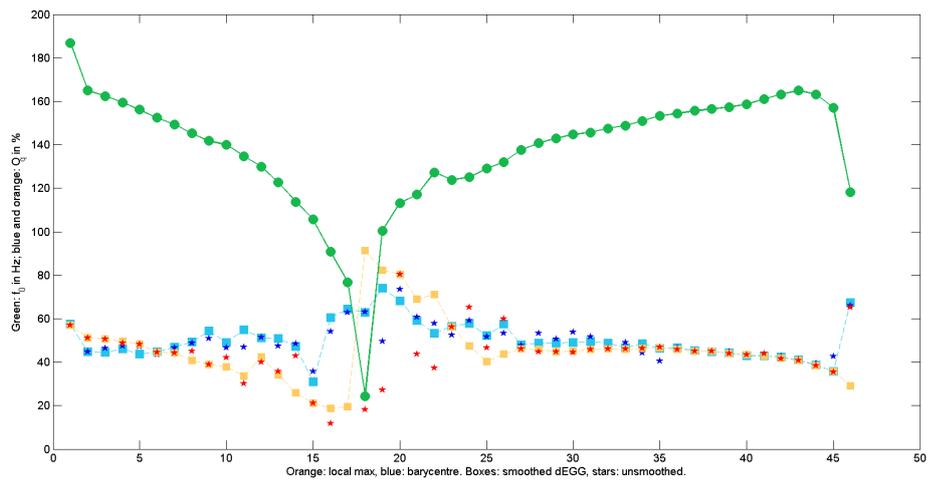


(b) Two parameters automatically calculated by Peakdet:  $f_{0 \text{ dEGG}}$  (green) and  $O_{q \text{ dEGG}}$  (blue and orange).

Figure 3.20: Example of a token where the first option (type o) of the  $f_{0 \text{ dEGG}}$  modification should be applied. Data from speaker M1, token: o3, UID: o111, syllable /ja<sup>2</sup>/ (first frame word).



(a) dEGG signal. Two undetected peaks inside the 18<sup>th</sup> cycle.



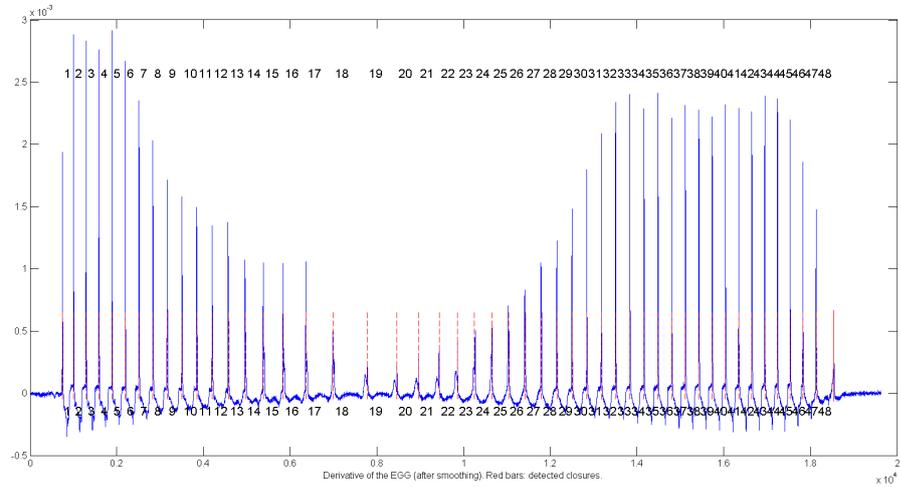
(b) Resulting values of  $f_{0 \text{ dEGG}}$  (green) and  $O_{q \text{ dEGG}}$  (blue and orange). A sudden drop in the 18<sup>th</sup> value is caused by undetected peaks.

Figure 3.21: Example of a token where the second option (*enter 1*) of the  $f_{0 \text{ dEGG}}$  modification should be applied. Data from speaker F21, token UID: 0501, syllable /paj<sup>4</sup>/.

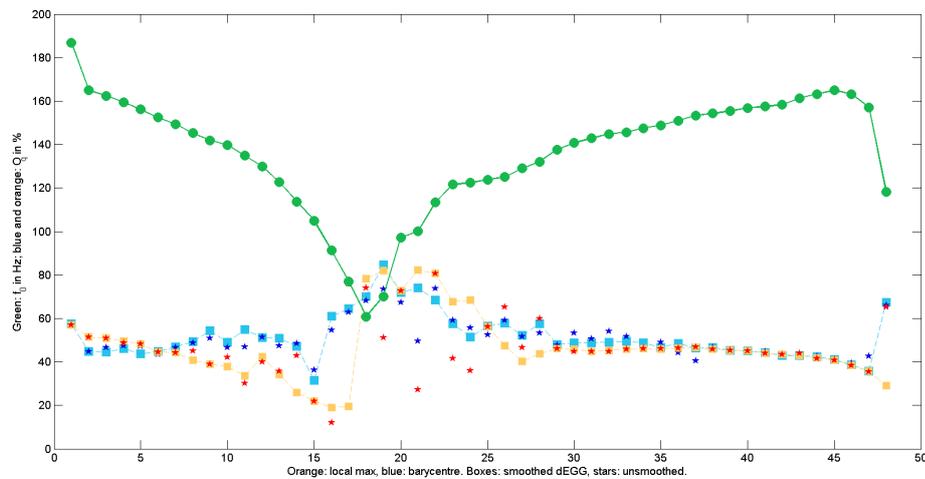
1	If all the $f_0$ values are correct, type 0 (zero).	
2		
3	The red lines on figures 2 and 3 indicate the first and last detected glottal cycles.	
4	If some of the glottal cycles went undetected, or extra cycles were erroneously detected:	
5	- enter 1 (one) to change the settings for automatic detection, or	
6	- enter 2 to split one of the automatically detected cycles into two	
7	(by visual detection of a cycle not detected by the script)	
8	- enter 3 to merge two automatically detected cycles	
9	(if visual detection reveals a spurious cycle).	
10		
11	If you wish to correct some of the $f_0$ values manually, enter 4.	
12	If the coefficient is correct but the initial/final cycle(s) must be suppressed, enter 5. > 1	#1: Select 1 to use the option of changing the settings for automatic detection.
13		
14	If too many glottal cycles were detected, you may change the threshold for maximum $f_0$ .	
15	The present threshold is: 500	
16	New value for the threshold (in Hz): > 500	#2: Type 500 to not change the threshold of the maximum $f_0$
17		
18	If too few glottal cycles were detected, you may change the threshold for peak detection.	
19	The present threshold is set (by default) at 0.5 of the maximum in this portion of the signal.	
20	- Enter a new value (absolute value; refer to figure to choose) for the threshold; or	
21	- press RETURN to leave threshold unchanged;	
22	- type 0 in case the syllable needs to be analyzed as several distinct portions, i.e.	#3: Indicate an absolute value for the peak amplitude threshold where it crosses the micro closing peak(s), allowing them to be detected.
23	if the discrepancy in peak amplitude is such that no setting gives satisfactory result	
24	for the entire syllable. > 0.0002	
25		

- (c) Three actions in the COMMAND WINDOW to use the option of changing threshold to detect undetected peaks.

Figure 3.21: Example of a token where the second option (*enter 1*) of the  $f_0$  dEGG modification should be applied. Data from speaker F21, token UID: 0501, syllable /**paj**<sup>4</sup>/.

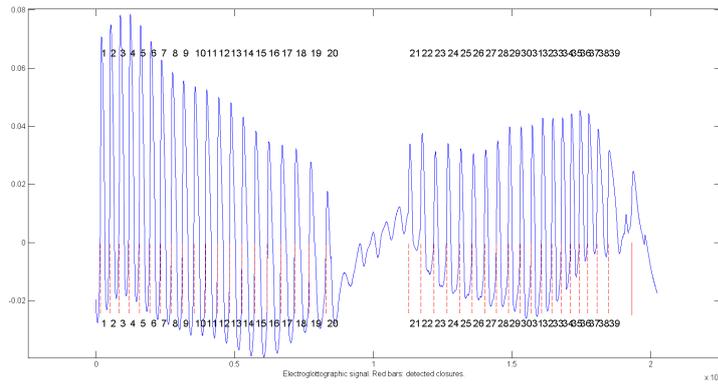


(d) dEGG signal. The two undetected peaks are caught by using the second option and indicating the exact amplitude of these micro-peaks (at 0.0002).

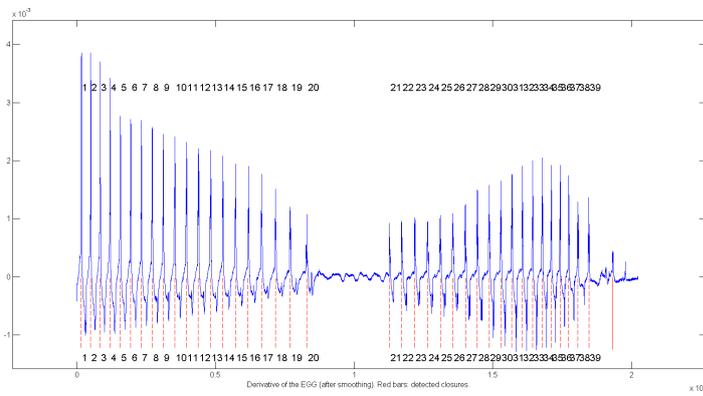


(e) Result of  $f_{0 \text{ dEGG}}$  (green) and  $O_{q \text{ dEGG}}$  (blue and orange). **PEAKDET** re-calculate these two parameters after the detection of closing peaks is corrected

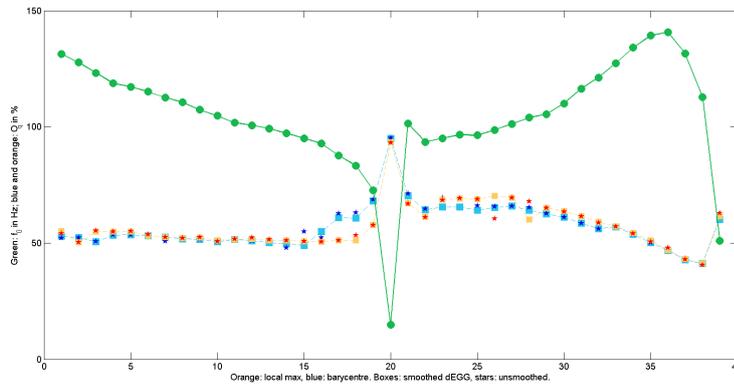
Figure 3.21: Example of a token where the second option (enter 1) of the  $f_{0 \text{ dEGG}}$  modification should be applied. Data from speaker F21, token UID: 0501, syllable /paj<sup>4</sup>/.



(a) EGG signal.



(b) dEGG signal.



(c) Result of  $f_{0 \text{ dEGG}}$  (green) and  $O_{q \text{ dEGG}}$  (blue and orange). A sudden drop of  $f_{0 \text{ dEGG}}$  in the 20<sup>th</sup> value is caused by undetected peaks.

Figure 3.22: Example of a token where the third option (enter 2) of the  $f_{0 \text{ dEGG}}$  modification should be applied. Data from speaker M14, token UID: 2001, target syllable /*aj*<sup>4</sup>/ in isolation, first performance.

1	If all the $f_0$ values are correct, type 0 (zero).	
2		
3	The red lines on figures 2 and 3 indicate the first and last detected glottal cycles.	
4	If some of the glottal cycles went undetected, or extra cycles were erroneously detected:	
5	- enter 1 (one) to change the settings for automatic detection, or	
6	- enter 2 to split one of the automatically detected cycles into two	
7	(by visual detection of a cycle not detected by the script)	
8	- enter 3 to merge two automatically detected cycles	
9	(if visual detection reveals a spurious cycle).	
10		
11	If you wish to correct some of the $f_0$ values manually, enter 4.	
12	If the coefficient is correct but the initial/final cycle(s) must be suppressed, enter 5. > 2	#1: Enter 2 to use the option of splitting a cycle into two or more.
13		
14		
15	Which glottal cycle needs to be split? Enter its number. > 20	#2: Indicate which cycle needs to be split.
16		
17		
18	How many cycles can you detect (by eye) in this portion of the signal? > 6	#3: Specify the number of cycles to be split from this erroneous cycle.
19		
20		
21	Is this the result you wanted? Enter 1 if yes, 0 if no. > 1	#4: Enter 1 if the modified result is satisfactory and to end this treatment. Otherwise, enter 0 to start over.
22		
23		

(d) Three actions in the COMMAND WINDOW to use the option of changing threshold to detect undetected peak(s).

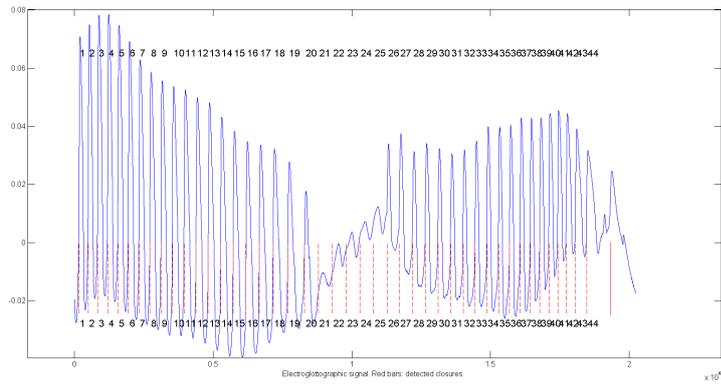
Figure 3.22: Example of a token where the third option (enter 2) of the  $f_0$  <sub>dEGG</sub> modification should be applied. Data from speaker M14, token UID: 2001, target syllable /**laj**<sup>4</sup>/ in isolation, first performance.

small oscillations. Therefore, it was somewhat difficult for me at first to decide whether these oscillations should be divided into cycles or not. It is the case that there is strong glottalization in the middle of the rhyme; seen in this light, changing the parameters in the script in a way that makes the  $f_0$  <sub>dEGG</sub> curve look more continuous can appear as a bad idea, as it irons out, as it were, a specificity in laryngeal behavior which the medial dent in the curve brought out forcibly. However, finally the choice was to split them. Since periodicity is not lost altogether, it appears more appropriate that the  $f_0$  <sub>dEGG</sub> curve should look continuous as in Figure 3.22g, than artificially assigning the entire span of low-amplitude electroglottographic signal to one extremely long glottal cycle (which would clearly be an artefact). Furthermore, the  $O_q$  <sub>dEGG</sub> result confirms that the phonation type does not switch to creak (phonation mechanism o), since most values are above 50%, that is, in the range of modal phonation. The  $O_q$  <sub>dEGG</sub> values surrounding the unusual  $f_0$  <sub>dEGG</sub> are not lower (if anything, they are higher than what surrounds them). Therefore, it would be very implausible for the fundamental frequency to drop below 10 Hz for just one cycle.

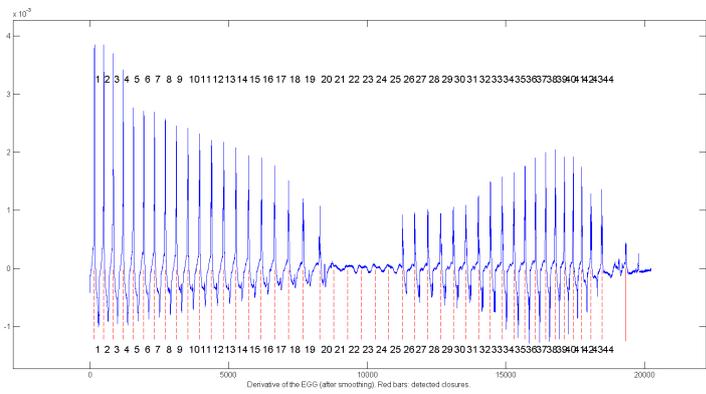
**The fourth option: “enter 3 to merge two automatically detected cycles (if visual detection reveals a spurious cycle)”.** Example taken from data of speaker M14, token UID: 3502, syllable /**kieŋ**<sup>4</sup>/ in isolation, second performance.

**The fifth option: “If you wish to correct some of the  $f_0$  values manually, enter 4”.** Example taken from data of speaker F3, token: 152, UID: 1602, the target syllable /**laj**<sup>5</sup>/ in isolation, second performance.

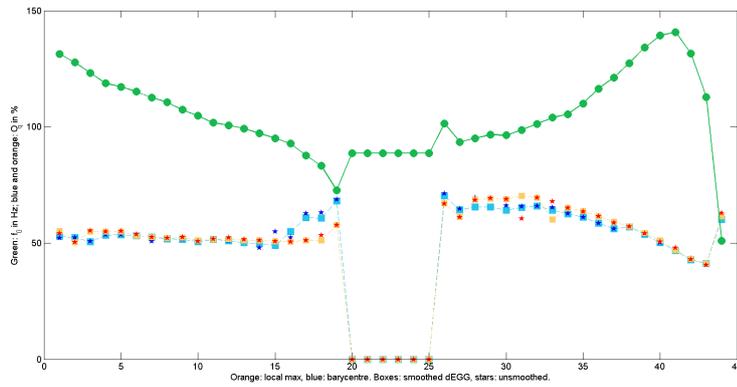
This is a dangerous option that I would not recommend in cases where one or more



(e) EGG signal. After the six small oscillations are split into six cycles.

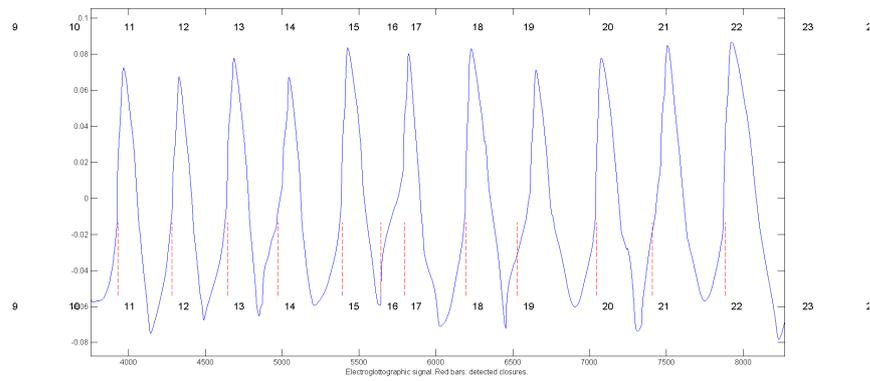


(f) dEGG signal. After the six small oscillations are split into six cycles.

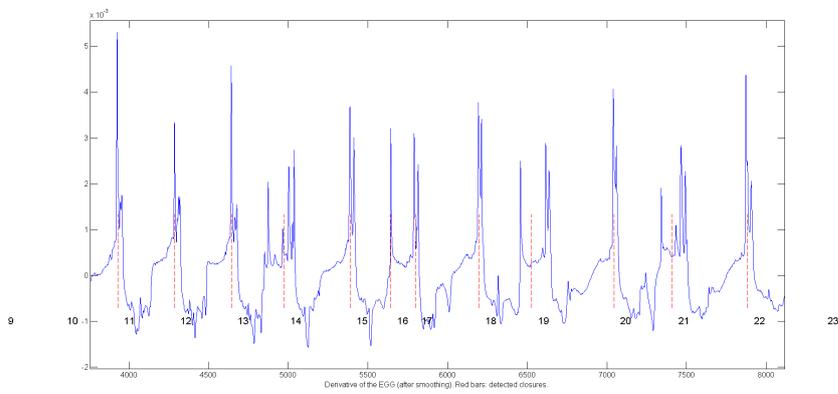


(g) Result of  $f_{0 \text{ dEGG}}$  (green) and  $O_{q \text{ dEGG}}$  (blue and orange). After the six small oscillations are split into six cycles.

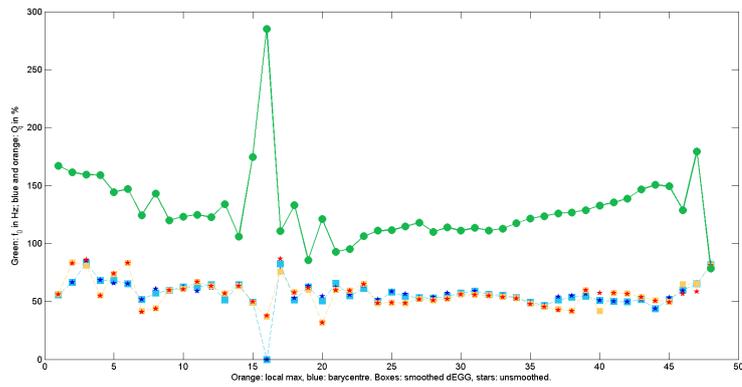
Figure 3.22: Example of a token where the third option (enter 2) of the  $f_{0 \text{ dEGG}}$  modification should be applied. Data from speaker M14, token UID: 2001, target syllable /laj<sup>4</sup>/ in isolation, first performance.



(a) EGG signal.



(b) dEGG signal.



(c) Result of  $f_0$  dEGG (green) and  $O_q$  dEGG (blue and orange).

Figure 3.23: Example of a token where the fourth option (enter 3) of the  $f_0$  dEGG modification should be applied. Data from speaker M14, token UID: 3502, syllable /kɪŋ<sup>4</sup>/ in isolation, second performance.

```

4 If all the f0 values are correct, type 0 (zero).
5
6 The red lines on figures 2 and 3 indicate the first and last detected glottal cycles.
7 If some of the glottal cycles went undetected, or extra cycles were erroneously detected:
8 - enter 1 (one) to change the settings for automatic detection, or
9 - enter 2 to split one of the automatically detected cycles into two
10 (by visual detection of a cycle not detected by the script)
11 - enter 3 to merge two automatically detected cycles
12 (if visual detection reveals a spurious cycle).
13
14 If you wish to correct some of the f0 values manually, enter 4.
15 If the coefficient is correct but the initial/final cycle(s) must be suppressed, enter 5. > 3 #1: Enter 3 to use the option of merging two cycles into one.
16
17 Which glottal cycles need to be merged? Enter their numbers, separated by a colon. #2: Indicate which two cycles should be merged.
18 For instance, to merge cycles 4 and 5, type: > 4:5
19 > 15:16
20
21 Is this the result you wanted? Enter 1 if yes, 0 if no. > 1 #3: Enter 1 if the modified result is satisfactory and to finish
22 this treatment. If not, enter 0 to start over.

```

(d) Three actions in the COMMAND WINDOW to use the option of changing threshold to detect undetected peak(s).

Figure 3.23: Example of a token where the fourth option (enter 3) of the  $f_{0\text{ dEGG}}$  modification should be applied. Data from speaker M14, token UID: 3502, syllable /kɪŋ<sup>4</sup>/ in isolation, second performance.

other options can be chosen instead. I used this option for one token in data of speaker F3 (token: 152, UID: 1602) which has a very serious signal error. Instead of splitting the small oscillations, I simply manually modified the unusual  $f_{0\text{ dEGG}}$  values in some cycles so that they were close to the surrounding values. However, when plotting the results, there was an error in this token, so I suppose it may be because I chose this option when treating  $f_{0\text{ dEGG}}$ .

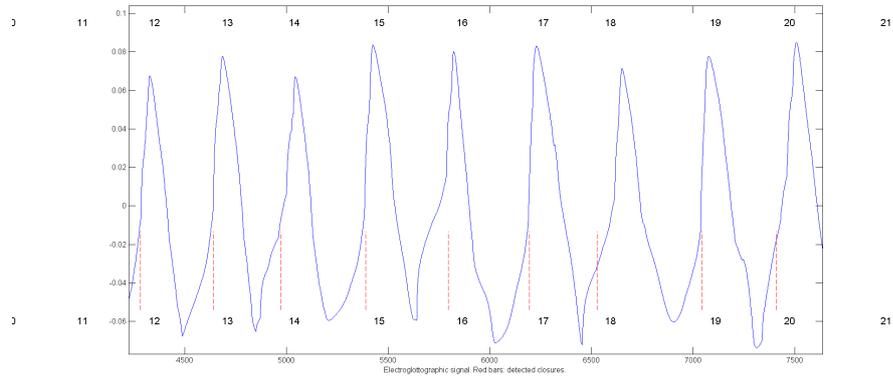
**The sixth option: “If the coefficient is correct but the initial/final cycle(s) must be suppressed, enter 5.”** Example taken from data of speaker M14, token: 10, UID: 0142, syllable /tǎŋ<sup>3</sup>/ the fourth word of carrier sentence, second performance.

This option is employed when the first or last period(s) are out of the range found in the rest of the rhyme: they have widely different value(s) from the others. This case can easily be corrected by suppressing that/those period(s).

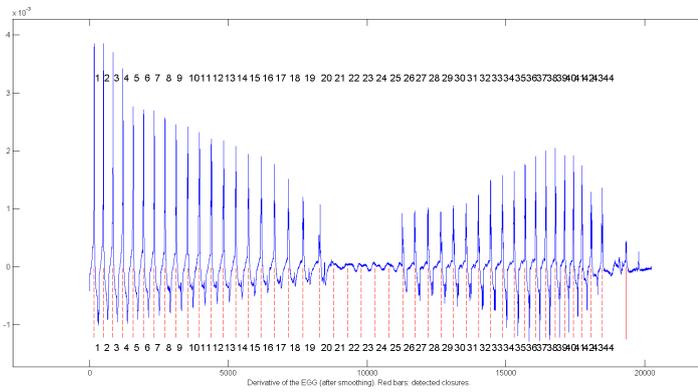
The token shown in Figure 3.24 has non-continuous  $f_{0\text{ dEGG}}$  toward the end at position (3.24b) because PEAKDET detected three closing peaks after the rhyme had ended (3.24a).

This is not an artefact: the shape of the dEGG signal indicates clearly that four small closing peaks are present, i.e. the glottis closes at these points, and there is quasi-periodic fluctuation in vocal fold contact area. Three of these four periods were detected by PEAKDET, i.e. the 31<sup>st</sup> to 33<sup>rd</sup> cycles.

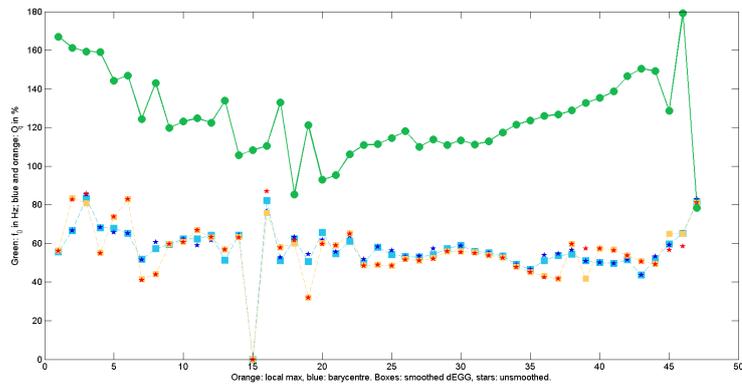
However, for the purpose of the present study, a crucial point is that these small closing peaks take place after phonation has ended: the four symmetrical oscillations that occur at the end are typical of vestigial vocal fold oscillation at ‘soft’ (nonglottalized) offset of voicing. It would be wrong to include these small oscillations along with the other normal glottal cycles. The unexpected  $f_{0\text{ dEGG}}$  value at cycles 31<sup>st</sup>, at about 220 Hz (3.21b), is clearly an artifact, and should be removed. Therefore, I suppressed the last three values. The actions to treat this case are illustrated in Figure 3.24c, and the corresponding result after correction is shown in Figures 3.24d and 3.24e.



(e) EGG signal. After merging cycles 15 and 16.

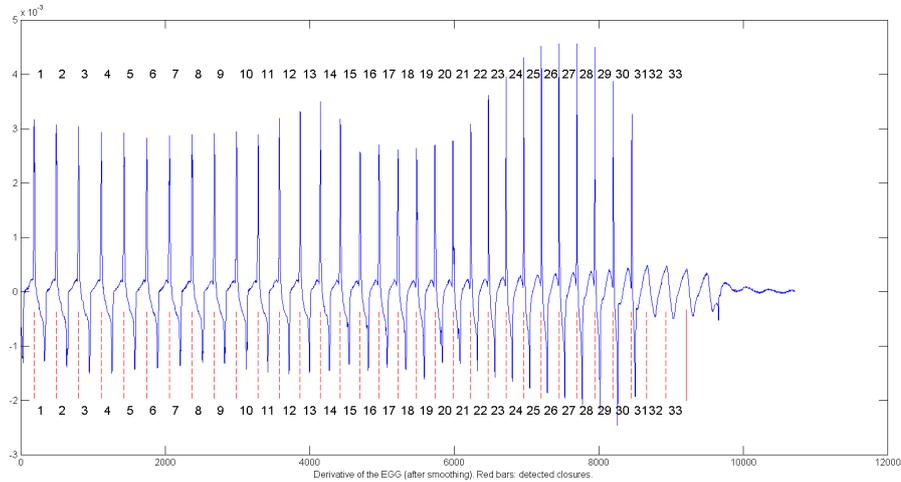


(f) dEGG signal. After merging cycles 15 and 16.

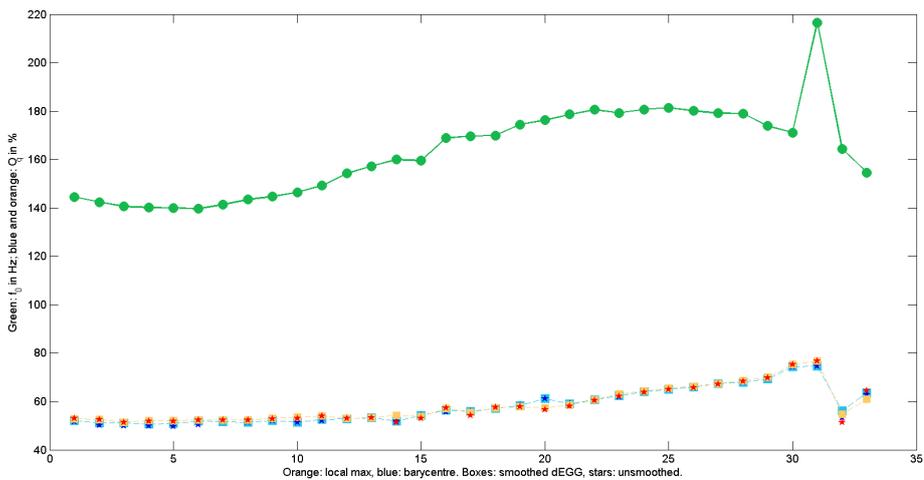


(g) Result of  $f_0$  dEGG (green) and  $O_q$  dEGG (blue and orange).

Figure 3.23: Example of a token where the fourth option (enter 3) of the  $f_0$  dEGG modification should be applied. Data from speaker M14, token: 339, UID: 3502, syllable /kɪɛŋ<sup>4</sup>/ in isolation, second performance.



(a) dEGG signal. Three extra peaks detected at the end of the syllable, at cycles 31<sup>st</sup> - 33<sup>rd</sup>.



(b) Result of  $f_0$  dEGG (green) and  $O_q$  dEGG (blue and orange). As a consequence of the extra peaks detected, an unusual shift of the  $f_0$  dEGG towards the end.

Figure 3.24: Example of a token where the sixth option (enter 5) of the  $f_0$  dEGG modification should be applied. Data from speaker M14, token: 10, UID: 0142, syllable /tʰŋ<sup>3</sup>/ the fourth word of carrier sentence, second performance.

25	The red lines on figures 2 and 3 indicate the first and last detected glottal cycles.	
26	If some of the glottal cycles went undetected, or extra cycles were erroneously detected:	
27	- enter 1 (one) to change the settings for automatic detection, or	
28	- enter 2 to split one of the automatically detected cycles into two	
29	(by visual detection of a cycle not detected by the script)	
30	- enter 3 to merge two automatically detected cycles	
31	(if visual detection reveals a spurious cycle).	
32		
33	If you wish to correct some of the $f_0$ values manually, enter 4.	
34	If the coefficient is correct but the initial/final cycle(s) must be suppressed, enter 5. > 5	#1: Enter 5 to use the option of suppressing the initial/final cycle(s)
35		
36	To suppress first cycle, enter 1. To suppress last cycle, enter 9.	
37	If no cycle suppression is needed, enter 0. > 9	
38		
39	To suppress first cycle, enter 1. To suppress last cycle, enter 9.	#2: Enter 9 to suppress the final cycle. To suppress as many cycles as you wish, type 9 as many times. Do the same with 1 to exclude the initial cycle(s).
40	If no cycle suppression is needed, enter 0. > 9	
41		
42	To suppress first cycle, enter 1. To suppress last cycle, enter 9.	
43	If no cycle suppression is needed, enter 0. > 9	
44		
45	To suppress first cycle, enter 1. To suppress last cycle, enter 9.	#3: Enter 0 to terminate this option after a satisfactory result is obtained.
46	If no cycle suppression is needed, enter 0. > 0	
47		

(c) Three actions in the COMMAND WINDOW to use the option of suppressing initial/final peak(s).

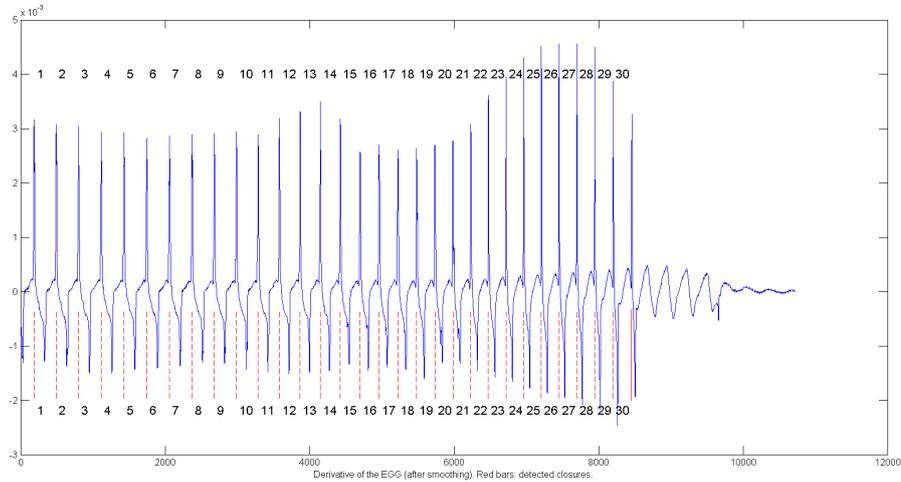
Figure 3.24: Example of a token where the sixth option (enter 5) of the  $f_0$  dEgg modification should be applied. Data from speaker M14, token: 10, UID: 0142, syllable /t͡ʰaŋ<sup>3</sup>/ the fourth word of carrier sentence, second performance.

To deal with this case, the second possible way is to use the option of changing the threshold of peak amplitude for automatic detection so that PEAKDET will not detect these extra peaks.

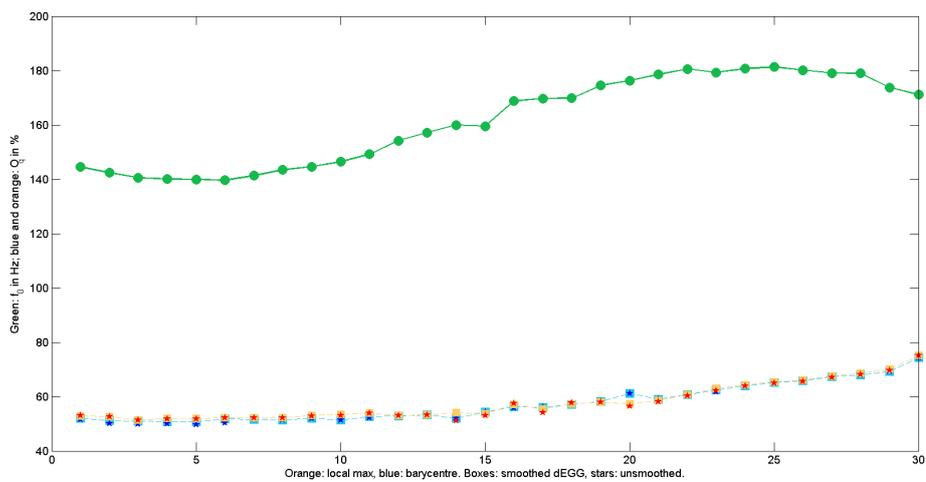
A final clarification concerning  $f_0$  dEgg estimation by PEAKDET is that this script makes no “sanity check” on the  $f_0$  dEgg results: no check of quasi-periodicity of the signal is performed. The script detects closing peaks, and the value offered as “fundamental frequency” is the inverse of the period. This can be very misleading in cases of period doubling (referred to in Section 5.2 as multiply-pulsed creak): the pattern that repeats itself contains three closing peaks on the electroglottographic signal, but the values reported in the “ $f_0$  dEgg” column wrongly suggest that there are two simultaneous pitches (as in diplophonia), one high and the other lowish. This peculiarity requires care in the interpretation of the obtained data: thus, a speaker’s average fundamental frequency as obtained by averaging all  $f_0$  dEgg values obtained through PEAKDET is likely to be slightly higher than that obtained using autocorrelation-based tools. This limitation is not likely to be of any consequence to the results reported here but needs to be kept in mind in case the results reported in the present study are re-used in future work (as one hopes may happen, in an Open Science perspective).

### User verification of $O_q$ dEgg

Whereas  $f_0$  dEgg is calculated based on closing peaks, which in the great majority of cases are well-defined (often with a unique peak), estimating  $O_q$  dEgg requires the



(d) dEGG signal. The three extra final peaks detected were excluded.



(e) Result of  $f_{0 \text{ dEGG}}$  (green) and  $O_{q \text{ dEGG}}$  (blue and orange). `PEAKDET` re-calculate these two parameters after the three extra final peaks detected were excluded.

Figure 3.24: Example of a token where the sixth option (`enter 5`) of the  $f_{0 \text{ dEGG}}$  modification should be applied. Data from speaker M14, token: 10, UID: 0142, syllable /tǎŋ<sup>3</sup>/ the fourth word of carrier sentence, second performance.

detection of opening peaks, which often runs into difficulties due to imprecise peaks: either cases where no peak stands out clearly, or cases where two or more peaks are present (multiple peaks). The search for opening peaks is even more difficult in the case of nonmodal phonation, such as when voicing transitions into creaky voice. This makes user verification of  $O_{q \text{ dEGG}}$  a delicate business, which is not so similar with verification of  $f_{o \text{ dEGG}}$ : it requires more than just a few adjustments for peculiar situation.

**PEAKDET** will ask the verification of  $O_{q \text{ dEGG}}$  after the verification of  $f_{o \text{ dEGG}}$  has finished. At first, it will process automatically and offer  $O_{q \text{ dEGG}}$  calculated in four different ways:

1. maxima<sup>8</sup> on unsmoothed dEGG signal (displayed as orange squares);
2. maxima on smoothed dEGG signal (displayed as orange stars);
3. barycentre of peak on unsmoothed dEGG signal (displayed as blue squares);
4. barycentre of peak on smoothed dEGG signal (displayed as blue stars)

The methods are divided into two sets:

- Detection of the local minimum on the signal in-between two closure peaks. This method is applied twice: on the unsmoothed dEGG signal, and on the smoothed dEGG
- Analysis of the shape of opening peaks and calculation of a barycentre of the detected ‘peaks-within-the-peak’, giving each of the peaks a coefficient proportional to its amplitude. Again, this method is applied twice: on the unsmoothed dEGG signal, and on the smoothed dEGG.

The display of the 2021 version (1.0.5), which can be seen in many figures in the  $f_{o \text{ dEGG}}$  section such as Figure 3.21, is better than that of the version I had used previously (in 2016), as shown in Figure 3.25. It uses fewer colors and fewer marking symbols, and they stand in a logical relationship: the colors show different methods (orange for local maxima, blue for barycenter) and the marking symbols distinguish results from the unsmoothed signal (stars) and smoothed signal (squares).

The verification of  $O_{q \text{ dEGG}}$  involves two steps: (i) selecting among the outputs of the four methods, by choosing one in four, and (ii) visually verifying the  $O_{q \text{ dEGG}}$  result of that method.

In simple cases, all open peaks of the item are well-defined peaks, which means that there is a single, well-defined negative peak inside each period. As a result, the four methods yield almost identical results of  $O_{q \text{ dEGG}}$ , as shown by four coincident curves. In such non-problematic cases, we can choose any method among the four, and there is no need for further correction of results before moving on to the next item. The example of good closing peaks in Figure 3.20 also presents the case where all opening peaks are well-defined and should be kept by type ‘o’ in the `COMMAND WINDOW`.

However the simple cases are few. The majority of my data required more complex verification, due to imprecise peaks in dEGG signal. ‘Imprecise peaks’ can mean several things, as can be seen in Figure 3.28: (i) there can be several negative peaks during the opening phase, with none of them standing out as a main peak; or (ii) there can be no well-marked opening peak at all.

---

<sup>8</sup>Technically, ‘maxima’ here should be referred to as ‘minima’, since the peak is a negative peak.

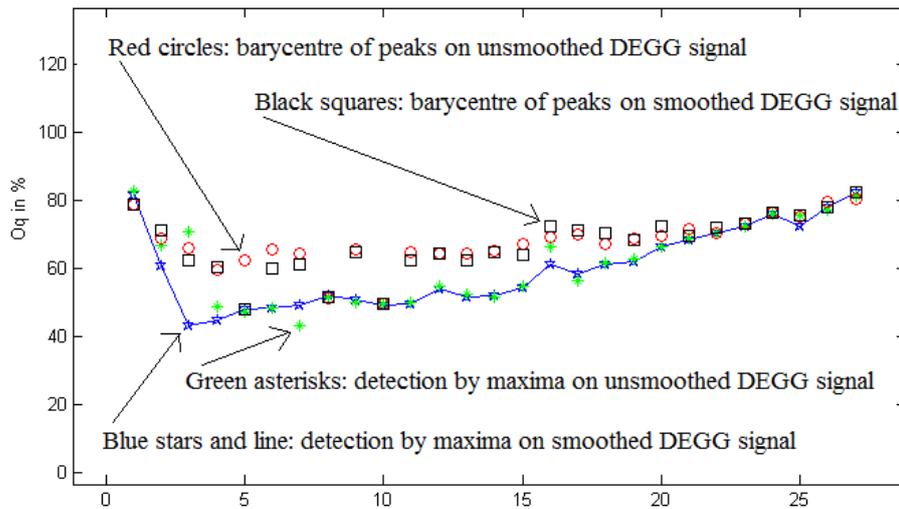


Figure 3.25: Former display (in the 2016 version) of  $O_q$  dEGG calculated in four different ways: (i) maxima on unsmoothed dEGG signal (in green), (ii) maxima on smoothed dEGG signal (in blue), (iii) barycentre of peak on unsmoothed dEGG signal (in red), (iv) barycentre of peak on smoothed dEGG signal (in black).

Consequently, the result of four methods will be different. The difference is often conspicuously visible as a broad scatter (a clear mess) of  $O_q$  dEGG curves, as in Figure 3.26. A careful decision needs to be made as to which method is appropriate, and which values among the set yielded by this method should be retained. By a visual verification of opening peaks in the dEGG signal, we can choose a method which accurately reflects the well-defined peaks. Then, the values corresponding to imprecise peaks need to be suppressed.

This does not appear unfeasible, but requires great care, and an investment of time. For instance, the analysis of total 660 items for data of each speaker, it took at least three days for careful examination and decision-taking. A practical tip to facilitate this stage is to use an extended screen (as in Figure 3.27) so that we can have two windows at the same time: (i) one screen shows four plots for verification on the signals, and (ii) one screen shows the COMMAND WINDOW for executing selections on  $f_0$  dEGG and  $O_q$  dEGG after verifying on the signals. This is very convenient and time-saving, especially when processing a large amount of data.

Figure 3.28 is a good example illustrating a case where there are both precise and imprecise opening peaks, and thus I will be able to explain how the verification step of  $O_q$  dEGG carried out in general.

In this example, the opening peaks in the middle (from cycles 11 to 24) are precise with one prominent peak in each cycle, as can be seen in the zoom in of Figure 3.28a.

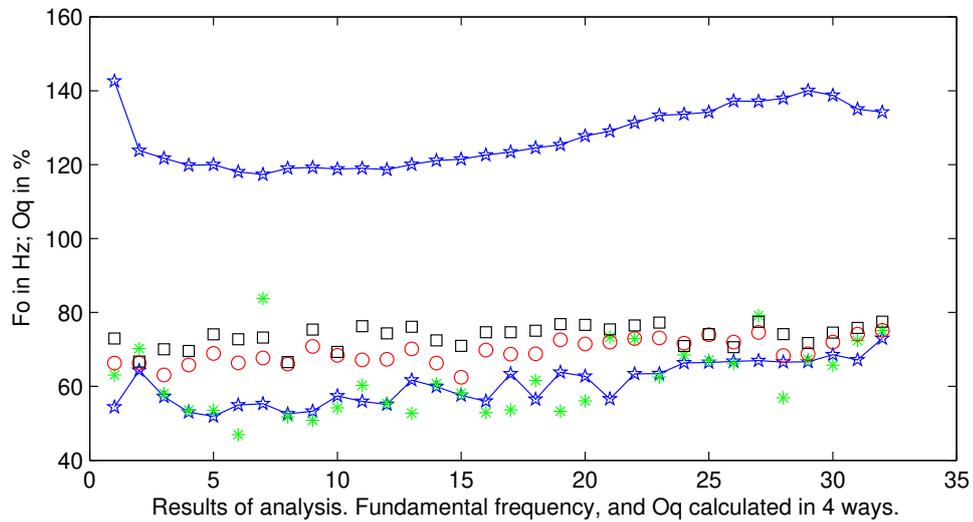


Figure 3.26: An example of  $O_q$  dREGG curves with large differences across methods as a result of imprecise opening peaks. Reproduced from M.-C. Nguyễn (2016): data of speaker M1, experiment 2, item 1331.

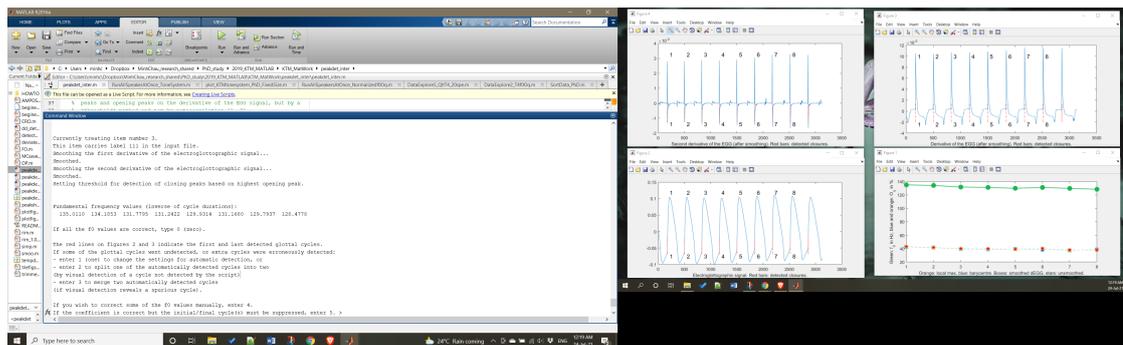
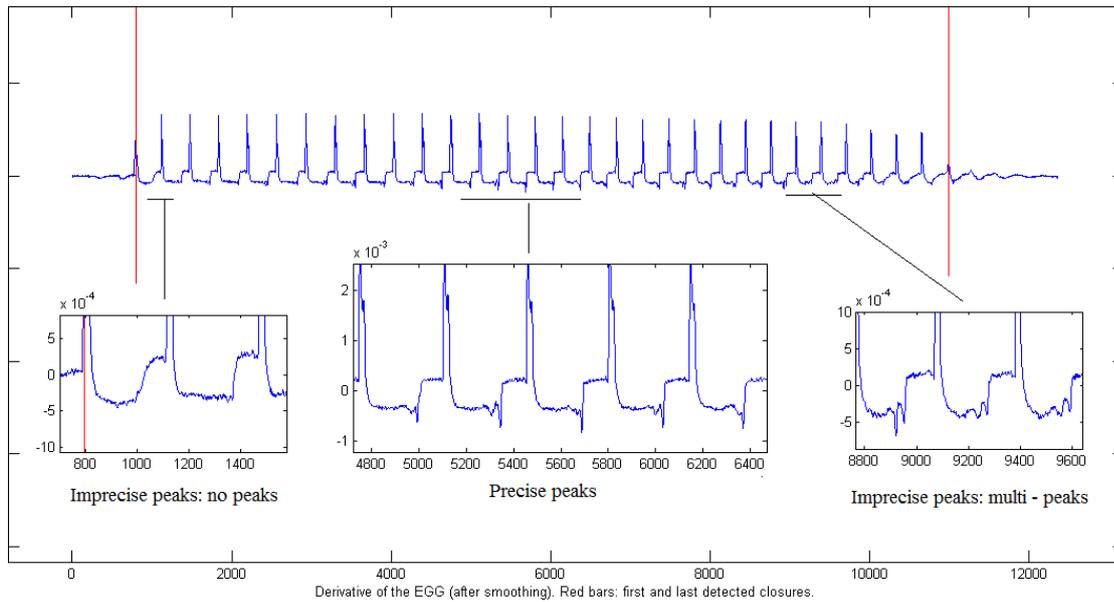


Figure 3.27: An important tip when processing with PEAKDET: use two screens for a convenient view and time-saving operation.



(a) A verification on dEGG signal

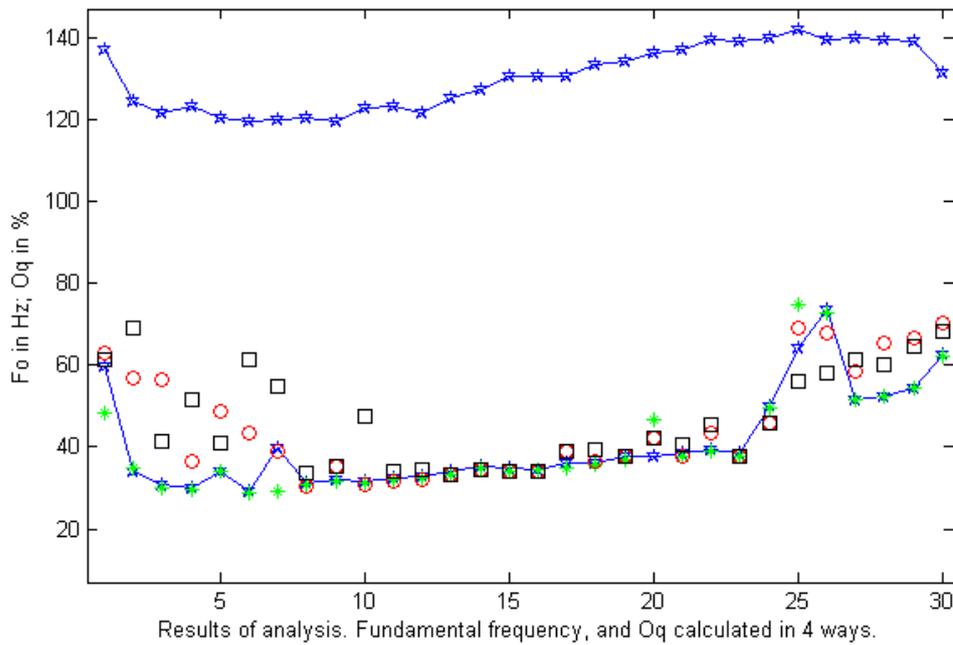
(b)  $f_{o \text{ dEGG}}$  (top) and  $O_{q \text{ dEGG}}$  (bottom) results

Figure 3.28: An example of item with imprecise opening peaks. Reproduction of (M.-C. Nguyễn, 2016): item carrying UID 1331, speaker M1, experiment 2. Abscissa: in samples (1 sample =  $1/44,100$  second).

This corresponds to good consistency across the four methods of calculating  $O_{q \text{ dEGG}}$  at the bottom of the figure 3.28b in these cycles.

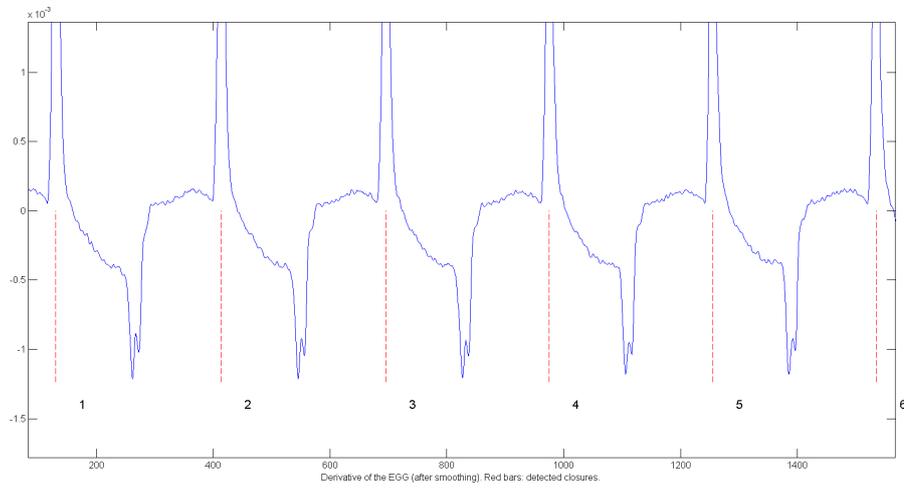
On the other hand, the fit across methods is lost at both ends (beginning and end of the rhyme). This is caused by imprecise opening peaks that can be classified into two kinds: (i) the first two peaks at the beginning without any precise opening peaks and (ii) the rest of the beginning at cycles 3 to 10 and cycles 25 to 30 at the end with more than one peak standing out, which is impossible to identify which is the real opening peak.

The treatment for this case is as follows: First, I decided to select the method of maxima on the smoothed signal by entering '1' in the COMMAND WINDOW. Then, all values from 1 to 10 at the beginning and from 25 to 30 at the end were suppressed by indicating '1:10' and '25:30' also in the COMMAND WINDOW.

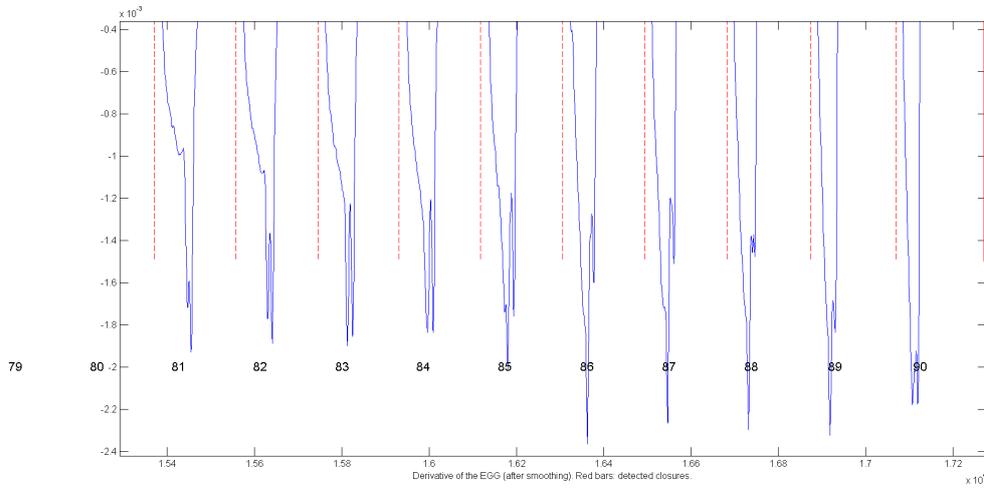
More generally, the basic processing in this step includes:

- The interface of **PEAKDET**: is shown as Figure 3.27 with four plots of (i)  $f_{o \text{ dEGG}}$  and  $O_{q \text{ dEGG}}$ , (ii) EGG, (iii) dEGG, and (iv) ddEGG in four separate figures, and simultaneously in the COMMAND WINDOW the instructions and the requests for  $O_{q \text{ dEGG}}$  verification are displayed as Figure 3.30.
- User's action #1: First, consider the figure of the results of  $f_{o \text{ dEGG}}$  and  $O_{q \text{ dEGG}}$  to check whether the calculation of the four methods is consistent. Second, select and zoom in on the figure of dEGG signal to verify all the opening peaks and pay particular attention to the part(s) where the methods do not match, or even differ widely. The next step is to decide which method to choose by considering the one that can detect the majority of the precise peaks. Usually, I have chosen either (i) the method of maxima on smoothed signal (typing 1) in the case where there is one and only one precise opening peak in each period; or (ii) the method of barycentre of peaks on smoothed signal (typing 3) in cases where there are multiple peaks during the opening phase but one really prominent peak can be detected and the others are close. Therefore, the results of the four methods will not be too different and justified to be close to the exact result. The acceptable difference used in this study is 5 %. Figure 3.29 provides some specific cases where selecting the method of the barycenter of the peaks on the smoothed signal (typing 3) is highly recommended.
- User's action #2: after a method has been selected (in command #1 of Figure 3.30, the next step is to suppress all imprecise peaks that are visually detected in the figure of dEGG by specifying in the command #2 the order number(s) of these cycles that are marked at the top and bottom of each cycle. Those suppressed values are set to zero, by convention, and displayed in the command window as well as in a new plot of only the  $O_{q \text{ dEGG}}$  values for the selected method. When all imprecise peaks are suppressed, type 'o' to finish and move on to the next token.

In this semi-automatic procedure, decisions are left to the user's appreciation. This may appear as a less safe path than fully automatic measurement, as it raises issues of reproducibility and replicability. Whether another researcher takes up the same data

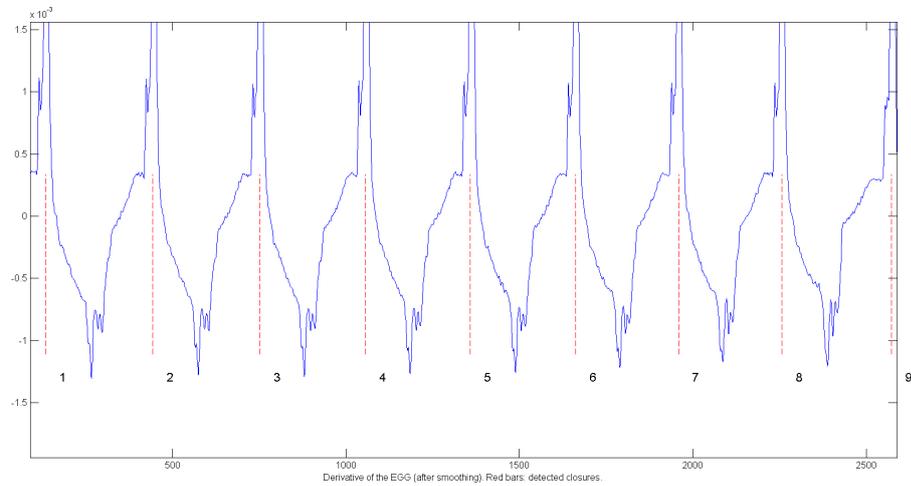


(a) Case #1: the top of the peak is divided into two small peaks that are really close to each other.

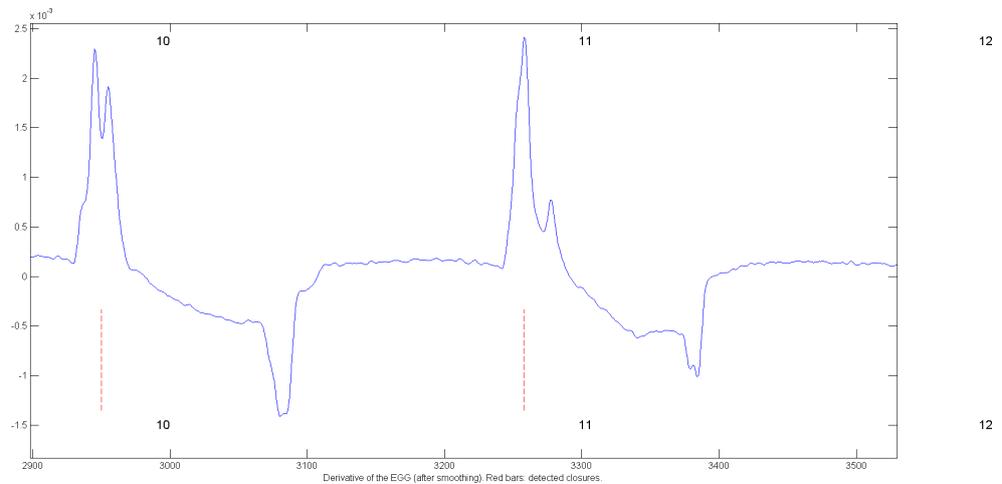


(b) Case #2: similar to case #1 with the top of the peak are split but more severe.

Figure 3.29: Some specific cases where selecting the method of the barycenter of the peaks on the smoothed signal (typing 3) is highly recommended.



(c) Case #3: there are multiple peaks during the opening phase but one peak really stands out and the other micro-peaks are nearby.



(d) Case #4: there is only one opening peak per period but the top of the peak is rather flat instead of being a clear spike.

Figure 3.29: Some specific cases where selecting the method of the barycenter of the peaks on the smoothed signal (typing 3) is highly recommended.

<pre> 1 Please select Oq results among the four sets obtained by different methods: 2 - by maxima on unsmoothed signal (shown as red stars): enter 0. 3 - by maxima on smoothed signal (shown as orange squares): enter 1. 4 - as a barycentre of peaks on unsmoothed signal (shown as blue stars): enter 2. 5 - as a barycentre of peaks on smoothed signal (shown as blue squares): enter 3. 6 To exclude all Oq values for this item, enter 4. 7 Your choice : 1 8 Open quotient values : 9 43.1193 41.7683 40.0000 39.8810 39.2330 40.3561 38.0531 37.7907 10 11 If values need to be suppressed, enter their index in vector: 12 for instance, 2 for 2nd value, 5:15 for values from 5 to 15. 13 If all the values are correct now, type 0. Your choice : 1 14 The specified value was 43.1193. 15 It is now set at zero, and will be excluded from the calculations. 16 Refer to Figure 5 to see modified curve. 17 Open quotient values : 18 0 41.7683 40.0000 39.8810 39.2330 40.3561 38.0531 37.7907 19 20 If values need to be suppressed, enter their index in vector: 21 for instance, 2 for 2nd value, 5:15 for values from 5 to 15. 22 If all the values are correct now, type 0. Your choice : </pre>	<p># 1: Choose a method that best reflects real and precise opening peaks.</p> <p># 2: Indicate the erroneous calculation cycle(s) because of imprecise peaks.</p> <p># 3 - # n: Repeat the above step (#2) until all erroneous values are suppressed. Select 0 to move on to the next token.</p>
--	---

Figure 3.30: Actions in the COMMAND WINDOW for the step of  $O_q$  dEGG verification

set (reproducing the study) or collects new data (replicating the study), they will not arrive at the same exact results, since there is a stage of decision-making that relies on subjective appreciation. Importantly, the values yielded by the four methods are all stored in the results matrix, and none is deleted during manual verification: instead, the values selected by the user are stored in a different column in the results matrix (the tenth and last column). It therefore remains possible, at any point in data processing further down the line, to compare the values chosen by the user to those yielded by fully automatic detection, or to compare the four methods. In other words, manual verification adds information to the results, and does not remove any.

**Output:** After all the tokens have been thus inspected and processed, i.e. when  $f_{0\text{dEGG}}$  and  $O_{q\text{dEGG}}$  have been calculated for all items in my data, **PEAKDET** saves the entire workspace in a “.mat” file, as a final output. The results of analysis are in a three-dimensional matrix called <data>, containing one two-dimensional matrix (hereafter referred to, for short, as a *sheet*) per token analyzed. It means that the <data> matrix for each speaker should contain 660 sheets, corresponding to 660 rhymes.

Each sheet presents the main results of a token in a single matrix including 10 columns x 100 lines. The contents of the 10 columns are explained in Table 3.6. The 100 lines are the default space corresponding to a maximum of 100 glottal cycles of one syllable rhyme. This length is automatically extended in cases where a syllable rhyme contains more than 100 cycles (Muong, like Vietnamese, can have really long syllables). As presented in Table 3.7, six out of twenty data files have more than 100 lines, which means that six of the twenty speakers produced at least one syllable rhyme with more than 100 cycles. The lines are extended up to 178 in the data of speaker F10 (who thereby wins the title of the speaker with the longest rhyme).

## 3.4 Plotting and data visualization

### 3.4.1 Visualization of results

After obtaining the final result (in .mat file) from **PEAKDET** for 20 speakers, the last step of this data processing is to plot the data in different ways to serve for the visualization and analysis of the result. Thanks to the available plotting scripts in my

Table 3.6: `PEAKDET` output information: .mat file containing a 10×100 matrix

Column	Value	Unit
1st	the beginning of cycle	ms
2nd	the end of cycle	ms
3th	f <sub>0</sub>	Hz
4th	DECPA: Derivative-Electroglottographic Closure Peak Amplitude	
5th	O <sub>q</sub> determined from raw maximum without smoothing	%
6th	DEOPA	
7th	O <sub>q</sub> determined from maximum after smoothing	%
8th	O <sub>q</sub> determined from peak detection without smoothing	%
9th	O <sub>q</sub> determined from peak detection after smoothing	%
10th	The open quotient values retained after user's verification	%

master study, I could easily and quickly use the same scripts (as can be consulted here, in M.-C. Nguyễn (2016, pp. 117–126)) with a few modifications so that they could be applied to the current data and run for all speaker data at once.

The available scripts were served for plotting the curves average f<sub>0 dEGG</sub> and O<sub>q dEGG</sub> of the tone system in a graph by several ways, including 5 types of graph:

1. average f<sub>0 dEGG</sub> curves with standard deviation
2. average O<sub>q dEGG</sub> curves with standard deviation
3. average f<sub>0 dEGG</sub> curves with no standard deviation
4. average O<sub>q dEGG</sub> curves with no standard deviation
5. average f<sub>0 dEGG</sub> curves in semitone.

All these types of graphs are plotted and available for the current data. However, due to the huge amount of figures, for the sake of a simple and efficient observation, we only provide in the Result chapter the figures of the average f<sub>0 dEGG</sub> and O<sub>q dEGG</sub> curves with standard deviation by 20 speakers, as can be seen in Figure 4.1. In addition, new plottings of data normalization for were added. The tone system (the five smooth tones) normalized by gender (one graph grouping data from the 10 men and one for the 10 women) and normalized over the entire group: 20 speakers (Figure 4.2). The formulas used to obtain the relative values of f<sub>0 dEGG</sub> and O<sub>q dEGG</sub> are taken from the 2008 study of the Tamang language by Mazaudon and Michaud (2008, p. 238). The following two (simple) equations are therefore reproduced from this study with permission.

The formula used for obtaining relative f<sub>0</sub> values (in semitones) is the following in Equation 3.1, where F<sub>REL</sub> is the relative value (in semitones), F<sub>TARGET</sub> the measurement on the target syllable (in Hertz) and F<sub>FRAME</sub> the measurement over the frame (in Hertz):

$$F_{REL} = 12 \times \frac{\log\left(\frac{F_{TARGET}}{F_{FRAME}}\right)}{\log(2)} \quad (3.1)$$

O<sub>q</sub> values were also recalculated, relative to a mean value across speakers. In view

of the strong cross-speaker differences in mean  $O_q$ , shown in Table 4.1,  $O_q$  values were converted using the following formula in Equation 3.2, where  $O_q$  TARGET is the measurement for the glottal cycle at issue, and  $O_q$  MEAN the mean  $O_q$  value across speakers, obtained by averaging across all the syllable rhymes in the corpus, carrier sentences included.

$$O_q \text{ REL} = 100 \times \left( \frac{O_q \text{ TARGET}}{O_q \text{ MEAN}} - 1 \right) \quad (3.2)$$

The same formulas are also applied in figure 4.3 not only for 5 smooth tones but also for 2 checkered tones which are plotted separately to clarify the two sub-systems.

In addition to these sets of figures for the entire tone system, we also have three other types of graphs to examine the glottalized tone in particular and in more detail: (i) a raw display, i.e. one curve corresponds to one item (Figure 4.4); (ii) a comparison on the distribution (Figure D.2) and the correlation (Figure D.3) of  $f_o$  dEGG and  $O_q$  dEGG across speakers. These figures are simply produced by using the plotting functions available in MATLAB such as [boxplot](#) and [scatter plot](#).

### 3.5 Data archiving and publishing

Long-term data preservation is central to the work reported in this dissertation. The underlying principles are currently grouped under the notion of Open Science, and gather increasingly strong support from policy makers and academic institutions, but they were familiar to linguists long before the notion of Open Science entered common usage. The long quotations below recapitulate decisive reasons to give data curation a central place in the context of field research.

Primary outputs of field research (lexicon, transcripts and interlinear glossed text collections, and their associated media) need to be coded and preserved. Long-term access to these data is addressed by the establishment of archives that also act as the locus for training and advocacy for well-formed data. (Thieberger and Jacobson, 2010, p. 147)

The field records linguists produce are meant to endure and to be available to the people we record and their communities, as well as to fellow researchers well into the future. Archiving is no longer something we do at the end of our fieldwork. It is apparent now that it should be integrated into everyday language documentation work and that it is a crucial aspect of documentary linguistics. (...) Recent technological advances have pointed to the importance of planning data management and workflow for ethnographic recording. (...) [W]e must, right from the moment of recording, be concerning with making good documents and placing them into a suitable archive for storage and discovery. Thus, we can distinguish archival practice, a process resulting in well-formed archival data, from archival storage in a repository. (Thieberger and Jacobson, 2010, p. 148)

In a perspective of *progressive archiving*, initiated since the early stages of data collection, the data were curated with a view to an early deposit in an archive ensuring long-term preservation: the Pangloss Collection, itself a part of a nationwide archival setup (Michailovsky et al., 2014). Data were deposited in 2020-2021, and thus received a Digital Object Identifier (DOI) in time for me to use these identifiers to propose links from the PDF version of the present work to the data. (On the process of assigning DOIs to linguistic data, and their relevance to linguistic research, see Vasile et al. 2020.)

### 3.6 A bird's eye view of the full data set for the main experiment

As a conclusion to the present methodological chapter, this section is devoted to a brief summary of the full data set for the main experiment in this study: recapitulating its design and providing an inventory of the collected materials. This will give the reader an overview of the basis that is available for this study.

**Speech materials:** This study is based on the recording (acoustic signal recorded simultaneously with EGG signal) of twelve minimal sets that distinguish five tones in smooth syllables (including eight complete minimal sets and four near-minimal sets) and three checked minimal pairs which distinguish two tones in checked syllables. The list of these sets and pairs is provided in Tables 3.1 and 3.2. Each item in this list has an associated target syllable that is a monosyllable. Therefore, we have a total of 66 target syllables.

**Carrier sentence:** In order to stabilize the phonetic context, a standard method is to ask speakers to pronounce the target words in a carrier sentence. As set out in Section 3.1.2, the carrier sentence is a short sentence constituted of 4 words: three frame words and one target word. The process of a recording is that the speaker says all the target words of each set in isolation, then repeats them with the carrier sentence, and so on from the first to the twelfth set and from the first to the third pair. The whole procedure is then repeated a second time.

**Sequence of an experimental session:** In the real performance of recording sessions, speakers were required to speak aloud from the minimal set N<sup>o</sup>1 to N<sup>o</sup>12, from minimal pair N<sup>o</sup>1 to N<sup>o</sup>3. For each minimal set, they first speak the five target syllables in isolation before repeating them once more but in the carrier sentence. This is exactly the same with 3 minimal pairs.

**The total corpus (per speaker):** Figure 3.31 recapitulates the total corpus of this study. Not only the target words but also the three frame words of the carrier sentence are annotated and processed. Thus, for each speaker, we have a total of 660 items, of which 264 items are target syllables and 396 items are frame syllables. A more detailed list of the amount of materials is given in Table 3.32. In some cases, the maximum number of items is not reached because some frame words are missing, as speakers tend to shorten the carrier sentence during a series of repetitions. The most serious case is in the data of the speaker F10. For some technical reason, we made a pause but mistakenly did not press the record button to resume, so the last part of

3.6 A bird's eye view of the full data set for the main experiment

$$((5 \times 12) + (2 \times 3) + (5 \times 12 \times 4) + (2 \times 3 \times 4)) \times 2 = 660 \text{ (items)}$$

Elements	Meaning
5	5 tones in smooth syllables
12	12 minimal sets (of 5 smooth tones)
2	2 tones in checked syllables
3	3 minimal pairs (of 2 checked tones)
4	4 syllables of the carrier sentence (1 target word + 3 frame words)
2	2 repetitions

Figure 3.31: Calculation of the total corpus

Total corpus: 660 items		
Target syllables: 264 items		Frame syllables: 396 items
<b>In isolation:</b> 132 items	<b>In carrier sentence:</b> 132 items	
- 24 tokens each smooth tone (x 5 tones) - 6 tokens each checked tone (x 2 tones)	- 24 tokens each smooth tone (x 5 tones) - 6 tokens each checked tone (x 2 tones)	- /ja <sup>2</sup> / : 132 tokens - /mãt <sup>6</sup> / : 132 tokens - /cãŋ <sup>3</sup> / : 132 tokens

Figure 3.32: A brief summary view of the corpus

the experiment was missed on the first run. In particular, the minimal set N<sup>o</sup>11 in carrier sentence, the minimal set N<sup>o</sup>12 and all three minimal pairs both in isolation and in carrier sentence were not recorded. As a consequence, this data lacks 75 items, including 11 target words in isolation and 16 target words in carrier sentence, which leads to the sorely felt absence of 48 frame words at all three positions (i.e. 16 items for each).

The actual status of the data of each speaker is summarized in Table 3.7. There are a total of 26 participants, 28 data files (F1 and M12 perform the experiment 2 times) twenty of which have been processed.

Beyond this quick inventory focusing on the main experiment used in this study, a list of recorded files is provided in Appendix B at the end of this volume.

Table 3.7: Current status of corpus: 20/28 data files have been annotated with SOUND FORGE and processed with MATLAB.

N <sup>o</sup>	Speaker	Quality of EGG signal	Data status	Size of .mat file
1	F1	Crackling noise	No annotation	No analysis

2	F1	Crackling noise	No annotation	No analysis
3	F3	Good	660/660 items	100×10×660
4	F7	OK	660/660 items	119×10×660
5	F8	Weak EGG	No annotation	No analysis
6	F9	Good	660/660 items	119×10×660
7	F10	Good	585/660 items Missing 75 items - 11 target words in isolation - 16 target words in carrier sentence - 16 frame words at 1st position - 16 frame words at 3rd position - 16 frame words at 4th position	178×10×585
8	F11	Weak EGG	No annotation	No analysis
9	F12	Good	660/660 items	111×10×660
10	F13	Good	646/660 items Missing 14 frame words at 1st position	133×10×646
11	F14	OK but there are a few flat segments	No annotation	No analysis
12	F16	Weak EGG	No annotation	No analysis
13	F17	Good	660/660 items	100×10×660
14	F18	Weak EGG	No annotation	No analysis
15	F19	Good	660/660 items	100×10×660
16	F20	OK	660/660 items	100×10×660
17	F21	OK	660/660 items	111×10×660
18	M1	Good	660/660 items	100×10×660
19	M5	OK	660/660 items	100×10×660
20	M7	Good	660/660 items	100×10×660
21	M8	OK	660/660 items	100×10×660
22	M9	Good	660/660 items	100×10×660
23	M10	Good	660/660 items	100×10×660
24	M11	Good	660/660 items	100×10×660
25	M12	Signal out of range	No annotation	No analysis
26	M12	Good	660/660 items	100×10×660
27	M13	Good	660/660 items	100×10×660

3.6 A bird's eye view of the full data set for the main experiment

28	M14	OK	656/660 items Missing 4 frame words: 3 at first position and 1 at 4th position	100×10×656
----	-----	----	---	------------



# Chapter 4

---

## Results

### 4.1 The Kim Thuong Muong tone system: observations based on fundamental frequency and open quotient curves averaged by speaker

The experimental setup and analysis method set out in the previous chapter, involving an experimental routine performed by 20 speakers, now puts us in a position to set out quantitative results about the tone system of Kim Thuong Muong.<sup>1</sup>

In this chapter, the figures representing the results will be set out in the following order:

- The tone system (the five smooth tones) as realized by the twenty speakers, shown speaker by speaker in the H z scale (Figure 4.1) (and the figures in semitones (Figure D.1) for  $f_{0 \text{ dEGG}}$  parameter are provided in Appendix D as a reference);
- The tone system (the five smooth tones) averaged across speakers. First normalized by gender, with one graph grouping data from the 10 men and one for the 10 women. Then normalized over the entire set of speakers: a third graph averaged over all twenty speakers (Figure 4.2);
- A comprehensive look at the full tonal system, including the two checked tones alongside the five smooth tones<sup>2</sup> (Figure 4.3);
- A particular description on the glottalized tone of each speaker, shown in two ways: (i) as a raw display, where one curve corresponds to one item (Figure 4.4), and (ii) a comparison on the distribution (Figure D.2).

Figure 4.1 provides curves of  $f_{0 \text{ dEGG}}$  and  $O_{q \text{ dEGG}}$  averaged by speaker, laying a basis for a large part of the analyses that follow. The data from the ten female speakers are provided first, in sub-figures (a) to (j), followed by data from the ten

---

<sup>1</sup>In this long chapter, I unpack a great deal of minor details and observations across tones and speakers, and thus the organization and division of sections and subsections is somewhat relative and, in some respects, a bit messy. Apologies if the reader finds it a bit difficult to follow along here.

<sup>2</sup>Remember that ‘smooth syllables’ are those without final stops, and hence, the tones that appear on these syllables are commonly referred to in Southeast Asian linguistics as ‘smooth tones’. Syllables with final stops are referred to as ‘stopped syllables’ or ‘checked syllables’, and accordingly, their tones are referred to as ‘stopped tones’ or ‘checked tones’. This distinction relates to syllable structure, and not to the phonetics of the tones themselves. Thus, paradoxical as it may seem, **Tone 4** appears on ‘smooth’ syllables and hence is one of the five ‘smooth’ tones, despite being, in impressionistic terms, not at all smooth to the ear, due to its creaky portion.

male speakers in sub-figures (k) to (t). Let us now look at these results, taking a quick tour of the individual results of all twenty speakers while keeping an eye on the cross-speaker averages in Figure 4.2. The averaged  $f_{o \text{ dEGG}}$  curves of each individual speaker are displayed in absolute values (i.e., on the Hertz scale). For producing cross-speaker averages, on the other hand, the raw  $f_{o \text{ dEGG}}$  values are converted into semitones (relative values), as a simple means to achieve normalization across speakers. The formulas used to obtain the relative  $f_{o \text{ dEGG}}$  values (in semitones) and the  $O_q \text{ dEGG}$  values were adopted from Mazaudon and Michaud (2008, p. 238). This method is also similar in its essentials to that used in our earlier study of the Naxi language (Michaud, Vaissière, and M.-C. Nguyen, 2015).

The representations shown in Figure 4.1 for twenty speakers were obtained from the analysis of twelve minimal sets illustrating the five tones (full detail on data collection and processing was provided in Chapter 3). Within the sub-figure for each speaker, the left-hand side represents  $f_{o \text{ dEGG}}$  curves, and the right-hand side represents the  $O_q \text{ dEGG}$  measurements at the same time points.

The average values of  $f_{o \text{ dEGG}}$  and  $O_q \text{ dEGG}$  shown in Table 4.1 are used as reference values to describe and evaluate all the results in this chapter. The values were obtained by averaging all values for syllable rhymes (target syllables and carrier sentences) for a given speaker, as a rule-of-thumb reference value of the speaker's mean  $f_o$  and  $O_q$ , for purposes of elementary normalization. The ratio of excluded  $O_q$  reflects the proportion of  $O_q$  values that were manually excluded at the stage of semi-automatic data processing, following a procedure set out in Chapter 3. It provides an indication on the overall representativity of the  $O_q$  values shown in the figures: thus, the fact that 88% of all measurements had to be excluded for speaker F20 means that estimation of  $O_q$  was exceptionally difficult for this speaker, and the curves cannot be considered as accurate as those for the other speakers, with exclusion rates ranging from 4.7% to 36.3%.

Let us start out from observations based on examining Figure 4.1. In general, the  $f_{o \text{ dEGG}}$  and  $O_q \text{ dEGG}$  curves of 20 speakers demonstrate agreement on the basic tonal patterns briefly summarized in Table 2.6 of Chapter 2.

#### 4.1.1 *Fundamental frequency*

In terms of fundamental frequency, the tone system appears relatively symmetrical. **Tone 1** is phonetically flat, located around the middle of the speaker's range. In contrast to this tone, **Tone 5** is another flat tone but at a higher pitch level. Indeed, it is the highest tone in the system, with  $f_{o \text{ dEGG}}$  values consistently at the top of the speaker's range. **Tone 2** and **Tone 3** are characterized by falling and rising contours respectively. The amplitude of the rise in Tone 3 is sometimes not as great as that of the fall in Tone 2, as one would expect if one assumes a small effect of declination in  $f_o$  (an effect found in declarative utterances). Last but not least, **Tone 4** has a pattern that looks symmetrical along the time scale. It has a descending-ascending contour with offset values close to the onset values, and a glottalized portion in-between.

Table 4.1: Some general information about the speakers' data: The values of the mean and standard variation of  $f_{o \text{ dEGG}}$  and  $O_{q \text{ dEGG}}$  and the ratio of the excluded  $O_{q \text{ dEGG}}$  values obtained by the quantitative analysis of the electroglottographic signal on twenty speakers.

Speaker	Mean $f_{o \text{ dEGG}}$ (Hz)	Std $f_{o \text{ dEGG}}$	Mean $O_{q \text{ dEGG}}$ (%)	Std $O_{q \text{ dEGG}}$	Ratio of excluded $O_{q \text{ dEGG}}$ values (%)
F3	185	35	59	9.9	25
F7	220	35.6	65.5	7.8	25
F9	251	48.8	52.5	10.8	36.3
F10	215	37.4	56.2	9.7	26.2
F12	192	43.4	52.4	8.8	12
F13	260	53.3	52.2	8	12.3
F17	182	40.6	45	9.7	14.5
F19	185	29.8	56.4	11.5	21
F20	224	46.1	57.5	13.7	88
F21	213	41.7	61.1	8.8	7.42
M1	142	26.6	44.5	11.8	17.3
M5	163	32.3	52.9	8	8.7
M7	131	25.6	48.9	9.1	6.9
M8	116	21.3	51.5	7.6	24
M9	165	37.7	58.2	10	4.7
M10	101	19.5	49.1	10.9	17.4
M11	118	24.9	49.4	9.8	13.4
M12	226	47.9	54	6.4	23.3
M13	90	15	55.4	8.6	31.4
M14	151	24.7	54.8	6.3	23.5

Chapter 4 Results

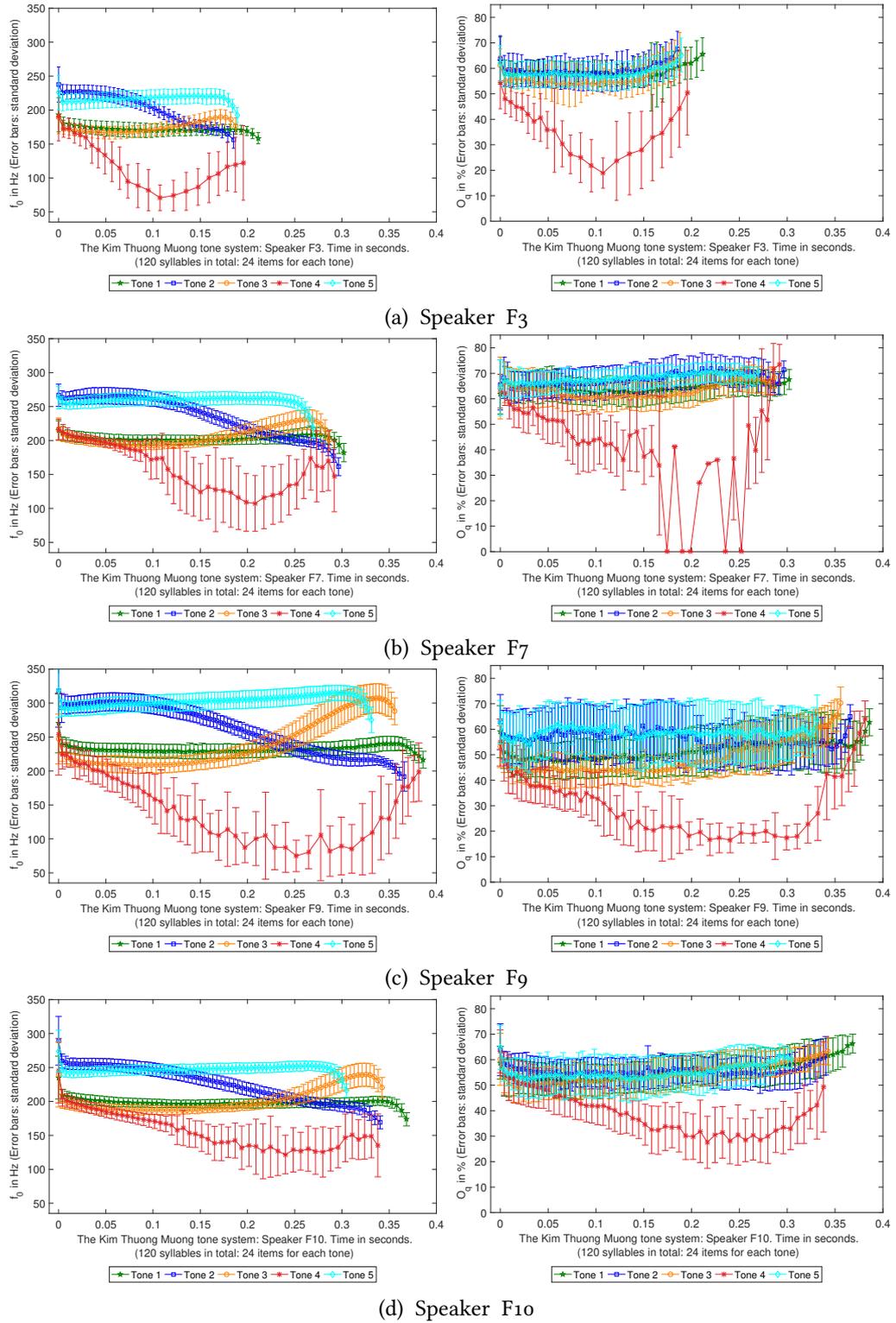
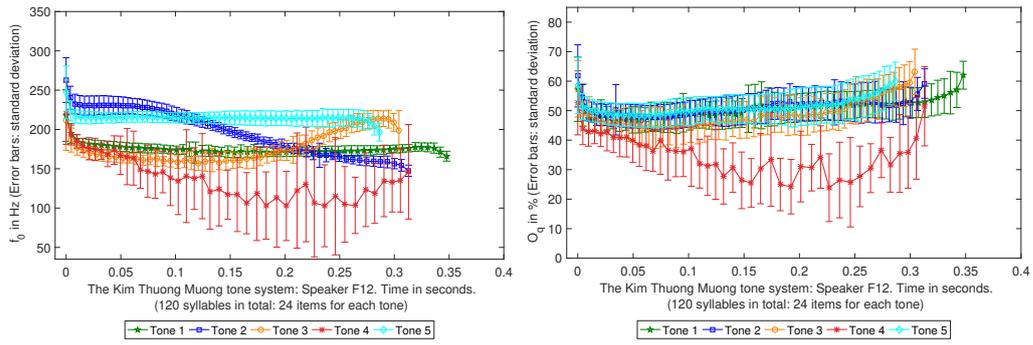
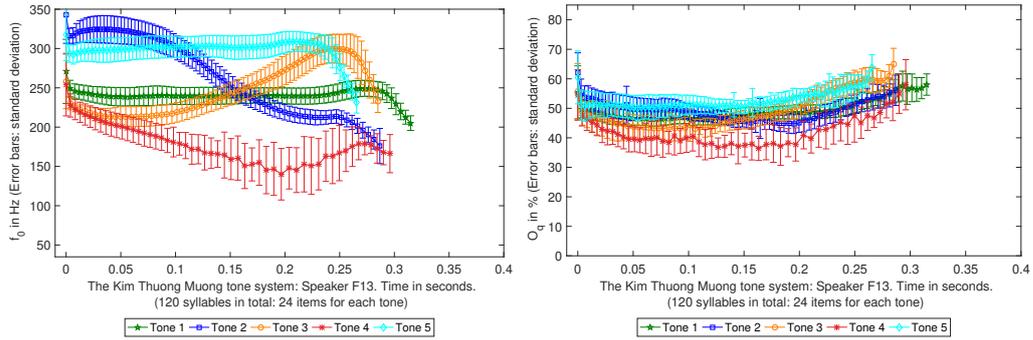


Figure 4.1: The tone system of Kim Thuong Muong: speaker by speaker.  $f_0$  dEGG on the left and  $O_q$  dEGG on the right.

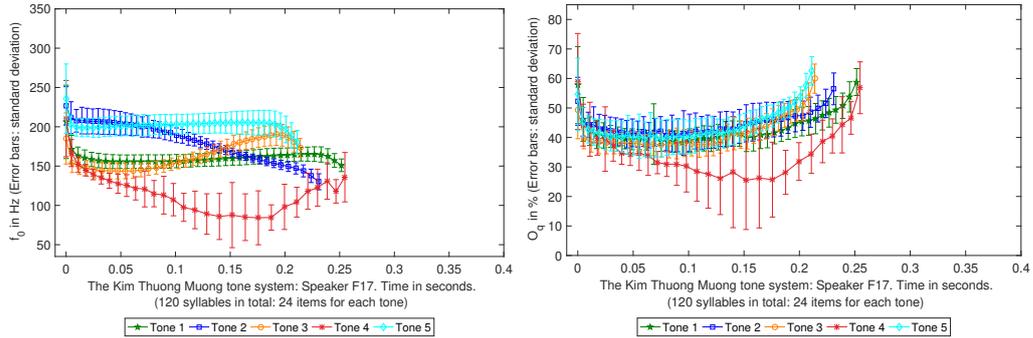
4.1 The Kim Thuong Muong tone system: observations based on fundamental frequency and open quotient curves averaged by speaker



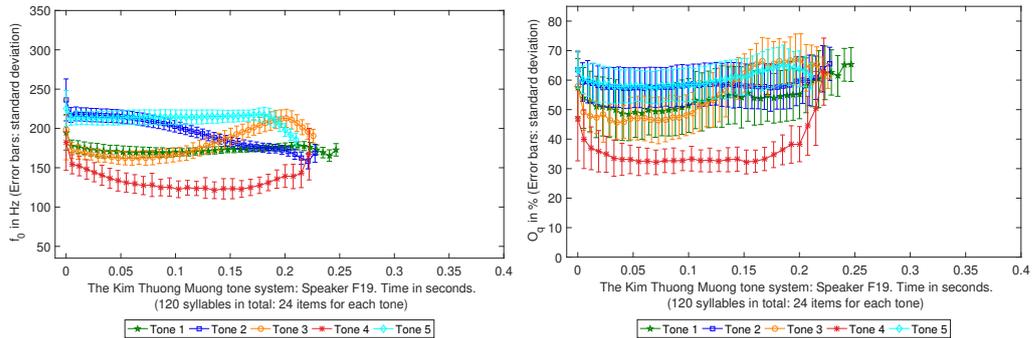
(e) Speaker F12



(f) Speaker F13



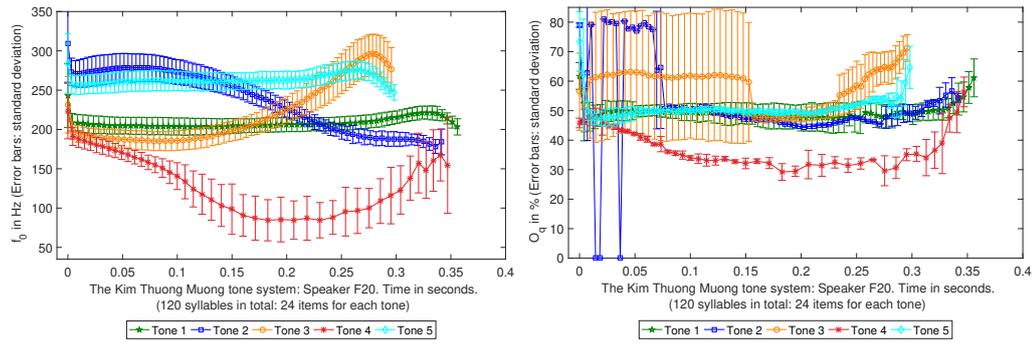
(g) Speaker F17



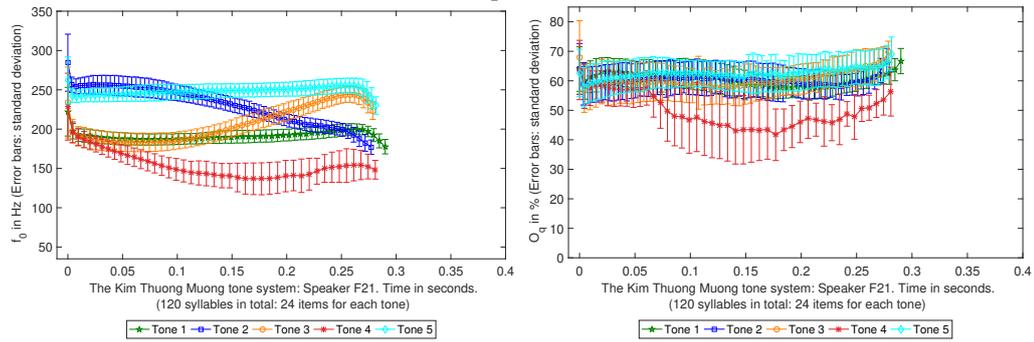
(h) Speaker F19

Figure 4.1: The tone system of Kim Thuong Muong: speaker by speaker.  $f_0$  dEGG on the left and  $O_q$  dEGG on the right.

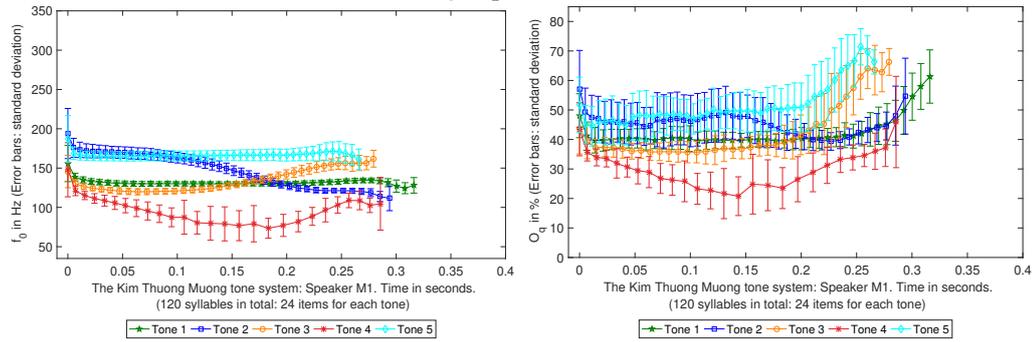
Chapter 4 Results



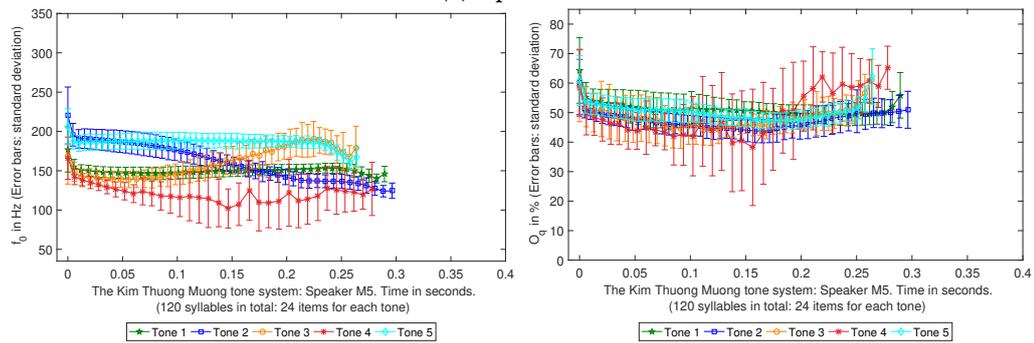
(i) Speaker F20



(j) Speaker F21



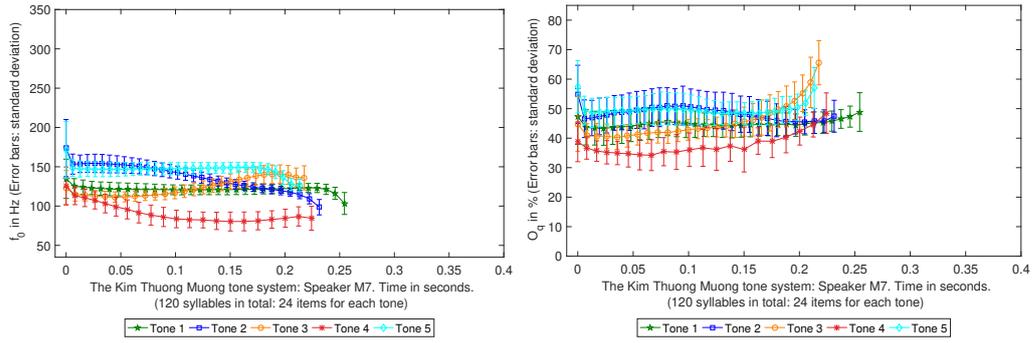
(k) Speaker M1



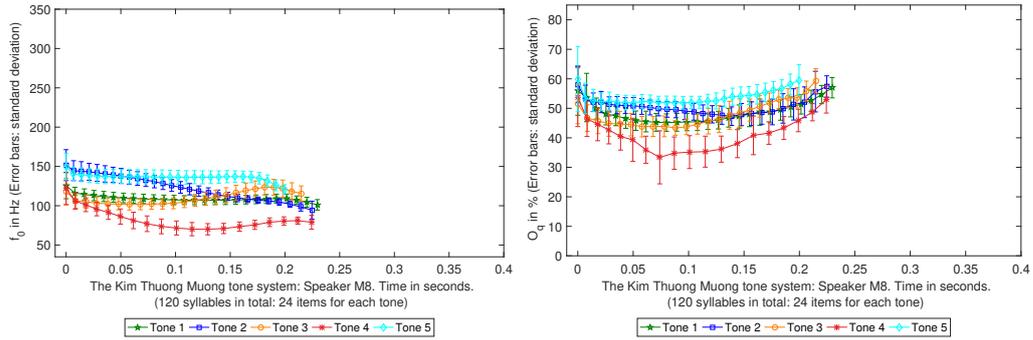
(l) Speaker M5

Figure 4.1: The tone system of Kim Thuong Muong: speaker by speaker.  $f_0$  dEGG on the left and  $O_q$  dEGG on the right.

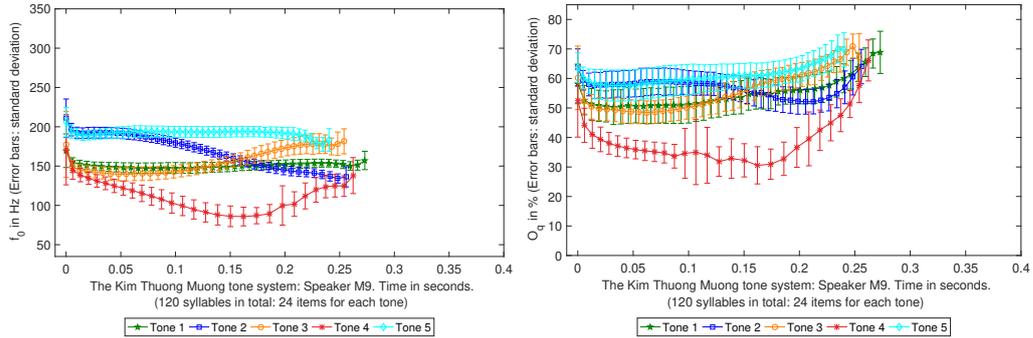
4.1 The Kim Thuong Muong tone system: observations based on fundamental frequency and open quotient curves averaged by speaker



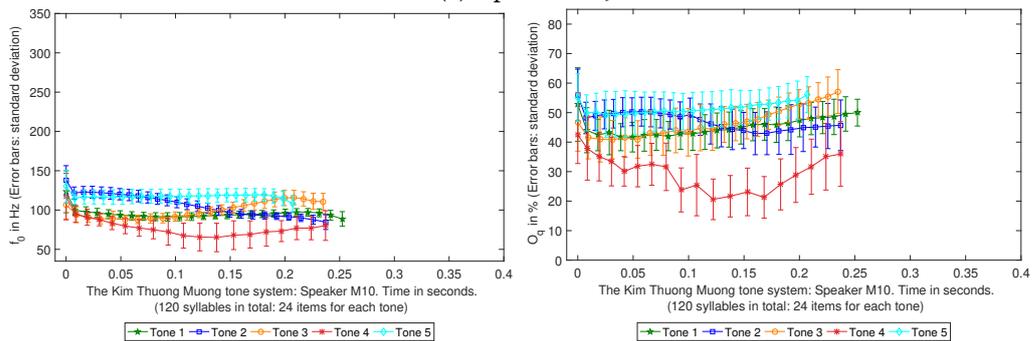
(m) Speaker M7



(n) Speaker M8



(o) Speaker M9



(p) Speaker M10

Figure 4.1: The tone system of Kim Thuong Muong: speaker by speaker.  $f_0$  dEGG on the left and  $O_q$  dEGG on the right.

Chapter 4 Results

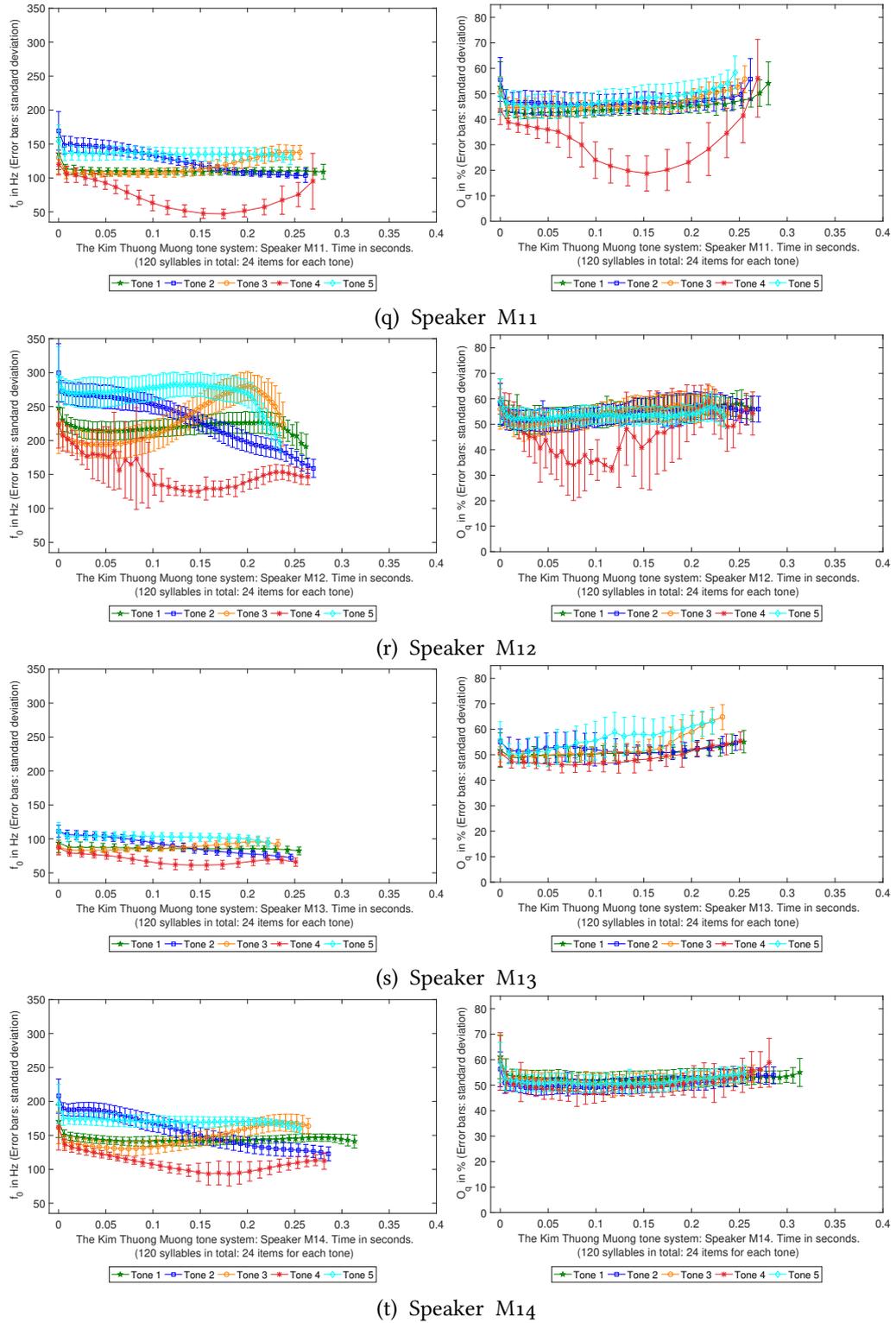


Figure 4.1: The tone system of Kim Thuong Muong: speaker by speaker.  $f_0$  dEGG on the left and  $O_q$  dEGG on the right.

### 4.1.2 Open quotient

In terms of open quotient, according to my studies on this tone system so far (M.-C. Nguyễn, 2016; M.-C. Nguyễn et al., 2019), it appears safe to say that the typical patterns of  $O_{q \text{ dEGG}}$  curves can be separated into two parts. In the upper part, the  $O_{q \text{ dEGG}}$  curves of four tones (Tone 1, Tone 2, Tone 3, and Tone 5) overlap. They are consistently above 40-50%. In contrast, in the lower part of  $O_{q \text{ dEGG}}$  range, we have only Tone 4, on its own. The  $O_{q \text{ dEGG}}$  curve of this tone plunges from initial mid-range values (already lower than the four other tones) to the very bottom of range. A large amount of  $O_{q \text{ dEGG}}$  values of Tone 4 tokens had to be excluded, due to the problem of unclear opening peaks on the derivative of the electroglottographic signal (an issue dealt with in the *Method* chapter (3)). The retained values are lower than 30% and can reach extremely low values: below 15% in some speakers. These values offer telltale evidence of the presence of creaky voice (*phonation mechanism zero* according to the classification of Roubeau, Henrich, and Castellengo 2009).

### 4.1.3 Duration

In terms of duration, there is no great difference among the five smooth tones: less than 0.05 seconds. Tone 1 is the longest tone in most speakers, reflected in a jut-out of the green line on many figures, which is most noticeable in the cases of F3 (4.1a), F10 (4.1d), F12 (4.1e), F19 (4.1h), M7 (4.1m) and M14 (4.1t). In contrast, the other flat tone, Tone 5, appears as the shortest among smooth tones. The average curve for Tone 5 is shorter than that of the other tones in many speakers, although only by a handful of milliseconds. A fairly clear difference can be observed in the cases of speakers F7 (4.1b), F9 (4.1c), F10 (4.1d) and F12 (4.1e), where Tone 5 is shorter than all other tones.

It seems as if there were some kind of relationship between pitch level and duration here. The rising tone (Tone 3) is generally shorter than the falling tone (Tone 2). In the case of the three speakers F20 (4.1i), F21 (4.1j) and M12 (4.1r), there is a pairwise difference in the offset of Tone 1 and Tone 2 versus Tone 3 and Tone 5. Each pair terminates at the same  $f_0$  level and also at the same point in its time course.

It can be noticed that the length of the glottalized tone is, on average, very close to that of the other tones: in this specific experimental condition, Tone 4 cannot be told apart from the others on the basis of its duration. This is different from Vietnamese tone B2, also a glottalized tone, which is about half shorter than the other tones in its system. In Kim Thuong Muong's system, glottalization does not appear to affect the duration of the tones, so Tone 4 ends up at the same offset time as the other tones. Were one to look for fine phonetic detail in patterns of duration, Tone 4 should probably be placed in the same category as Tone 1 and Tone 2: a set of tones that are a bit longer in comparison with Tone 3 and Tone 5.

The relatively longer duration of Muong Tone 4 as compared with the Northern Vietnamese tone B2 can be accounted for in at least two (complementary) ways. First, there is a difference in the location of the glottalization. Whereas Vietnamese B2 has

a final glottalization that constrains duration (by sharply interrupting the syllable), Kim Thuong Muong has a medial glottalization, followed by a return towards modal voice at the end, so the glottalization event does not interrupt the syllable. Second, the nature of the glottalization is different. Whereas Vietnamese B2 is characterized by a *glottal constriction* (Michaud, 2004b; Kirby, 2011), the phonation-type specificity of Kim Thuong Muong **Tone 4** is an event that has *creaky voice* as its canonical characteristic. The constriction in Vietnamese B2 is physiologically a gesture of strong adduction of the vocal folds; it is definitely the factor that makes this tone shorter than the others. By contrast, the Kim Thuong Muong **Tone 4** is not shortened by its glottalization, which does not have a strong degree of constriction.

Due to the time required for a shift to creaky voice (the time for the irregular pulsation pattern to set in), it could even be that the presence of creaky voice in **Tone 4** would tend to lengthen the syllable carrying that tone, rather than shorten it. In some cases, such as speakers F9 (4.1c) and F17 (4.1g), **Tone 4** is the longest tone. But to put the hypothesis to the test, one would need to examine the tone system in a greater diversity of contexts than was done in the present work: in the context under study, the target item is in focus, all syllables are really long, and syllables carrying **Tone 4** do not stand out systematically in terms of duration, as was pointed out above.

Thus, in view of the present experimental results, duration by itself does not appear to play an important role in contrasts among the five tones of smooth syllables. On the other hand, duration differs greatly between smooth and stopped syllables. The two checked tones, i.e. **Tone 6** and **Tone 7**, are only about half as long as tones in smooth syllables, as can be observed in Figure 4.3. This can be ascribed to the influence of obstruent codas. Final /p/, /t/, /c/ and /k/ are phonologically unspecified for voicing (there is no voicing opposition among final consonants in Kim Thuong Muong), but these codas are phonetically unvoiced, and appear to exert the full shortening effect that is cross-linguistically associated with final voiceless stops.

The duration of syllable rhymes is diverse across the 20 speakers, with an average of 0.28 seconds. Speaker F3 (4.1a) has the shortest syllable rhymes, with only 0.20 seconds. This value is roughly equal to just half the duration of the tones of F9 (4.1c), the speaker who produced the longest rhymes in this experiment. The rest range from 0.25 seconds to 0.30 seconds. Interestingly, there is no straightforward influence of rhyme duration on the clarity of the general tonal picture: the figure showing the tone system of F3 (who produced the shortest rhymes) is one of those that present the clearest tonal contrasts. Admittedly, 0.20 seconds is already a fairly sizeable duration for a syllable rhyme: not an extremely brief duration, and not a particular challenging setting for realizing a phonetically complex tone.

Diversity across speakers has been mentioned at various points in this paragraph (devoted to duration), as also in the two preceding paragraphs (devoted to fundamental frequency and open quotient, respectively). Let us now turn our attention to cross-speaker similarities and differences in tonal spaces, confronting headlong the rich patterns of language variation found even in controlled, elicited materials as studied here.

#### 4.1.4 Cross-speaker similarities and differences in tonal spaces

The separate plots for the twenty speakers effectively show a highly rich and diverse range of realizations of the tonal system, against a clear background of common phonological characteristics, reflected both in  $f_{0 \text{ dEGG}}$  and in  $O_{\text{q dEGG}}$ . It would of course be inappropriate to say that some speakers have a better tonal system than others, but there is no harm in making subjective comments so long as those are clearly labeled as such. With this precaution, one can point out that, based on the criteria of clear spacing between tones and neat modulation of contour tones, the tonal systems of speakers F3 (4.1a), F9 (4.1c), F20 (4.1i), M9 (4.10) and M11 (4.1q) give an impression of being *well behaved*. Any of these five plots could serve as a textbook example of the Kim Thuong Muong tonal system, in the unlikely event of this language variety ever being taught in a language school.

Among male speakers, M14 has somewhat similar patterns with M9 (the latter being one of the group identified as the ‘well-behaved five’). Their glottal behavior verges on timidity (by its limited range), but nonetheless draws with full consistency a clear, smallish, delicate-looking tonal picture in which **Tone 4** stands out by some consistent irregularity indicative of the presence of creak. To the extent that creak, which is a departure from periodicity, can ever be described as being *controlled*, these speakers’ **Tone 4** can be said to have cogently controlled creak (under experimental conditions which, it should be noted, are by themselves conducive to consistent behavior).

Among female speakers, F7 and F10 also have curves with similarities to the ‘well-behaved’ patterns, although within somewhat broadened ranges: longer rhymes (especially for F9), and more noticeable creak for F7, which does not fully abate by the end of the rhyme.

In general, most speakers have their  $f_{0 \text{ dEGG}}$  tracings well separated across tones. There are nonetheless several speakers whose tonal spaces are really narrow and small compared to the others. This is the case for speakers M7 (4.1m), M8 (4.1n), M10 (4.1p) and M13 (4.1s). Their tones only occupy a small part of the overall space in the figure – which embodies the fullest range used by the set of twenty speakers, so as to be able to plot all the subfigures within the same ranges.

How these observations on differences in overall  $f_{0 \text{ dEGG}}$  range found in these figures relate to the speakers’ actual use of their personal ranges (their laryngeal possibilities) remains to be investigated in future. The figures for average standard deviation for  $f_{0 \text{ dEGG}}$  in Table 4.1, which include the syllable rhymes in the carrier sentence, are fairly abstract and general, and not straightforward to interpret. The speakers listed above (M7, M8, M10 and M13), whose tonal spaces are the smallest, also have lowest standard deviation in  $f_{0 \text{ dEGG}}$ . But the rough estimate provided by pooling all the data together and calculating mean standard deviation in  $f_{0 \text{ dEGG}}$  also places speaker M14 in the same range (standard deviation of about 25 Hz). More fine-grained tools, teasing out the effects of parameters such as the amount of glottalization in a speaker’s data, will be required for future work relating the speakers’  $f_{0 \text{ dEGG}}$  ranges in this experiment with broader patterns used in these speakers’ prosodic strategies in actual communication.

In terms of tonal space, compared to women, men tend to have a narrower range of

pitch. Among 20 speakers of this study, it can be observed that the males' ranges only span an average of about 100 Hz, usually from 70-80 Hz to 170-200 Hz. The range is narrowest in the cases of speakers M10 (4.1p) and M13 (4.1s): from 50 Hz to 110-120 Hz only. An exception is speaker M12 (4.1r): his range of pitch is wide (from about 120 Hz to 300 Hz).

The range for women is much wider, with an average of about 220 Hz. It is convenient, as a first approximation, to distinguish two subsets for women's tonal systems. The first subset groups cases where the range is from about 50 Hz to 250 Hz: such is the case for most female speakers. The range is maximal in the cases of speakers F9 (4.1c) and F20 (4.1i): from 50 Hz to 300 Hz (i.e. a range of roughly 250 Hz). A second subset consists of cases where top values are also at about 250 Hz, but bottom values are much less extreme: around 100 Hz, so that the overall range is lower. Needless to say, this rough grouping into two sets does not tell the full story: speaker F13 (4.1f) is an outlier, in that her bottom end of range is above 100 Hz but her upper end of range is at about 350 Hz, which makes for an overall wide range (a 250 Hz span).

The overall difference between men and women is predictable since male vocal frequency is lower than female vocal frequency (Laver, 1980; Catford, 1977), with average  $f_{0\text{ dEGG}}$  on the order of 100 Hz and 200 Hz respectively.

One way to look at these gender differences is to consider that women have more space to "play" with their tones, with generally wider spaces and clearer modulations. The figures for speakers F9 (4.1c), F13 (4.1f) and F20 (4.1i) look like cases in point: all of the tones are visually very clear, well distinguished from one another, and with salient, clear modulations for glissandos.

However, when examining these differences in speaker ranges, it should be kept in mind that absolute frequencies in Hz may not always be the most suitable way to represent fundamental frequency in speech. Conversion to a logarithmic scale has been argued to correspond best to listeners' intuitions (Nolan, 2003). When converted to relative values (specifically, to semitones, the interval between two adjacent notes in a 12-tone scale), differences in Hz come out fairly differently: the same distance in Hz will be twice larger, in semitones, if the frequencies at issue are on the order of 100 Hz, than if they are on the order of 200 Hz.

Thus, conversion to semitones facilitates cross-speaker comparison, especially across groups with different overall ranges of  $f_{0\text{ dEGG}}$ : here, women and men. An additional reason for looking at the data through the lens of the semitone scale in the present study is that, as compared with the Hz scale, it brings out the lower part of the fundamental frequency range, as compared to the higher part. That is an attractive property for a phonetic study that pays special attention to **Tone 4**, the lowest in the tone system under investigation. Therefore, Figure D.1, which can be observed in Appendix D, plots the same data as in Figure 4.1 ( $f_{0\text{ dEGG}}$  for all speakers), but in semitones instead of Hz.

Now returning to the description of individual speakers' curves, and of patterns of cross-speaker similarities: there are at least two things that make the tone systems of

speakers F9 (4.1c), F13 (4.1f) and F20 (4.1i) appear especially neat and clear.

The first thing is the rise of **Tone 3**. In these three speakers, **Tone 3** can be observed with an exemplary falling-rising contour. The fall of the first half puts into sharp relief the steepness of the rise of the second half, which is not as salient in the other speakers. Especially in comparison with the **Tone 3** of speakers F3 (4.1a) and M13 (4.1s), which almost overlaps with their **Tone 1**. Indeed, although I have not yet done a formal perception test, in my experience, I have no difficulty distinguishing **Tone 1** from **Tone 3**: what is difficult for me is to distinguish **Tone 1** from **Tone 2**. A possible explanation for this, I think, is that in terms of pitch perception, the mid-low pitch of **Tone 1** brings a perception that is more similar to a fall than to a rise. In what I feel to be a parallel situation, I usually perceive a global rise in **Tone 5** (the highest tone), although what can be observed in the  $f_{0 \text{ dEGG}}$  contour of this tone does not substantiate the perception: it is usually a flat tracing (with a fall at the very end).

The second point which makes for a display that is particularly clear visually in these speakers' data is the sharp rise of **Tone 3**. That observation raises the issue of the correlation between  $f_0$  and  $O_q$ . Open quotient and fundamental frequency are not linked by a straightforward relationship of correlation: for instance, at the end of **Tone 2**,  $f_{0 \text{ dEGG}}$  decreases while the  $O_q \text{ dEGG}$  tends to increase; and for **Tone 1**, the slight rise in open quotient towards the end does not correspond to an increase in  $f_{0 \text{ dEGG}}$ : instead, it is interpreted as reflecting the gradual decrease in the degree of vocal fold adduction towards the end of phonation. In the case of **Tone 3**, the sharp rise in  $O_q \text{ dEGG}$  over the last third of the syllable corresponds to the rising part of the  $f_{0 \text{ dEGG}}$  curve. It could be that, as  $f_0$  rises, the phonation mechanism changes to mechanism II ('head voice') and the vibrating part of the vocal folds becomes smaller, with the higher open quotient values characteristic of phonation mechanism II. This observation will be taken up again when discussing **Tone 5**.

The second, and equally important, thing that affects the clarity of the tonal space is that the  $f_{0 \text{ dEGG}}$  curve of **Tone 4** stands out by a clear dip in the middle, which is more U-shaped than V-shaped. This determines the bottom values set in the display (the y axis of figures), since this is the lowest tone of the system. In most cases, among the 20 speakers present, it is easy to tell apart the red line from the rest of the system, since it has half or more of the pitch range to itself. It alone occupies a separate space: even the highest points of this tone (located in positions of onset and offset) are lower than all the other tones.

By visual inspection, the seven speakers who possess the best-distinguished **Tone 4** are F3 (4.1a), F7 (4.1b), F9 (4.1c), F19 (4.1h), F20 (4.1i), M9 (4.1o), and M11 (4.1q). Impressionistically, those can be considered as ideal contrasts for **Tone 4**. The other cases are less distinct: the range of pitch is narrower. This is the case of speakers F10 (4.1d), F12 (4.1e), M5 (4.1l), M7 (4.1m), M8 (4.1n), M10 (4.1p), M13 (4.1s) and M14 (4.1t). In these cases, the  $f_{0 \text{ dEGG}}$  curve of **Tone 4** is still located at bottom, but much less distant from the other tones, and the V-shape or U-shape is less steep.

The discussion of this issue will be continued and detailed in the section on glottalization (4.3) later in this chapter, where the glottalized tone is the object of close

scrutiny. In general, it can be observed that there is great diversity in terms of tonal space across the 20 speakers. The greatest difference is found between speakers F9 (4.1c) and M13 (4.1s) with a range of 250 Hz versus 60 Hz.

## 4.2 An overview of the full tone system from the results of normalizing 20 speakers, comparing males and females, and comparing smooth and checked tones

To provide a convenient overview, Figure 4.2 shows the trajectories of  $f_{0 \text{ dEGG}}$  and  $O_{q \text{ dEGG}}$  averaged for each gender (10 males and 10 females), and for all 20 speakers. In addition, for a complete view of the tone system, Figure 4.3 provides a side-by-side plot for two sub-systems of five smooth tones (on the left) and two checked tones (on the right). The plot for smooth tones is actually the same as the one for 20 speakers in Figure 4.2c.

Keeping in mind that averaged curves only offer a rough approximation of the data, Figures 4.2 and 4.3 allow for a recapitulation of some salient characteristics of Kim Thuong Muong tones.

From these normalized results,<sup>3</sup> some general observations about the tone system can be proposed – with apologies for the large amount of overlap with observations made above when discussing patterns of similarity across speakers.

### 4.2.1 Smooth tones

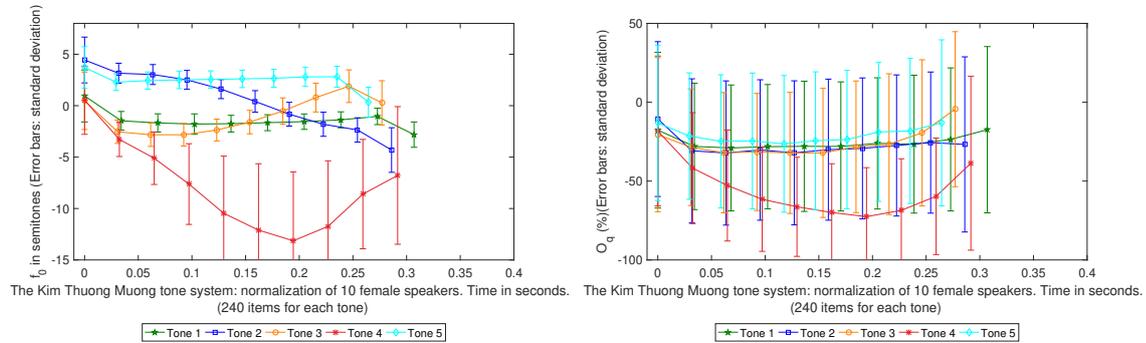
Let us first consider the main sub-system of smooth tones.

**Tone 1** is the longest of the five tones, although the difference is only less than 0.05 seconds, thus it makes no real distinction in terms of duration characteristics. This tone is remarkably level. It has a moderate final dip, also found in other tones, corresponding to the offset of voicing. It is a little below the speaker's average  $f_{0 \text{ dEGG}}$ . In opposition to **Tone 1**, **Tone 5** is another flat tone but in a higher pitch level. In fact, it is the highest tone, with  $f_{0 \text{ dEGG}}$  values always at the top of the speaker's range, so it can be called a 'top-high' tone.

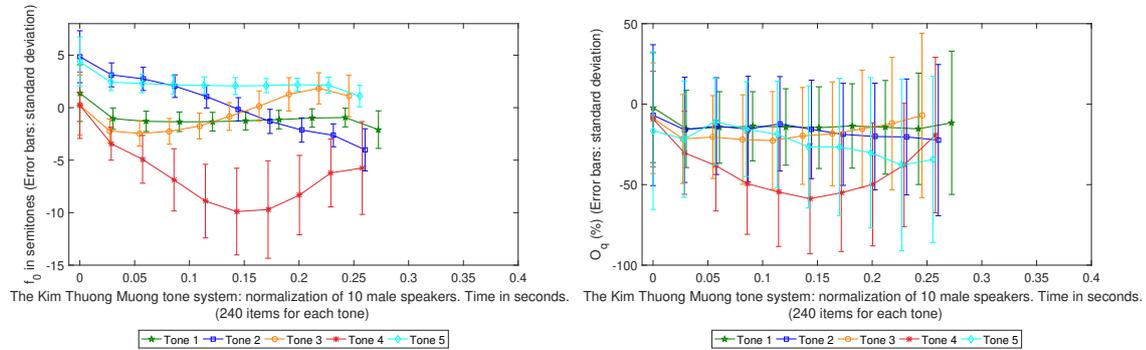
**Tone 2** and **Tone 3** are characterized by falling and rising curves, respectively; the amplitude of the rise of **Tone 3** is not quite as great as that of the fall of **Tone 2**. If taking two level tones (i.e., **Tone 1** and **Tone 5**) as two measures, it can be noticed on all the  $f_{0 \text{ dEGG}}$  graphs (on the left) of Figure 4.2 that while the rising tone (**Tone 3**) has a lower onset than the mid-level tone (**Tone 1**) and also a lower offset than the top-level tone (**Tone 5**), the falling tone (**Tone 2**) has a higher onset than the top-level tone and a distinctively lower offset than the mid-level tone. This phonetic asymmetry is to be expected if one takes into account a measure of declination in the course of affirmative utterances (statements).

<sup>3</sup>Recall that the formulas used to obtain these relative values of  $f_{0 \text{ dEGG}}$  and  $O_{q \text{ dEGG}}$  are the same as in Mazaudon and Michaud (2008, p. 238): see Section 3.4.

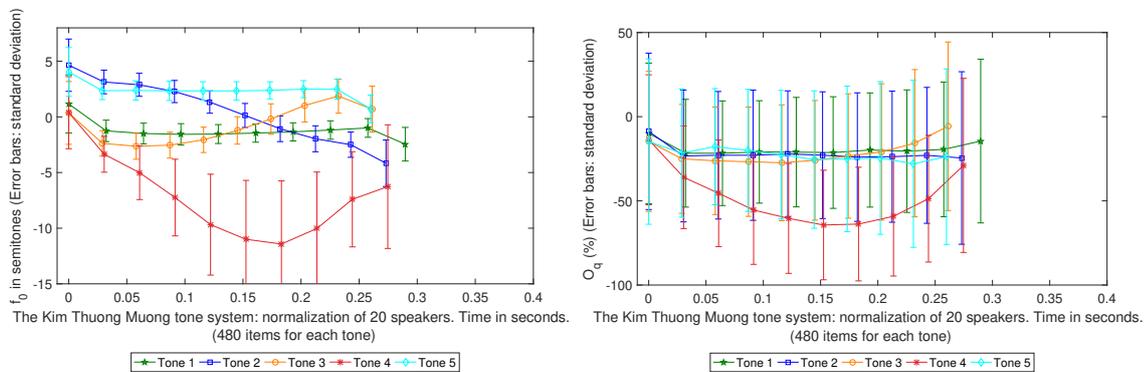
4.2 An overview of the full tone system from the results of normalizing 20 speakers



(a) Normalization of 10 female speakers.



(b) Normalization of 10 male speakers.



(c) Normalization of 20 speakers.

Figure 4.2: The tone system of Kim Thuong Muong: normalization across speakers.  $f_0$  dEGG on the left and  $O_q$  dEGG on the right.

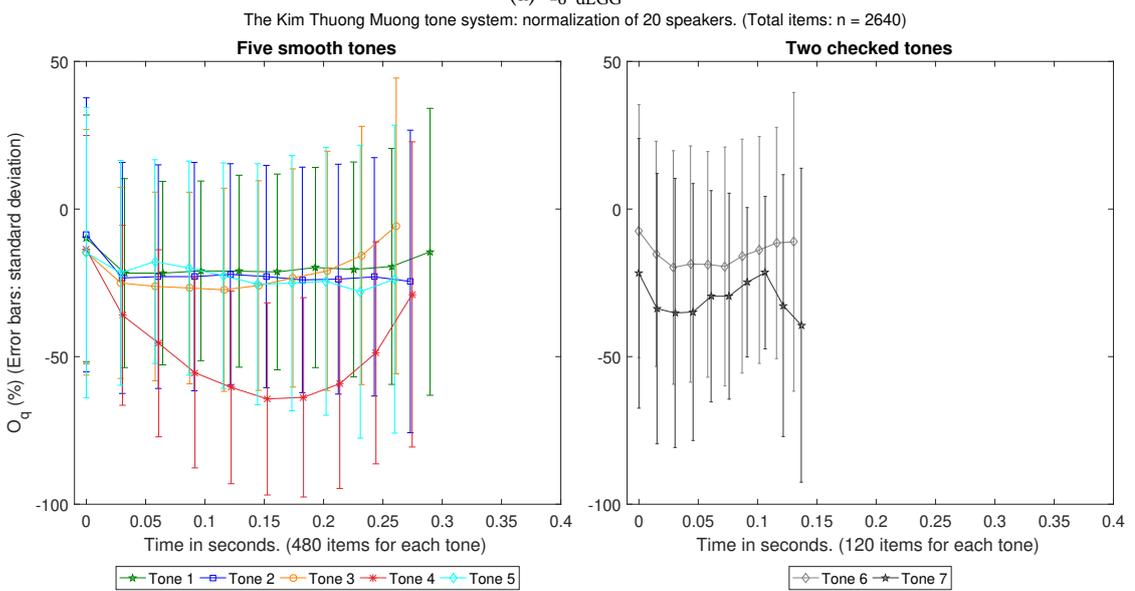
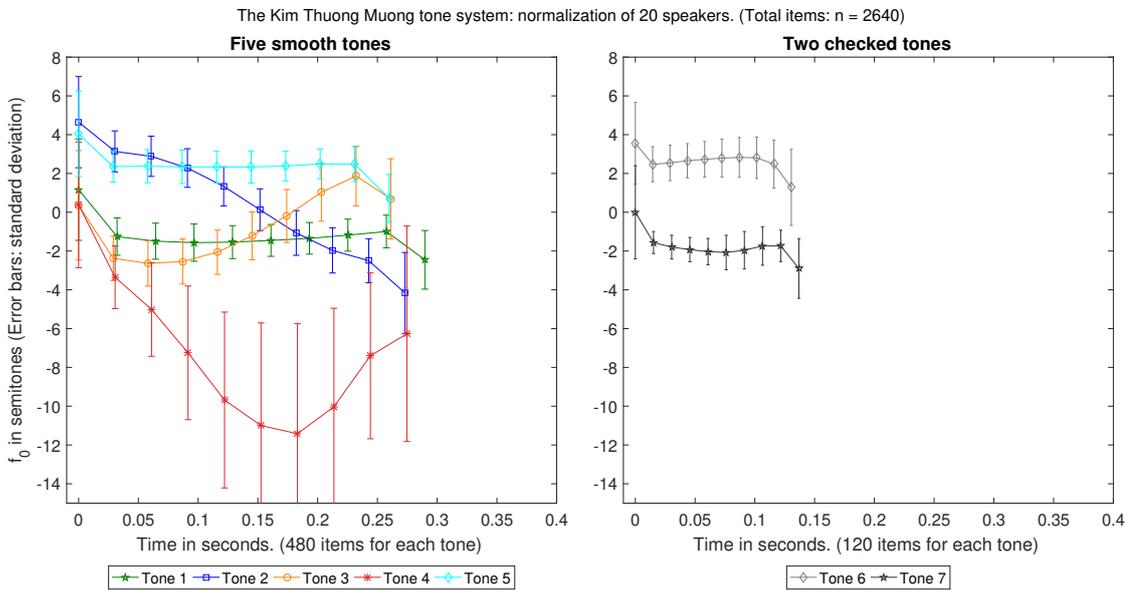


Figure 4.3: The Kim Thuong Muong's tone system: smooth tones (left) and checked tones (right), normalized on 20 speakers.

The most remarkable feature of **Tone 4** is glottalization. The results from twenty speakers allow for correcting my earlier statement that glottalization “does not appear at the beginning but in the mid-part of the rhyme”. On the 10-point normalized averaged tracings, the variability in  $f_{0 \text{ dEGG}}$  increases noticeably as early as the third data point, and is unmistakable at the fourth data point, as is the rapid dip in  $f_{0 \text{ dEGG}}$ , both providing evidence of glottalization.

The open quotient patterns show a strong overlap across tones, with two salient characteristics. First, the curves share a similar overall shape: rather flat, but slightly modulated at both ends (the onset and offset), with **Tone 4** an extreme case – a smooth V-shaped pattern. Secondly, there is a shared tendency for standard deviation to increase from beginning to end of the syllable rhyme.

The largest departure from these overall trends is the  $O_{q \text{ dEGG}}$  tracing of the women’s average **Tone 5**, as it is less stable than the others, and clearly drops to a low position that is even lower than the offset of **Tone 4**. Since the values of  $f_{0 \text{ dEGG}}$  of **Tone 5** do not reveal anything out of the ordinary (it is always the highest tone, until the end), it can be hypothesized that this tone in females tends to be increasingly tense towards the end. Note, however, that the decrease in  $O_{q \text{ dEGG}}$  in the second half of the rhyme is clearly milder than the medial dip in  $O_{q \text{ dEGG}}$  for **Tone 4**, and that standard deviation is very high, so that the more pressed phonation type associated with **Tone 5** in female speakers is nowhere as consistent, as a phonation type, as the creaky voice found in the medial part of **Tone 4**.

Although the difference in trajectory of  $O_{q \text{ dEGG}}$  between the tones is very small, it is still possible to notice a pattern: the highest tone (**Tone 5**) has highest  $O_{q \text{ dEGG}}$ , and the lowest tone (**Tone 4**) has lowest  $O_{q \text{ dEGG}}$ . This generalization must not hide from view the considerable asymmetry between the two, however.  $O_{q \text{ dEGG}}$  for **Tone 5** is barely higher than for the other nonglottalized tones (**Tone 1**, **Tone 2** and **Tone 3**) – not to mention its gradual *decrease* in female speakers, leading it down to values lower than for **Tone 1**, **Tone 2** and **Tone 3**–, whereas **Tone 4** is way below all the others, with a consistent shape reaching lowest values from fourth to eighth data points out of ten.

The question of the correlation between  $f_{0 \text{ dEGG}}$  and  $O_{q \text{ dEGG}}$  springs to mind again when noting that the  $O_{q \text{ dEGG}}$  curve of **Tone 4** seems to follow its  $f_{0 \text{ dEGG}}$  curve (as a V-shape), but the other tones do not have this neat correlation. Whatever the tone’s shape, flat or rising or falling, their  $O_{q \text{ dEGG}}$  tracings are pretty flat and overlap with one another. To find clearer clues on this question, Figure D.3 in Appendix D directly displays the correlation between  $f_{0 \text{ dEGG}}$  and  $O_{q \text{ dEGG}}$ . It confirms that there is generally no clear correlation between these two parameters for twenty speakers, with a large dispersion and no perfect positive or negative line. However, there does appear to be some correlation in the case of **Tone 4** in not all speakers but several speakers such as: F3 (D.3a), F12 (D.3e), F17 (D.3g), F20 (D.3i), F21 (D.3j), particularly clear in the cases of F10 (D.3d), M1 (D.3k) and M11 (D.3q). The red dots in these cases form some sort of positive trend while the others show no trend and overlap. One possible explanation for this distribution is that it reflects the transition of different mechanisms. **Tone 4**

is the only tone that shows a rapid transition from a modal voice at the beginning to a creaky voice and back to the modal voice at the end. In cases where speakers have made this transition clearly in **Tone 4**, we can see a good correlation between  $f_{o\text{ dEGG}}$  and  $O_{q\text{ dEGG}}$ . Otherwise, in the case of speakers F13 (D.3f), M5 (D.3l), M8 (D.3n), M12 (D.3r), the red dots are still in the bottom position of the but no clear correlation. The other tones remain in modal voice, so there is no trend in their  $f_{o\text{ dEGG}}$  and  $O_{q\text{ dEGG}}$  distribution neither.

Overall, the plots are remarkably consistent between women and men, so that it does not come as a surprise that the figure showing normalized results across all 20 speakers also has a very similar outlook. The offset of voicing in women is clearer, though, with the last point of four modal tones falling more steeply. The duration differs by only 0.02 seconds, in a mild tendency for females to produce slightly longer syllable rhymes than males in this data set.

#### 4.2.2 Checked tones

As noted above, duration by itself does not appear to play an important role in contrasts among the five tones of smooth syllables. On the other hand, duration differs greatly between smooth and stopped syllables. The tones in stopped syllables are only about half as long as tones in smooth syllables; this can be ascribed to the influence of obstruent codas. Final /p/, /t/, /c/ and /k/ are phonologically unspecified for voicing (there are no voicing opposition among final consonants in Kim Thuong Muong), but these codas are phonetically unvoiced, and appear to exert the full shortening effect that is cross-linguistically associated with final voiceless stops. Among stopped tones, there appears to be a relatively neat parallel between a mid-level-checked tone (**Tone 6**) and a high-level-checked tone (**Tone 7**). This is consistent with the well-established symmetry of the tonal system. Visually, looking at the  $f_{o\text{ dEGG}}$  tracings in Figure 4.3, **Tone 6** and **Tone 7** look like shorter versions of two smooth level tones, **Tone 5** and **Tone 1**, respectively, since the pitch heights and the shapes of the contours look strikingly alike. Perceptually, according to my Vietnamese ears, these two tones are somewhat similar to the two checked tones D<sub>1</sub> and D<sub>2</sub> in Vietnamese, which are a rising-checked tone and a low-checked tone.

Considering the parameter of  $O_{q\text{ dEGG}}$ , unlike the tones of the smooth sub-system, the  $O_{q\text{ dEGG}}$  curves of **Tone 6** and **Tone 7** are clearly apart, and follow the distinction in  $f_{o\text{ dEGG}}$ , i.e. the higher tone (**Tone 6**) has higher  $O_{q\text{ dEGG}}$ — which is also higher than the  $O_{q\text{ dEGG}}$  values of all modal tones in the smooth system. The lower tone (**Tone 7**) has lower  $O_{q\text{ dEGG}}$ — which is lower than the  $O_{q\text{ dEGG}}$  values of all the tones in the smooth subsystem except **Tone 4**. The three points at the end show a sudden drop that breaks the previous upward trend. This decline also differs from the general pattern of an offset rise suggestive of syllable-final decrease in degree of vocal fold adduction in all the other tones. This would suggest that **Tone 7** has a relatively tense offset, although this point needs to be verified with more data as here we only have 3 minimal pairs of checked tones (performed twice).

### 4.3 Glottalization as a component of lexical tone: A closer look at the glottalized tone across 20 speakers

To take a closer look at the glottalized tone, we will combine here observations about the entire tone system (Figure 4.1) and about **Tone 4** in Figure 4.4 which shows raw data (i.e., one curve for each item) of  $f_{0 \text{ dEGG}}$  and  $O_{q \text{ dEGG}}$  tracings for only this tone. In this figure, the color of the lines changes gradually from red to blue to reflect the order of items in the experiment. Red curves are used for items at the beginning of set, maroon curves at the middle of set, and blue curves at the end of set. This color code is also applied for all figures which show one curve for each item. The aim is (i) to make the lines easier to tell apart from one another and (ii) to provide a way to visualize possible changes in the speakers's behavior in the course of the recording (commonly observed changes include gradually decreasing  $f_0$ : see Niebuhr and Michaud 2015). The same range of  $f_0$  and  $O_q$  is used in all the figures (men and women) to facilitate comparison.

#### 4.3.1 A general look at twenty speakers

As already mentioned in several places in this thesis, it has been well noticed that **Tone 4** is the most striking tone of the system with both  $f_{0 \text{ dEGG}}$  and  $O_{q \text{ dEGG}}$  parameters well-distinguished from all the other tones.

##### 4.3.1.1 The ideal cases of **Tone 4**

**Tone 4** determines the bottom base of the speaker's range since this is the lowest tone in the system. In data from most of the twenty speakers, it is easy to separate the red line from the rest of the system. This tone usually has half of the  $f_{0 \text{ dEGG}}$  range to itself. It alone occupies a separate space in the lower part: the highest points at two ends (in positions of onset and offset) of this tone are lower than all the other tones, and the curve plunges to bottom values in the middle part. **Tone 4** is the only tone that has a complex trajectory combining descending and ascending components, while the other tones have only a straight trajectory (either flat, rising or falling).  $f_{0 \text{ dEGG}}$  curves of **Tone 4** stand out by a clear dip in the middle, which is more V-shaped than U-shaped. The speakers who possess this well-distinguished shape of **Tone 4** are F3 (4.1a), F7 (4.1b), F9 (4.1c), F19 (4.1h), F20 (4.1i), M9 (4.1o), and M11 (4.1q). These are the speakers that have the lowest  $f_{0 \text{ dEGG}}$  values of **Tone 4**, which can reach down to about 50 Hz. Those  $f_{0 \text{ dEGG}}$  trajectories go down from the low-mid part of the speaker's range to the bottom of range, at least 15 to 20 semitones lower.

The  $O_{q \text{ dEGG}}$  curves of these speakers likewise tell the same story. It is possible to draw a line to set apart **Tone 4** (in red) from the rest of the system. In the upper part, the  $O_{q \text{ dEGG}}$  curves of four tones overlap, and they are consistently around the mean of 50%, and the bottom part of the standard deviation usually around 40%. Whereas, in the lower part, there is only **Tone 4**: the  $O_{q \text{ dEGG}}$  curve of this tone plummets from initial mid-range values (which are already lower than all the other

tones) to the very bottom of range. In some speakers (most saliently for speakers F7 (4.1b) and F20 (4.1i)), most of the  $O_{q \text{ dEGG}}$  values of **Tone 4** were excluded, due to the issue of unclear opening peaks. The values retained for 20 speakers in general reflect the consistent presence of very low values of  $O_{q \text{ dEGG}}$ , especially in the middle of the rhyme, which are commonly below 30%, and can even reach strikingly low values, on the order of 15%: for speakers F3 (4.1a), F7 (4.1b), F17 (4.1g) and M11 (4.1q). These values constitute compelling evidence of the presence of creaky voice.

#### 4.3.1.2 The cases of less distinct **Tone 4**

Apart from these ideal contrasts for **Tone 4**, we have other cases which are less distinct, and accordingly, with a narrower range of  $f_{o \text{ dEGG}}$ . Considering the parameters  $f_{o \text{ dEGG}}$  and  $O_{q \text{ dEGG}}$  tracings in Figure 4.1, we can propose a rough classification into three sets of cases:

- The case of speaker F10 (4.1d), F12 (4.1e) and M10 (4.1p): the distinction of the  $f_{o \text{ dEGG}}$  curve of **Tone 4** from the rest is less striking but the  $O_{q \text{ dEGG}}$  curve still maintains a neat distance from the others. In particular, in the case of M10 (4.1p), the  $f_{o \text{ dEGG}}$  curve of **Tone 4** is not strikingly apart from that of the other tones, but its  $O_{q \text{ dEGG}}$  curve is still well isolated.
- The cases of speakers M7 (4.1m), M8 (4.1n), M10 (4.1p) and F21 (4.1j): the  $f_{o \text{ dEGG}}$  curve of **Tone 4** is kept a little separate from the other tones, with its V-shape less marked. The  $O_{q \text{ dEGG}}$  tracing is still at the bottom of the system but just next to the other tones. A borderline case for this category is speaker M13 (4.1s) with a minimal distinction in  $f_{o \text{ dEGG}}$  and no salient distinction in  $O_{q \text{ dEGG}}$ .
- A special case is F13 (4.1f) with clear  $f_{o \text{ dEGG}}$  tracings for all the tones, which even make it seem like she is overdoing the contrasts: exaggerating them towards an ideal model, with a nice rise of **Tone 3** and a clear V-shape of **Tone 4**. However, the  $O_{q \text{ dEGG}}$  parameter does not bring out a massive distinction between the glottalized tone and the rest, although **Tone 4** still has the lowest  $O_{q \text{ dEGG}}$  values. Similar to this case, M14 (4.1t) also presents a nice tone system according to  $f_{o \text{ dEGG}}$ , but the  $O_{q \text{ dEGG}}$  values are not distinct: all the curves of five tones overlap.

The tonal space is used with much more abandonment by speaker F9 (4.1c), whose long rhymes and high  $f_{o \text{ dEGG}}$  set upper bounds for the entire data set. Variability in her data is somewhat higher, too (reflected in wider standard deviation ranges), suggesting that the speaker avails herself of the comfortable distance that she leaves in-between tones to allow herself somewhat more freedom in the exact pitch at which she sets individual tokens of the tones that she produces. **Tone 4** shows ample phonetic evidence of creakiness, starting as early as the first quarter of the rhyme. Conversely, speakers with the smallest tonal spaces and with no striking visual evidence of using glottalization for **Tone 4** have very small standard deviation across tokens of the same tones. Speaker M13 is clearly an epitome of a super-economical (Malthusian?) approach to tonal contrasts, making up (to an extent that remains to be verified perceptually) for the small amplitude of the difference across tones through remarkable consistency across

tokens. Chance has it that M12 and M13 received adjacent numbers when speakers were assigned numbers, resulting in the juxtaposition, as Figures 4.1r and 4.1s, of two tonal spaces of remarkably different proportions. Once the initial shock wears off, a phonologist's eye can nonetheless acknowledge strong structural similarities: **Tone 3** is rising, whether dramatically or in the most subdued way, and so on.

Looking at the  $O_q$  dEGG tracings alone, there is a majority of speakers in whose data **Tone 4** stands out by a curve that is unmistakably, obviously apart from that of all the other tones (F3 (4.1a), F7 (4.1b), F9 (4.1c), F10 (4.1d), F12 (4.1e), F17 (4.1g), F19 (4.1h), F20 (4.1i), F21 (4.1j), M1 (4.1k), M8 (4.1n), M9 (4.1o), M10 (4.1p) and M11 (4.1q)). Interestingly, this leaves about one third of speakers for whom the picture is not so obvious. It should however be remembered here that the estimation of the glottal open quotient from electroglottography can be difficult to carry out for some types of signals. If the  $O_q$  dEGG curve for speaker M13's **Tone 4** is not as clearly set apart from that of the other tones at the point where the  $f_o$  dEGG curve shows clear hints of irregularity in vocal fold vibration, it is not unlikely to be due to the absence of some data points in the matrix of  $O_q$  dEGG results for this tone.

This difference is more salient for speaker F3 (4.1a), since the  $O_q$  dEGG values for F13 (4.1f) are overall lower and thus less distant between **Tone 4** and the rest. But the shape of the  $O_q$  dEGG curves for speaker F13 contributes to keeping **Tone 4** apart from the others, since its  $O_q$  dEGG curve only begins to rise after a steady decrease that lasts until the middle of the averaged curve, whereas an overall increase in  $O_q$  dEGG is found for all the other tones starting no later than one-third into the rhyme.

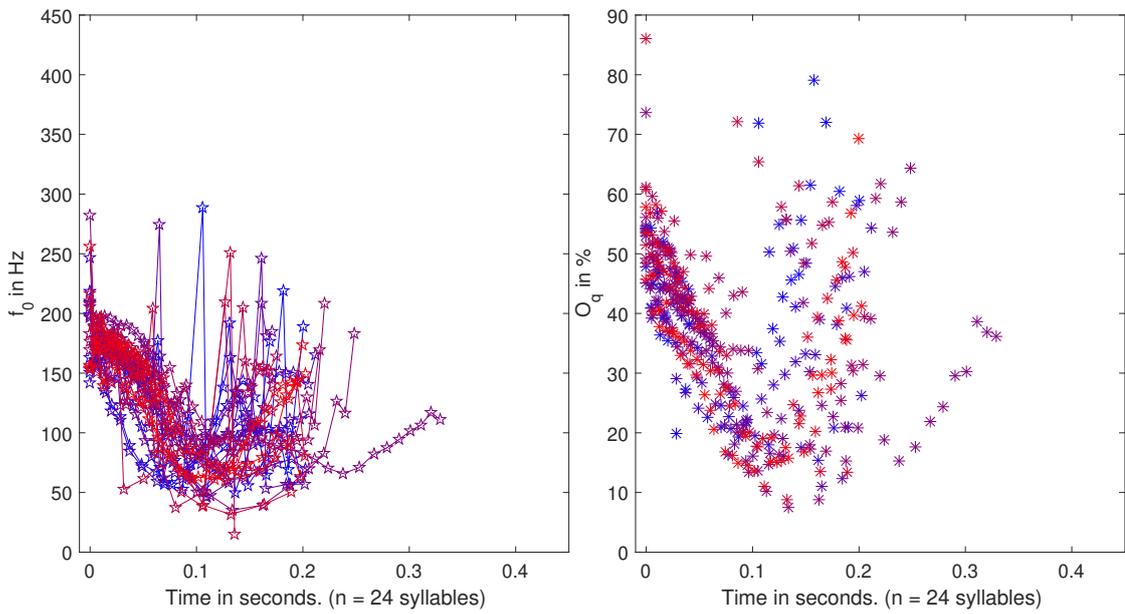
The above remarks allow for several generalizations. Some of these generalizations are negative: thus, a large amount of creakiness does not necessarily go hand in hand with a boisterous realization of all the tones. Speaker F12 (4.1e) has outstanding amounts of creakiness in **Tone 4** but an otherwise very moderate spacing of tones.

Other generalizations are positive, in the sense that they bring out trends. Thus, even though the phasing of glottalization is highly different across speakers, patterns do emerge. To categorize into four rough categories depending on whether creak occurs early, medially, late, or not noticeably (in the curves in Figure 4.1), it appears that there are more *late creakers* (eight: {F7 (4.1b), F9 (4.1c), F10 (4.1d), F12 (4.1e), F13 (4.1f), F20 (4.1i), F21 (4.1j), M11 (4.1a)}) and *medial creakers* (seven: {F3 (4.1a), F17 (4.1g), M1 (4.1k), M5 (4.1l), M9 (4.1o), M10 (4.1p), M14 (4.1t)}) than early creakers (only M8 (4.1n) and M12 (4.1r)) and (apparent) *noncreakers* (only F19 (4.1h), M7 (4.1m) and M13 (4.1s)).

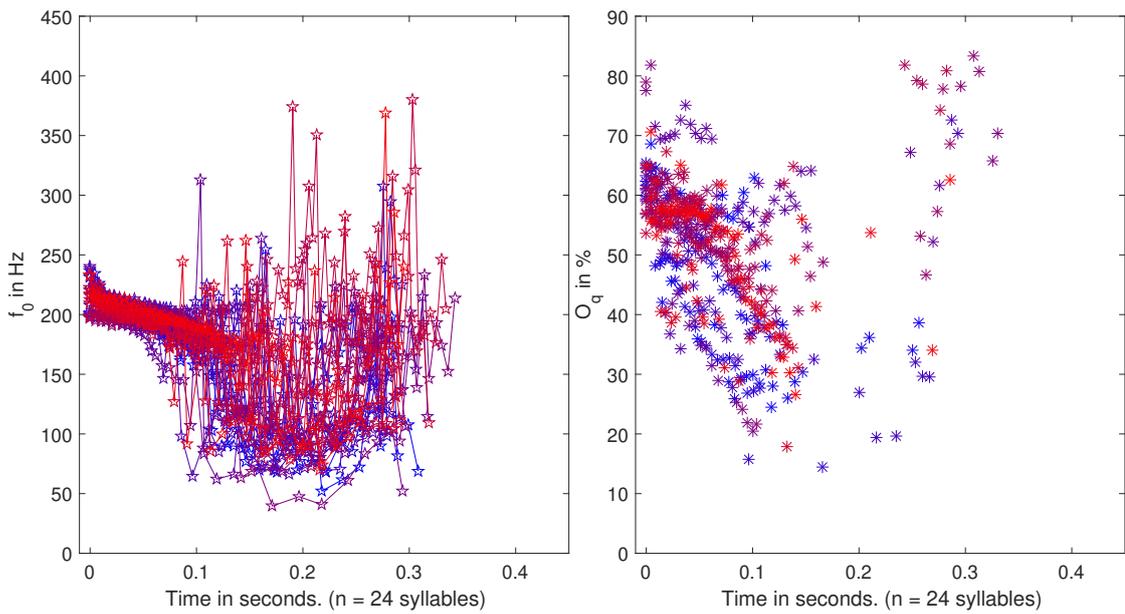
#### 4.3.2 A comparative look at gender differences

The shape of **Tone 4** is less V-shaped for men, which does not tell us that men creak less than women, but that they creak in a different way from women. Figure 4.4 reveals that women tend to perform **Tone 4** with a tremendous turbulence of  $f_o$  dEGG, while these curves in men are more stable and follow the V-shape as the averaged results.

Eight women out of ten have jittery  $f_o$  dEGG curves, with a large fluctuation generally with an amplitude from 30-50 Hz up to 250-350 Hz. In the case of speaker F19 (4.4h),



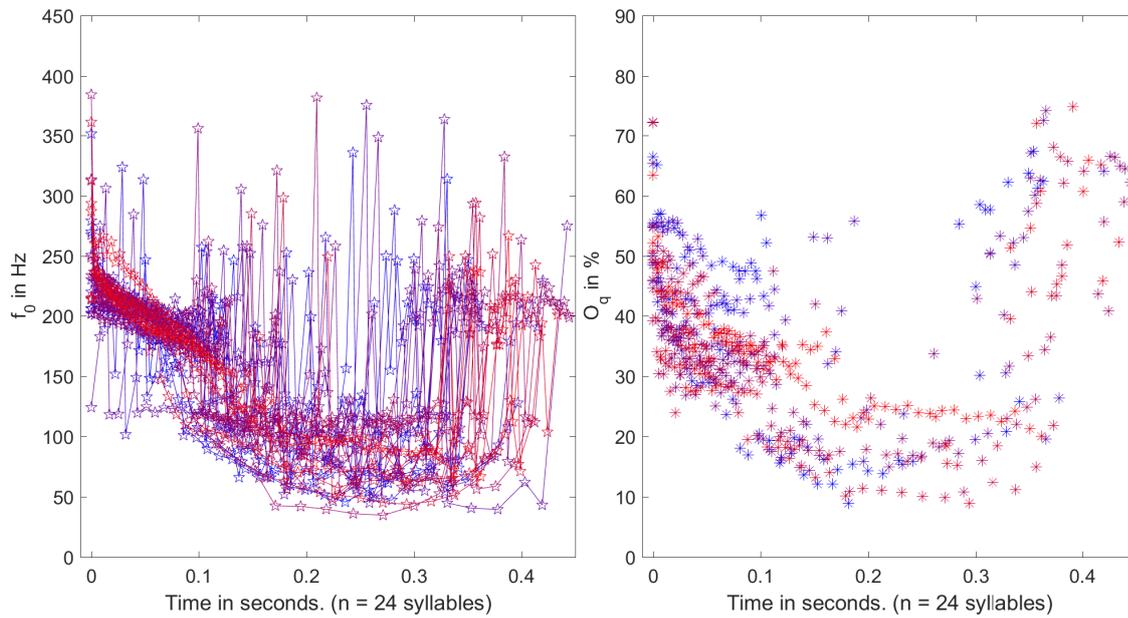
(a) Speaker F3



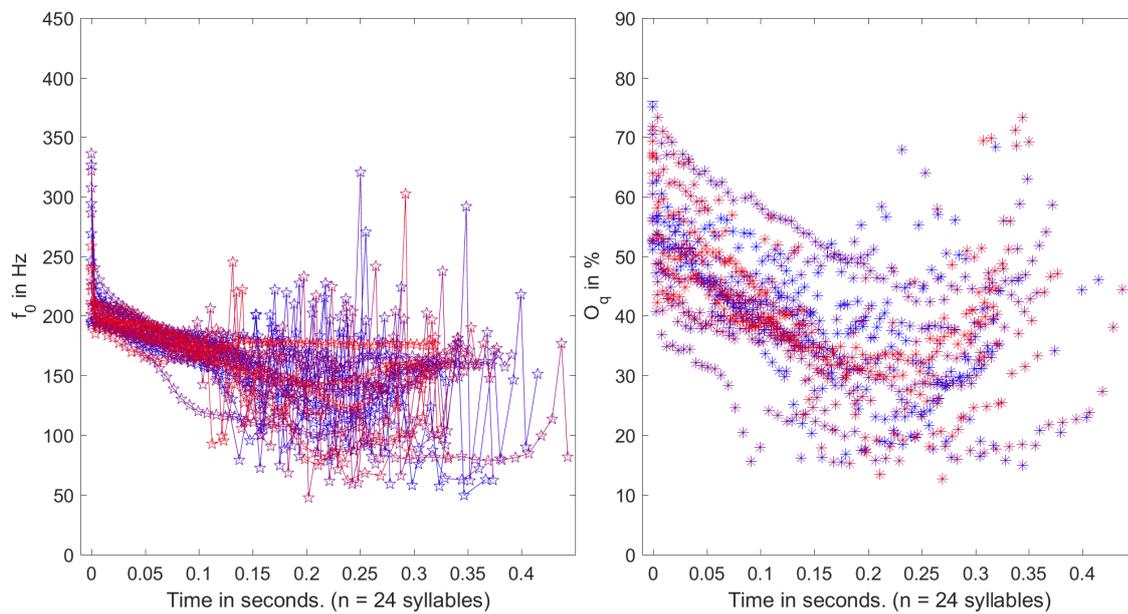
(b) Speaker F7

Figure 4.4: The glottalized tone of Kim Thuong Muong: speaker by speaker.  $f_0$   $\text{dEGG}$  on the left and  $O_q$   $\text{dEGG}$  on the right.

4.3 Glottalization as a component of lexical tone: A closer look at the glottalized tone across 20 speakers

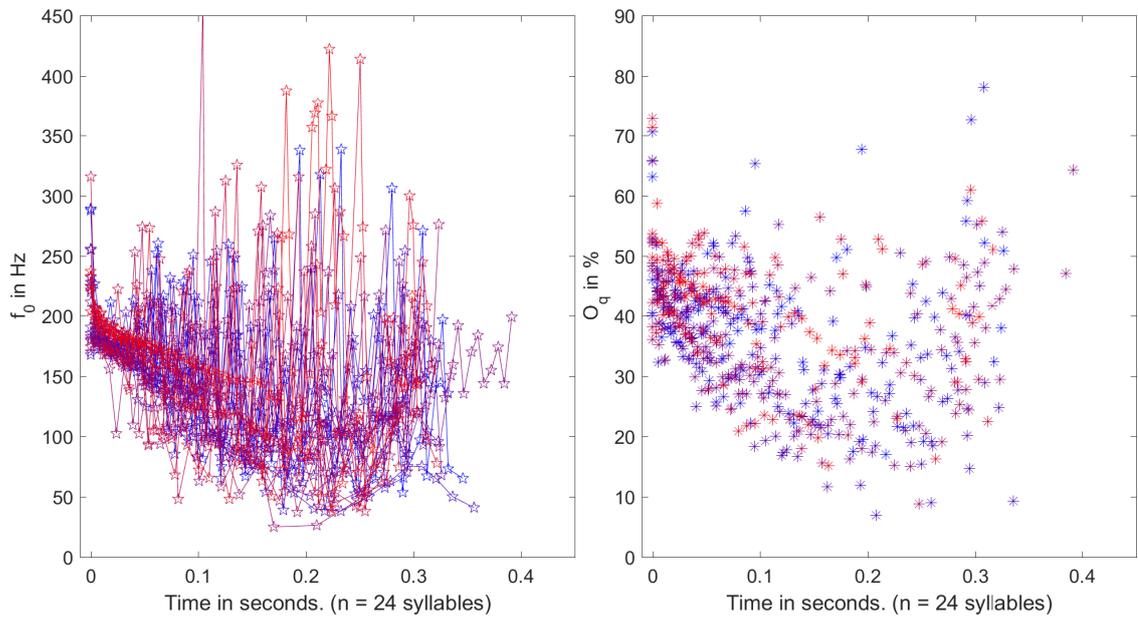


(c) Speaker F9

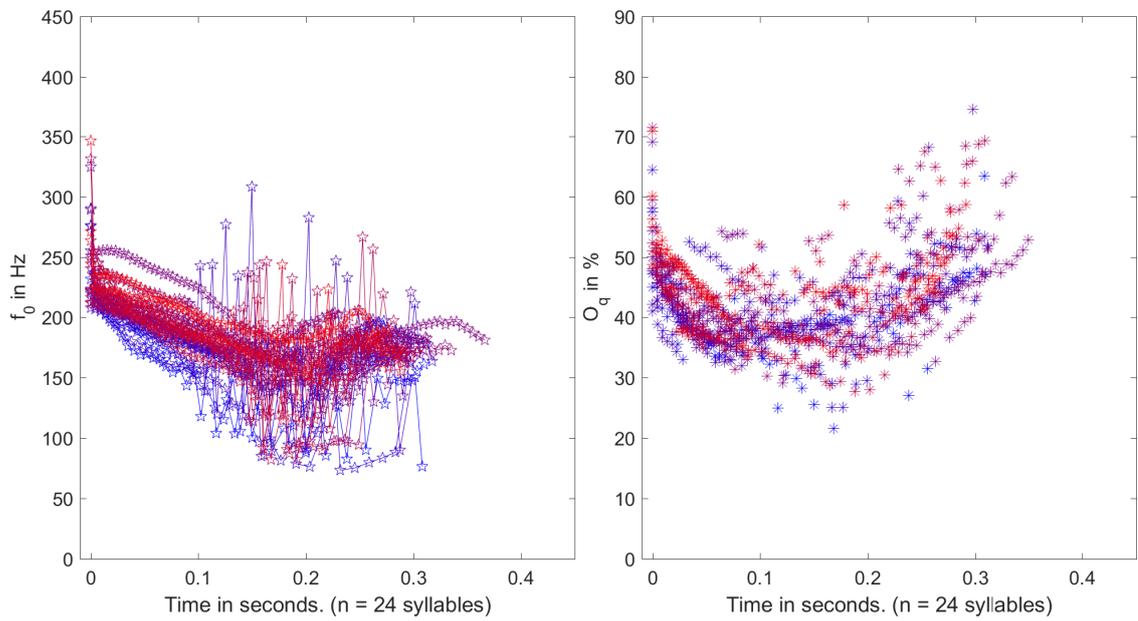


(d) Speaker F10

Figure 4.4: The glottalized tone of Kim Thuong Muong: speaker by speaker.  $f_0$   $\Delta$ EGG on the left and  $O_q$   $\Delta$ EGG on the right.



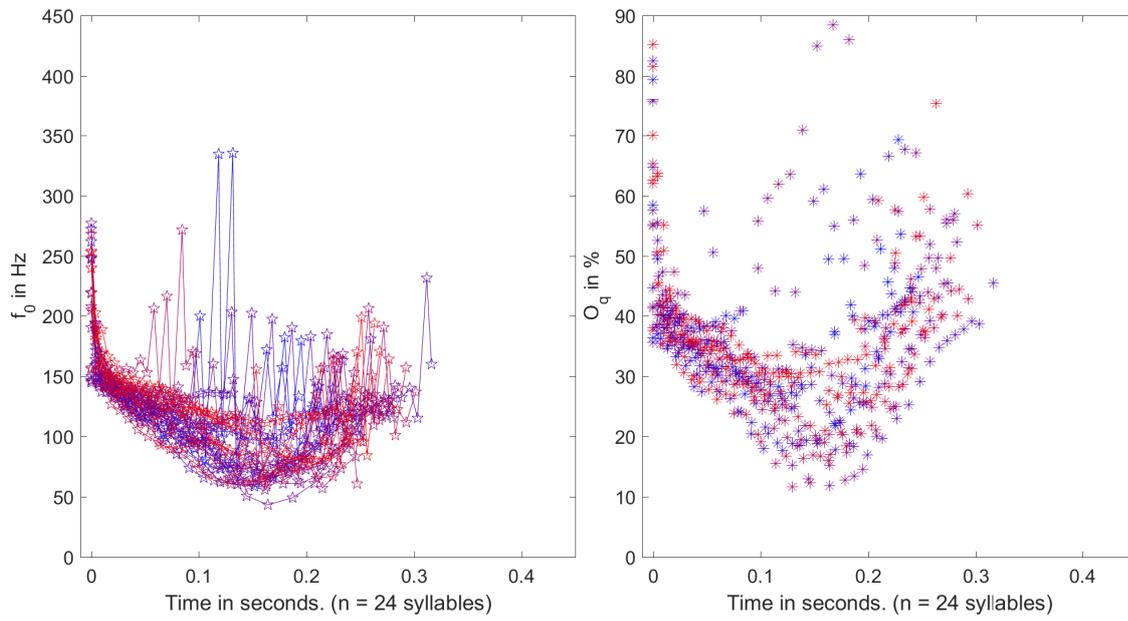
(e) Speaker F12



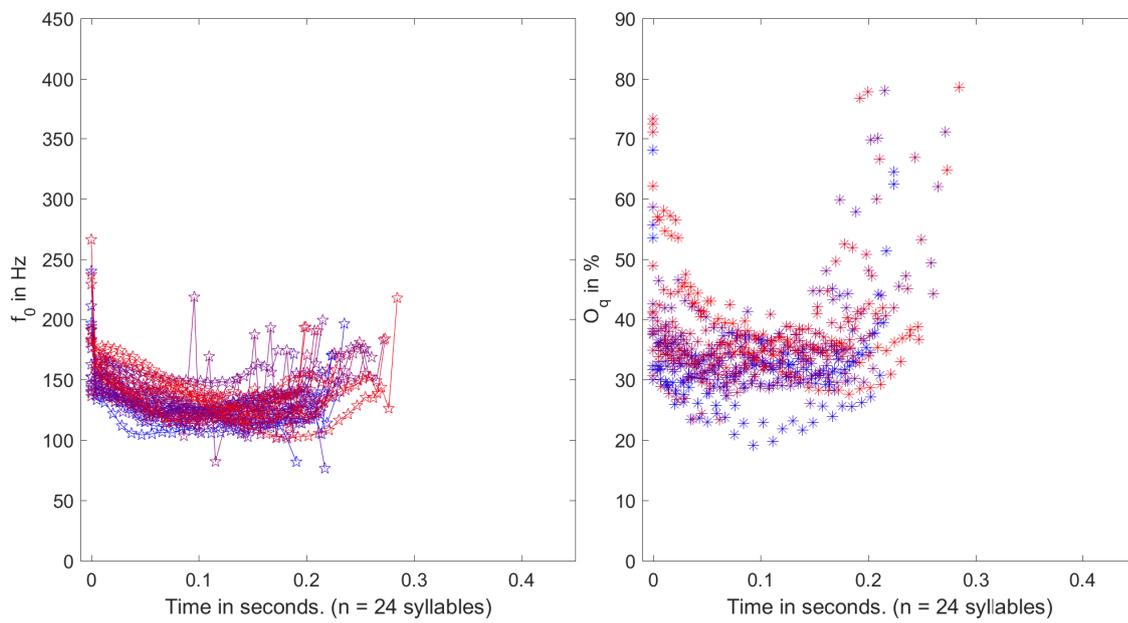
(f) Speaker F13

Figure 4.4: The glottalized tone of Kim Thuong Muong: speaker by speaker.  $f_0$  <sub>dEGG</sub> on the left and  $O_q$  <sub>dEGG</sub> on the right.

4.3 Glottalization as a component of lexical tone: A closer look at the glottalized tone across 20 speakers

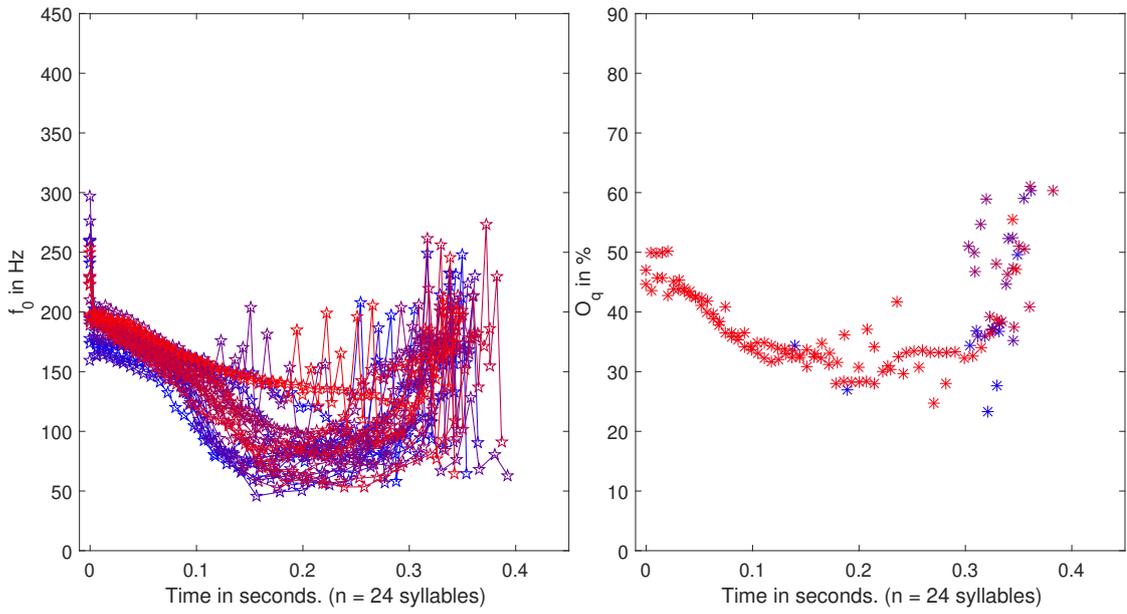


(g) Speaker F17

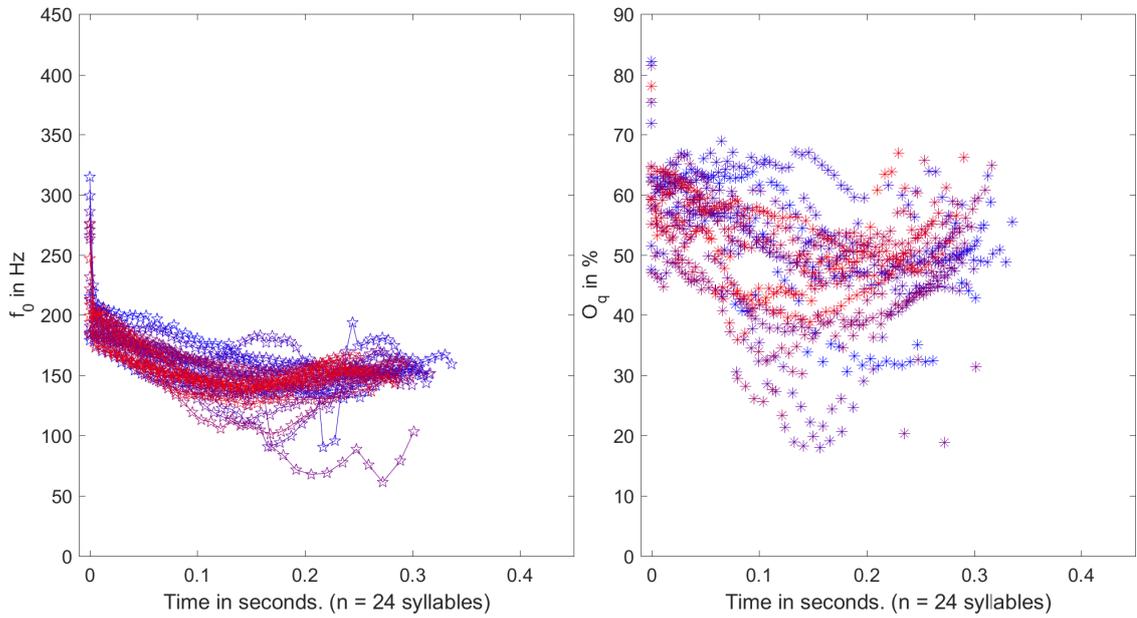


(h) Speaker F19

Figure 4.4: The glottalized tone of Kim Thuong Muong: speaker by speaker.  $f_0$  <sub>dEGG</sub> on the left and  $O_q$  <sub>dEGG</sub> on the right.



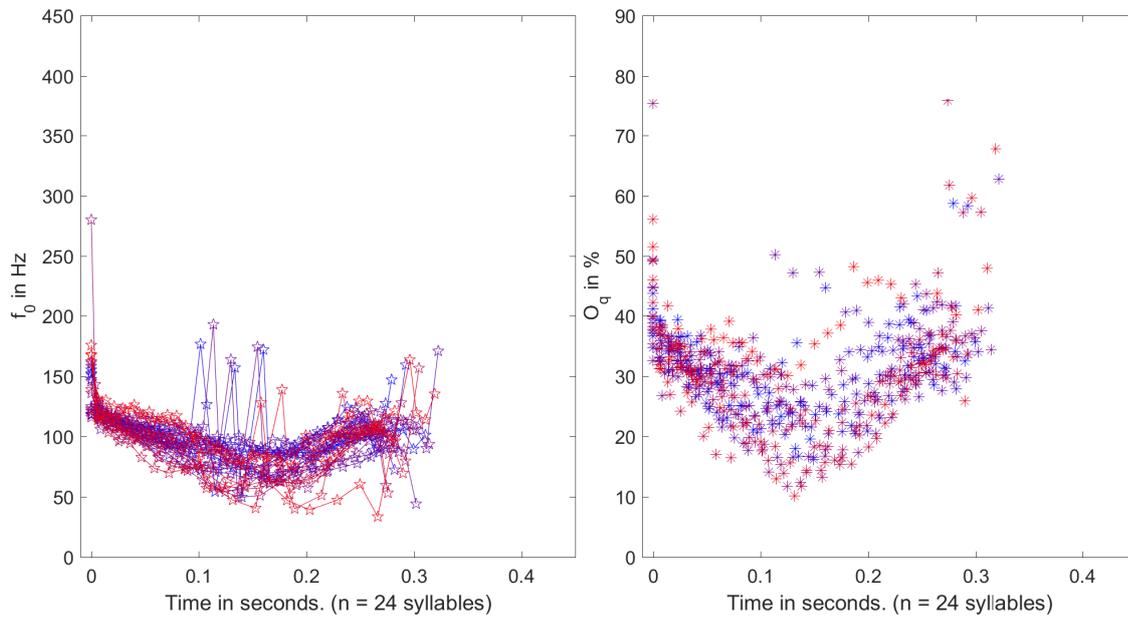
(i) Speaker F20



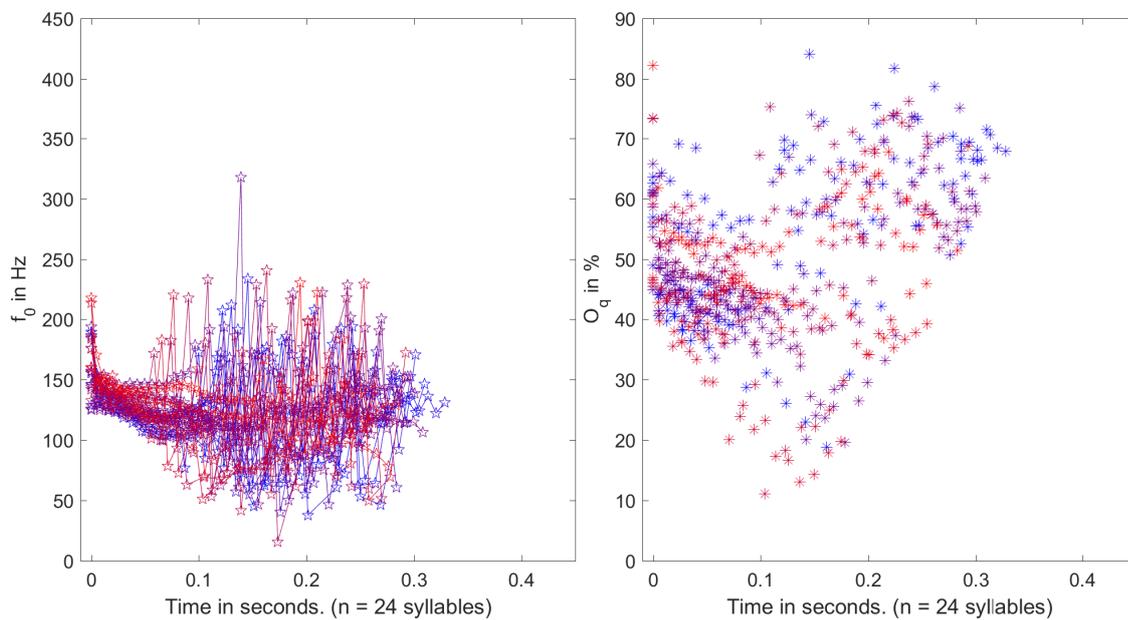
(j) Speaker F21

Figure 4.4: The glottalized tone of Kim Thuong Muong: speaker by speaker.  $f_0$   $_{dEGG}$  on the left and  $O_q$   $_{dEGG}$  on the right.

4.3 Glottalization as a component of lexical tone: A closer look at the glottalized tone across 20 speakers

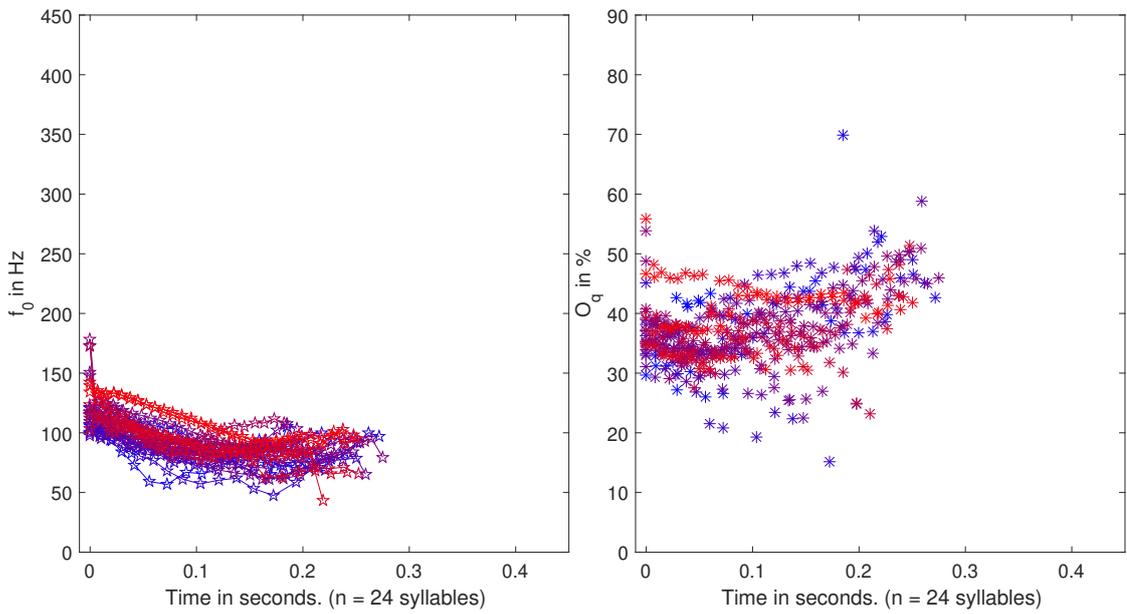


(k) Speaker M1

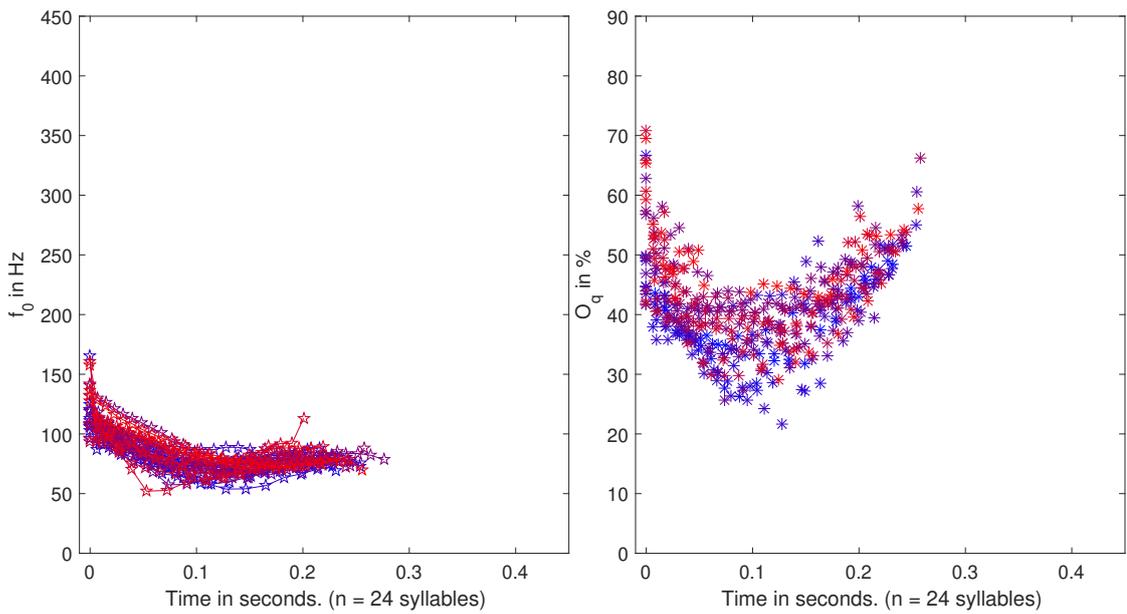


(l) Speaker M5

Figure 4.4: The glottalized tone of Kim Thuong Muong: speaker by speaker.  $f_0$  <sub>dEGG</sub> on the left and  $O_q$  <sub>dEGG</sub> on the right.



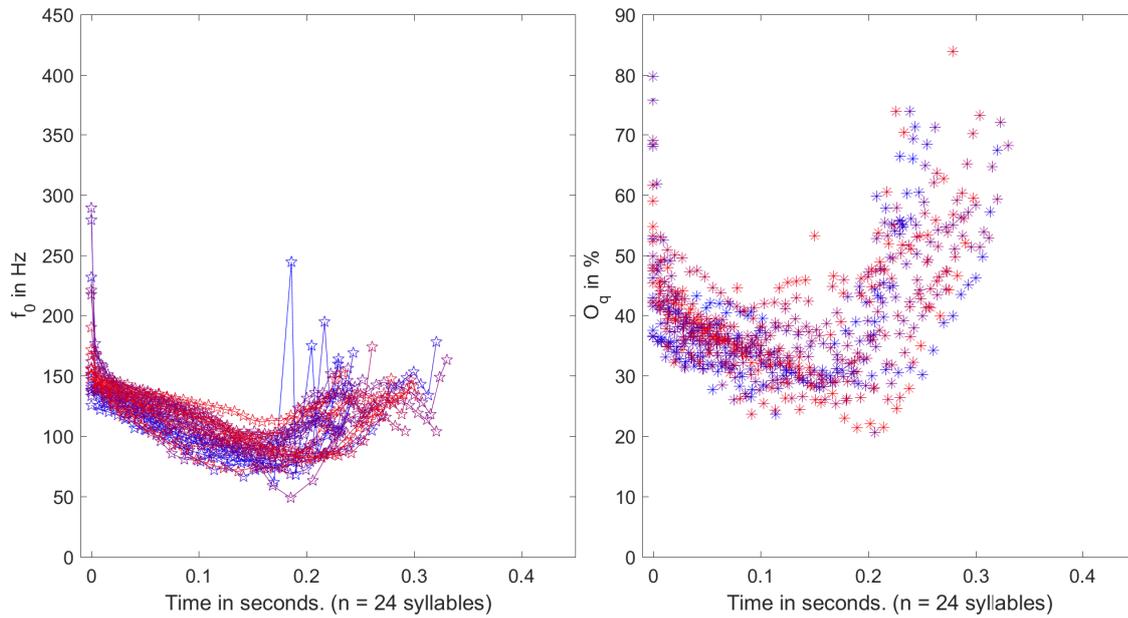
(m) Speaker M7



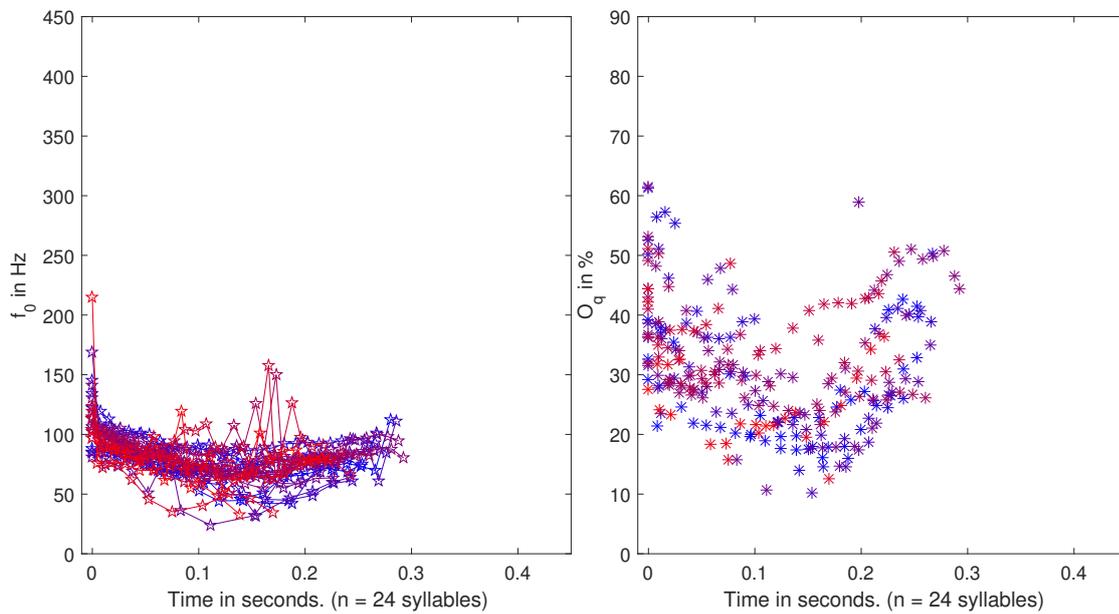
(n) Speaker M8

Figure 4.4: The glottalized tone of Kim Thuong Muong: speaker by speaker.  $f_0$  <sub>dEGG</sub> on the left and  $O_q$  <sub>dEGG</sub> on the right.

4.3 Glottalization as a component of lexical tone: A closer look at the glottalized tone across 20 speakers

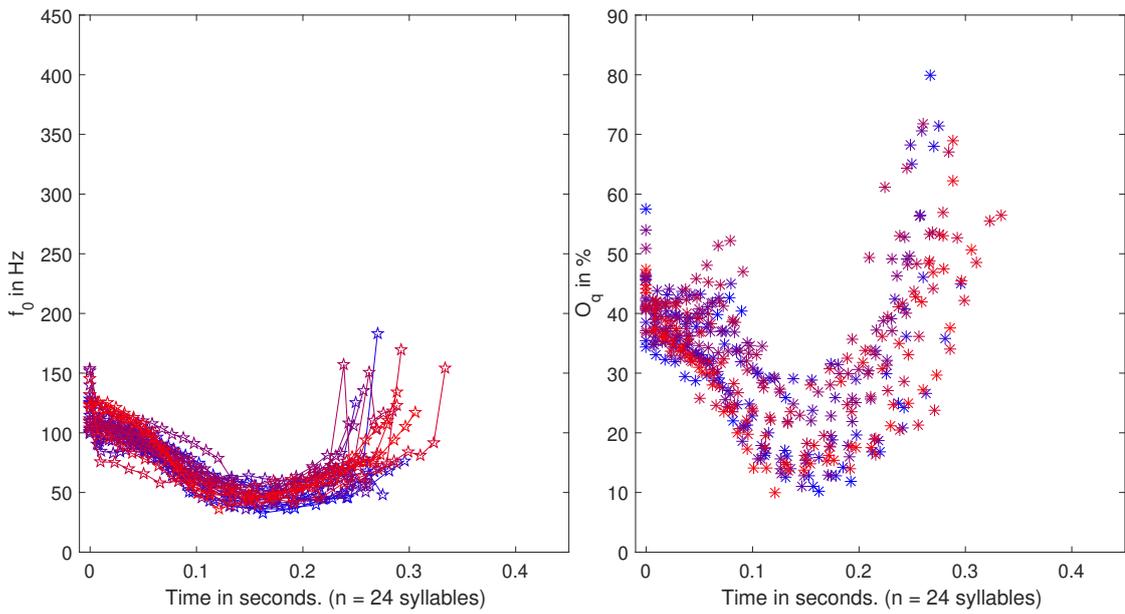


(o) Speaker M9

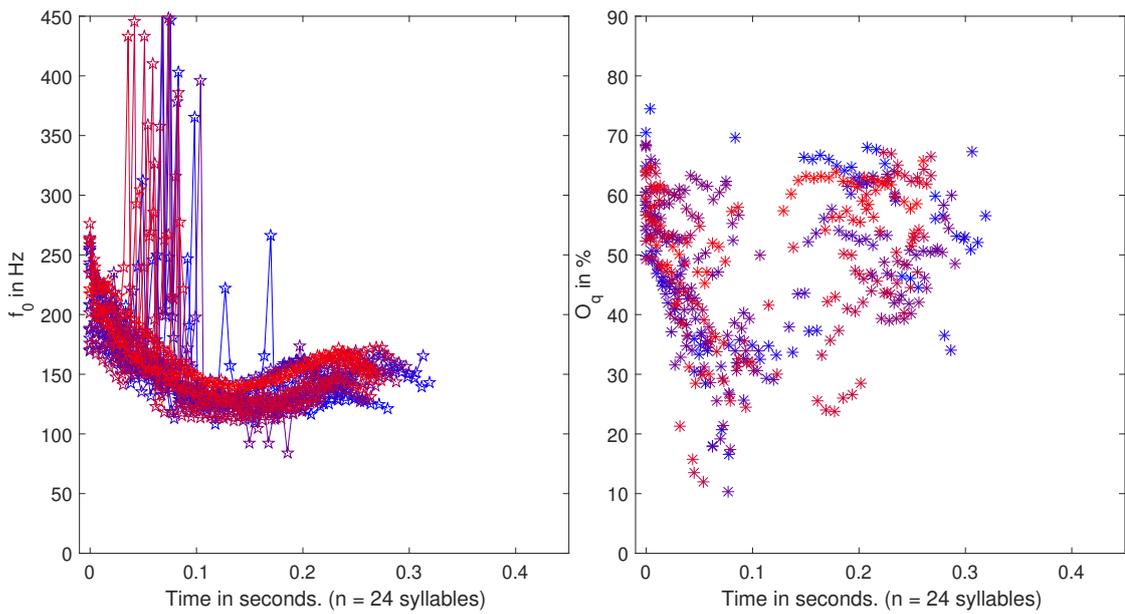


(p) Speaker M10

Figure 4.4: The glottalized tone of Kim Thuong Muong: speaker by speaker.  $f_0$  <sub>dEGG</sub> on the left and  $O_q$  <sub>dEGG</sub> on the right.



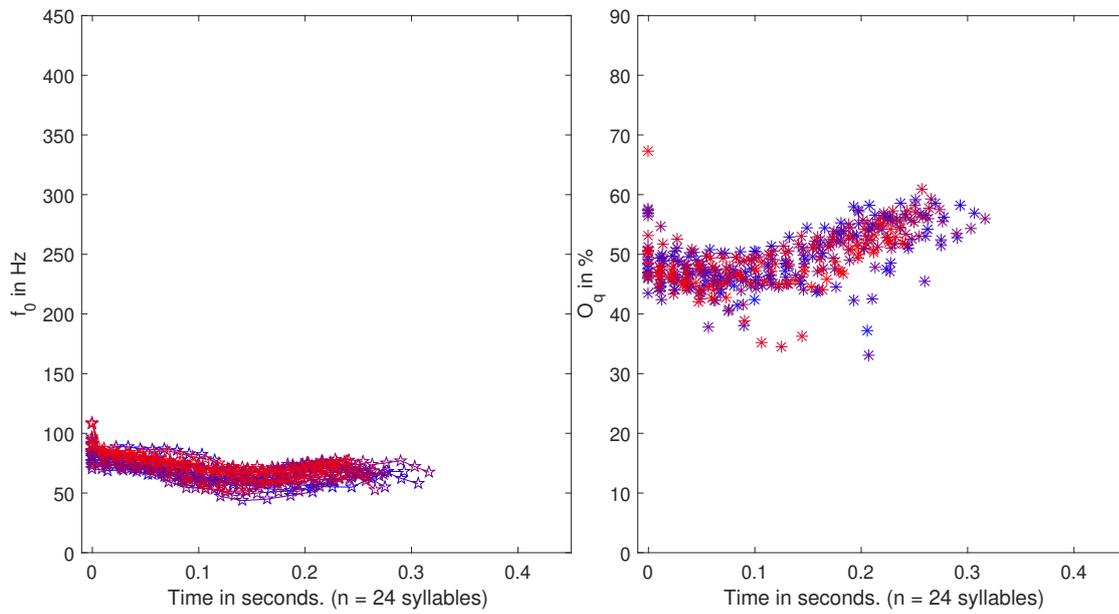
(q) Speaker M11



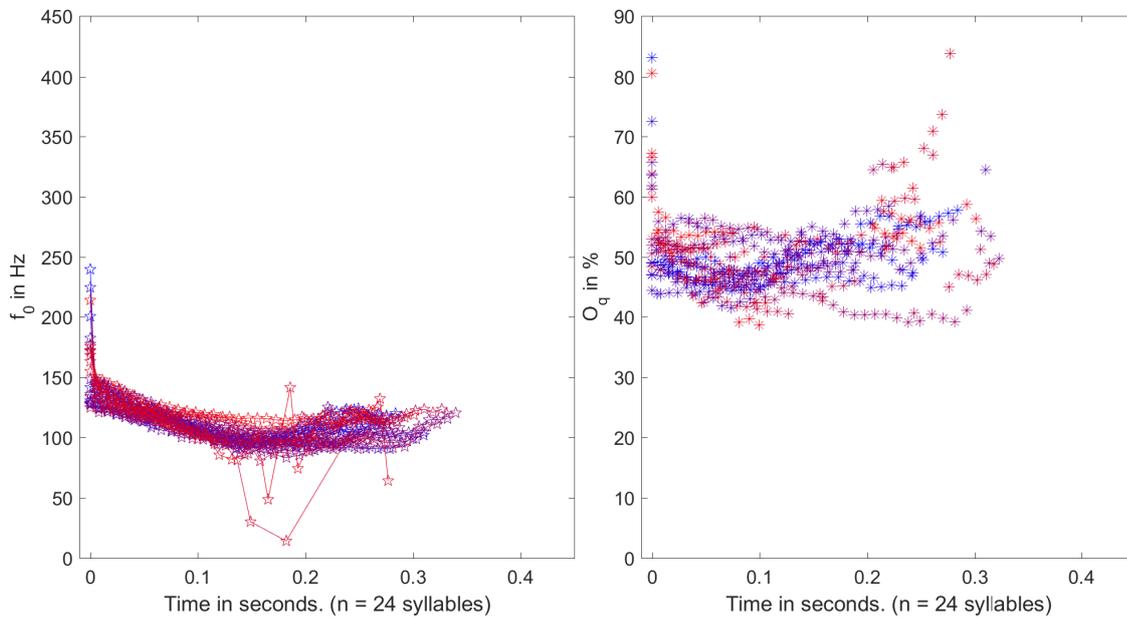
(r) Speaker M12

Figure 4.4: The glottalized tone of Kim Thuong Muong: speaker by speaker.  $f_0$  <sub>dEGG</sub> on the left and  $O_q$  <sub>dEGG</sub> on the right.

4.3 Glottalization as a component of lexical tone: A closer look at the glottalized tone across 20 speakers



(s) Speaker M13



(t) Speaker M14

Figure 4.4: The glottalized tone of Kim Thuong Muong: speaker by speaker.  $f_0$  <sub>dEGG</sub> on the left and  $O_q$  <sub>dEGG</sub> on the right.

jittery curves of  $f_{0\text{ dEGG}}$  only appear in a few items during the middle part of the experiment, as evidenced by the maroon-colored values at 200 Hz. Speaker F21 (4.4j) is the only one that does not have any item with jittery  $f_{0\text{ dEGG}}$ . This does not mean that women only do glottalization with jittery  $f_{0\text{ dEGG}}$ . In the cases of speakers F13, F17 and F20, it is easy to spot that there are clearly V-shaped curves that make the middle part (in F13) and the bottom part (in F17 and F20) more consistent and dense. This is much less obvious in the other speakers but from my logs during data processing, I can confirm that women mix different ways of creaking/glottalizing but apparently prefer to creak irregularly.

In an opposite situation, there is one man out of ten whose **Tone 4** has characteristics of jitter, namely speaker M5 (4.4l). Five other speakers have just a few items with jitter: M1 (4.4k), M9 (4.4o), M10 (4.4p), M12 (4.4r) and M14 (4.4t). And the remaining four speakers (M7 (4.4m), M8 (4.4n), M11 (4.4q) and M13 (4.4s)) have this tone without jitter.

This difference in glottalized characteristics between male and female speakers is reflected well in the distribution of  $f_{0\text{ dEGG}}$  in Figure D.2 in Appendix D. In general, the interquartile range of the female data is much broader than that of the male, especially in the case of speakers F30, F9, F12 and F20 with a spread of at least 70 Hz. The two speakers with negligible and no jitter in the  $f_{0\text{ dEGG}}$  curves, namely F19 and F21, have the narrowest interquartile range.

In the male data, the distribution of  $f_{0\text{ dEGG}}$  does not really reflect what we can observe in the raw data in Figure 4.4 since the only male speaker M5 reported having jittery  $f_{0\text{ dEGG}}$  as in females but his interquartile range is not the widest one, but rather M11. This is the speaker whose **Tone 4** I could say is the most creaky, both perceptually and signal-visually: the V-shape of **Tone 4** is the most plunging of the whole data set. Two of the six typical examples of sub-types of creaky voice in **Tone 4**, shown in Figures 5.1 and 5.1, are from his data.

Speaker M12 (4.4r) is noted as a special case where the first half of **Tone 4** appears unusual with extremely high values, highly implausible for a male voice, above 400 Hz. This is reflected in the box plot (D.2) in Appendix D with the most numerous and highest outliers above 200 up to 500 Hz to be found in this speaker's data. The cause of this anomaly comes from double closing peaks with an early closing micro-peak that occurs at the beginning of each cycle (as shown in Figure 4.5) resulting in detection errors.

This particular characteristic of M12 is clearly due to the fact that this speaker has glottalization in **Tone 4**: the detection errors frequently occur in this tone. As can be seen in Figures 4.1r and D.1r, the amplitude of the standard deviation at the beginning of **Tone 4** is large and irregular and spans at least five half tones for one side. While the other tones do not show these erratic fluctuations.

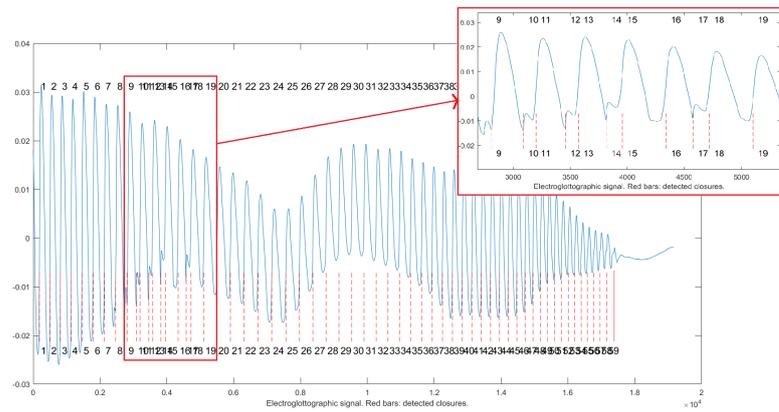
In fact, with this kind of signal, we can technically treat the micro-peak as part of a cycle since it is not a complete cycle. The method of re-setting the threshold to integrate the micro-cycles within a larger cycle will not work since their amplitude is almost equal to that of the normal cycles, as can be seen in the dEGG signal in

Figure 4.5b. It is possible to use the more manual method to combine the micro-cycle with its adjacent cycle, so we will have a smooth  $f_{o\ dEGG}$  curve without those unreasonable values. However, when processing the signal, I decided to keep the result that `PEAKDET` automatically detected, to keep in mind this fascinating phenomenon. There is a particular gesture that occurred during the glottalization of **Tone 4** in this speaker; it is perhaps no coincidence that he is the oldest participant (he was born in 1937).

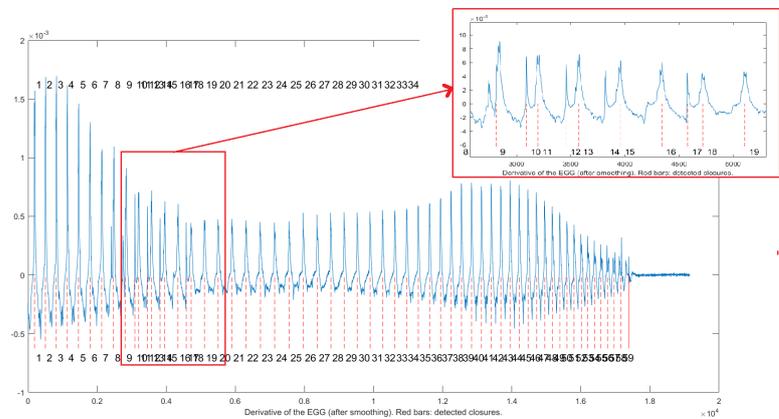
Regarding the  $O_{q\ dEGG}$  parameter, as can be seen in the graphs on the right of Figure 4.4, more  $O_{q\ dEGG}$  values had to be discarded in the cases where there is jitter, with a sparser density of values. Notably, in the case of speaker F20, only two elements retained their  $O_{q\ dEGG}$  curves, namely is the last target syllable of the list in Table 3.1, over syllable **ku**<sup>4</sup>.

In addition, very high  $O_{q\ dEGG}$  values, above 60%, are recorded much more often in cases of jittery  $f_{o\ dEGG}$ , which is actually easy to explain due to the smaller cycles. It seems unreasonable to have such high  $O_{q\ dEGG}$  values in a glottalized portion, and especially in creaky voice, since this phonation type is at the bottom of the speaker's voice range, below the modal register. On the other hand, I find that if we suppress all the precise opening peaks of the small cycles, it will create an artefact. There seemed to be no point removing high  $O_{q\ dEGG}$  values: instead, they are part of the results obtained by the method employed here, which follows vocal fold vibration very closely, and treats successive positive peaks as evidence of distinct glottal pulses, as a (debatable) matter of convention.

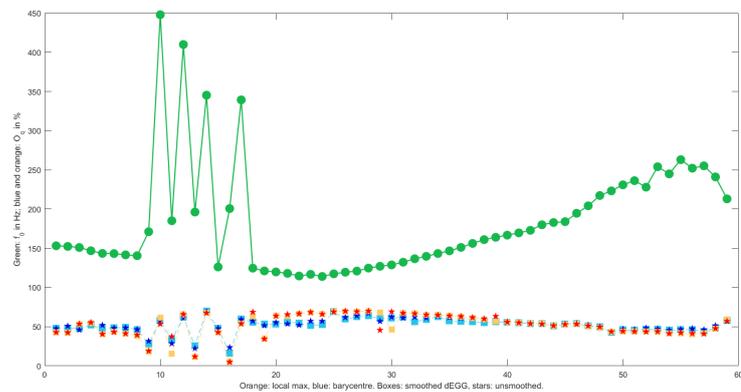
According to the characterization and sub-classification of creakiness in Section 5.2, we know that the jittery  $f_{o\ dEGG}$  curves here mean that women tend to perform multiply-pulsed creak with double complex-repetitive patterns, or aperiodic creak which complete irregularity in successive pulses. Men, on the other hand, generally have single pulsed creak or usually pressed voice.



(a) EGG signal with a zoom on cycles 9 to 19 where the micro-closing peaks appear.



(b) dEGG signal with a zoom on cycles 9 to 19 where the micro-closing peaks appear.



(c) Result of  $f_0$  dEGG (green) and  $O_q$  dEGG (blue and orange).

Figure 4.5: An example illustrates the unusual beginning of speaker M12's Tone 4 with unreasonably high  $f_0$  dEGG values caused by micro-closing peaks. Data from speaker M12, token UID: 2002, target syllable /laj<sup>4</sup>/ in isolation, second performance.

## Chapter 5

---

### Discussion

This chapter builds on the “ground-level” information provided in the previous one (*Results*) to propose various higher-level elaborations about the tonal contrasts found in Kim Thuong Muong. The reflections first hinge on the outlook of the tone system (in section 5.1). A second part of the chapter is devoted to glottalization as seen in the perspective of general phonetics: exploring in what sense, and to what extent, the facts observed in Muong can make a contribution to a fine-grained understanding of the linguistic roles of glottalization and creaky voice. From there, we progress towards the broader picture of Muong prosody: the interplay between tone and intonation (§5.5). As a final flourish, observations about a potential *intonational tone* in Muong are set out (§5.5.4): a prosodic pattern on two sentence-final particles (the one interrogative, the other affirmative) that illustrates the space for linguistic creativity opened up by the presence of glottalization in the tone systems of languages in contact (in this instance, Muong and Northern Vietnamese).

#### 5.1 General outlook of the Kim Thuong Muong tone system

A first thing to say about the Kim Thuong Muong tone system is that its general outlook is here confirmed to be as described in my earlier work (M.-C. Nguyễn, 2016). It has five tones, plus two: its tones are divided into two sub-systems, five contrasting tones on smooth syllables and two tones on checked syllables. No additional tones were turned up in the course of fieldwork since 2016 (although interesting phenomena at the border line between tone and intonation did come up, as will be elaborated on in §5.5). For the sake of simplicity, the tones are labeled with numbers from one to five for the *smooth* system, and six and seven for the *checked* system. There is no specific logic to this numerical labeling, which simply followed an order of discovery in early fieldwork. Readers can refer back to Table 2.6, which recapitulates each tone’s phonetic-phonological characteristics and provides shorthand names and corresponding labels using the five-level system taken up in the International Phonetic Alphabet (Chao, 1930).

As to the phonetic/phonological properties of each of the tones, there is, again, an overall good match between the generalizations proposed back in 2016 and the fresh observations on the tone system reported here, based on systematic implementation of the most basic procedure to study a tone system experimentally – recording minimal

sets. The results from twenty speakers confirm the organization of Kim Thuong Muong tones into what I find to be an elegant, neat and symmetrical space (as shown in Figures 4.1, 4.2 and 4.3 of the previous chapter). The system makes use of combinations among several phonetic dimensions to contrast between the tones: levels, contours, plus creaky voice.

Firstly, pitch levels suffice for the distinction of two level tones: **Tone 1** and **Tone 5**.

**Tone 1** is phonetically flat and typically located around the middle of the speaker's range, usually a little lower than the speaker's average, as can be confirmed by comparing the results in Figure 4.1 and the average values provided in Table 4.1. On the basis of this characteristic, I shall refer to this tone as a low-mid level tone.

**Tone 5**, in direct contrast to **Tone 1**, is another flat tone but at a higher pitch level. This is actually the highest tone in the system, with the  $f_{0 \text{ dEGG}}$  values held at the top of the speaker's range throughout its duration.

These two tones are generally well maintained in a moderate and effective pitch distinction, without the support of other phonetic parameters to enhance the  $f_0$  contrast or supplement it. Although, as can be seen in Figure D.1, most speakers have these two tones with only five semitones of difference, I have never had difficulty recognizing them. (On this and other points, readers can form an opinion for themselves by listening to the data set, which is available from the Pangloss Collection.<sup>1</sup>)

The second factor used to contrast the tones is modulation of *pitch*. Besides the two level tones that have distinct pitch levels, the other three tones each has a unique pitch contour.

**Tone 2** is a falling tone. It has a high beginning and descends to a low ending, as low as **Tone 1**, but not as low as **Tone 4** (i.e., the glottalized tone). This tone normally starts at the same onset as the top-high level tone (**Tone 5**) and ends near the offset of the low-mid level tone (**Tone 1**), with a shape similar to a children's slide: a mild slope at both ends and a steeper drop in the middle.

**Tone 3** goes in an opposite direction. It is characterized by a rising curve, although the extent of the rise in this tone is usually not as great as that of the fall in **Tone 2**. This is not surprising from a typological perspective, as a rise in  $f_0$  is perceptually more salient than a fall (all other things being equal).

Contrasted with all the others, **Tone 4** is the only tone to possess a complex contour: falling-rising with the onset and offset located at almost the same pitch level, below the low-mid level tone (**Tone 1**). As mentioned earlier, most speakers have an  $f_{0 \text{ dEGG}}$  curve for **Tone 4** that stands out by a clear dip in the middle, which is more U-shaped than V-shaped.

A further phonetic parameter, which sets **Tone 4** apart from all the others, is the contrast of phonation types between the modal voice and the creaky voice. It can be observed in the graphs of  $O_{q \text{ dEGG}}$  values (on the right of Figures 4.1 and 4.2), which provide insights into the degree of vocal fold adduction, that **Tone 4** is the only tone clearly distinguished from the others. In most cases, it would not be difficult to draw a line to separate **Tone 4** (red line) from all the rest of the system. The  $O_{q \text{ dEGG}}$

<sup>1</sup><https://pangloss.cnrs.fr/corpus/Muong>

curves of the four tones in the upper part overlap, and they are systematically located around a mean value of 50%, and the bottom part of the standard deviation is generally around 40%, i.e. these tones are squarely in the range of modal voice.

By contrast, in the lower part, only **Tone 4** is present: the  $O_{q \text{ dEGG}}$  curve of this tone plummets from initial mid-range values (which are already lower than those of all the other tones) to the very bottom of the range. Most of the  $O_{q \text{ dEGG}}$  values from **Tone 4** had to be excluded at the stage of data verification, reflecting the issue of unclear opening peaks on the derivative of the electroglottographic signal. The values retained are thus to be taken carefully. It is nonetheless safe to conclude, in view of the data from the twenty speakers, that there is a consistent presence of very low values of  $O_{q \text{ dEGG}}$ , especially midway through the rhyme, where the values are commonly below 30%. These values are strong evidence for the presence of creaky voice.

In Kim Thuong Muong, canonical realizations of the glottalized tone (**Tone 4**) show a lapse into creaky voice rather than a glottal stop or glottal constriction. Investigations into different kinds of materials (such as narratives, or elicited materials in which the syllables carrying **Tone 4** are not under focus) are likely to uncover a wider range of realizations, however. Since **Tone 4** is the only glottalized tone in the Kim Thuong Muong system, its glottalization can be expected to show a larger field of allophonic variation than in tone systems that contrast two or more glottalized tones (such as Hanoi Vietnamese, for instance). It stands out nonetheless as an important observation about **Tone 4** that its nonmodal phonation type is *consistently creaky*. A phonological interpretation is that this tone is specifically a creaky one: not simply a phonologically glottalized tone (in a general sense) that happens to have much creak for some sort of low-level phonetic reason.

It may come as a surprise that no mention was made so far of duration, often an important (potentially distinctive) characteristic in tone systems of East and Southeast Asia. It must be stated that duration does not play any detectable role in the contrasts between the five tones of the smooth syllables. Although there are some measurable average differences, such as that the two tones with a higher offset (**Tone 5** and **Tone 3**) are generally shorter than the two tones with a lower offset (**Tone 1** and **Tone 2**), and **Tone 1** is always the longest tone in the system, these differences are negligible: they are only a few milliseconds long.

On the other hand, the duration is markedly different between smooth and checked syllable rhymes. As can be seen in Figure 4.3, the checked tones are roughly half as long as the smooth tones. This bears a clear relationship to obstruent codas. The final consonants /p/, /t/, /c/, and /k/ are phonologically unspecified for voicing (there is no voicing opposition among the final consonants in Kim Thuong Muong), but these codas are phonetically unvoiced, and appear to exert the full shortening effect that is cross-linguistically associated with final voiceless stops.

Within the checked sub-system, **Tone 6** and **Tone 7** are both phonetically short and flat. There appears to be a relatively clear-cut parallel course between a mid-level tone (**Tone 6**) and a high-level tone (**Tone 7**).

Concerning the checked sub-system, there are two points that I would like to discuss

here. First, by comparing the current result of the checked sub-system (Figure 4.3) with the one I reported in the 2016 study (M.-C. Nguyễn, 2016, pp. 66–67), one can easily notice that the results are vastly different. The two checked tones in this latest study are both phonetically flat, whereas those in the previous study were clearly descending (Tone 6) and ascending (Tone 7), respectively.

This difference is caused by the different materials of the phonetic experiments. In this study, we used only minimal pairs of real words, which ensures that the result here is proper. It was a shortcoming of the 2016 experiment that I did not have material on real minimal pairs of checked syllables, but only relied on nonsense words combined from some phonemes that exist in Kim Thuong Muong, as explained in (M.-C. Nguyễn, 2016, pp. 22–26) – clearly not a reliable method, as it turned out. The method of using nonsense words in phonetic and phonological experiments is actually quite common (Michael S Vitevitch et al., 1997a; Michaud, 2004b; Hay, Drager, and Thomas, 2013b). However, this method turned out to be inappropriate in an unwritten language like Muong.

First of all, it is an issue how to show the phonemic combinations for nonsense words in an unwritten language. There is no other way but to borrow another writing system that all native speakers know and use fluently, in this case Vietnamese. This leads to an obvious drawback: the consultant’s experience in reading and writing is intimately tied to their practice of the Vietnamese language. All of the consultants had received training in Vietnamese writing throughout their elementary school years, whereas none had received training in Muong writing in school or later.

I thought it was possible to adapt the writing system of Vietnamese to the Muong, and, through a thorough training phase, to instruct consultants to respond to stimuli using Muong tones and not Vietnamese tones. But this was definitely a failure. The Muong language does not have an established writing system to encode the language systematically, and their native speakers use the language as an oral one and are mostly unaware of the specifics of their phonemic system. In this context, “classroom” behavior is likely to occur at various points: code-switching between the sound systems of Muong and Vietnamese, and perhaps mixing the two in varying proportions.

The task becomes even more impossible when even the researcher does not firmly grasp the difference between the two languages. With my nascent experience in the early stages of studying this dialect, I struggled to get rid of the inherent Vietnamese tonal system in order to learn Muong, which bears a clear overall family resemblance but is truly different in detail. Using the Vietnamese tone background (“this tone of Muong sounds like that Vietnamese tone”) caused me to lose a lot of time in the process of acquiring and perceiving another tone system – but at least it allowed me to clarify forever that those are truly two independent tone systems. The nonsense word experiment was conducted during that stage of my learning, and so I was not self-assured enough in my mastery of Muong tones to monitor code-switching problems as they were occurring.

However, even if the researcher is well versed in both languages, it is still not a reliable guarantee that such an incident will not occur when the phonemes of the two

languages are closely similar. My failure demonstrated that nonsense word experiments are unlikely to be possible in an unwritten language, and in any case, that they involve clear risks of code-switching between two languages.

The second point I would like to discuss here is the question of the relationship between two sub-systems in tonal languages, i.e., between smooth tones and checked tones.

In Vietnamese, the most closely related language to Muong, there have been opinions that the two checked tones (etymologically: the D-tones) could be allophones of two B-tones, due to their similar perception. Dating back to the 17<sup>th</sup> century up to now, the orthography used the same diacritics and also the same names: “*thanh sắc*” for B<sub>1</sub> and D<sub>1</sub>, and “*thanh nặng*” for B<sub>2</sub> and D<sub>2</sub>. However, the orthographic convention whereby they are written in the same way may initially have been a matter of convenience: the authors of the romanized Vietnamese alphabet had to use a wide set of diacritics for tones, and using seven tone diacritics rather than five would have been uneconomical, since (to use terminology that is anachronistic with reference to 17<sup>th</sup> century language work) there is a different distribution for the stopped tones and the smooth tones, which do not occur on the same rhymes. The stopped and smooth tones of Vietnamese are now generally recognized to be phonetically distinct (Kirby 2011, *pace* Earle 1975).

For Muong, the issue is whether the two checked (stopped) tones, **Tone 6** and **Tone 7**, are allophones of some smooth tones: for instance, of **Tone 5** and **Tone 1**, which are somewhat similar phonetically. Looking at the  $f_0$  <sub>dEGG</sub> tracings in Figure 4.3, **Tone 6** and **Tone 7** visually look like a shorter version of two smooth level tones, **Tone 5** and **Tone 1**, respectively: the pitch heights and the shape of the contours look very much alike.

However, from a structural point of view, it is clearly better to avoid use of the notion of allophony here, and consider the checked tones as a standalone system. Kim Thuong Muong’s tone system has 5+2 tones, with two sub-types of five smooth tones and two checked tones.

All in all, the tonal system (as viewed through  $f_0$  <sub>dEGG</sub> tracings) thus appears relatively symmetrical, consisting of two parts. In the upper part, there are a low-mid level tone (**Tone 1**) contrasting with a top-high level tone (**Tone 5**), and a falling tone (**Tone 2**: falling from the top to the low-mid of the speaker’s range) contrasting with a rising tone (**Tone 3**: rising from low-mid to the top of the speaker’s range). Meanwhile, in the lower part, the glottalized tone (**Tone 4**) looks symmetrical along the time scale, with offset values close to onset values, and glottalization (canonically, creaky voice) in-between. The checked sub-system also participate in this well-established synchronic pattern distinguishing higher and lower tones, with a clear and parallel contrast between a high-level-checked tone (**Tone 6**) and a mid-level-checked tone (**Tone 7**).

These observations hold for all twenty speakers, strongly suggesting that these are characteristics of the Kim Thuong Muong tone system, not idiosyncratic properties (specific to one or several speakers). In addition, since these characteristic are not effect of phonemic composition such as vowel quantity, vowel height, stop or fricative character of the initial consonant, or the like, this confirms that they are properties of

lexical tones, not other phonemic components.

This result is consistent with the suggestion that Kim Thuong Muong's tones are not realized by pitch alone, but by a complex combination of characteristics: most saliently the combination of pitch and nonmodal phonation types commonly found among Vietic languages. It reinforces the observation that all Vietnamese languages have at least one glottalized tone. Besides the well-known case of Vietnamese, we also have some studies on other languages and dialects, such as Arem (Ferlus, 2014), Ruc (Tấn Thành Tạ, 2021). The next step is to make comparisons of these languages to bring out the general pictures of glottalization, how they are similar or different in terms of characteristics and function, and to what extent their variations and their role in the tonal system. In this study, we will be able to make some comparison between Kim Thuong Muong and Vietnamese because the similar method and result of Vietnamese in (Michaud, 2004b; Brunelle, 2009b; Brunelle, D. D. Nguyễn, and K. H. Nguyễn, 2010; Kirby, 2011) that allows for a reasonable comparison here. Comparisons with other Vietic languages and dialects are worth continuing later.

## 5.2 A characterization and classification of sub-types of creaky voice based on audio and electroglottographic signals

Building on the framework of glottalization concepts and nomenclatures summarized in the previous section, I attempt here to suggest a sub-classification of creaky voice into subtypes attested among realizations of the glottalized tone (Tone 4) in Muong, illustrating each sub-type by an example.

By sifting through the data obtained, it is not difficult to notice that creaky voice in Tone 4 is realized with much variation – which can be described as *allotonic variation*, since creak is part of the tone's phonological specification. An obvious thing to do at this point is to clarify the nomenclature. A four-way classification will be proposed as a basis for further analysis. But before presenting and exemplifying these four types, two clarifications are in order here.

Firstly, since there is so far no consensus in the phonetic classification of creaky voice into sub-types, the name chosen here for each of the four proposed sub-types is essentially based on its salient phonetic characteristics: *single-pulsed creak*, *multiply-pulsed creak*, *aperiodic creak*, and *pressed voice*. This choice appeared better than borrowing from the list of commonly-used labels from existing nomenclatures (as discussed in §2.3.1).

A second clarification is that the representative examples chosen here for each type are admittedly cherry-picked: they constitute extreme cases – almost *ideal*, even though they are fully real examples from the experimental dataset collected for the present study. These examples are selected to bring out with greatest clarity the properties that characterize the categories as classified here. There are frequently less clear-cut cases, cases on the borderline between certain categories, and complex combinations of two or more categories. Those are definitely worthy of more attention than could be granted to them within the time frame of the present study. Addressing these cases

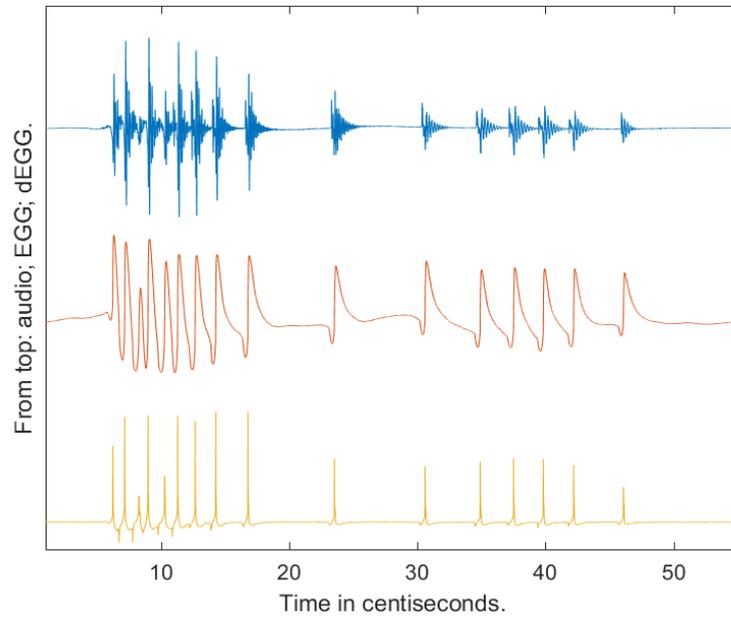
in full detail is a task that has to be deferred until future work; it is hoped that the present discussion can serve as a useful stepping-stone in that direction and there can be cumulative progress.

### ***Type 1: Single-pulsed creak***

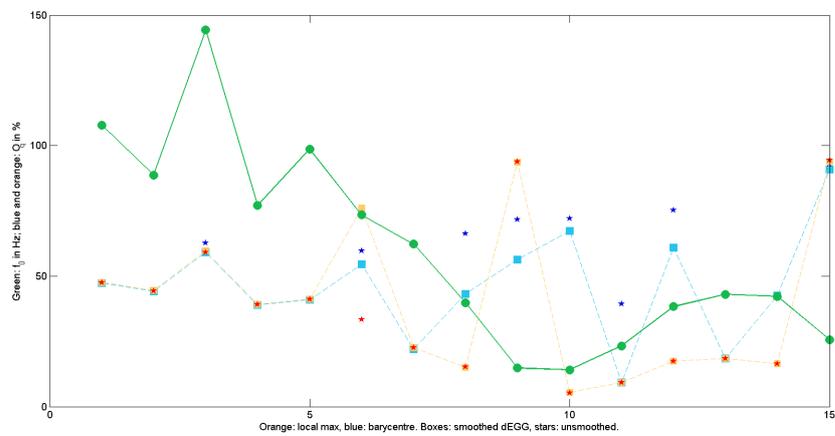
Going from simple to complex, the first sub-type that we identify and classify as a sub-category of creakiness is single-pulsed creak, exemplified in Figure 5.1. Three key properties of this sub-type are: (i) glottal cycles (periods) are long, and remain a train of single, discrete pulses, as in modal voice, and as opposed to the double-pulsed or multiple-pulsed patterns which will be described below; (ii) both  $f_{0 \text{ dEGG}}$  and  $O_{q \text{ dEGG}}$  reach low values (the lowest in comparison to the three other types of creak, and – conspicuously – also very low as compared with modal voice); and (iii) quasi-periodicity in a strict technical sense might become lost when the pulses get longest (i.e. there can be visually salient differences in the durations of successive glottal cycles) but these variations do not go so far as to interrupt voicing. This is an important difference from the perception of a telltale *interruption* of voicing in a glottal stop.

This sub-type is associated with different labels according to different classifications, such as *creak* or *vocal fry* in Hedelin and Huber (1990) (a classification which distinguishes *creak* from *creaky voice*), *damping* or *creak* in Batliner et al. (1993), *creak* in Redi and Shattuck-Hufnagel (2001), and *prototypical creaky voice* or *vocal fry* in Keating, Garellek, and Kreiman (2015).

Two examples of single-pulsed creak are provided in Figures 5.1 and 5.1 in order to show how wide the range of variation of this sub-type can be. Both examples are taken from exactly the same dataset of speaker M11, in the same first minimal set. They are from the syllable /paj<sup>4</sup>/. The only difference is that they are in two different contexts. The first example (in Figure 5.1) is spoken in isolation. This is the most extreme case of this type of creak that was encountered to date. It is clearly apparent from the signals and the extracted parameters that creakiness occurs right from the beginning of the rhyme. The pulses are stretched to a maximum duration in the 9th to 11th cycles. Quasi-periodicity is lost, as there is uneven duration across successive cycles, but visually and auditorily, there is no real *interruption* of voicing. The bottom pane shows the  $f_{0 \text{ dEGG}}$  and  $O_{q \text{ dEGG}}$  calculated automatically with **PEAKDET** (which, as explained in the Method chapter, operates cycle by cycle, not through autocorrelation). It can be seen that five or six cycles of jittery  $f_{0 \text{ dEGG}}$  are followed by rock-bottom values (below 40 Hz and even down to 20 Hz): extremely low-frequency voicing. Fundamental frequency and open quotient rise again after reaching the lowest values, but remain low (lower than they are early on in the rhyme). The opening peaks on the derivative of the electroglottographic signal are still (just barely) clear enough to allow for confident evaluation of the glottal open quotient. It is not unreasonable, in view of the shape of the electroglottographic signal, to consider that these cycles have an extremely short open phase, and that the lowest  $O_{q \text{ dEGG}}$  values (those in orange, correcting for two outliers at the 9th and 15th cycles) provide good estimates:  $O_{q \text{ dEGG}}$  is on the order of just 5% (i.e. rock-bottom values, like for  $f_{0 \text{ dEGG}}$ ) for the longest

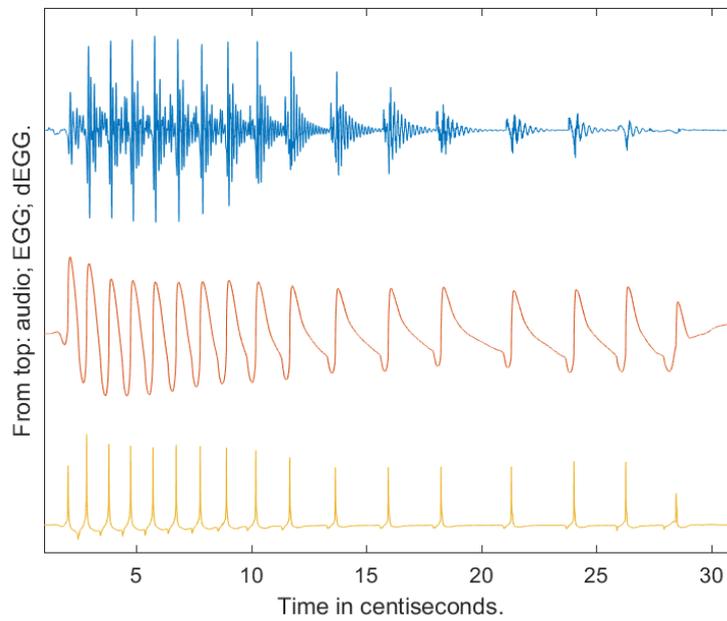


(a) Signals. From top to bottom: (i) acoustics, (ii) EGG, (iii) dEGG

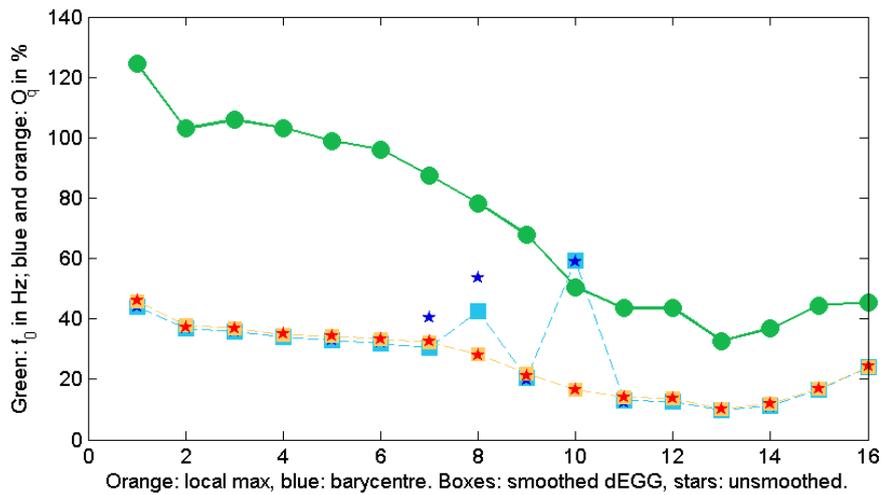


(b) Two parameters automatically calculated by `PEAKDET`:  $f_{0 \text{ dEGG}}$  (green) and  $O_{q \text{ dEGG}}$  (blue and orange).

Figure 5.1: Example of single-pulsed creak: extreme case. Data from speaker M11, token: 41, UID: 0501, over syllable /paj<sup>4</sup>/ in isolation.  
DOI: <https://doi.org/10.24397/pangloss-0006782#W9>



(a) Signals. From top to bottom: (i) acoustics, (ii) EGG, (iii) dEGG.



(b) Two parameters automatically calculated by `PEAKDET`:  $f_{0 \text{ dEGG}}$  (green) and  $O_{q \text{ dEGG}}$  (blue and orange).

Figure 5.1: Example of single-pulsed creak: common case. Data from speaker M11, token: 47, UID: 0531, over syllable /paj<sup>4</sup>/ in carrier sentence. DOI: <https://doi.org/10.24397/pangloss-0006782#W10>

cycles. This case undeniably has to be considered as creak (phonation mechanism zero in terms of the classification of Roubeau, Henrich, and Castellengo (2009)). Overall, this example can be described as an extreme sample of clear lapse into creaky voice, with the lowest possible  $f_{0 \text{ dEgg}}$  and  $O_{q \text{ dEgg}}$ , but still with a single pulse per cycle. A possible physiological interpretation is that phonation is almost arrested by the strong glottal constriction, and only continues ‘pulse by pulse’ as puffs of air find their way through the closed sphincter.

The second example of single-pulsed creak (Figure 5.1), on the other hand, is from a token spoken in the carrier sentence. We recognize this as the commonest case of single-pulsed creak because, compared to the previous case, this one occurs much more frequently, and across speakers. In other words, whereas the previous example is only found particularly in the case of speaker M11 when the words are spoken in isolation, and can hence be seen as a case of hyper-articulation, the sample in Figure 5.1 is a representative of this sub-type of creak as it appears in most of the cases and for most speakers.

The token shown in Figure 5.1 does not show stunningly long, discrete pulses as in the previous example. Instead, the length of successive cycles increases with some regularity, before decreasing again. There are no sudden, dramatic changes. The cycles are gradually stretched out (here: during the first 8-9 cycles), in a gentle transition towards the very low  $f_{0 \text{ dEgg}}$  values in the second half of the rhyme: a low range where  $f_{0 \text{ dEgg}}$  values remain until the end of voicing. As a result, the values of both  $f_{0 \text{ dEgg}}$  and  $O_{q \text{ dEgg}}$  do not reach the same extremes as in Figure 5.1b, but they are still very low. Fundamental frequency gets under 100 Hz, reaching down to 40 Hz towards the end (in the 13th cycle). The  $O_{q \text{ dEgg}}$  values are consistently below 40% and even drop to 20% in the second half, an indicator of the presence of strong glottalization. In consulting other works, this kind of signal seems closest to what is proposed as *creak (vocal fry)* in Hedelin and Huber (1990, p. 361) and as *damping* in Batliner et al. (1993, p. 7).

Despite the huge difference between the two samples examined above, we believe that they should not be considered as two distinct sub-types, but as two distant points along a continuum characterizing one and the same type, namely single-pulsed creak. The interpretation put forward here is that the actual phonetic realization will be closer to one end of the continuum or the other based on differences in prosodic environments. The more emphasis is placed on a (monosyllabic) word, the more care will be put into its articulation, and the more time will be spent on it (allowing for special cases such as phonologically – contrastively – short phonemes, of course, whose lengthening could jeopardize their proper identification, defeating the communicative purpose of emphasis). This intuition can be stated in terms of allocation of resources: when speakers focus on the syllable (including its tone) and hyper-articulate it, then the realization gets closer to the extreme side of the continuum, whereas when resources allocated to the tone are less abundant, then the realization is closer to the other end. Thus, the fewer constraints are placed on phonetic realization by a syllable’s context (coarticulatory constraints and temporal constraints), the longer the single-pulsed creak, and the more

discrete (far apart) the successive pulses.

On this basis, the detection of this kind of signal (single-pulsed creak) can be based on the rate of  $f_{o \text{ dEGG}}$  change (hereafter called *delta  $f_o$  rate*). Comparing the rate of change (in absolute values) of the two samples examined above with the rate found in tones other than **Tone 4** (tones devoid of phonological creak, and hence hypothesized to have non-creaky phonation: modal voice, for short) to set a threshold, a first-pass approximation is to set a threshold at 10%: this value works well for the two samples under consideration, as can be seen in Figure 5.6 (please mind the difference in y axis scales for the two sub-plots, (a) and (b)). This topic will be taken up further below, and the threshold adjusted accordingly, as we encounter further cases of single-pulsed creak.

### ***Type 2: Multiply-pulsed creak***

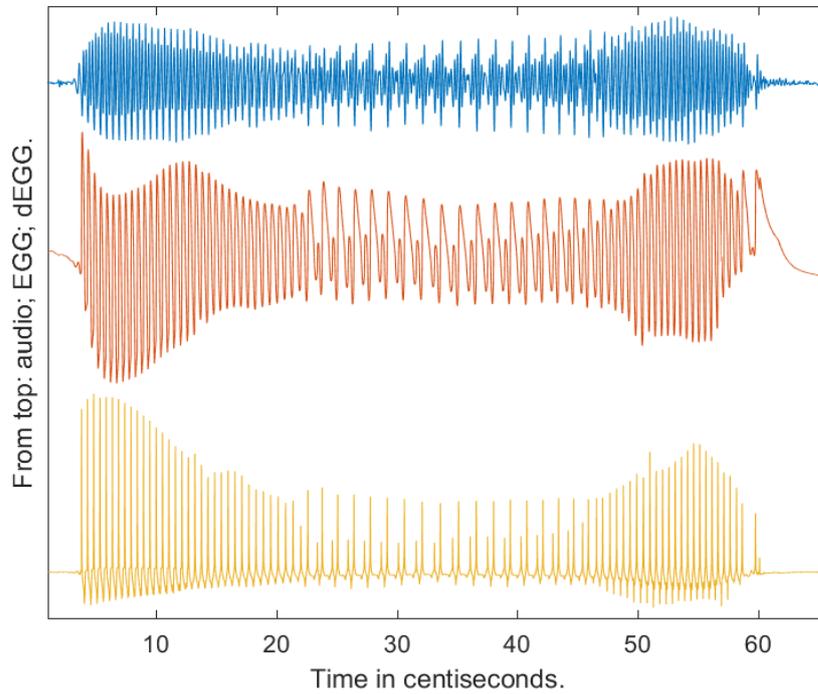
In the literature, the terminology for multiply-pulsed creak as a sub-category within creaky voice is relatively consistent, in keeping with the high degree of specificity of its characteristics. It is sometimes referred to as *diplophonia* or *diplophonic phonation*, as in the studies of Hedelin and Huber (1990), Batliner et al. (1993), and Redi and Shattuck-Hufnagel (2001). It is called *multiply pulsed voice* in the classification put forward by Keating, Garellek, and Kreiman (2015). We employ the term “multiply-pulsed creak” because, in contrast to the previous sub-type, the telltale sign of the present sub-type is the quasi-regular double repetitive pattern with alternating short and long glottal periods.

This pattern is prominently reflected in the electroglottographic signal and its derivative (dEGG), with the length and height of the signal in the creak portion alternately varying from wide to narrow and also from high to low.

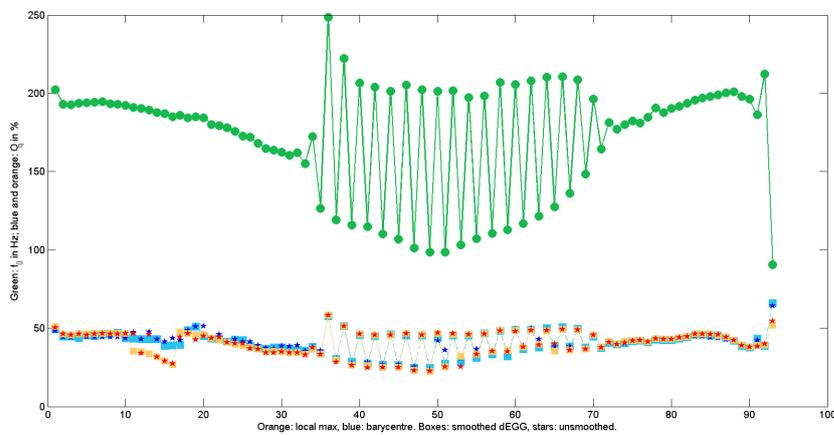
Triple-repetitive patterns, mentioned and illustrated in Blomgren et al. (1998, p. 2655), are not found in the dataset of the present study, which (so far as I am aware) only contains double-repetitive patterns. But since triple-repetitive patterns are attested from the point of view of general phonetics, it is clearly better to use the more general label “multiply pulsed creak”, rather than “double-repetitive pattern” or “double-repetitive creak”, even though the two coincide in the case of the data set at hand.

To take a close look at the example in Figure 5.1, a first, obvious observation is that there are three main parts in this example: non-creak, creak, then back to non-creak again. The double-repetitive pattern covers half of the syllable (specifically: from 34<sup>th</sup> cycle to 71<sup>st</sup> cycle). The complex-repetitive pattern is reflected in a saw-like shape in both  $f_{o \text{ dEGG}}$  and  $O_{q \text{ dEGG}}$  tracings.

A caveat is in order here, concerning the values of  $f_{o \text{ dEGG}}$ : these values, calculated cycle by cycle on the basis of peaks in the derivative of the electroglottographic signal (as explained in §3.3.2), must be taken with a grain of salt. Much lower values would be obtained if the complex-repetitive pattern as a whole were considered as the pattern that repeats itself in time, and hence as the basis for estimating periodicity (and hence fundamental frequency). Provisionally looking at the raw  $f_{o \text{ dEGG}}$  and  $O_{q \text{ dEGG}}$  values as estimated by the `PEAKDET` script nonetheless allows for bringing out patterns that help



(a) Signals. From top to bottom: (i) acoustic signal, (ii) EGG, (iii) dEGG.



(b) Two parameters estimated by **PEAKDET**:  $f_0$  dEGG (green) and  $Q_q$  dEGG (blue and orange).

Figure 5.1: Example of multiply-pulsed creak. Data from speaker F10, token: 551, UID: 6002, over syllable /ku<sup>4</sup>/ in isolation.

DOI: <https://doi.org/10.24397/pangloss-0006784#W224>

detect multiply-pulsed creak, keeping in mind that detection and phonetic/phonological analysis are two different things.

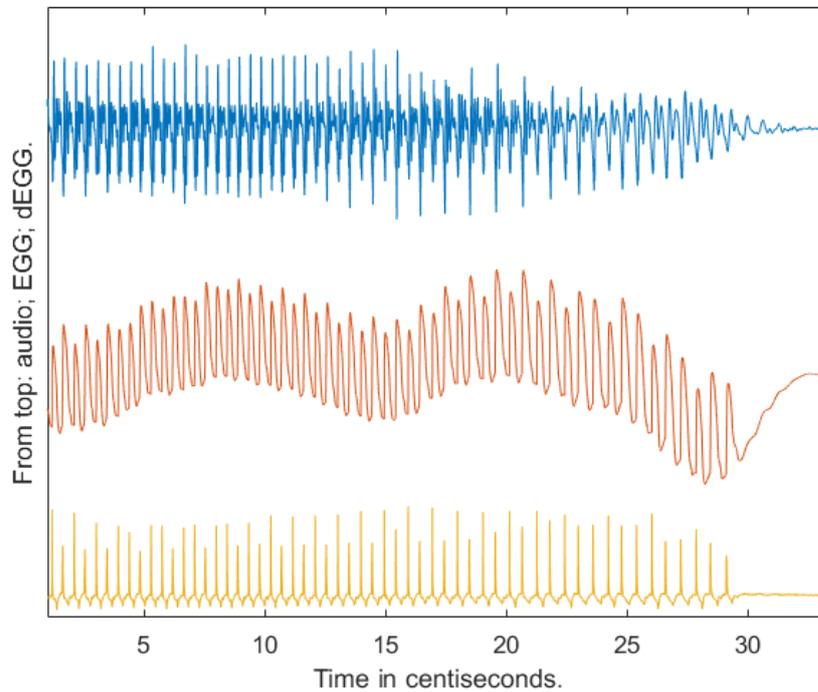
In term of  $f_{o\text{ dEGG}}$  curve, the onset and offset in this example are in the same register, in the vicinity of 200 Hz. The long complex-repetitive pattern in the middle has a range of oscillation between 100 and 220 Hz. The  $O_{q\text{ dEGG}}$  values of this token can be estimated with some precision. The glottal open quotient is in the range between 45% and 35% (i.e. clearly low values) in the initial and final portions (those that do not have multiply-pulsed voice). During the creaky portion,  $O_{q\text{ dEGG}}$  oscillates between markedly different values for the smaller pulses (on the order of 50%) and the larger ones (on the order of 25%: values clearly indicative of glottalization). This alternation in  $O_{q\text{ dEGG}}$  values, with the longer glottal cycles having lower  $O_{q\text{ dEGG}}$  than the shorter ones, offers an interesting insight into the strong differences between the main pulse (which, even taken by itself, clearly looks glottalized) and the secondary pulse (which, by itself, has neither of the two following indicators of glottalization: low  $f_{o\text{ dEGG}}$  and low  $O_{q\text{ dEGG}}$ ).

This sub-type of creak looks like one that will be easy to detect automatically, based on detection of the peaks in the dEGG signal corresponding to glottis-closure instants. Knowing the duration of each glottal pulse is enough to notice the alternation of short and long pulses, characteristic of double- (and triple-) repetitive patterns. The following paragraphs explain how this hypothesis was put to the test.

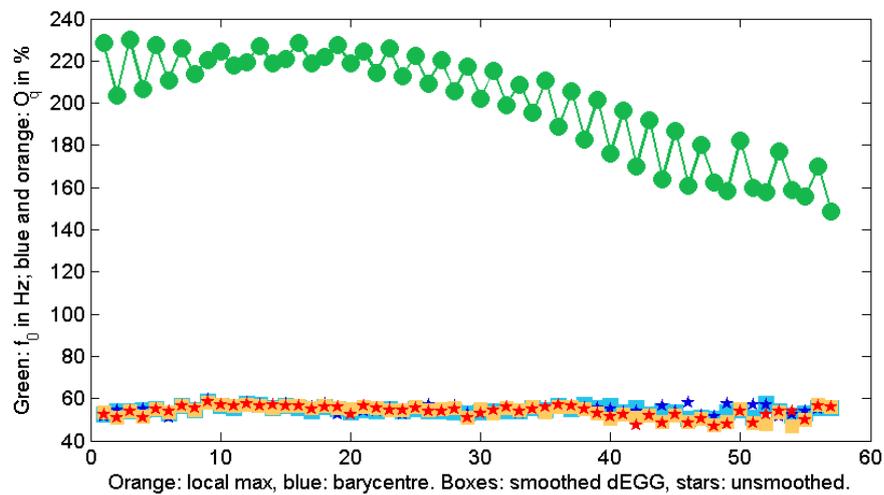
When conducting an automatic search, the two relevant variables are precision and recall rates. Precision is the proportion of correct identifications among the retrieved results. Recall is the proportion of relevant instances that were retrieved by the search. (To clarify with examples: if the search returns all and only the relevant results, precision and recall are both 100%; if the search returns results that are all correct but miss relevant instances, precision is 100% but recall is low; and if the search returns lots of results including all the relevant ones, precision is not so good but recall is at 100%.) In this instance, the aim is to get the multiply-pulsed-creak detector to detect from the 34<sup>th</sup> to the 71<sup>st</sup> cycle.

Although this sub-type is not difficult to detect visually by means of its specific quasi-regular pattern, it is impractical to detect it automatically in the present workflow. The most frequent and obvious limitation of automatic detection is that the larger patterns, containing one large peak and one or more smaller peaks, are not recognized as such by **PEAKDET**. The larger peaks are detected, usually without glitches, but the smaller pulses within the complex-repetitive pattern (hereafter ‘micro-peaks’) are often undetected, depending on the ratio of their closing peak to that of the highest closing peak in the token (used as a reference by **PEAKDET**). As a consequence, the complex-repetitive pattern can be detected in either of two ways: (i) if the micro-peaks go undetected, **PEAKDET** outputs values for one long cycle, as if it were single-pulsed creak; (ii) if at least one of the micro-peaks is detected, **PEAKDET** outputs a sequence of a moderately long cycle followed by a super-short cycle, incorrectly giving the visual impression (on a display of  $f_{o\text{ dEGG}}$  values) of considerable jitter or aperiodicity.

A semi-automatic workflow using **PEAKDET** (as was applied in the present work)



(a) Signals. From top to bottom: (i) acoustic signal, (ii) EGG, (iii) dEGG.



(b) Two parameters automatically calculated by `PEAKDET`:  $f_{0 \text{ dEGG}}$  (green) and  $O_{q \text{ dEGG}}$  (blue and orange).

Figure 5.1: Example of jitter in harshness, illustrating a technical difficulty in automatic detection of multiply-pulsed creak. Data from speaker F12, token: 178, UID: 1832, over syllable /*la*<sup>2</sup>/ in carrier sentence.  
DOI: <https://doi.org/10.24397/pangloss-0006790#W170>

might look like a solution in this case. The user can adjust the threshold for detection of closing peaks, so that the micro-peaks are detected. However, in practice this is not an optimal solution (maybe not even a feasible solution at all). Not only does it take a lot of time and effort to process the items: there are cases where this solution simply does not work technically. The closing peak amplitudes (DECPA values) for micro-peaks can be so low that lowering the threshold to capture them results in the detection of spurious peaks at other points within the glottal cycle: multiple closing peaks can be higher than the micro-peaks. A manual workaround consists in lowering the threshold, catching all micro-peaks along with spurious peaks, then merging the latter one after the other. That method is not only tedious: it also comes along with biases, since the more work is done manually, the more likelihood there is of user bias, artifacts and user errors creeping in. Clearly, `PEAKDET` was not devised as a tool to address such cases, and the way to go here would be to change the computer code, using a different algorithm.

Additionally, a challenge is encountered in terms of precision, as detection based on oscillation in  $f_o$  and  $O_q$  from one pulse to the next will also include phenomena that are not instances of creaky voice. Delta  $f_o$  is technically known as *jitter*: the cycle-to-cycle variation of fundamental frequency. Interestingly, jitter is not only found in multiply-pulsed creak, but also in a phonation type known as harsh voice. Harsh voice is described by Laver (1980, p. 122) as somewhat similar to creak but in higher fundamental frequency. Harshness is distinguished from creakiness in the doctoral dissertation of Michel (1964) (he use the term “vocal fry”). Figure 5.1 is an example of a syllable rhyme containing a jittery passage. The syllable is /laɪ<sup>2</sup>/ (bearing a phonologically non-creak tone) in blue curve. The speaker is F12; her signals frequently exhibits such jitter, in all tones. The signal throughout Figure 5.1 exhibits a slight long-short alternation in glottal cycles that resembles the complex-repetitive pattern of multiply-pulsed creak. However, the differences between successive pulses are much less salient than in multiply-pulsed creak. A saw-like shape can be observed in  $f_{o \text{ dEGG}}$  and  $O_{q \text{ dEGG}}$ , but again, with less salience (differences in  $f_{o \text{ dEGG}}$  below 30 Hz, and almost negligible in  $O_{q \text{ dEGG}}$ ). In addition, the values of these two parameters does not reach down into rock-bottom values as it does in creak. The contour of  $f_{o \text{ dEGG}}$  is within the range from 220 Hz to 160 Hz, and  $O_{q \text{ dEGG}}$  values are located around 55%. This agrees with the description of  $f_{o \text{ dEGG}}$  value in Michel (1964), Michel and Hollien (1968), and Michel (1968).

Thus, distinguishing jitter as exemplified in Figure 5.1 from the telltale complex-repetitive pattern found in multiply-pulsed creak need not be seen as an impossible task (drawing a line in the sand). A threshold in the rate of  $f_{o \text{ dEGG}}$  change has the potential to solve the technical issue, given that this parameter reaches higher values in multiply-pulsed creak.

However, since a threshold in  $f_{o \text{ dEGG}}$  change is also used as a common prerequisite condition for identifying creak (of any sub-type), a further distinction needs to be made in order to tell cases of multiply-pulsed creak apart from cases with less drastic change in  $f_{o \text{ dEGG}}$ : as in harsh voice, as has just been mentioned, and also in pressed voice

(about which more below), which has the smallest change of  $f_{o\_dEGG}$  in comparison with other sub-types of glottalization. Technically, this is a limitation of using a hierarchical tree structure in detection and classification (“if  $\Delta f_o$  is above threshold  $T$ , this is a case of creaky voice, otherwise not; further sub-classification depends on...”). For a well-rounded technical solution, the conditions should be set independently of each other. They are then combined to identify each sub-category: identifying a certain phenomenon, such as multiply-pulsed voice. These categories can be grouped later as appropriate, e.g. defining the higher-level category of *creak* as a case of *either multiply-pulsed voice or single-pulsed creak or pressed voice*. This topic will be taken up again in the section that discusses the detection algorithm (§5.3).

### **Type 3: Aperiodic creak**

This type corresponds to aperiodic voice in the terminology put forward by Keating, Garellek, and Kreiman (2015):

Another variant of  $F_0$  irregularity is when it is taken to the extreme -- vocal fold vibration is so irregular that there is no periodicity and thus no perceived pitch. Like multiply pulsed voice, aperiodic voice lacks the prototypical property of low  $F_0$ ; instead, the property of irregular  $F_0$  is enhanced, and the voice is therefore noisy.

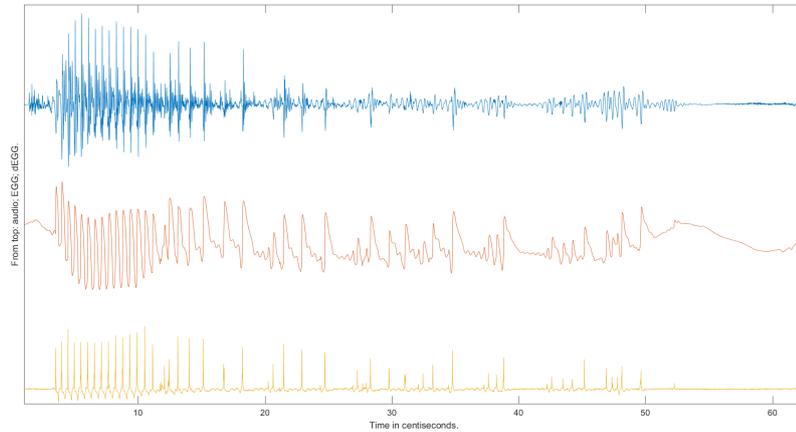
This also seems to match closely the characterization of aperiodicity proposed by Redi and Shattuck-Hufnagel (2001, p. 414): “irregularity in duration of glottal pulses from period to period”.

An example of what we call aperiodic creak is provided in Figure 5.1. It has a smooth start with a steady, linear decline in  $f_{o\_dEGG}$  from first to 15<sup>th</sup> cycle, and then it becomes as aperiodic as can be. The electroglottographic signal is especially welcome here as a complement to the audio, in order to understand what is happening in detail, glottal pulse after glottal pulse. From the electroglottographic signal, it is obvious that shape and duration differ greatly among successive cycles. No pattern stands out clearly: the longer and shorter cycles are arranged in no particular order, as opposed to their neat organization into alternating patterns in multiply-pulsed creak. Such electroglottographic signals constitute a challenge for the `PEAKDET` script (and even for human eyes) to detect precise closing and opening peaks, as glottalization plays out the full gamut of its jarring, aperiodic, chaos-like score. The measurement of glottal open quotient is barely applicable in this case: using the current state of `PEAKDET` (essentially unchanged since first release: version 1.3.2 through 1.4.2 in the Covarep repository,<sup>2</sup> corresponding to version 1.0 in the master repository),<sup>3</sup> cherry-picking any of the  $O_{q\_dEGG}$  values within the portion of aperiodic creak in Figure 5.1 would seem unreasonable. Judging from the shape of the electroglottographic signal, none of the cycles can safely be hypothesized

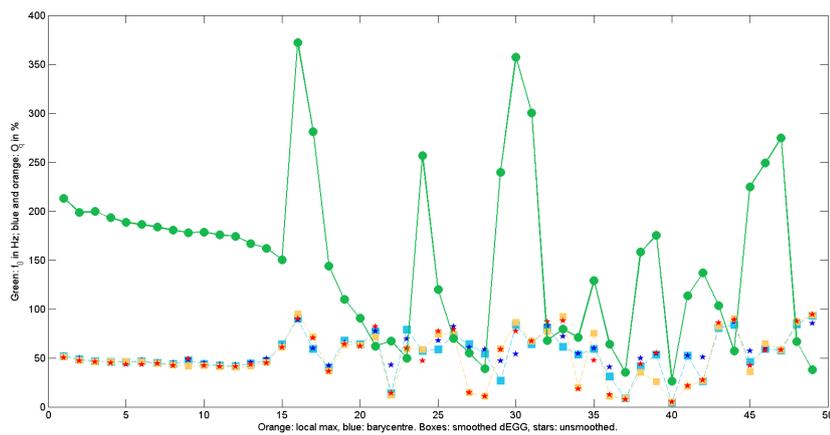
<sup>2</sup><https://github.com/covarep/covarep/releases>

<sup>3</sup><https://github.com/alexis-michaud/egg/releases>

5.2 A characterization and classification of sub-types of creaky voice based on audio and electroglottographic signals



(a) Signals. From top to bottom: (i) acoustic signal, (ii) EGG, (iii) dEGG.



(b) Two parameters automatically calculated by `PEAKDET`:  $f_0$  dEGG (green) and  $O_q$  dEGG (blue and orange).

Figure 5.1: Example of aperiodic creak. Data from speaker F12, token: 541, UID: 5501, over syllable /kaj<sup>4</sup>/ in isolation  
DOI: <https://doi.org/10.24397/pangloss-0006790#W109>

to be divided into a closed phase followed by an open phase – even if one were to loosen greatly the definition of what “closed phase” and “open phase” mean.

Theoretically, detection of this sub-type of creak could be based on the detection of micro-cycles. Following as closely as possible the chaotic patterns of glottal pulses would bring out the irregular, sudden increase or decrease of cycle lengths (still referred to as “ $f_{0\text{ dEGG}}$ ” for short, with the very explicit caveat that what is meant thereby is not really a *frequency*, since it is calculated cycle after cycle, with no condition on repetition in time). By bringing out those cycles, the portion during which aperiodic creak occurs would stand out. However, my budding programming skills are not up to the task of creating an algorithm for detecting irregularity (lack of periodicity). Instead, as a first pass I chose to use the simplest method: by exclusion. That is, when there are sudden and salient changes in  $f_{0\text{ dEGG}}$ , but no complex-repetitive pattern is captured, the item will be classified as belonging to the sub-category of *aperiodic creak*. Of course, other conditions also need to be added to increase the precision. For instance, it seems highly likely that high dispersion of  $O_{q\text{ dEGG}}$  (values of  $O_{q\text{ dEGG}}$  that are scattered all over the place in the raw output produced by [PEAKDET](#)) could in many cases serve as reliable evidence to help to classify this sub-type. Such explorations are left for later work.

#### **Type 4: Pressed voice**

Inclusion of pressed voice inside an inventory of types of glottalization is controversial. Fortunately, there is a handy reference here, which simultaneously lends some legitimacy to the proposal, and formulates it in a clear and cogent way, placing it in an area that is phonetically *adjacent* to creak, and phonologically (cross-linguistically) *within* the space of allophonic variation of phonological creak.

When the glottis is constricted, but the  $f_0$  is neither low nor irregular, a tense or pressed voice quality is heard. While not always considered a form of creaky voice, it can function phonologically as such in languages in which a creaky (or laryngealized) phonation can co-occur with high tone. Here the constricted glottis is criterial. (Keating, Garellek, and Kreiman, 2015)

Considering the simplicity of the signals (audio as well as electroglottography) for pressed voice, this phonation type should in all logic be discussed before multiply-pulsed creak and aperiodic creak: in fact, it could deserve to appear at top of list, as the simplest phonation type – only a modest departure from modal voicing. The reason why it appears last in the inventory proposed here is that the focus is on creaky phonation: from the point of view of general phonetics it would not make great sense to consider pressed voice as a sub-type of creaky voice. The logic followed in the present research, following the proposal by Keating, Garellek, and Kreiman (2015), is thus phonological (tonemic) rather than phonetic: pressed voice in the Muong dialect under investigation is among the frequently occurring phonation types in the phonologically creaky tone

(**Tone 4**). It is therefore listed here as a fourth type of “creaky voice”, in a perspective which admittedly departs from the perspective of general phonetic nomenclatures, but which, I hope, nonetheless makes an indirect contribution to a better understanding of creak as a general phonetic phenomenon, by helping shed light on the dynamics of a phonological system that uses glottalized phonation as part of a lexical tone.

As was done above for single-pulsed creak, two extreme examples are provided in order to show the extent of variation in this sub-type. Phonetically, it seems safe to hypothesize that there is a continuum between these two cherry-picked examples. They are examples of exactly the same target syllable /paj<sup>4</sup>/ spoken in isolation by the same speaker (F21) and in the same recording session, but the first was on the first repetition and the second on the second repetition.

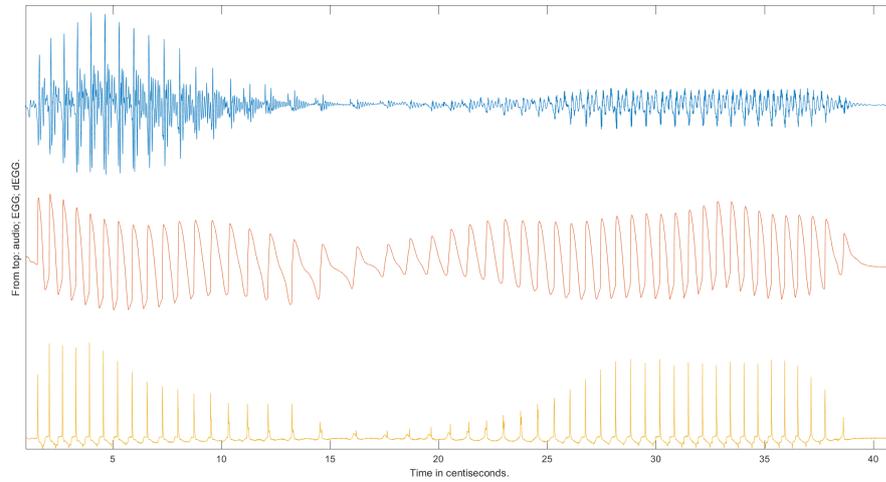
Auditorily, the first example sounds much more clearly pressed than the second. It is reasonable to assume that the consultant tended to pronounce the word in a hyper-articulated way in the first performance, then relaxed somewhat and said it with less hyper-articulation the second time.

The first example in Figure 5.1 can be described as a textbook sample of pressed voice since the glottal tension is clearly reflected in all signals in Figure 5.1a: the three signals all have a bottleneck (decreased amplitude) in the middle of their time course, from the 15<sup>th</sup> to the 20<sup>th</sup> centisecond. The amplitude of the audio signal is very notably reduced, which constitutes evidence that the vocal folds are strongly adducted, thus impeding the passage of airflow for a brief span during this strong adduction, before allowing a modest return in the second half. The return in the final portion is almost as great as that at the beginning in both the EGG and dEGG signals, but is much smaller in the acoustic signal. The strong constriction caused a challenge for **PEAKDET** to detect the closing and opening instants on the dEGG signal, since the amplitude of the corresponding peaks in the derivative of the EGG signal, which serves as a basis for detection, is so much lower (see bottom of Figure 5.1a). In this example, **PEAKDET** missed two cycles: it erroneously detects three cycles as a single cycle, the 18<sup>th</sup>, as can be seen in Figure 5.1c. The corresponding spurious  $f_{o\text{ dEGG}}$  value detected is at 20 Hz (as shown in Figure 5.1b). This problem is solved by visually locating a more accurate threshold and setting the threshold manually in **PEAKDET**, so that the algorithm detects the small, undetected peaks within the spurious 18<sup>th</sup> cycle. By modifying this threshold, **PEAKDET** will recalculate both parameters ( $f_{o\text{ dEGG}}$  and  $O_{q\text{ dEGG}}$ ), as shown in Figure 5.1d.<sup>4</sup> The  $f_{o\text{ dEGG}}$  curve now has less of a dagger shape at its lowest point, although it retains a remarkable, pointed trough around a lowest value of 60 Hz.

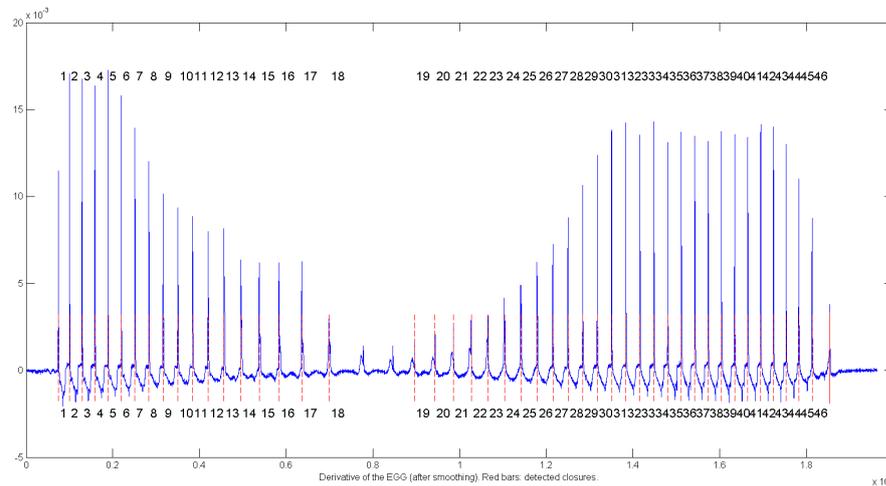
There was an interesting discussion about this example between me and my supervisor.

---

<sup>4</sup>This user modification, carried out at the stage of semi-automatic processing the electroglottographic signals to extract  $f_{o\text{ dEGG}}$  and  $O_{q\text{ dEGG}}$  for all tokens, will then be stored in the final results matrix for the recording at issue. Therefore, when carrying out automatic creak detection down the line of the research process, the values taken into account are those in Figure 5.1d and not those in Figure 5.1c. In case the measurements were re-run using **PEAKDET** in full automatic mode, without user verification, the values obtained would be those in Figure 5.1c. These facts go without saying, as they result straightforwardly from the description of the workflow provided in section 3.3.3, but it seemed useful to state so nonetheless for the sake of explicitness and clarity.



(a) Signals. From top to bottom: (i) acoustic signal, (ii) EGG, (iii) dEGG.



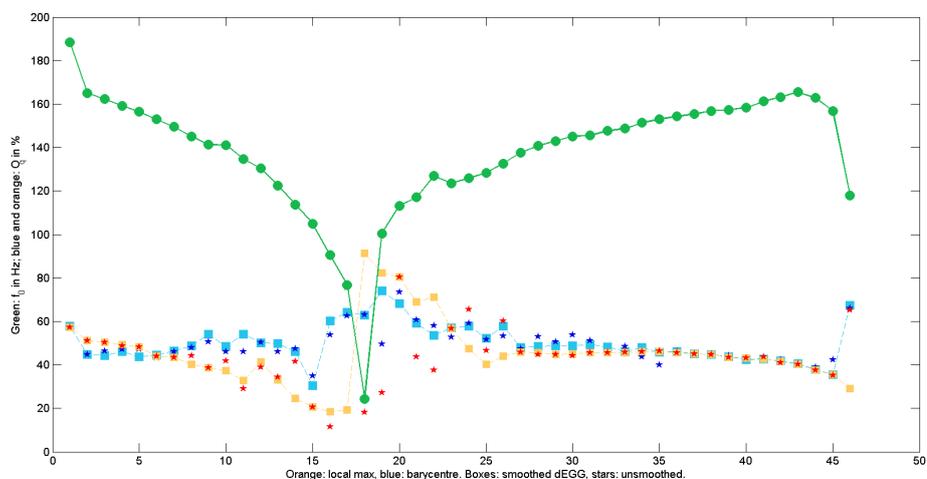
(b) Two closing peaks are not detected at the 18<sup>th</sup> cycle by the automatic `PEAKDET`.

Figure 5.1: Example of a maximum pressed voice. Data from speaker F21, token: 41, UID: 0501, over syllable /paj<sup>4</sup>/ in isolation.

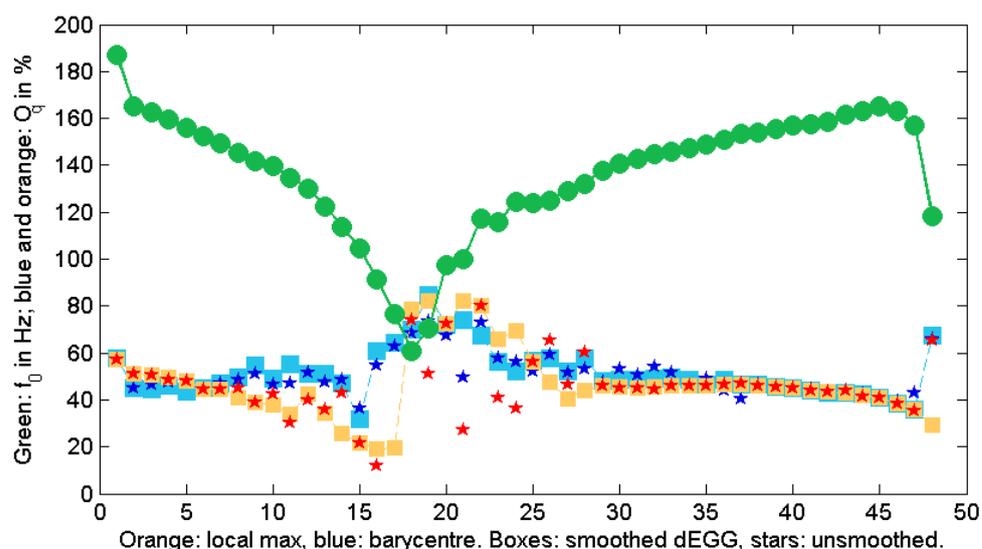
DOI: <https://doi.org/10.24397/pangloss-0006812#w9>

He thought that he would tend to classify this example as a single-pulse creak rather than a presser voice as my opinion because the glottal cycles are lengthy and irregular during the glottalized portion. This is indeed true and appropriate from a theoretical perspective based on signal observation.

However, from a perceptual standpoint, although I have not done a perceptual test so



(c) Two parameters automatically calculated by `PEAKDET`:  $f_{0 \text{ dEGG}}$  (green) and  $O_{q \text{ dEGG}}$  (blue and orange).



(d) Two parameters after resetting the threshold at 0.001 to catch two undetected closing peaks.

Figure 5.1: Example of a maximum pressed voice. Data from speaker F21, token: 41, UID: 0501, over syllable /paj<sup>4</sup>/ in isolation.

DOI: <https://doi.org/10.24397/pangloss-0006812#W9> (cont.).

far but rely on my ear with the experience I have had in this dialect, I could not perceive the creak in this example (DOI: <https://doi.org/10.24397/pangloss-0006812#W9>). It's far from what I can perceive in two examples of single-pulsed creak as provided in example 5.1 (DOI: <https://doi.org/10.24397/pangloss-0006782#W9>) and example

5.1 (DOI: <https://doi.org/10.24397/pangloss-0006782#W10>).

In addition, I think the medial bottleneck in all the signals in Figure 5.1a must reflect something related to a sudden, brief and severe constriction of the glottis, which is not encountered in the other sub-types classified here.

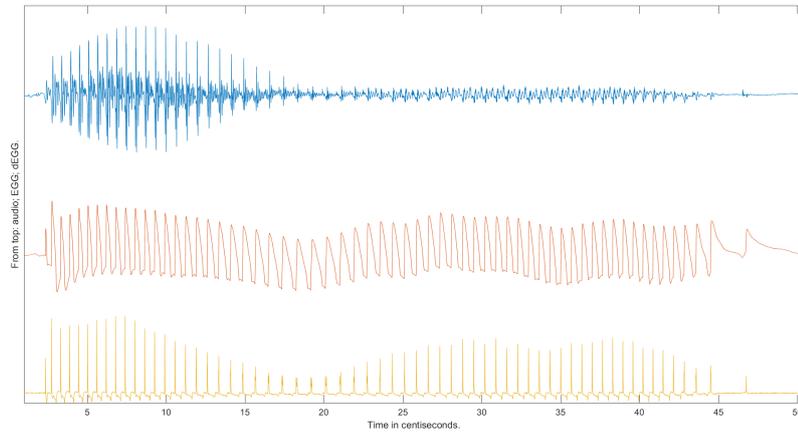
In fact, several possibilities for classification are open here: this example compounds characteristics from single-pulsed creak and pressed voice. Not all cases fit into sharply delineated categories like single-pulsed creak and multiply-pulsed creak. There is a continuum in at least three criteria, which, together, make for a complex space (rather than one linear continuum): (i) the degree of glottal constriction, (ii) the regularity of damped pulses, and (iii) the duration of the constriction. Some cases like this one are best described in terms of a set of parameters, rather than just in terms of belonging to one category.

This explains my decision to have a place for pressed voice in the sub-classification of creak. Although perceptually different, pressed voice is closer to glottal constriction and even glottal stop than to creaky voice, but in its essence of glottis setting which reflected in acoustic and electroglottographic signal, they are closely related and difficult to separate in some cases. In other words, it could be said that pressed voice *borders on creak*, in the sense that they share certain phonetic characteristics and pressed voice can function phonologically in which a creaky phonation can co-occur. This is indeed the case in Kim Thuong Muong when the creak voice and the pressed voice are considered as two allotonic variations occur in **Tone 4**.

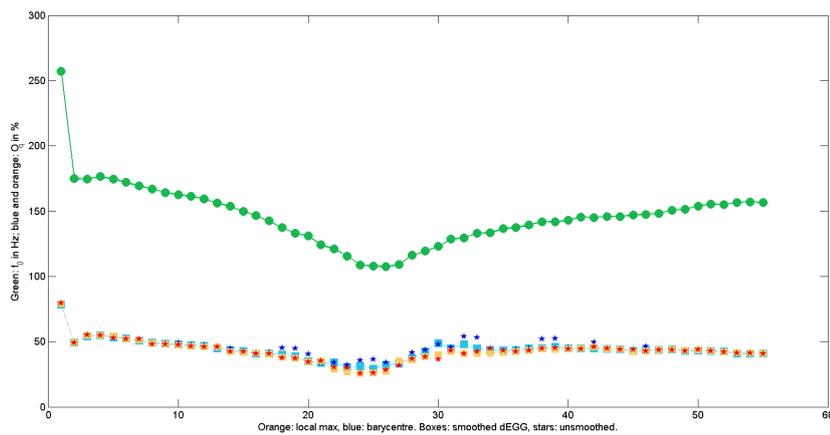
The second example of pressed voice is shown in Figure 5.2. It constitutes a case where the strength of vocal fold compression looks much less than in the previous example. If the first example is a maximal case of pressed voice, this second example is in contrast a minimal case. The evidence of pressed voice can be seen most clearly in the acoustic signal with a maximum reduce early after a short start in the state of modal voice. It is maintained until the end. The dEGG signal also presents a bottleneck corresponding to the most constricted part. The amplitude of EGG signal on the other hand does reflect any obvious change but some longer cycles can be noticed. The calculated fundamental frequency does not drop to very low values as in the ideal example. The bottom of the shallow valley in the  $f_{0 \text{ dEGG}}$  curve is above 100 Hz. The parameter  $O_{q \text{ dEGG}}$  is measurable around 50%. These parameters do not reveal a glottalization phenomenon. In other words, they do not differ from those parameters normally measured in the modal voice. The clearest evidence to argue that this is an item of pressed voice is the damping of the acoustic signal.

We have also provided two other examples of pressed voices in a Github repository ([here](#)) that offers a gallery of examples of glottalization. The repo, which was elaborated in parallel with the work on the present thesis, offers a brief summary of materials set out in this section: it shows electroglottographic and acoustic signals with a discussion of various phenomena of glottalization. All the examples are tokens of Kim Thuong Muong **Tone 4**. The reason why I chose to present two different examples for pressed voice here, instead of choosing the two examples from the online gallery (reproduced here as Figures 5.3 and 5.4), is that I find them more consensual: phoneticians are

5.2 A characterization and classification of sub-types of creaky voice based on audio and electroglottographic signals



(a) Signals. From top to bottom: (i) acoustic signal, (ii) EGG, (iii) dEGG.



(b) Two parameters automatically calculated by `PEAKDET`:  $f_{0 \text{ dEGG}}$  (green) and  $O_{q \text{ dEGG}}$  (blue and orange).

Figure 5.2: Example of a minimally pressed voice. Data from speaker F21, token: 42, UID: 0502, over syllable /paj<sup>4</sup>/ in isolation.  
DOI: <https://doi.org/10.24397/pangloss-0006812#W141>

likely to agree more readily to their description as instances of pressed voice, than for the signals in Figures 5.3 and 5.4 (reproduced from the GitHub repository). The latter signals could easily be categorized using labels other than *pressed voice*.

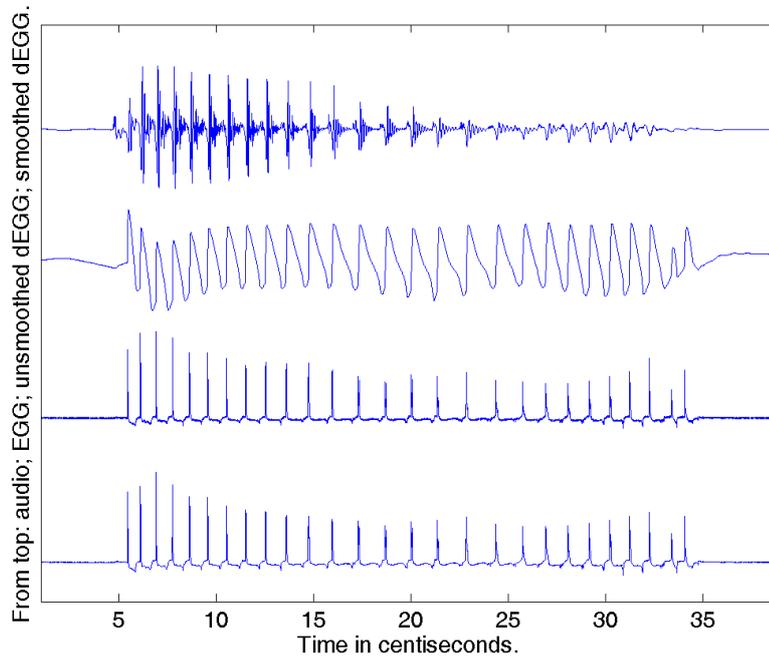
Admittedly, the same point could be made about the token shown in Figure 5.1: authors who prefer to use the label “pressed voice” for a lasting voice quality setting may describe the glottal event as an instance of glottal constriction, rather than pressed voice. Clearly, there is no hard-and-fast divide between pressed voice and single-pulsed creak, and terminological choices will continue to vary. An advantage I see in the examples in Figure 5.1 and Figure 5.2 is that they clarify the range along the glottal stricture continuum that is referred to in the present work as *pressed voice*. These two examples represent the two poles of a continuum from maximum to minimally pressed voice. Needless to say, it needs to be kept in mind that signal shapes are not the full story of phonation types, and some less clear-cut cases are encountered: this topic will be taken up again later when discussing the topic of boundaries (junctures in the utterance) and their ties to phonation types.

As mentioned above regarding the continuum of single-pulsed creak, the relative similarities in the shape of the electroglottographic signal create difficulties in attempts to distinguish pressed voice from single-pulsed creak. Both appear among phonetic realizations of the glottalized tone; for this phonological reason, as also for phonetic reasons mentioned above, it appears reasonable to add pressed voice to a vast continuum that extends from strongly damped pressed voice (close to single-pulsed creak), to light pressed voice which, in turn, is close to modal voice. Therefore, we intentionally classify pressed voice as a sub-type of creak, although some pressed voice items are more close to modal voice than to typical creak (single-pulsed or multiply-pulsed). It could be more accurate to describe pressed voice as one of several (phonetic) variants in the realization of (phonological) creak – and we hereby repeat our adhesion to the proposal by Keating et al. mentioned at the outset of this section.

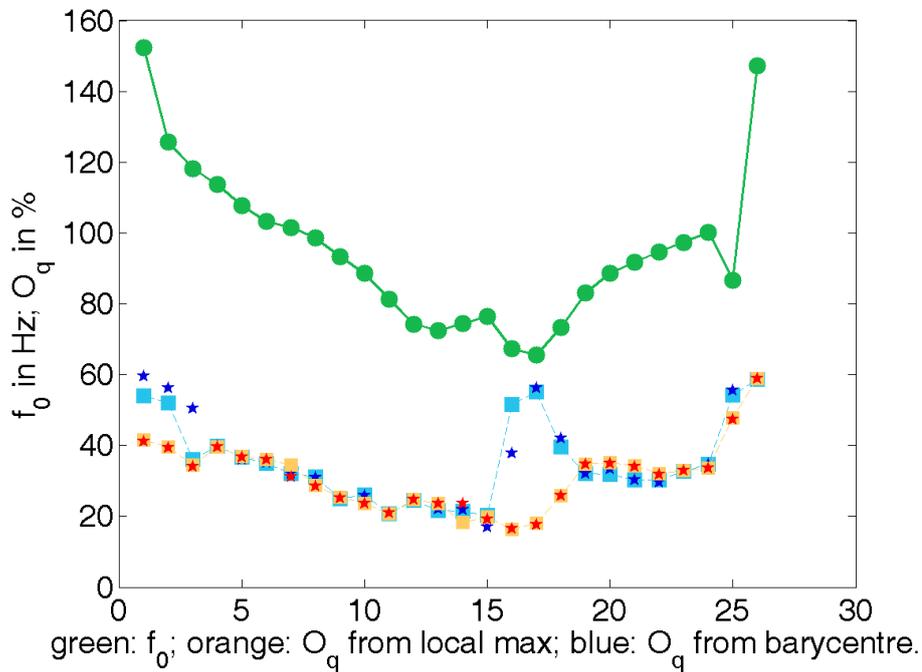
### 5.3 Detection of creaky voice from the electroglottographic signal

The above developments about nomenclature relative to creaky voice now make it possible to turn to the issue of the detection of creaky voice from the electroglottographic signal, to take a quantitative view of creak phenomena in the data set under investigation. The  $f_0$  dEGG tracings in Figure 4.1 only provided an indirect view into creak, as a perturbatory phenomenon detectable through jagged (jittery) shape of averaged curves for **Tone 4** as opposed to the four others. It appears well worth teasing out the creaky portion from the rest, instead of damping the creaky signal through across-the-board averaging with non-creaky portions. The classification provided above into four basic sub-types is used here as a reference.

5.3 Detection of creaky voice from the electroglottographic signal

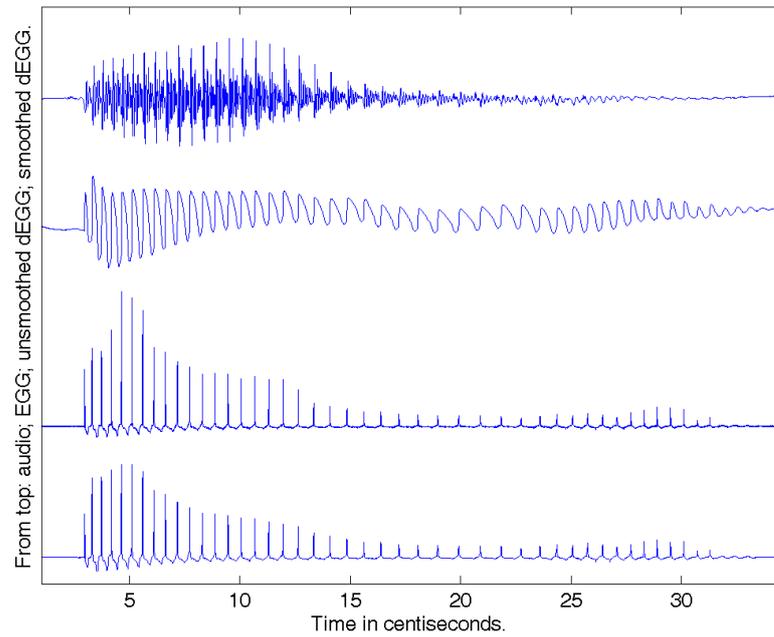


(a) Signals. From top to bottom: (i) acoustic signal, (ii) EGG, (iii) dEGG, (iv) smoothed dEGG.

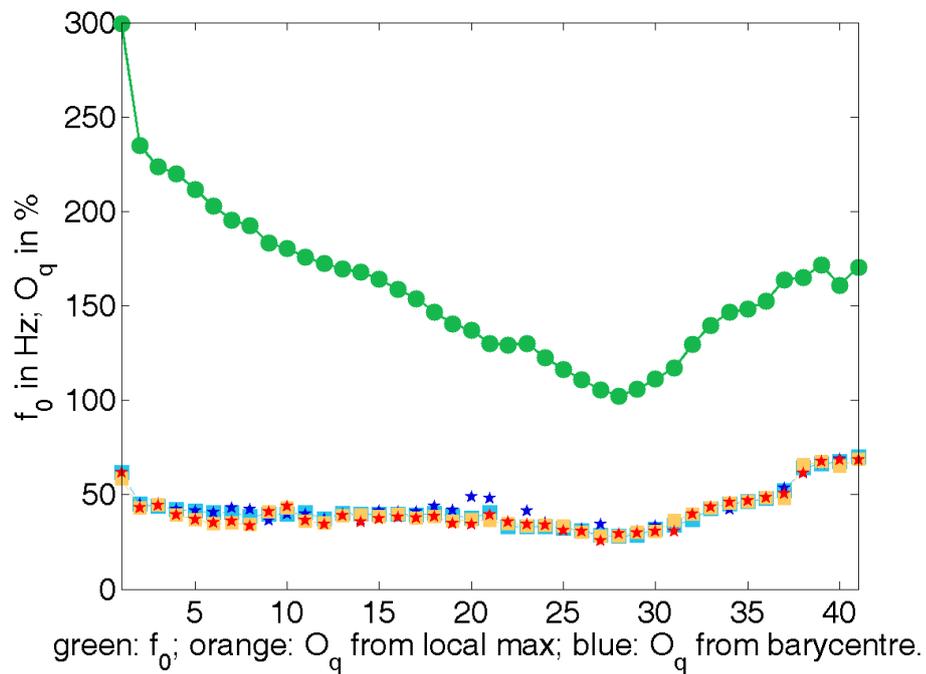


(b) Two parameters automatically calculated by `PEAKDET`:  $f_{0 \text{ dEGG}}$  (green) and  $O_{q \text{ dEGG}}$  (blue and orange).

Figure 5.3: First example of pressed voice reproduced from the Github gallery of glottalized signals.



(a) Signals. From top to bottom: (i) acoustic signal, (ii) EGG, (iii) dEGG, (iv) smoothed dEGG.



(b) Two parameters automatically calculated by `PEAKDET`:  $f_{0 \text{ dEGG}}$  (green) and  $O_{q \text{ dEGG}}$  (blue and orange).

Figure 5.4: Second example of pressed voice reproduced from the Github gallery of glottalized signals.

### 5.3.1 *An easy problem, but paradoxically without an off-the-shelf solution*

Creak detection from the electroglottographic signal is, in an important sense, an easy problem, yet there is no easy and off-the-shelf solution. The electroglottographic signal appears especially suitable for the detection of creaky voice (Gao, 2015), since creaky voice involves a lot of rapid changes in vocal fold contact area, and moreover there is a large amount of vocal fold contact in creaky voice. The electroglottographic signal is less directly interpretable in the case of breathy voice, which, by its very nature, involves less adduction of the vocal folds, and – in typical cases – incomplete contact between the vocal folds even at the point in time where the vocal fold contact area is greatest. Creaky voice is first and foremost a matter of patterns of vocal fold contact, whereas whispery/breathy voice has stronger ties to airflow and aerodynamics. Visually, creaky voice is often not too difficult to make out in audio signals, and it is even much clearer in electroglottographic signals, as exemplified in Figures 5.1, 5.1 and 5.1.

One would therefore expect that software packages for the detection of creaky voice from the electroglottographic signal would be available from one of the world's research centres in phonetics. But while there are available toolkits for the detection of creak from the audio signal (Drugman, Kane, and Gobl, 2014; Dallaston and Docherty, 2019), at the time when the experiments reported here were conducted, there was, to the best of my knowledge, no available script for automatic detection of creak from electroglottographic signals. This may well be due to the fact that audio signals are much more abundant than electroglottographic signals, and industry applications requiring phonation-type identification deal with an audio input, without the luxury of having an associated electroglottographic signal.

There were several ways to go here: one was statistical modelling (machine learning), and another was 'deterministic' modeling: classical programming by rule. The latter may seem outdated in the age of Artificial Intelligence. However, it has advantages in terms of explicitness, which are suitable for linguistic exploration. (Not to mention the practical fact that it was clearly more within my reach to write computer code in a few lines over which I had control than to put together machine-learning software 'bricks' that are of great internal complexity.) This is not to say that I do not have an eye on the exciting prospects opened up by end-to-end toolkits in Natural Language Processing (as exemplified by Adams, Galliot, et al. 2021 for automatic speech transcription applied to language documentation): these prospects are considered as a central topic of further studies in the wake of the present dissertation.

### 5.3.2 *Step 1: list of specifications*

The first step when writing a computer script is obviously to outline specifications. In this case, here is what we aim at:

1. Detecting which tokens have the presence of creak (phonetically)
2. Distinguishing phonetic sub-types within creak (specifically: single-pulsed voice, multiply-pulsed voice, aperiodic creak, and pressed voice, as distinguished above)

3. Providing detailed temporal information on creak: its duration, and its beginning and end relative to the beginning and end of the rhyme (allowing for the calculation of a range of indicators such as the ratio of creaky portion to duration of the entire rhyme)

The output can then be used for analyzing linguistic patterns of creaky voice in the entire data sets.

### 5.3.3 Step 2: writing and testing the Creakdet script

The algorithm in Figure 5.5 presents the process for detecting and classifying the different sub-types of creak described above.

In order to facilitate the understanding of the algorithm presented here, we will elaborate in turn on: (i) the input to CreakDet, (ii) the three initial conditions for detecting types of creak, and (iii) the output as viewed through simple descriptive statistics.

**The input to CreakDet:** The input material is the three-dimensional matrix yielded by application of **PEAKDET** to electroglottographic recordings, as described in Section 3.3.2. Out of the wealth of information in the results matrix, the first-pass version of the creak detector only uses  $f_{0 \text{ dEGG}}$ , DECPA and time codes (beginning and end of each cycle). In future work, the information of  $O_{q \text{ dEGG}}$  is expected to contribute much to refining the detector.  $O_{q \text{ dEGG}}$  is very revealing about phonation types – that is, when it can be reliably estimated. As pointed out in section data processing (3.3.2, there are cases when  $O_{q \text{ dEGG}}$  can by no means be reliably estimated, either because the signal is too noisy or because the glottal cycle is not neatly divided into a closed phase followed by an open phase. Use of  $O_{q \text{ dEGG}}$  is therefore avoided for detection of those configurations that can safely be detected on the basis of  $f_{0 \text{ dEGG}}$  alone. On the other hand,  $O_{q \text{ dEGG}}$  is central to the detection of pressed voice, for obvious reasons (recall from Section 5.2 that  $O_{q \text{ dEGG}}$  relates straightforwardly to the degree of glottal fold abduction).

Thus, in the first version of Creakdet (short for ‘creak detection’), the parameter  $f_{0 \text{ dEGG}}$  is primarily employed. The first and second conditions are defined based on the rate of  $f_{0 \text{ dEGG}}$  change (or also known as the rate of delta  $f_0$ ) which is calculated by the function in the listing 5.1 named ‘fochange’. The output of this function can be seen to include three variables:

1. **delta  $f_0$**  in Hz (MATLAB script in the listing 5.1 from 2<sup>nd</sup> line to 5<sup>th</sup> line): it is simply a calculation of the difference between the  $f_{0 \text{ dEGG}}$  values of successive glottal cycles. The  $f_{0 \text{ dEGG}}$  value of the current cycle is subtracted from that of the next cycle. The result of this variable can be positive or negative.
2. **rate\_delta  $f_0$**  (in %): after measuring the delta  $f_0$  as a basis for further measurements, we move on to measure the rate of this change by dividing it by the  $f_{0 \text{ dEGG}}$  of the current cycle, and multiplying it by 100 to convert the result into a percentage.
3. **smoo\_delta  $f_0$**  (in %): the name of this variable is kept short for convenience, and does not tell the full story. Two operations are performed here. First, in order

5.3 Detection of creaky voice from the electroglottographic signal

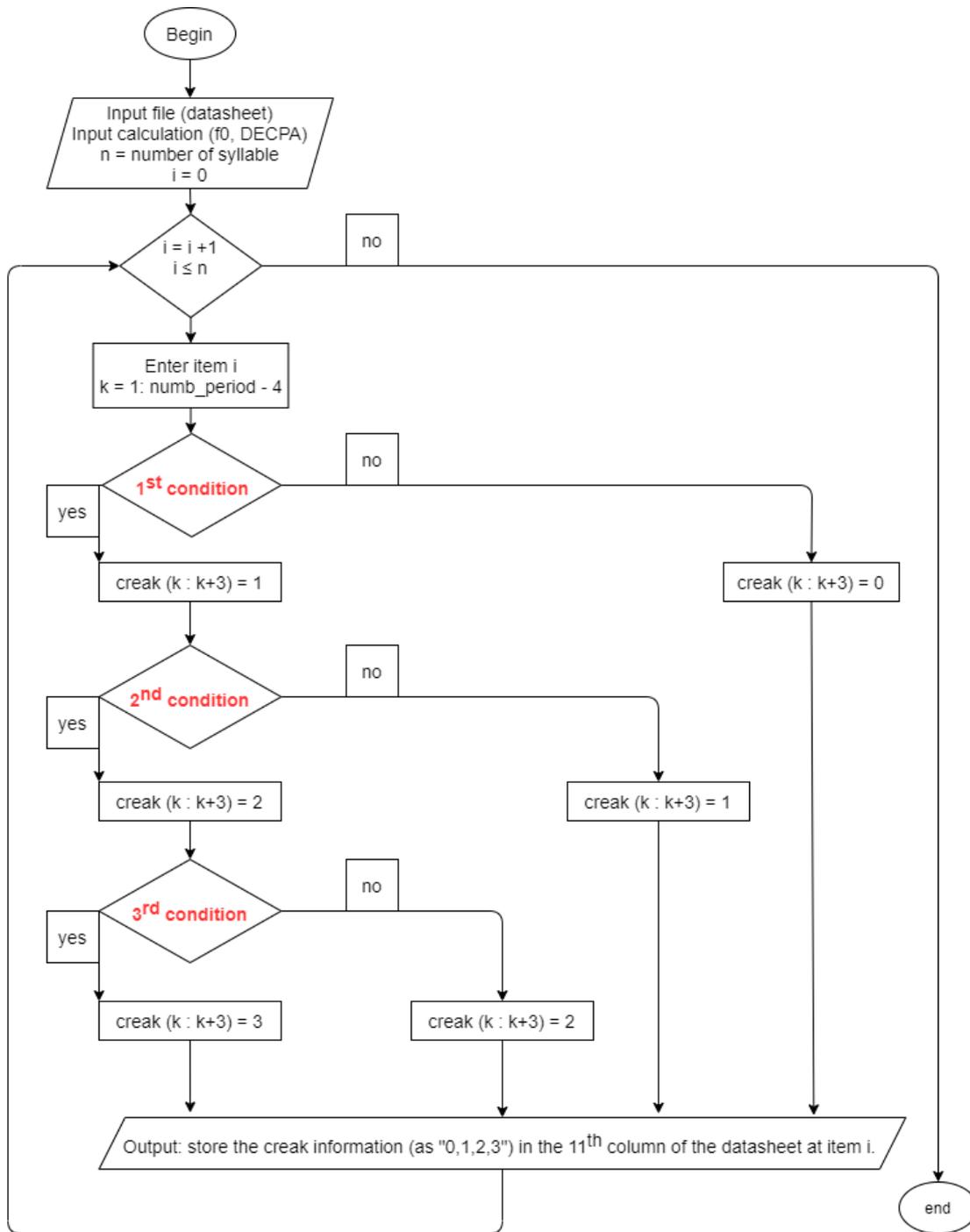


Figure 5.5: Algorithm of CreakDet version 1

to smooth out the rate of  $f_{0 \text{ dEGG}}$  change (to avoid hypersensitivity to local jitter incidents), we average the delta  $f_0$  rate over four successive glottal cycles. Second, in order to facilitate comparison, we have absolutized these values, making them all positive.

Listing 5.1: function ‘fochange’

```

1 function [deltafo , rate_deltafo , smoo_deltafo] = fochange (fo)
2 % Calculating delta fo
3 for i = 1:length(fo) - 1
4     % Calculate deltafo
5     deltafo(i) = fo(i+1) - fo(i);
6     % The rate of deltafo
7     rate_deltafo(i) =(deltafo(i)/fo(i)) * 100;
8 end
9 % Smoothing the delta fo
10 for k = 1:length(rate_deltafo) - 3 %Window of 4 cycles
11     smoo_deltafo(k) = mean(abs(rate_deltafo(k:k+3)));
12 end

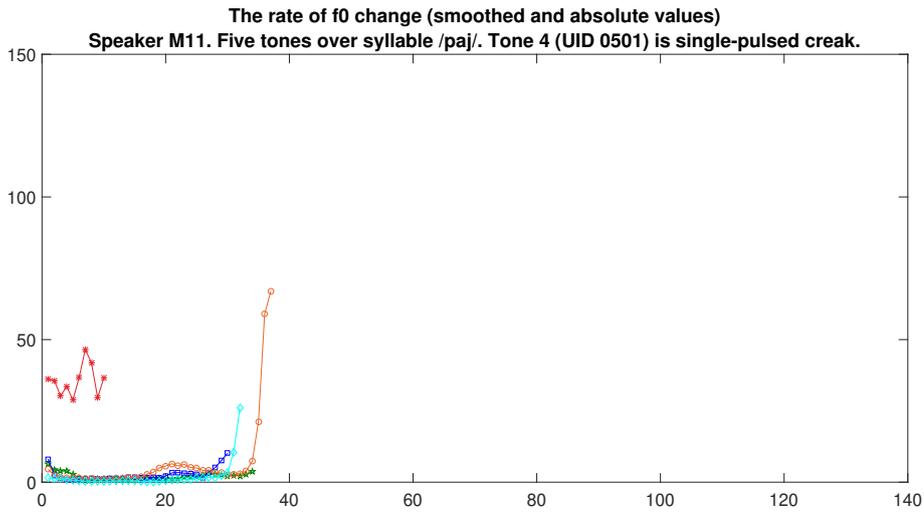
```

In addition to  $f_{0 \text{ dEGG}}$ , which provides information about the duration of glottal cycles (in visual terms: the width of cycles along the x axis of displays of the audio and electroglottographic signals), the output of the `PEAKDET` script also includes another parameter: “Derivative- Electroglottographic Closure Peak Amplitude” (DECPA for short), which provides information about their height (i.e. the y axis). DECPA is the amplitude of the peak on the derivative of the electroglottographic signal at glottal closure (Michaud, 2004a). DECPA is also referred to in the literature as PIC, for Peak Increase in Contact (Keating, Esposito, et al., 2010). By using the same ‘fochange’ function but changing the input variable to DECPA, we obtained similar information to  $f_{0 \text{ dEGG}}$  but for DECPA: the three variables become deltaDECPA, rate\_deltaDECPA, and smoo\_deltaDECPA. The combination of  $f_{0 \text{ dEGG}}$  and DECPA allows for a detection of complex-repetitive patterns (in multiply-pulsed creak) that is based on two dimensions of the deviated electroglottographic signal. In a nutshell: the input data for `CreakDet` includes the parameters calculated and stored by `PEAKDET` (in particular the times,  $f_{0 \text{ dEGG}}$ ,  $O_{q \text{ dEGG}}$ , and DECPA) as well as some parameters derived from these: calculations over delta  $f_0$  and delta DECPA.

**The setup of the 1<sup>st</sup> and 2<sup>nd</sup> conditions on the rate of  $f_{0 \text{ dEGG}}$  change.** The rate of  $f_{0 \text{ dEGG}}$  change is studied through the variable `smoo_delta fo`. The values of this variable in the **Tone 4** examples provided as prototypes (typical examples) in Section 5.2 are compared with those found on other target syllables carrying other tones (in the same minimal sets) to set initial thresholds for detecting creaky voice. The Figure 5.6 is provided for this purpose. The `smoo_delta fo` values of each tone are represented using the color assigned to the tone as mentioned in Appendix C.

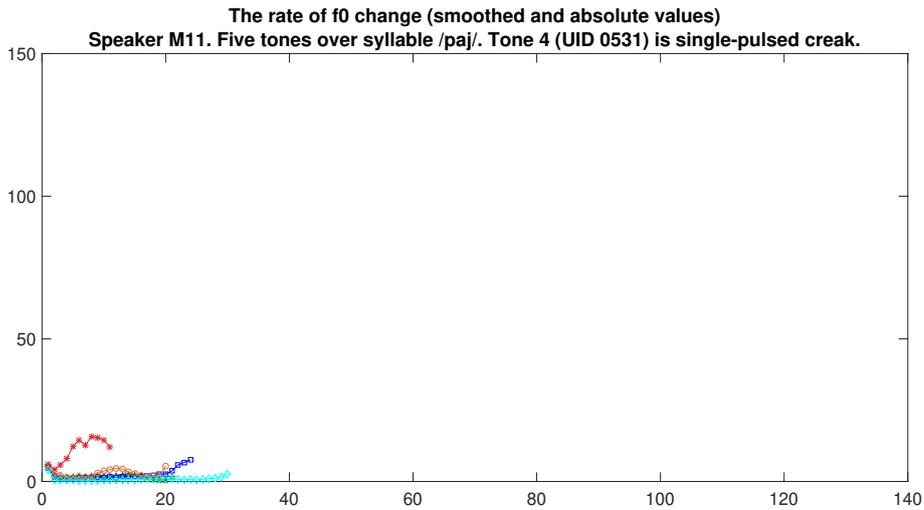
Since our focus is on creaky voice, a consistent characteristic of **Tone 4** and only that tone (whereas the other tones are all modally voiced), the experimental approach consists in paying attention to differences between **Tone 4** and the other tones. However, to avoid reasoning within a narrowly dialect-specific perspective (and thus,

5.3 Detection of creaky voice from the electroglottographic signal



(a) Minimal set containing the ‘ideal’ sample of single-pulsed creak in Fig. 5.1.

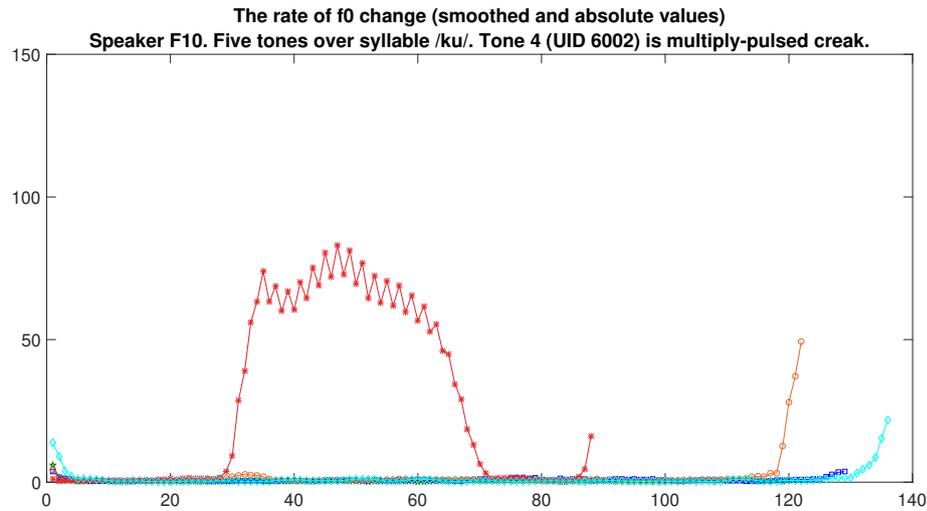
- DOI Tone **1**: <https://doi.org/10.24397/pangloss-0006782#W7>
- DOI Tone **2**: <https://doi.org/10.24397/pangloss-0006782#W5>
- DOI Tone **3**: <https://doi.org/10.24397/pangloss-0006782#W3>
- DOI Tone **4**: <https://doi.org/10.24397/pangloss-0006782#W9>
- DOI Tone **5**: <https://doi.org/10.24397/pangloss-0006782#W1>



(b) Minimal set containing the ‘normal’ sample of single-pulsed creak in Fig. 5.1.

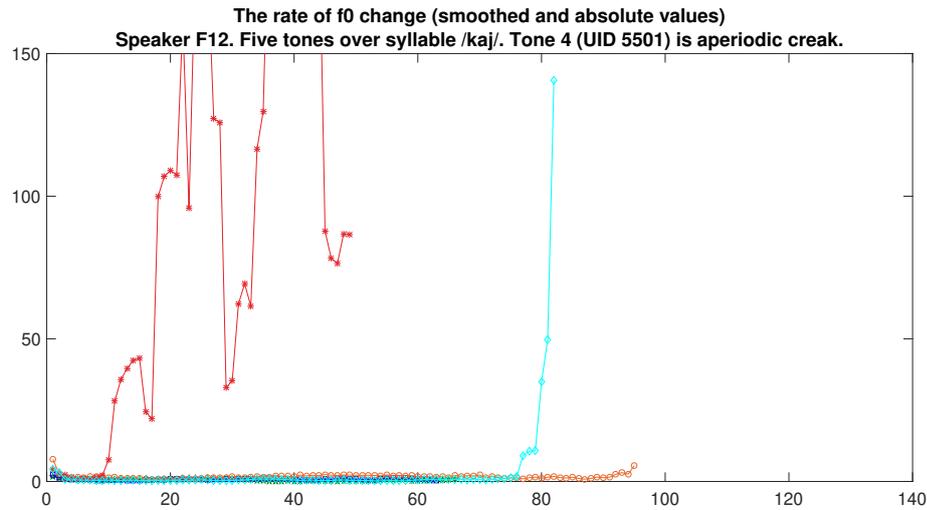
- DOI Tone **1**: <https://doi.org/10.24397/pangloss-0006782#W8>
- DOI Tone **2**: <https://doi.org/10.24397/pangloss-0006782#W6>
- DOI Tone **3**: <https://doi.org/10.24397/pangloss-0006782#W4>
- DOI Tone **4**: <https://doi.org/10.24397/pangloss-0006782#W10>
- DOI Tone **5**: <https://doi.org/10.24397/pangloss-0006782#W2>

Figure 5.6: Comparison of the  $\text{smoo\_delta}f_0$   $_{\text{dEGG}}$ : an illustration for the consideration of the threshold established in condition 1 and 2.



(c) Minimal set containing the sample of multiply-pulsed creak (in Fig. 5.1).

- DOI Tone 1: <https://doi.org/10.24397/pangloss-0006784#W222>
- DOI Tone 2: <https://doi.org/10.24397/pangloss-0006784#W220>
- DOI Tone 3: <https://doi.org/10.24397/pangloss-0006784#W218>
- DOI Tone 4: <https://doi.org/10.24397/pangloss-0006784#W224>
- DOI Tone 5: <https://doi.org/10.24397/pangloss-0006784#W216>

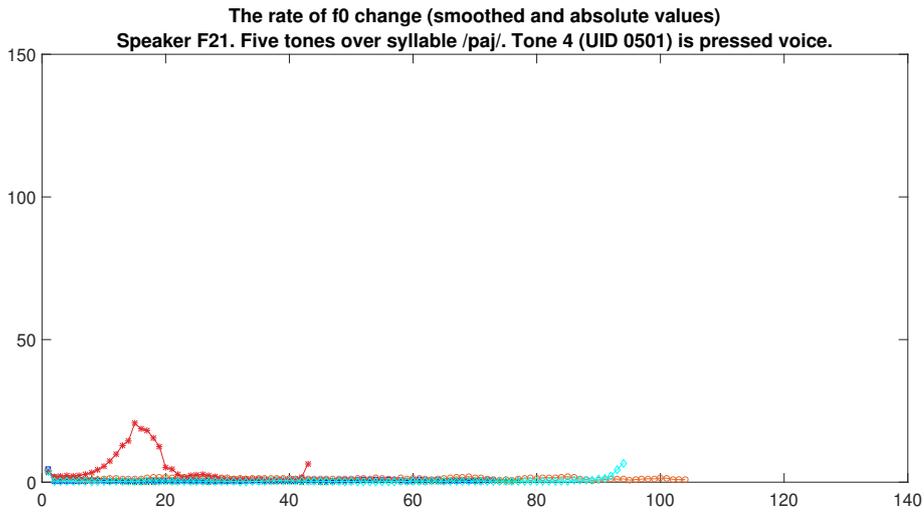


(d) Minimal set containing the sample of aperiodic creak (in Fig. 5.1).

- DOI Tone 1: <https://doi.org/10.24397/pangloss-0006790#W107>
- DOI Tone 2: <https://doi.org/10.24397/pangloss-0006790#W105>
- DOI Tone 3: <https://doi.org/10.24397/pangloss-0006790#W103>
- DOI Tone 4: <https://doi.org/10.24397/pangloss-0006790#W109>
- DOI Tone 5: <https://doi.org/10.24397/pangloss-0006790#W101>

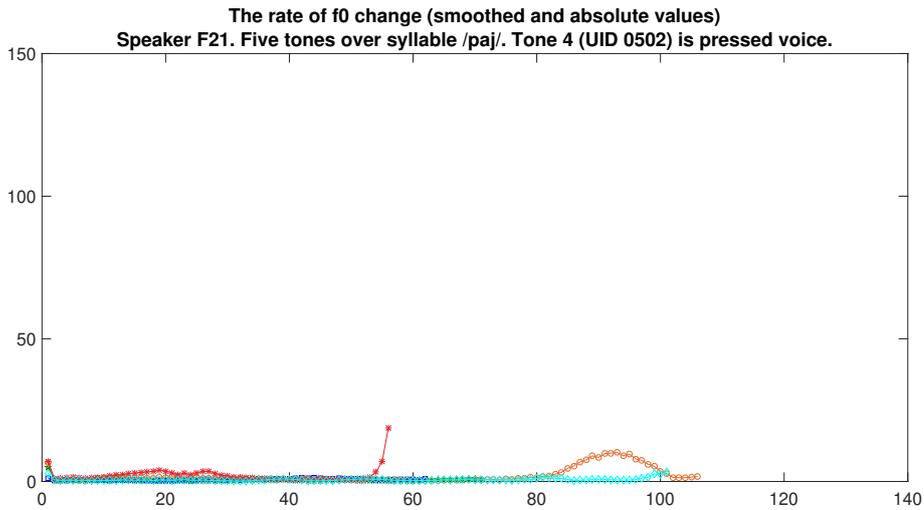
Figure 5.6: Comparison of the  $\text{smoo\_delta}f_0$   $_{\text{dEGG}}$ : an illustration for the consideration of the threshold established in condition 1 and 2. (cont. 1)

5.3 Detection of creaky voice from the electroglottographic signal



(e) Minimal set containing the sample of maximum case of pressed voice in Fig. 5.1.

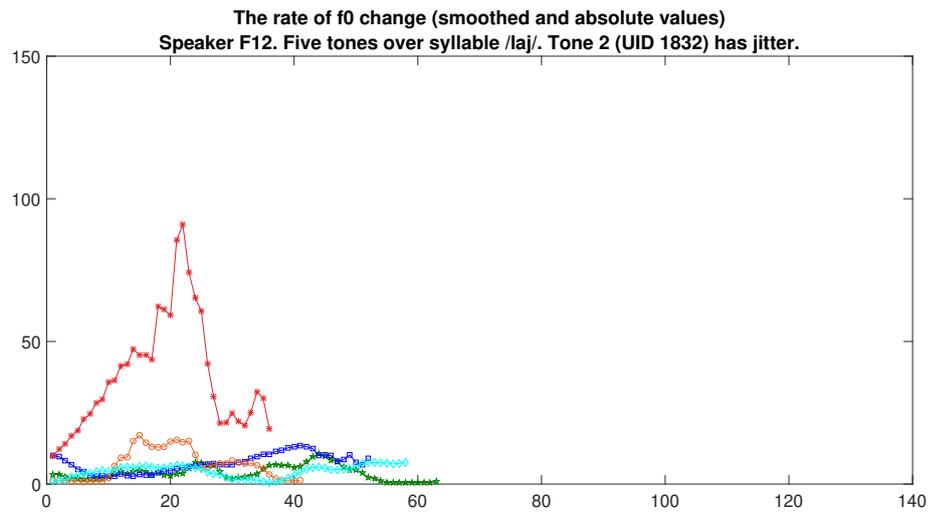
- DOI Tone **1**: <https://doi.org/10.24397/pangloss-0006812#W7>
- DOI Tone **2**: <https://doi.org/10.24397/pangloss-0006812#W5>
- DOI Tone **3**: <https://doi.org/10.24397/pangloss-0006812#W3>
- DOI Tone **4**: <https://doi.org/10.24397/pangloss-0006812#W9>
- DOI Tone **5**: <https://doi.org/10.24397/pangloss-0006812#W1>



(f) Minimal set containing the sample of minimally case of pressed voice in Fig. 5.2.

- DOI Tone **1**: <https://doi.org/10.24397/pangloss-0006812#W139>
- DOI Tone **2**: <https://doi.org/10.24397/pangloss-0006812#W137>
- DOI Tone **3**: <https://doi.org/10.24397/pangloss-0006812#W135>
- DOI Tone **4**: <https://doi.org/10.24397/pangloss-0006812#W141>
- DOI Tone **5**: <https://doi.org/10.24397/pangloss-0006812#W133>

Figure 5.6: Comparison of the  $\text{smoo\_deltaf}_0$   $_{\text{dEGG}}$ : an illustration for the consideration of the threshold established in condition 1 and 2. (cont. 2)



(g) Minimal set containing the sample of jitter in Fig. 5.1.

- DOI Tone 1: <https://doi.org/10.24397/pangloss-0006790#W170>
- DOI Tone 2: <https://doi.org/10.24397/pangloss-0006790#W168>
- DOI Tone 3: <https://doi.org/10.24397/pangloss-0006790#W166>
- DOI Tone 4: <https://doi.org/10.24397/pangloss-0006790#W172>
- DOI Tone 5: <https://doi.org/10.24397/pangloss-0006790#W164>

Figure 5.6: Comparison of the  $\text{smoo\_delta}f_o$   $_{dECC}$ : an illustration for the consideration of the threshold established in condition 1 and 2. (end)

ultimately, within a circle, as there is a risk that creakiness and **Tone 4** gradually come to be used with the same extension), it is constantly kept in mind that the ultimate goal is to distinguish creaky voice from modal voice *from a phonetic point of view*. That is to say, one needs to remember that, even though creaky voice is a prevalent phonological and phonetic feature of **Tone 4**, that does not entail that every syllable carrying **Tone 4** has creaky voice, nor that creaky phonation never appears in syllables with other tones.

Considering the smoo\_delta  $f_0$  of the tones in those seven examples, it is obvious that this variable fluctuates the most in **Tone 4** (as one would have predicted, because creak is associated with disruption of quasi-periodic phonation). The first example of extreme single-pulsed creak (Figure 5.6a) is the only one where the smoo\_delta  $f_0$  curve for **Tone 4** is constantly far apart from the curves for the four other tones, which are snugly close to one another (except for the very first detected glottal cycle and the last three or four cycles). The curve for **Tone 4** is in a space of its own, as it were, with values above 20%.

In the other figures, the onset and offset parts of **Tone 4** overlap, or at least are not too distant from the other tones (near bottom values, close to 0%  $f_0$  dEGG change). It demonstrates that the creaky voice does not usually occur right away at the beginning of the carrier syllable's rhyme, and that it also tends to taper off at the end. However, once the creaky voice sets in, the change of  $f_0$  dEGG is quick and remarkable. In single-pulsed creak and pressed voice, this change is generally in the range between 10% and 25%, but can be as high as 50% in the case of extreme single-pulsed creak or just less than 5% in the case of minimally pressed voice. In the remaining two types, this variable is recorded at much higher values: 80% in the case of multiply-pulsed creak and even reaching values such as 250% in the case of aperiodic creak. This makes sense: the sudden change between long and short cycles (also called micro-cycles) in these two latter types result in extreme changes in the derivative (delta).

Other tones, by contrast, have relatively flat smoo\_delta  $f_0$  dEGG curves, often in the range from 0% to 5%, as can be observed in all the following figures: 5.6a, 5.6b, 5.6c, 5.6d, 5.6e, and 5.6f. However, noticeable changes of  $f_0$  dEGG can occasionally be spotted in tones other than **Tone 4**. For instance, the second half of **Tone 3** in Figure 5.6a has delta values at about 8%. Another token of **Tone 3**, shown in Figure 5.6f, has a short portion which reaches 10%, just before the offset of voicing. Notably, in Figure 5.6g, besides **Tone 4** which has a remarkable  $f_0$  dEGG change at nearly 100%, all other tones also show some fluctuations, in the range of 5% to 15%. Once again, **Tone 3** is the second most 'jittery' tone (after **Tone 4**): a change on the order of 20% is found in the middle part of this token of **Tone 3**. This is common in data by speaker F7, a speaker who has an electroglottographic signal with great jitter in all the tones.

Another observation can be made that the final glottalization is prevalent in both elicitation contexts: in isolation as well as inside the carrier sentence, although it is more frequent in the former context. This phenomenon is revealed by a great and rapid rise in instability towards offset of syllables. The duration and the magnitude of the rise varies, but there are general trends nonetheless. It frequently happens over the

course of just a few final glottal cycles. The change can remain below 10% (as in the case of tokens of **Tone 4** and **Tone 5** in Figure 5.6e), but it is usually above 20% (for instance, such is the case of all the tones in Figure 5.6g), and even getting above 150% as in the token of **Tone 5** in Figure 5.6d.

By examining all the sub-types of creak that have been characterized here, in comparison with other tones in non-creaky voice, the key characteristic noticed is that the loss of periodicity in creaky voice causes rapid and abrupt changes in the fundamental frequency. Therefore, the first and the second conditions based on the rate of  $f_{0 \text{ dEGG}}$  change (i.e. in the variable of  $\text{smoo\_delta}f_{0 \text{ dEGG}}$ ) are established to catch the earliest signs of this phonation type.

**The 1<sup>st</sup> condition:** threshold of  $\text{smoo\_delta}f_{0 \text{ dEGG}}$  is set at 10%. In fact, the value initially considered for this threshold was 5%, because it allows the detection of most types of creaky voice, even pressed voice, while still excluding modal voice, where this parameter mostly remains under 5%. However, the precision and recall of this threshold are poor because there is a great deal of false positives: jitter in  $f_{0 \text{ dEGG}}$  is common, and the fact that the script used for estimating  $f_{0 \text{ dEGG}}$  is cycle-based, not autocorrelation-based, makes the presence of mild jitter especially prevalent in its output. It is not highly uncommon for jitter to reach rates of up to 15% (without a lapse into creak), as in the example shown in Figure 5.6g. Besides, some cases that constitute instances of light pressed voice (and thus, of "creak" in the sense used here) are not detected. The case of minimally pressed voice in **Tone 4** (Figure 5.6f) will not be detected using a 5% threshold since the rate of change is only from 2 to 4%. The decision to raise the threshold to 10% instead of 5% means that such cases of pressed voice need to be detected by other means. This makes sense since we can really prioritize the detection of 'real' creak at the first stage and continue later on the further detection of the pressed voice which is relevant but not classified in the package of sub-type of creak. Setting the threshold at 10% leads to greater precision. An undesired consequence is that more cases of pressed voice will go undetected by this method (and this will need to be compensated by using other tools), but a desirable consequence is that fewer cases of mild jitter will be caught in the net. All in all, the threshold at 10% makes good sense for detecting phonetic phenomena of creaky voice. Pressed voice, with a change in  $f_{0 \text{ dEGG}}$  less than 10%, is closer to modal voice in the continuum of single-pulsed creak – pressed voice – modal voice; and strong jitter, with a change in  $f_{0 \text{ dEGG}}$  greater than 10%, is really close to double-pulsed creak in terms of signals (both acoustics and electroglottography) and perception.

In short, this condition preliminarily separates creaky voice from non-creaky voice, with two clear and allows two existences: (i) the pressed voice close to the modal voice will not be included, (ii) the great jitters will be included.

**The 2<sup>nd</sup> condition:** threshold on  $\text{smoo\_delta}f_{0 \text{ dEGG}}$  at 20%. This condition aims at continuing to separate multi-pulsed creak and aperiodic creak from the four sub-types of creak which were initially filtered out by the first condition. From the observations in Figure 5.6, it can be noticed that generally, the more irregular the periodicity is in the sub-types of creaky voice, the more remarkable is the  $\text{smoo\_delta}f_{0 \text{ dEGG}}$  values. Indeed,

in two examples of Figures 5.6c and 5.6d, the  $\text{smoo\_delta}f_{o\_dEGG}$  values of **Tone 4** are recorded prominently at 80% and even up to 250% in cases of double-pulsed creak and aperiodic creak, respectively. The 80% to 90% values of those two syllables are higher than 20%. Whereas, the other examples, this variable generally does not exceed 20%. The example of maximum pressed voice in 5.6e has only one value at cycle 16<sup>th</sup> located above 16%. The extreme single-pulsed creak in Figure 5.6a is an exception, with all values in the range of 20% to 50%. Technically, this example will be detected and classified along with the second basket of multi-pulsed creak and aperiodic creak at this stage of the second condition. Thus, it shows that there are still flaws in the separation of the sub-types of creak. As a result, the recall of CreakDet version 1 will not be high. Apparently, the condition on the threshold of the rate of  $f_{o\_dEGG}$  change basically distinguishes the creaky voice from the modal voice, but it is not sufficient to enable the classification of all the different sub-types. Therefore, the third condition on the direction of  $f_{o\_dEGG}$  change and DECPA change is added to continue this task. Note that while DECPA does not have a linear relationship to acoustic intensity, there is still a strong relationship between the two.

**The 3<sup>rd</sup> condition:** based on the particular characteristics of multiply-pulsed creak to set apart this sub-type from aperiodic creak. The most obvious feature of double-pulsed creak is the alternation of short and long cycles (the starting cycle can be either a % short or a long cycle) in a great change of fundamental frequency. This characteristic has been broadly recognized by numerous previous works such as: “double pulsation” in (Paul Moore and Leden, 1958), “double glottic pulses” in (Hollien and Ronald W. Wendahl, 1968) and (Hedelin and Huber, 1990), (Roubeau, Henrich, and Castellengo, 2009), etc. The other also mentioned the possibilities of triplet cycles (Blomgren et al., 1998) or higher multiples (Keating, Garellek, and Kreiman, 2015). In the current study, triplet cycle patterns were not found, hence we only address here the case which we call double complex-repetitive pattern.

The third condition is actually a combined condition containing three sub-conditions that are applied simultaneously by the “and” command, including: (i) condition on the zigzag pattern; (ii) condition on the similarities between second nearest cycles; and (iii) condition on dissimilarity of adjacent cycles.

*Sub-condition #1: condition on the zigzag pattern (Line 82 in List 5.3).* The name of this sub-condition reflects the shape of  $f_{o\_dEGG}$  modulation in multiply-pulsed creak, as shown in the example in the Figure 5.1. The alternation of short and long glottal cycles corresponds to the alternation of high and low  $f_{o\_dEGG}$  values. As a consequence, it leads to alternating positive and negative delta  $f_o$  values. To catch this kind of pattern at each syllable, a loop is created to consider delta  $f_o$  across windows of four successive cycles, testing whether there are alternations of positive and negative delta  $f_o$ . Such alternations are marked as positives of zigzag pattern (see line 82 in 5.3).

*Sub-condition #2: condition on the similarities between second nearest cycles, in terms of  $f_{o\_dEGG}$  and DECPA (Lines 83-90 in List 5.3).* This condition is based on the rate of change of  $f_{o\_dEGG}$  and DECPA for a two-dimensional verification of each current cycle with its second nearest one. The calculation is similar to the

equation of  $\text{rate\_delta}f_{o\_dEGG}$  in the “ $f_{o\_dEGG}$  change” function since it is in fact an opposite relationship: the more similar, the less different. In the same window of each four successive cycles as the first sub-condition, the cycle at the position  $k$  is subtracted from the cycle at  $k+2$  in absolute value for an easier comparison, then the result is divided for the smaller value between  $k$  and  $k+2$ . The same calculations are applied to  $k+1$  and  $k+4$ .

The idea of this condition is to catch the telltale characteristic of aperiodic creak. The hypothesis is that aperiodic creak has much higher values in this variability index than all the other types, as a consequence of completely irregular glottal pulses. The simple detection script might have a few false positives corresponding to big jumps found in transitions into (and out of) multiply-pulsed creak, and also in some cases of single-pulsed creak (as in the example in Figure 5.1), but generally they are nowhere as salient as aperiodic creak. Therefore, in order to exclude aperiodic creak and keep only multiply-pulsed creak, we set the threshold on this condition at less than 0.2 for all  $f_{o\_dEGG}$  and DECPA change of second nearest cycles.

*Sub-condition #3: Condition on dissimilarity of adjacent cycles (Lines 91-96 in List 5.3).*

The journey towards the creation of the third condition had actually begun before the CreakDet function was written out. In the initial tests, in order to detect the complex-repetitive pattern, we created a function called “CRPdet.m” (as can be seen in List 5.2) with two conditions applied simultaneously. If there are alternations of positive and negative  $\text{delta}f_o$  (i.e. a zigzag pattern) and the change of  $\text{smoo\_delta}f_{o\_dEGG}$  in these cycles is equal to or greater than 20% (i.e. the 2<sup>nd</sup> condition is met), they are detected as positives of complex-repetitive pattern in double-pulsed creak. This function is tested in all the samples we took in the first step of identifying the sub-types of creak. The focus is on three samples, respectively illustrating (i) double-pulsed creak, (ii) jitter, which has similar characteristic of zigzag pattern but has a much smaller change in  $\text{smoo\_delta}f_{o\_dEGG}$ , and (iii) aperiodic creak with an unexpected change in glottal cycles. Figure 5.7 provides the demonstration of how CRPdet function worked on those three samples.

In order to understand that figure, it should be noted that the main demonstration is displayed in the top sub-figure with two conditions (zigzag pattern and threshold on  $\text{rate\_delta}f_{o\_dEGG}$ ) and the electroglottographic signal below to reflect the detection of these conditions on corresponding glottal periods. The two conditions applied in “CRPdet.m” are reflecting in coloring, to illustrate which values are positive or negative, i.e., whether they meet the conditions or not. In particular, the condition of zigzag pattern is presented in magenta and green circles. It turns green if the window of four successive  $\text{delta}f_{o\_dEGG}$  values meets the condition, otherwise it remains magenta. The result of detection is stored in the first cycle of the window. This means that for any green cycle in the figure, there is a positive zigzag pattern of four successive values, starting from that value. At the same time, the detection of a positive zigzag pattern is reflected in the electroglottographic signal below by five green dots located near the beginning (closure instants) of the corresponding glottal periods. This is due to the fact

that each  $\text{delta}f_{0 \text{ dEGG}}$  value is calculated by two values of  $f_{0 \text{ dEGG}}$ , a window of four successive  $\text{delta}f_{0 \text{ dEGG}}$  values therefore involves five values of  $f_{0 \text{ dEGG}}$ . If one or more values in the window do not meet the condition, the window will be detected as a negative zigzag, shown by a magenta cycle at the first  $\text{delta}f_{0 \text{ dEGG}}$  value, and will not be marked in the electroglottographic signal.

Similarly, the condition on  $\text{smoo\_delta}f_{0 \text{ dEGG}}$  variable is represented by red and blue asterisks for positive and negative, respectively. The red line at 20 indicates the threshold that is set for this condition. All values of  $\text{smoo\_delta}f_{0 \text{ dEGG}}$  that reach or cross this line are marked in red, otherwise they are in blue. Since a value of  $\text{smoo\_delta}f_{0 \text{ dEGG}}$  is calculated from five  $f_{0 \text{ dEGG}}$  values as the formula in 5.1, a positive value corresponds to five  $f_{0 \text{ dEGG}}$  values, i.e., five glottal period is detected, as they are marked in blues dots in the electroglottographic signal. The negatives, on the other hand, are marked with blue asterisks in the display part of this variable but are not marked in the electroglottographic signal.

In the electroglottographic signal, if red and blue dots are marked in the same cycle, it means that the cycle meets both conditions of the zigzag pattern and that the  $\text{smoo\_delta}f_{0 \text{ dEGG}}$  is equal to or greater than 20%, and is therefore detected as a cycle belonging to the complex-repetitive pattern. At least five successive cycles will be detected for a minimum interval.

As can be seen in Figure 5.7, CRPdet.m works pretty well on the sample of multiply-pulsed creak since it caught the long span of double-complex repetitive pattern in the middle. There are six cycles early in the beginning (from 15<sup>th</sup> to 20<sup>th</sup>) and four cycles at transition positions (two before and two after at the 30<sup>th</sup> and 31<sup>st</sup> cycles, and 73<sup>th</sup> and 74<sup>th</sup> cycles, respectively), which are positive with the zigzag pattern but negative with the  $\text{smoo\_delta}f_{0 \text{ dEGG}}$  condition. Therefore, they are not counted as instances of complex-repetitive patterns. This is reasonable because the first six cycles are just a very short and light jitter at the beginning of modal voice before entering into creak. Four cycles in transitional positions between modal voice and creaky voice are in a borderline situation.

In the case of jitter in Figure 5.7b, the function also did well in not detecting this type of signal as a complex-repetitive pattern, with none of the cycles meeting both conditions. It is predictable that some portions would be positive with the zigzag condition since the shared characteristic of jitter and multiply-pulsed creak is the alternation of long-short glottal cycles but to varying degrees. Indeed, in the jitter sample here, a short portion right at the beginning (at cycles 1<sup>st</sup> to 9<sup>th</sup>) and a long portion (at cycles 18<sup>th</sup> to 48<sup>th</sup>) are marked as positives of zigzag pattern, but both of them and even the entire syllable are negative with second condition, i.e., their change of  $\text{smoo\_delta}f_{0 \text{ dEGG}}$  is always less than 20%.

The sample of aperiodic creak in Figure 5.7c, on the other hand, highlights an issue: that during the presence of creak event, some discrete portions are detected as positive of complex-repetitive pattern. Among them, the first one consists of two cycles, the 9<sup>th</sup> and 10<sup>th</sup>. It seems that these are points of intersection between two different portions: (i) cycles from the 5<sup>th</sup> to the 8<sup>th</sup> meet the condition of zigzag pattern but

do not cross the threshold of  $\text{smoo\_delta}f_{o\_dEGG}$  condition, whereas (ii) cycles from the 11<sup>th</sup> to the 16<sup>th</sup> are negative with zigzag pattern but positive over 20% change in the  $\text{smoo\_delta}f_{o\_dEGG}$  condition. A minimally complex-repetitive pattern (double case) must occur in at least four successive cycles for there to be a repetition, so it could be said that this brief portion of two cycles is a false positive. The next positive portions from the cycles from 17<sup>th</sup> to 28<sup>th</sup>, and from 30<sup>th</sup> to 35<sup>th</sup> are separated by only one cycle 29<sup>th</sup>. Technically, it is almost a detection of a long span of complex-repetitive pattern, but in fact, what can be observed from the signals in Figure 5.7c does not support this. It is far from a proper case of complex-repetitive pattern in multiply-pulsed creak. The alternation between long and short glottal cycles in this case is not quasi-periodic, causing the change of  $f_{o\_dEGG}$  to be formed as an irregular zigzag line. In short, with the initial CRPdet.m, the two conditions on the zigzag pattern and the  $\text{smoo\_delta}f_{o\_dEGG}$  variable are intended to distinguish multiply-pulsed creak (case of double pulses) from jitter, through the different in magnitude of  $\text{smoo\_delta}f_{o\_dEGG}$ , as well as from aperiodic creak, through the regular zigzag pattern. However, the case of aperiodic creak here raised the issue of a possible false positive case with detected irregular zigzag patterns. Therefore, more conditions must be set to improve the accuracy of the function. Two sub-conditions of the third condition of the CreakDet.m function are provided for this purpose.

Listing 5.2: function “CRPdet.m”

```

1 function [CRP_fo,CRP_DECPA] = CRPdet(datasheet)
2 %%%%%%%%%%
3 %%%%%%%%%% Detection of complex-repetitive patterns (CRP)%%%%%%%%%
4 % the distinctive characteristics of multi-pulsed creak.%
5 %%%%%%%%%%
6 % Input: a datasheet from Peakdet.m
7 % Output: CRP is a vector that contains as many values as there are cycles
8 % in the data sheet. It is set at 0 or 1 depending on whether the cycle at
9 % issue is part of a detected CRP span.
10 %%%%%%%%%%
11
12 % Log:
13 % 11/2020: the function was created
14 % 15/2/2021: Separate CRP (complex - repetitive patterns) from ZZZ (Zig zag
15 % pattern). CRP is preserved as the identification of double-multi pulsed
16 % creak, whereas ZZZ detects all other phenomena (mostly jitter) which also
17 % have alternation of long and short periods (faux CRP).
18 % This script detects complex - repetitive patterns which is the identified
19 % condition of multi-pulsed creak (case of double-pulsed creak,
20 % the case of triple-pulses is not found in my data, hence it isn't take in
21 % to account here). The identifier of this pattern includes three conditions:
22 % (i)condition #1 - ZZZ: the alternation of short and long cycles (the
23 % starting cycle can be either a short or a long cycle). To detect this
24 % kind of signal, we consider delta fo of each 4 successive cycles, if
25 % there are alternation of positive and negative delta fo.
26 % (ii) condition #2: set a threshold for soo_abs_rate_deltafo >=20%.
27 % This condition helps to eliminate the case of jitter. (to be verified)
28 % (iii) condition #3 : The similarity of "combined cycles".

```

### 5.3 Detection of creaky voice from the electroglottographic signal

```

29 % (a combined cycle = a long cycle + a short cycle)
30 % The combined cycles need to be analogous to each other. This condition
31 % helps to eliminate the case of aperiodic creak which has faux CRP
32 % pattern, and the glottalization usually occurs at the end of syllables.
33 % The expected output of this function is: it will tell us at each token,
34 % whether there are CRP or not. If the answer is yes, it will indicate
35 % exactly which cycles are.
36 %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
37
38 % Trimming any empty lines in the datasheet
39 datasheet = cleanzeros (datasheet);
40 % Initializing the variable.
41 CRP_fo = []; fo = [];
42 % Store input values
43 fo = datasheet(:,3);
44 DECPA = datasheet(:,4);
45 % Calculating the number of periods of token
46 numb_period = length(datasheet(:,1));
47 % Calculating two input variables for the detection: (i) deltafo and (ii)
48 % rate of deltafo:
49 [deltafo , rate_deltafo] = fochange (fo);
50 [deltaDECPA , rate_deltaDECPA] = fochange (DECPA);
51 % For debugging, uncomment the code below
52 % disp (['deltafo: ', num2str(deltafo)])
53 % disp (['rate_deltafo: ', num2str(rate_deltafo)])
54 % for i = 1:numb_period - 1
55 %     % Calculate deltafo
56 %     deltafo(i) = fo(i+1) - fo(i);
57 %     % The rate of deltafo (semitone)
58 %     rate_deltafo(i) =(deltafo(i)/fo(i)) * 100;
59 % end
60
61 % For counting number of complex - repetitive patterns (CRP)
62 % To calculate the ratio of CRP portion in whole syllable.
63 CRP_count=0;
64
65 % To detect the cycles belong to "real" CRP we need an "if" statement
66 % containing three conditions:
67 % (i) The main condition (Condition on sign inversion):
68 % 4 successive deltafo values have to be alternating between negative
69 % and positive values;
70 % (ii) The additional condition #1: compare the similarities of two
71 % successive "huge" cycles which combined by one CRP component
72 % (1 big cycle + 1 small cycle).
73 % (To do this, setting a threshold for the rate_deltafo of (k+k1) and (k2+k3)
74 % (iii) The additional condition #2: Condition on amount of fo change (%).
75 % Unit tests:
76 % 1. Typical real CRP in a multi-pulsed creak (token 551, UID 6002, F10,
77 % syllable /ku4/): min = , -> safe guarantee threshold at 20%
78
79 for k = 1:numb_period - 4 % Window of 4 cycles. NOTE: The length of deltafo
80 % is 1 token less than the number of total cycles.
81 % Hence, we need k = 1:numb_period - 4

```

```

82         % instead of k = 1:numb_period - 3
83
84     % Smooth deltafo;
85 %     smooth_delta(k) = mean(abs(rate_deltafo(k:k+3))); %Why we need this
86 %     ???
87     if and(deltafo(k)*deltafo(k+1)<0 & deltafo(k+1)*deltafo(k+2)<0 & deltafo(k+2)*deltafo(k+3) < 0 ,...
88         mean(abs(rate_deltafo(k:k+3))) >=20)
89     %     The third condition
90 %     mean(rate_deltafo(k:k+3)) <=???
91 % % The next line is for debugging to follow the progress of detection
92 %     disp(['mean of rate deltafo over a window of 4 cycles detected
93 %     as CRP pattern: ',num2str(mean(abs(rate_deltafo(k:k+3)))]);
94
95     CRP_count=CRP_count+1;
96     % Storing the cycles in CRP span. The span extends from k to k+3 in
97     % terms of delta fo. So, in terms of glottal cycles, the span
98     % extends over one more cycle: from k to k+4.
99     CRP_fo(k:k+4)=1;
100 else
101     CRP_fo(k:k+4)=0;
102 end
103
104 % Same with DECPA.
105 if and(deltaDECPA(k)*deltaDECPA(k+1)<0 & deltaDECPA(k+1)*deltaDECPA(k+2)
106 <0 & deltaDECPA(k+2)*deltaDECPA(k+3) < 0 ,...
107     mean(abs(rate_deltaDECPA(k:k+3))) >20)
108 % If the condition is met: declare all four successive cycles as
109 % belonging in a complex-repetitive pattern.
110 CRP_DECPA(k:k+4)=1;
111 else
112 % For explicitness:
113 CRP_DECPA(k:k+4)=0;
114 end
115 end
116
117 % Bringing the two sources together: match and check.
118 % Condition on identity. Careful MC: check the length of the 2 vectors
119 % and take care of any rim effects.
120 for m = 1:length(CRP_fo)
121     if and(CRP_fo(m)==1,CRP_DECPA(m)==1)
122         disp('Presence of CRP indicated by tests on both fo and DECPA.')
123         CRP_conf(m) = 1;
124     elseif and(CRP_fo(m)==1,CRP_DECPA(m)==0)
125         disp('Presence of CRP indicated by tests on fo but not DECPA.')
126         CRP_conf(m) = 0;
127     elseif and(CRP_fo(m)==0,CRP_DECPA(m)==1)
128         disp('Presence of CRP indicated by tests on DECPA but not fo.')
129         CRP_conf(m) = 0;
130     elseif and(CRP_fo(m)==0,CRP_DECPA(m)==0)
131         disp('No CRP detected by either fo or DECPA.')
132         CRP_conf(m) = 0;
133     else

```

### 5.3 Detection of creaky voice from the electroglottographic signal

```

132     error('You should never see this text. Something has gone badly wrong
133     .')
133     end
134 end

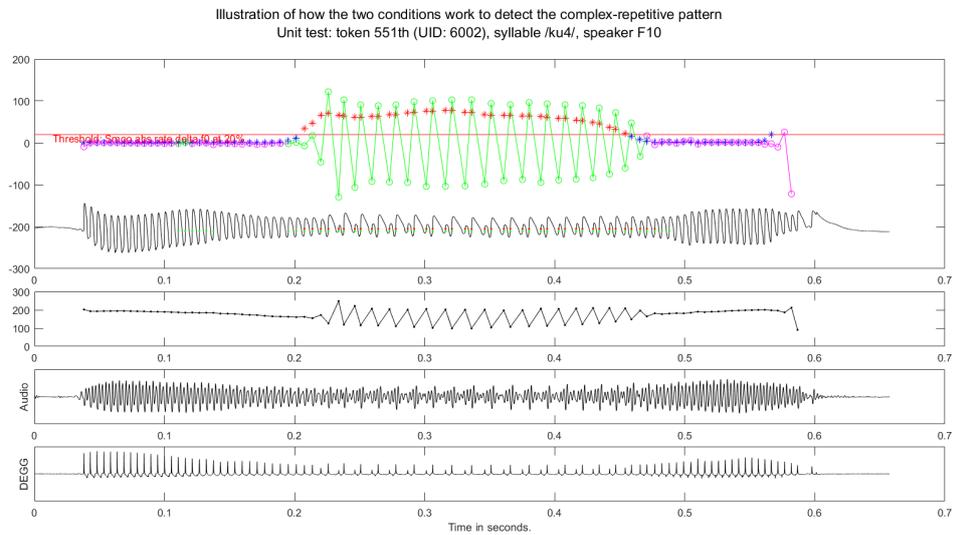
```

Listing 5.3: function 'CreakDet.m'

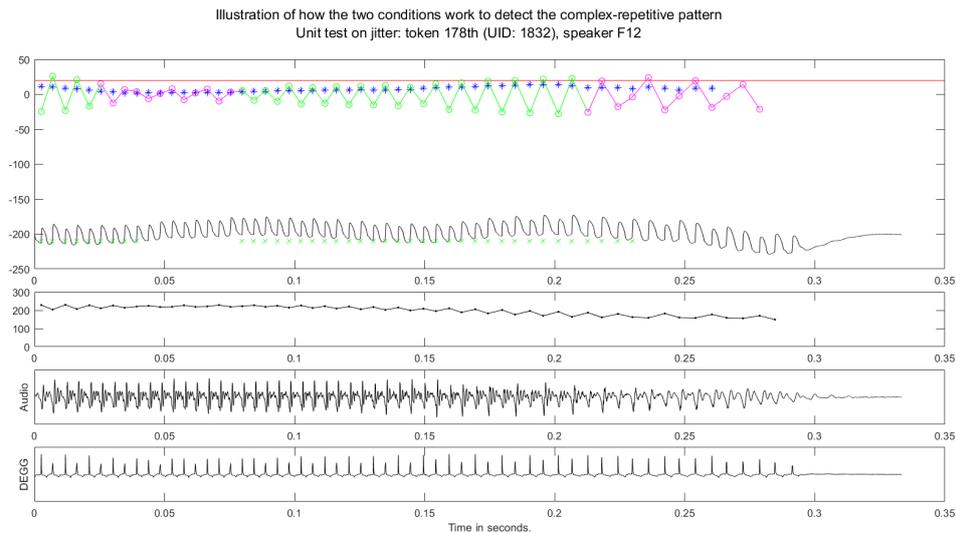
```

1 function [creak] = CreakDet(datasheet , windowlength)
2 %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
3 % Detection of creak from electroglottographic signals.
4 % Three types are distinguished:
5 % (i) complex-repetitive patterns (CRP): the distinctive characteristics of
6 % multi-pulsed creak
7 % (ii) aperiodic creak
8 % (iii) (to an extent to be verified) single-pulse creak.
9 %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
10 % Input: a datasheet from Peakdet.m (and was sorted by tones)
11 % Output: <creak>, a vector with one value for each line in the input <
12 %         datasheet >.
13 %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
14 % No creak detected: set at 0
15 % Single-pulsed creak: set at 1 (based on a measurement of irregularity over
16 % a given time window)
17 % Aperiodic creak: set at 2 (based on previous condition plus presence of a '
18 % zig-zag' pattern)
19 % Double-pulsed creak: set at 3 (based on previous conditions plus
20 % similarities between a cycle and the second nearest cycle)
21 %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
22 % Log:
23 % 31/3/2021: the function was created using materials from CRPdet_test.m
24 % 15/2/2021: Separate CRP (complex - repetitive patterns) from ZZP (Zig zag
25 % pattern). CRP is preserved as the identification of double-multi pulsed
26 % creak, whereas ZZP detects all other phenomena (mostly jitter) which also
27 % have alternation of long and short periods (faux CRP).
28 % This script detects complex - repetitive patterns which is the identified
29 % condition of multi-pulsed creak (case of double-pulsed creak,
30 % the case of triple-pulses is not found in my data, hence it isn't take in
31 % to account here). The identifier of this pattern includes three conditions:
32 % (i) condition #1 - ZZP: the alternation of short and long cycles (the
33 % starting cycle can be either a short or a long cycle). To detect this
34 % kind of signal, we consider delta fo of each 4 successive cycles, if
35 % there are alternation of positive and negative delta fo.
36 % (ii) condition #2: set a threshold for soo_abs_rate_deltafo >=20%.
37 % This condition helps to eliminate the case of jitter. (to be verified)
38 % (iii) condition #3 : The similarity of "combined cycles".
39 % (a combined cycle = a long cycle + a short cycle)
40 % The combined cycles need to be analogous to each other. This condition
41 % helps to eliminate the case of aperiodic creak which has faux CRP
42 % pattern, and the glottalization usually occurs at the end of syllables.
43 %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
44 % Trimming any empty lines in the datasheet

```



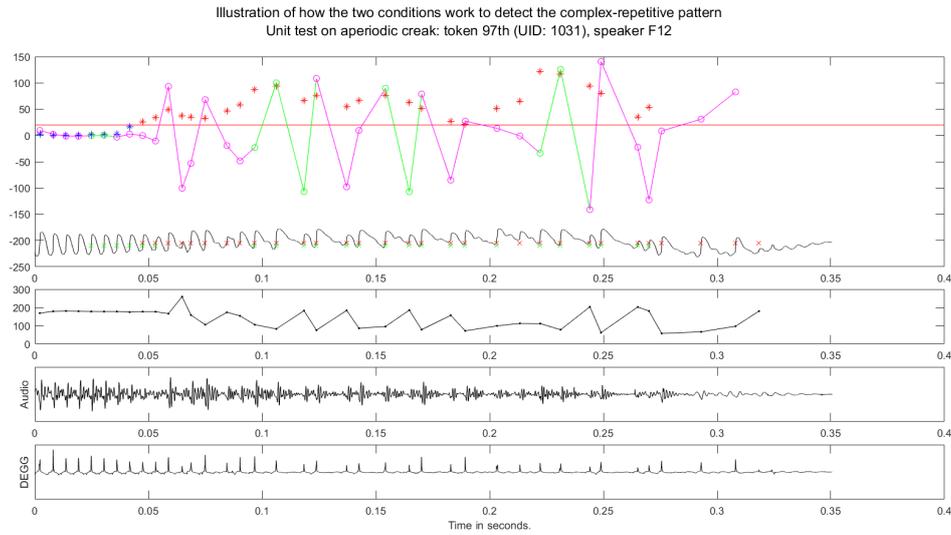
(a) Double-pulsed creak. Speaker F10, token 551, UID 6002. (Also in Figs. 5.1 and 5.6c.)



(b) Jitter. Speaker F12, token 178, UID 1832. (Also in Figs. 5.1 and 5.6g.)

Figure 5.7: Illustration of how the “CRPdet.m” function works, through examples of (a) double-pulsed creak, (b) jitter, and (c) aperiodic creak. In each sub-figure, from top to bottom: (i) two variables:  $\Delta f_0$  (in magenta and green) and  $\text{smoo\_}\Delta f_0$  (in blue and red) with EGG signal; (ii)  $f_{0\_dEGG}$ ; (iii) acoustic signal; and (iv) dEGG.

5.3 Detection of creaky voice from the electroglottographic signal



(c) Aperiodic creak. Speaker F12, token 97, UID 1031. (Also in Figs. 5.1 and 5.6d.)

Figure 5.7: Illustration of how the “CRPdet.m” function works, through examples of (a) double-pulsed creak, (b) jitter, and (c) aperiodic creak. In each sub-figure, from top to bottom: (i) two variables:  $\Delta f_0$  (in magenta and green) and  $\text{smoo\_}\Delta f_0$  (in blue and red) with EGG signal; (ii)  $f_{0\_dEGG}$ ; (iii) acoustic signal; and (iv) dEGG.

```

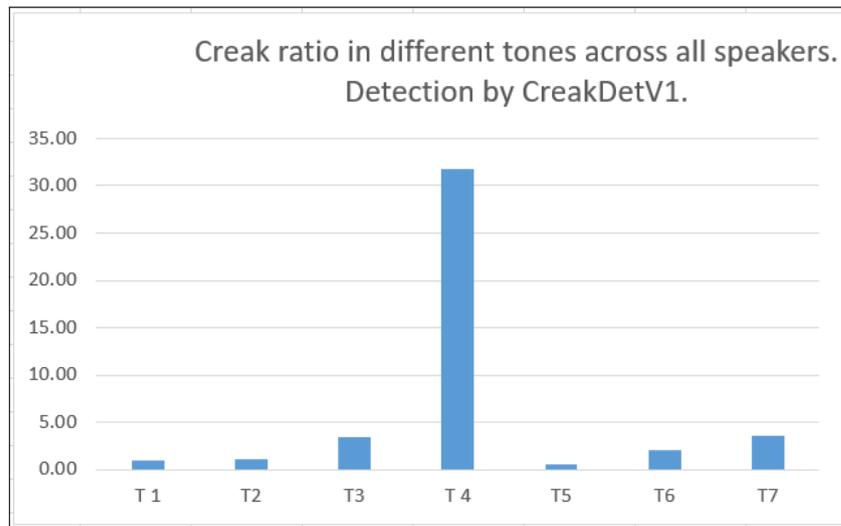
42 datasheet = cleanzeros (datasheet);
43 % Initializing the variables.
44 % Output variable:
45 creak = []; fo = []; DECPA = [];
46 % Store input values
47 fo = datasheet(:,3);
48 DECPA = datasheet(:,4);
49 % Calculating the number of periods of token
50 numb_period = length(datasheet(:,1));
51 % Calculating two input variables for the detection: (i) deltafo and (ii)
52 % rate of deltafo:
53 [deltafo, rate_deltafo, smoo_deltafo] = fochange (fo,windowlength);
54 [deltaDECPA, rate_deltaDECPA, smoo_deltaDECPA] = fochange (DECPA,windowlength
55 );
56 % For debugging, uncomment the code below
57 % disp (['deltafo: ', num2str(deltafo)])
58 % disp (['rate_deltafo: ', num2str(rate_deltafo)])
59 for k = 1:numb_period - (windowlength + 1) % Window of 4 cycles. NOTE: The
60     length of deltafo
61         % is 1 token less than the number of total cycles.
62         % Hence, we need k = 1:numb_period - 4
63         % instead of k = 1:numb_period - 3
64         if smoo_deltafo(k) >= 10
65             % smooth delta fo is calculated as: mean(abs(rate_deltafo(k:k+3)))
66             >=10
67             % If the first condition is met: <creak> is set at 1 to begin with.
68             This value will later be overwritten in case
69             % it is found that more specific types (aperiodic creak or double-pulsed
70             creak) are detected.
71             creak(k:k+windowlength-1) = 1;
72
73             % Detection of aperiodic creak: first pass: set a higher threshold
74             if smoo_deltafo(k) >=20
75                 creak(k:k+windowlength-1) = 2;
76             end
77         end
78     end
79 % Detection of CRP: two conditions: 'zig-zag' pattern and similarity among
80 % second nearest cycles.
81 for k = 1:numb_period - 4 % Window of 4 cycles. NOTE: The length of deltafo
82     % is 1 token less than the number of total cycles.
83     % Hence, we need k = 1:numb_period - 4
84     % instead of k = 1:numb_period - 3
85     % First condition:
86     if deltafo(k)*deltafo(k+1)<0 & deltafo(k+1)*deltafo(k+2)<0 & deltafo(k+2)*
87         deltafo(k+3)<0
88         % add condition on similarities between second nearest cycles, in
89         terms of fo and DECPA.
90         % Cycles 1 and 3
91         similfo_firstpair = abs( fo(k+2) - fo(k) ) / min(fo(k),fo(k+2));
92         similDECPA_firstpair = ( DECPA(k+2) - DECPA(k) ) / DECPA(k);

```

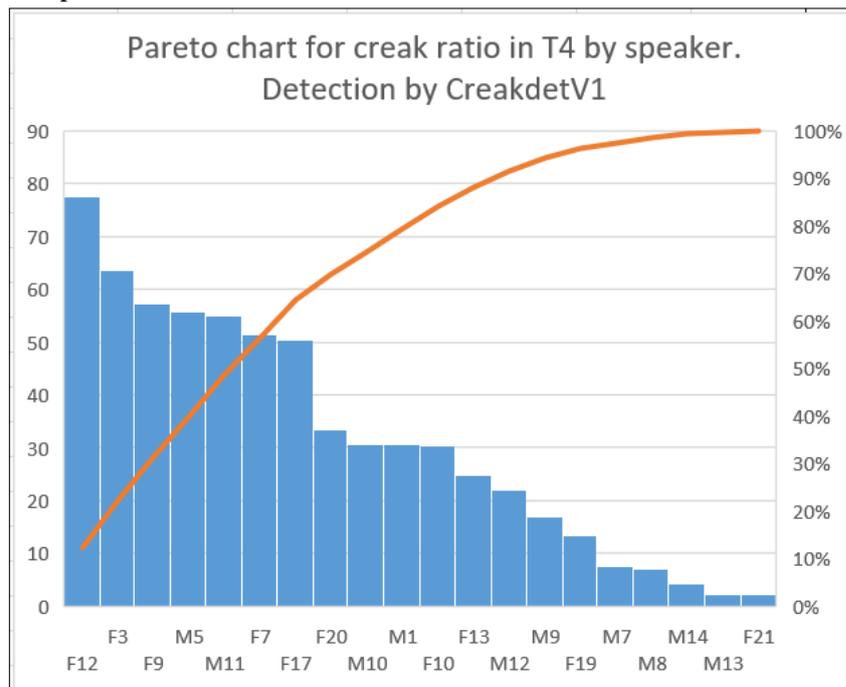
### 5.3 Detection of creaky voice from the electroglottographic signal

```
87 % Then cycles 2 and 4
88 similfo_secondpair = abs( fo(k+3) - fo(k+1) ) / min(fo(k+1),fo(k+3));
89 similDECPA_secondpair = ( DECPA(k+3) - DECPA(k+1) ) / DECPA(k+1);
90 if similfo_firstpair < 0.2 & similDECPA_firstpair < 0.2 &
similfo_secondpair < 0.2 & similDECPA_secondpair < 0.2
91 % add a condition on dissimilarity of adjacent cycles
92 adj1 = abs( fo(k+1) - fo(k) ) / min(fo(k),fo(k+1));
93 adj2 = abs( fo(k+2) - fo(k+1) ) / min(fo(k+1),fo(k+2));
94 adj3 = abs( fo(k+3) - fo(k+2) ) / min(fo(k+2),fo(k+3));
95 if adj1 > 0.3 & adj2 > 0.3 & adj3 > 0.3
96 creak(k:k+3) = 3;
97 end
98 end
99 end
100 end
```

Overall results of Creakdet version 1



(a) Ratio of creak (all sub-types) detected in seven tones across twenty speakers



(b) Ratio of creak (all sub-types) detected in Tone 4 across twenty speakers

Figure 5.8: Overall results of CreakDet version 1.



Figure 5.8a shows the ratio of cycles detected as creaky (any sub-types) by CreakDet (version 1), to the total number of glottal cycles ( $n = 426,676$ ), sorted by tone. For testing at this stage, we initially analyze tokens spoken in a carrier sentence (i.e. the target syllables extracted from the carrier sentence). Later, isolated tokens can be similarly analyzed, for comparison. Tokens are pooled across all speakers, and hence also across genders. It is obvious to notice that the creak ratio thus computed is by far higher in **Tone 4** (the creaky tone) compared to all other tones. While no other tone has a creak rate greater than 5%, **Tone 4** alone exhibits this extremely prominent rate above 30%, providing further evidence of the close association of phonetic creak (as detected through the procedure explained above) with **Tone 4** as a lexical tonal category.

However, there is clearly still further work to be done to improve the precision of these results. If blue color on Figure 5.8b were *only* found for **Tone 4**, it would be a reliable guide to identification of the tone. But since blue color is not found exclusively on **Tone 4**, creak in the sense of that figure (detection by CreakDet) is not specific to **Tone 4**, and it is an issue how it can function as a major cue in tonal identification. Obviously, the next question is how to progress from detection of creak in the most general phonetic sense to the patterns of creak specific to **Tone 4**— assuming that such patterns can be identified, of course.

Temporal information is likely to matter in the move that consists in teasing apart different phenomena within the blue-colored cases in Figure 5.8a. The blue color pools together longer creaky stretches and short glottalization phenomena, because both share the characteristic of having dramatic change in fundamental frequency. The conditions of  $\Delta f_0$  rate, setting a threshold at 10% and 20%, does not eliminate the cases of onset and offset glottalization. In Muong as in many (all?) other languages, glottalization can occur at junctures in the utterance (such as phrase boundaries). Functionally, this intonational phenomenon is on a different level from phonological specifications of the lexical tones. Luckily for my study, the difference is fairly neat from a phonetic point of view, too. I hypothesize on the basis of memories from processing all the syllables (as part of the semi-automatic workflow) that the marking of junctures is located with some precision at the end of the last syllable before the juncture (at the juncture), whereas the lapse into creaky voice typical of canonical **Tone 4** occurs much earlier in the syllable rhyme. Under this hypothesis, it should be possible to tease apart two sharply different subsets within CreakDet's *creak* results. One is final glottalization, occasionally observed on any of the tones, and the other is the typical lapse into creaky voice found in **Tone 4**. (Cases of short, initial glottalization as part of coarticulatory effects after certain initial consonants will need to be dealt with separately.)

Now moving on to clarify my intuitive sense that there are salient differences (obvious to the trained eye) in terms of position and duration. The stretch of initial or final glottalization due to the marking of junctures is brief; as an initial test, it seems reasonable to try a temporal dividing line at five cycles at the onset or offset and consider that longer stretches belong to *bona fide* creaky voice. One of the bases for this initial choice is that, as a rule of thumb, “phonologically real” creaky voice in

**Tone 4** lasts from about 10 to 15 cycles (with an average around 13 cycles).

Version 1 of CreakDet still needs to resolve some remaining issues. Besides the issues of glottalized onset and offset just mentioned above, the next thing that will need to be noticed and addressed is that pressed voice, which is known from the sifting of examples to be well attested among allotones of **Tone 4**, is not treated adequately by CreakDet version 1 (whose main focus is on phonetic creak). Already in the first condition on  $\text{smoo\_delta}f_{o\_dEGG}$ , the threshold of 10% excluded many cases of mild to minimally pressed voice as in the sample shown in Figure 5.2. Therefore, a further solution will be needed to improve the detection of pressed voice and provide a full picture of glottalization.

In a nutshell, at the current stage of the script for creak detection, CreakDet (version 1: April 2021), the detection is based on three simple conditions listed in Section 5.3.3. There is still a range of issues that need to be addressed further down the road: more conditions will be added and existing conditions will be modified in the next versions. The final goal is to obtain a result that contains the information about the creaks detected from the input data as precisely and elaborately as possible. Work that cannot be done in this ongoing study will be opened for further continuation.

## 5.4 Glottalization in Kim Thuong Muong: phonetics and phonology

In light of the above developments, the time now seems ripe for summarizing key facts on the phonetics and phonology of glottalization in Kim Thuong Muong.

As has been mentioned at several places in this thesis, and as can be seen clearly in Figures 4.1, 4.2 and 4.3, glottalization makes **Tone 4** well-distinguished from the other four tones of smooth syllables. Glottalization sets apart **Tone 4** by two acoustic properties. Firstly, fundamental frequency: **Tone 4** is very different from the other tones, as its  $f_0$  values usually drop from under 100 Hz to minimum values below 30 Hz whereas the four remaining tones are consistently inside a range between 100 and 150 Hz. Even without information of open quotient, **Tone 4** can be identified as being the lowest tone, at the bottom of speaker's range, contrasting with the four remaining tones (which are distributed in the area from the middle to the top of speaker's range). Secondly, open quotient also tells apart **Tone 4** from the others, with values of **Tone 4** reaching below 30% whereas all the other tones have  $O_q$  values above 40%.

The evidence examined here suggests that  $f_{o\_dEGG}$  ensures the bulk of the tonal distinctions in Kim Thuong Muong, whereas  $O_{q\_dEGG}$  is only relevant insofar as it discriminates the glottalized tone from the remaining four tones (modal tones). Moreover, the glottalized tone is also the lowest of all five tones in terms of pitch, so that pitch and phonation could be said to stand in a simple relationship of co-occurrence or correlation. The correspondence between pitch and phonation in this tone system is as expected from a typological point of view: higher pitch corresponding to modal (non-creaky) voice, and lower pitch (bottom values) corresponding to glottalized/creaky

voice.

In this light, one could even wonder whether glottalization needs to be granted phonological status at all. The fact that glottalization occurs on the phonetically lowest tone in Kim Thuong Muong raises the issue whether this glottalization may not be a low-level phonetic phenomenon associated to a low (or extra-low) phonological target.

But a clear answer in the negative can be provided. The data set out in Chapter 4 reveals that the phasing of glottalization in Kim Thuong Muong **Tone 4** is precise – much more precise than would be expected if this were simply a low-level phonetic phenomenon. Moreover, the association of glottalization to Kim Thuong Muong **Tone 4** is highly consistent: glottalization is a robust feature of Kim Thuong Muong **Tone 4**, also found in contexts where speakers raise their voice. Qualitative data observation suggests that the phonetic realization of glottalization is slightly different when the utterance's overall pitch is higher, with less creaky voice and more constriction, but glottalization is still present.

Furthermore, if the glottalized tone were underlyingly (phonologically) specified simply as having an extra-low pitch target, and if glottalization were only a low-level phonetic consequence of this target, then one would expect the lowest tone to be sometimes modal, sometimes creaky, *and also sometimes breathy*. Breathiness is opposite to creaky voice along the continuum of vocal-fold adduction degree, but lapse into breathy voice could be seen as a low-level phonetic consequence of a low pitch target, in the same way as creaky voice.<sup>5</sup>

Thus, glottalization phenomena in Kim Thuong Muong are consistent with the typological tendency whereby creak is associated with lower pitch, but this typological tendency does not have the power to 'predict' glottalization phenomena in Kim Thuong Muong in full phonetic detail. The phasing of glottalization inside rhymes carrying **Tone 4** does not result from low-level phonetic phenomena. The association of creaky voice to Kim Thuong Muong **Tone 4** can safely be interpreted as due to a *phonological* specification.

Said differently, pitch plays the main role in the Kim Thuong Muong tone system, but it is not the only phonologically relevant parameter.

To what extent these two acoustic properties – fundamental frequency and open quotient – contribute to the listeners' perceptual impression of the tones is a matter for future studies of speech perception to investigate. As an example of hypotheses to be tested in future work, it seems that creak with complex-repetitive patterns (to which Keating et al. refer as 'multiply pulsed') occurs frequently in canonical or hyperarticulated realizations of Kim Thuong Muong **Tone 4**. It is a safe guess that reduction to less canonical creak is frequent in casual speech – a hypothesis which will later be tested through examination of other materials, such as narratives.

---

<sup>5</sup>This argument could also be extended to Beijing Mandarin. It has often been reported that the Beijing Mandarin third tone is occasionally accompanied by creaky voice. Some authors interpret this as a low-level phonetic consequence of a low pitch target for the third tone. But if this tone's phonation type were phonologically unspecified, one would expect various types of nonmodal phonation at low pitch in Mandarin, including cases of breathy phonation. It seems that observed cases are tilted towards creak, suggesting that appealing to universal phonetic factors does not tell the full story.

#### 5.4.1 Vietnamese dialectology as a source of inspiration for the study of Muong in general and Kim Thuong Muong in particular

Kim Thuong Muong is closely related with Vietnamese from a diachronic point of view; moreover, Muong dialects have undergone wave after wave of influence from Vietnamese, strengthening similarities between these language varieties. Vietnamese dialectology, which is overall much more advanced than the study of Muong dialects, can serve as a source of inspiration for the study of Muong in general, and of Kim Thuong Muong in particular.

For instance, it appears interesting to see to what extent glottalization in Kim Thuong Muong **Tone 4** resembles glottalization phenomena in Vietnamese dialects (and in other languages of the Vietic group). In Vietnamese, tone B2 (*nặng*) and C2 (*ngã*) are glottalized tones. In a comparison between Kim Thuong Muong **Tone 4** (based on the results of this study) and Vietnamese B2 and C2 (based on the results and figures of Kirby, 2011, reproduced here as Figure 5.9),<sup>6</sup> it can be confirmed that Kim Thuong Muong **Tone 4** differs from the glottalized tones of Vietnamese by at least the following characteristics:

- (i) The position of glottalization:
  - Kim Thuong Muong **Tone 4** has *medial* glottalization: the glottalization of **Tone 4** happens in the middle part of the rhyme. The tone starts in the lowest part of the speaker's  $f_0$  range: its onset  $f_0$  is the lowest of the five tones;  $f_0$  then decreases sharply, revealing early glottalization; finally, glottalization is relaxed towards the end of the syllable, returning to  $f_0$  values suggestive of phonation mechanism I (chest voice).
  - Vietnamese B2 has final glottalization: whereas  $f_0$  curves of Kim Thuong Muong **Tone 4** can be analyzed into three portions (initial modal portion; glottalized portion; and final modal portion), Vietnamese B2 can be analyzed as consisting of two portions, with glottalization strongest at the *end* of the rhyme.
  - Vietnamese C2, similar to Kim Thuong Muong **Tone 4**, has medial glottalization, but the onset and offset  $f_0$  values of tone C2 are not at the same level, whereas those of Kim Thuong Muong **Tone 4** are close to each other. The figure from Kirby (2011) reproduced here as Figure 5.9 shows that the beginning of Vietnamese C2 is high, and its end clearly much higher, at the top of the speaker's range.
- (ii) The values of  $f_0$ :

Differences in  $f_0$  are relative, and are affected by many factors, such as differences across speakers and experiments. In order to get a feel for this difference, the evaluation of  $f_0$  values of glottalized tones need to be conducted against the background of a comparison with the other tones in the system.

- Kim Thuong Muong **Tone 4** reaches very low values: it is consistently at the bottom of the speaker's range. Sometimes it can go down to values under 30 Hz

<sup>6</sup>I also consulted (Michaud and Tuan Vu-Ngoc, 2004) and (Brunelle, Hung, and Duong, 2010) to have a good understanding of Vietnamese B2 and C2.

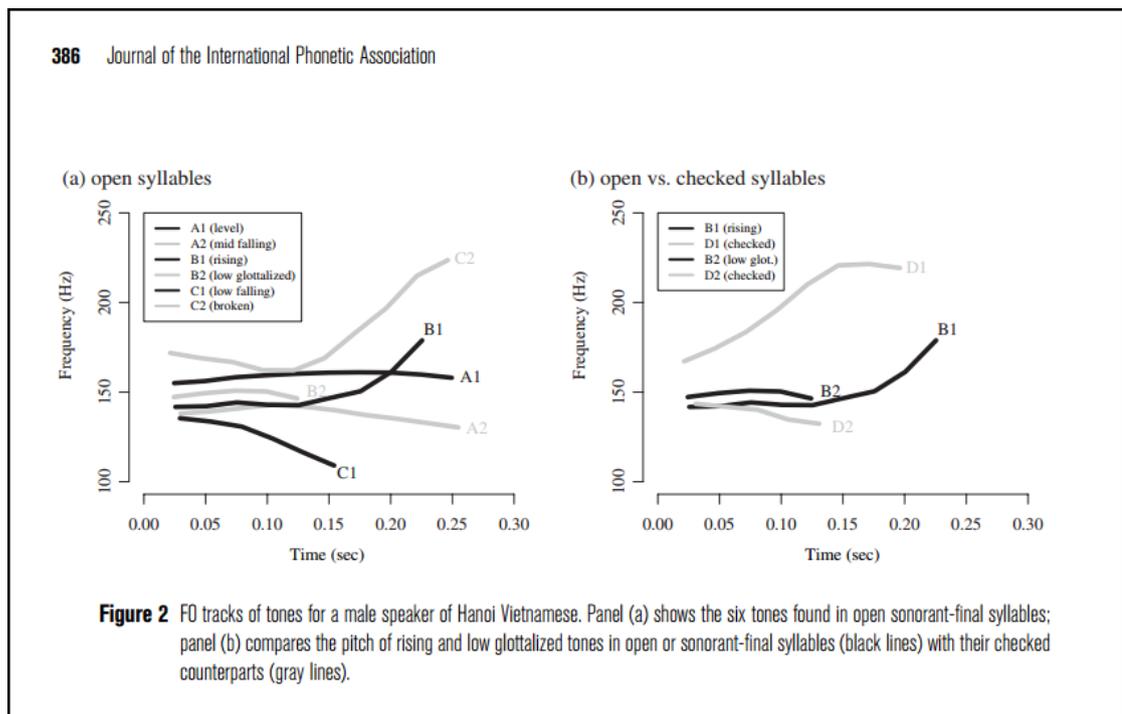


Figure 5.9: Figure of  $f_0$  curves of Hanoi Vietnamese tones, reproduced from Kirby, 2011, p. 386.

in the case of most speakers.

- Vietnamese B2 as seen on Figure 5.9 has an  $f_0$  curve in the middle of the speaker's range, lower than A1 and C2, and higher than A2 and C1. The  $f_0$  values at the beginning of the rhyme are thus clearly higher than those of Kim Thuong MuongTone 4.
- The difference between Vietnamese C2 and Kim Thuong MuongTone 4 is even greater: C2 is in the top of the speaker's range at both endpoints of its course, in contrast with Kim Thuong MuongTone 4 which is close to the bottom of the speaker's range throughout. The highest  $f_0$  point of C2 reaches up to levels never seen for Kim Thuong MuongTone 4 (at least in the type of contexts examined here: careful, hyperarticulated realizations).

(iii) Duration

- In Kim Thuong Muong, as explained above, pitch and phonation type are both important cues to distinguish between the tones, whereas duration by itself does not appear to play any major role to distinguish between the five tones of smooth rhymes. The duration of Kim Thuong MuongTone 4 (the glottalized tone) is similar to that of the four others; under the experimental conditions used here, the average duration is on the order of 25 centiseconds.
- Vietnamese B2 is *short*. James Kirby's figure of the Vietnamese tones shows that the duration of B2 is only about half that of the other tones. The author comments that "[a]lthough duration has not been shown to be a salient perceptual cue to Vietnamese tone, syllables bearing tones B2 and C1 are often shorter than syllables bearing other tones due to the effects of final glottalization" Kirby, 2011, p. 386.
- The duration of Vietnamese C2 is similar to that of Kim Thuong MuongTone 4.

Overall, although the phonation types of Vietnamese B2 and C2 and of Kim Thuong MuongTone 4 can all be described as glottalized, there are multiple differences between Kim Thuong MuongTone 4, on the one hand, and Vietnamese B2 and C2 on the other. In terms of the tones' position inside the tone system as a whole, the viewpoint defended here for Kim Thuong MuongTone 4 – that glottalization sets it apart from the other four tones of smooth syllables – could not apply to Vietnamese. Vietnamese B2 and C2 appear to have very different ties with the rest of the system. (This point need not be labored here, in view of the wealth of publications about the tone system of the Hanoi dialect of Vietnamese.)

The existence of deep differences does not mean that Vietnamese studies cannot provide inspiration for Muong studies and vice versa. The point that Northern Vietnamese tones "are realized by a complex of pitch and voice quality features" has been made eloquently by specialists of Vietnamese Kirby, 2011, p. 386; this also applies to Muong, saving specialists of Muong the trouble of laboring a point that has already been made for the closest relative of the Muong language. No less importantly, the issue of which specific phonation type (if any) is actually associated to each tone remains a topical issue in Vietnamese studies; this situation points to the difficulty of reaching definitive conclusions about which properties of a tone can be considered to be phonological, and

how the tone should ultimately be modeled.

[G]lottalization plays an important role in the production and perception of the broken (C2) and glottalized (B2) tones. The falling tones (A2, C1) have been described by some researchers as accompanied by a breathy voice quality (Thompson 1965; Phạm 2001, 2003); the low falling tone (C1) has also been described as accompanied by light final laryngealization (Nguyễn Văn Lợi & Edmondson 1998; Michaud 2004; Kirby 2010). Kirby, 2011, p. 386

These discussions about Vietnamese stimulate the search for other phonation types beside glottalization in Muong dialects. While this search did not prove fruitful for Kim Thuong Muong (although a few gender-specific trends were brought out in the Results chapter), it will definitely need to be kept in mind in future work on tone systems across Muong dialects.

An important point brought out by the comparison of Muong and Northern Vietnamese is that Kim Thuong Muong **Tone 4** is to be recognized specifically as a *creaky tone*, a label that is more specific than just *glottalized tone*. As amply exemplified above, various sub-types of glottalization are attested in realizations of **Tone 4** in Muong. The question raised here is whether this tone should be recognized as a “creaky tone” or a “glottalized tone”. Concerning the distribution of allotonic variants: in deliberate speech (tending towards hyper-articulation), **Tone 4** is canonically produced with creaky voice, although glottal constriction is also attested. This helps place Muong in typological perspective. Muong is provisionally proposed as an example of a language having a lexical tone that includes creak as part of its phonetic/phonological template. Typologically, this canonical realization appears sufficiently distinct from the glottal constriction of Northern Vietnamese tones (final in Vietnamese tone B2, medial in Vietnamese tone C2: Brunelle, D. D. Nguyễn, and K. H. Nguyễn 2010) to warrant recognition as a separate type of glottalized tone. For this reason, we suggest to call **Tone 4** as creaky tone, to distinguish it from constricted tones of Northern Vietnamese.

#### 5.4.2 Some free associations around the glottalized tone (Tone 4)

As a concluding note about **Tone 4**, this paragraph tries to leverage methods that are not part of the standard toolbox of the phonetician. To justify the use of arcane symbols like the dollar sign and the pound symbol as diacritics for High tones in Yongning Na, Alexis Michaud argued that “desperate tones call for desperate measures” (Michaud, 2017). The simile does not really hold, as there is nothing *desperate* about Muong tones, unlike Na tone: the phonological patterns are essentially as expected within this branch of Austroasiatic (whereas Na is tonally very much unlike its otherwise close sibling Naxi: Na has morpho-tonology not found in Naxi). But the presence of phonation-type specifications within tone is nonetheless a situation that stands some phonological concepts on their head. Hence, a variety of approaches can still be useful, including some “freestyle” reflections of a somewhat literary (as opposed to technical) nature. Perhaps such reflections are even more useful since there are few phonological alternations to shed light on the phonological nature of the tones.

What is **Tone 4** when all is said and done? So far, it has been referred to by a number, placing emphasis on the fact that the choice is conventional and could have been different. “**Tone 4**” means nothing specific, beyond that the tone is part of a system. How to progress beyond this label? How to look at **Tone 4** in a different way, calling it by its “real name”?

In Shakespeare’s *Romeo and Juliet*, Juliet’s passionate denial of the importance of a name (“What’s in a name? that which we call a rose / By any other name would smell as sweet”) highlights even more sharply the real (sometimes deadly) importance of names.

Tis but thy name that is my enemy; Thou art thyself, though not a Montague. What’s Montague? it is nor hand, nor foot, Nor arm, nor face, nor any other part Belonging to a man. O, be some other name! What’s in a name? that which we call a rose By any other name would smell as sweet; So Romeo would, were he not Romeo call’d, Retain that dear perfection which he owes Without that title. Romeo, doff thy name, And for that name which is no part of thee Take all myself.

Names can be very telling, if one pays sufficient attention. In the case of persons, the relationship to names is especially vital and complex, but the names of tones too can bear an important relationship to the entity for which they serve as labels. Names given to tones (as to other linguistic entities) are intended to serve as handles to a certain linguistic reality (or fiction: a social creation). Why are the tones the way they are? That is not a fair question to ask speakers (although any question that opens an opportunity for an interesting answer can look legitimate in retrospect). But that is a question that can lurk in the consciousness of speakers of any language. Looking at names, and looking for names, can be interesting in this respect: names are intended to reflect the familiar ‘feel’ of tones.

The Vietnamese language encourages us to go in this direction, as it has names for all its tones. They serve as mnemonic, being instances of the intended tone (as also in traditional Chinese phonology), but they also convey a feel for the tone’s identity. In historical Chinese phonology, tones have names, too, and these have been adduced in 21st-century research by Michel Ferlus: he uses the labels chosen by authors in medieval times to explore hypotheses about the phonetic realization of tones at the time, specifically suggesting that *qù* may have been a breathy tone (Ferlus, 2009). In this tentative exploration of Middle Chinese and its Old Chinese roots, the exercise consists in reflecting on the options among which the labels were chosen, and thereby retrieving (hypothetical) characteristics of the historical tones. For Muong, the exercise proposed here is to think about possible choices for the tones as synchronically attested. We now know what **Tone 4** sounds like, and we can choose a name for it, as one of various means to approach this phonological entity.

The names of Vietnamese glottalized tones are commonly known to speakers, from schooling. The glottalized tones are *thanh nặng* (etymological category: B2) and *thanh*

*ngã* (C<sub>2</sub>). *Nặng* literally means heavy ‘heavy’, and such is the label adopted in English-language literature about this tone (i.e. translating the Vietnamese label). How is the tone heavy? What is its heaviness? What is the ‘stuff’ that its heaviness is made of? Is the low register in which tone *nặng* terminates ‘heavy’ by itself? Arguably not: the lowest tone in the system, *huyền* does not feel heavy, although it is defined by a low register. Instead, the heaviness of tone *nặng* has connotations of *solidity*. That tone is short (hence a feeling of smallness) but has strong glottalization, which makes it feel *solid*, somewhat like a ball of iron, which is small but heavy. Clearly, the connotation of *solid weight*, and hence of heaviness, comes from glottalization, and specifically, from glottalization that goes low and ends fast. The glottal gesture presses the tone to the floor, fast, and keeps it pinned there.

As for *thanh ngã* (C<sub>1</sub>), it is often referred to in the English-language literature as *broken* (e.g. by Kirby, 2011). As a speaker of Vietnamese, this label appears counter-intuitive to me, as I do not feel that the syllable is actually broken. The tone is what makes it a complete syllable, fitting snugly inside the tone system: quite the opposite of *breaking* the rhyme, it makes it complete. Here we hit upon one of the paradoxes of glottalization: *ngã* is a phonological unit (a lexical tone) defined by *discontinuity* – by a lack of internal unity, a breach of what speakers of English, French... feel as continuity.

Interestingly, the Vietnamese language does not label this tone as ‘broken’. A handy term here would be *gãy*, which, as desired, exemplifies the intended tonal category. Instead, the Vietnamese name *ngã* means ‘to fall’. (Part of the reason why it was not translated as ‘falling’ in English is because a falling tone, in English, is a tone with a falling contour of  $f_0$ , drawing attention away from glottalization.) What kind of fall? It is not just an act of tripping, but actually a fall right to the ground (phonetically: the medial glottal constriction), followed by the act of rising up again (phonetically: the final, clearly rising part of the tone).

In Muong, colors were assigned to tones, and their feel described. Now is the time to elaborate on these intuitions, and make the most of them, also putting words on them. Ideally, the exercise would be carried out in the Muong language itself, finding words that exemplify the tonal category at issue. My command of the Muong language is still rather frail, but it is no reason not to try to play a game of *Who-am-I*, using various languages.

As a first go: **Tone 4** feels like a boomerang, in that one throws it away, it goes *out of hand* (out of modal phonation, into creak), where it has a course of its own – which feels rather long –, then it comes back in again.

**Tone 4** also feels like a sickle, due to its grating feel (again, due to creaky voice), like the feel of a sickle cutting through a handful of rice plants.

If it were a kind of stone, **Tone 4** would be lava stone: light rather than heavy (and thereby unlike Vietnamese tone B<sub>2</sub>, the ‘heavy’ tone), and rough and grating with its solidified bubbles – like the bubbles of air coming through the closed glottis in creaky voice. (Remember from C that such was the reason why a red color was selected for plotting **Tone 4**.)

In Muong, a plausible label would be **kuəj<sup>4</sup>** ‘bottom’. An advantage of this label is that it does not focus only on creaky voice, thereby avoiding a potential bias of the present dissertation: paying so much attention to phonation type as to become somewhat obsessed by it and losing sight of the broader tonal picture. ‘Bottom’ refers to the position of **Tone 4** at the bottom of the system. Thereby, this label also includes a reference to creak, which is what makes this tone a *bottom* tone, not just a *low* tone. I also like the familiar ring of ‘bottom’, suitable for a language spoken at the village by people who are not conceited and have a healthy sense of humor (including occasional overtones of self-deprecation). With this final comment, discussion of the tones is provisionally considered closed, and we move on to considerations about the interplay of tone and intonation.

## 5.5 Intonation in Kim Thuong Muong: the view from sentence-final particles

Some hints about intonation in Muong could be gleaned in the course of the study of monosyllables in a carrier sentence: even when the materials are ‘well-behaved’ tones studied in a highly controlled way, some insights into intonation can be gained, but in a fairly indirect way since the main focus is on the contrasts among lexical tones. To offer further insights into the intonation of Muong, the present section adduces evidence from sentence-final particles: the one that appears in the carrier sentence, but also others that appear in fieldwork data. Sentence-final particles are a borderline case in terms of tone, so that their study can arguably offer insights into the interplay of tone and intonation.

### 5.5.1 *Final particles and intonation in Southeast Asian languages, and the case of Vietnamese and Muong*

Various languages of Southeast Asia, including Vietnamese and Muong, make abundant use of sentence-final particles, a marginal class of expressive words indicating speech act types, evidential/epistemic nuances, and affective/emotional coloring. There are about ten sentence-final particles in Mandarin, thirty in Cantonese (Kwok, 1984), and about the same number in Vietnamese (Thi Hue Tran, 2010). Sentence-final particles are ubiquitous in casual, conversational speech. Sentence-final particles “often carry much of the meaning and function that intonation does in non-tone languages” (Chan, 1998, p. 117). The relationship is not simply one of functional equivalence between intonation and sentence-final particles, however. Studies by Seitz (1986) and Do, Thien Hường Tran, and Boulakia (1998) on Vietnamese indicate that if a final particle is removed, it is possible to compensate in part through intonation, but they also note the artificiality of the exercise. In both Muong and Vietnamese, the absence of particles in daily speech would clearly be abnormal, as particles play an important role in spontaneous speech.

Moreover, sentence-final particles also carry intonational information, since sentence-level intonational phenomena are known to cluster on sentence-final particles. One

and the same sentence-final particle can take on different nuances (creating different sense-effects) depending on the intonational realization of the sentence-final particle itself (the ‘tune’ that it carries) and of the sentence as a whole. Sentence-final particles are so tightly linked to intonation that it is not immediately obvious whether they have a tone of their own or not: sentence-final particles are bound morphemes that can only appear sentence-finally, i.e. in a position where intonational phenomena are known to abound. This makes it difficult to test experimentally for inclusion of a sentence-final particle in one given tonal category or other: tone, if present at all, seems hopelessly tangled with intonation. Despite this methodological difficulty, it now appears well-established that there exist different situations across languages in this respect.

#### 5.5.1.1 Do sentence-final particles have lexical tones?

A first-level question concerning sentence-final particles is whether they have tone at all.

Chinese languages (Sinitic languages) are extensively researched in this respect, as in many others. In Mandarin, sentence-final particles belong among toneless syllables (C. N. Li and Thompson, 1981, p. 238) – a set that also includes some other grammatical morphemes, and some final syllables within lexical disyllables. Cantonese does not have a parallel set of toneless syllables, and the status of sentence-final particles in Cantonese is less clear-cut than in Mandarin. The view that Cantonese sentence-final particles carry lexical tones is relatively widespread. However, descriptions that assign a lexical tone to Cantonese sentence-final particles only use a subset (three to five, depending on the analysis) of the six tones found on sonorant-final syllables, which constitutes evidence that Cantonese sentence-final particles are not fully integrated into the language’s system of lexical tones. Cantonese sentence-final particles have been described as possessing segmental and tonal variants, e.g. *je* vs. *jek* (Chan, 1998); Leung (2009) distinguishes two variants of *wo* (唔), *wo<sub>3</sub>* vs. *wo<sub>5</sub>*, and Ding (2013) distinguishes three. Cantonese sentence-final particles possess an especially strong expressive dimension. There have even been attempts to dissect them into a dozen of sub-syllabic semantic units: “4 initials (l, z, l/n, m), 2 rhymes (aa and o), 3 tones (1, 4, 5), 1 coda (k) and 2 such elements incorporating a tone (g<sub>3</sub> and aa<sub>4</sub>)” (Sybesma and B. Li, 2007). A phonetic comparison of Cantonese sentence-final particles with (near-)homophonous syllables in the same position within a carrier sentence (W. W. Li, 2009, p. 2293) concludes that sentence-final particles have lexical tone, in view of the fact that the *f<sub>0</sub>* curves over sentence-final particles and those of lexical words with the same (hypothesized) tone are “more similar than different”. Chance similarity between the *f<sub>0</sub>* curves of a lexical tone and of an intonation pattern realized over a sentence-final particle cannot be ruled out, however, especially for a language with as many as six lexical tones. It may therefore well be the case that the pitch of Cantonese sentence-final particles only bears the most distant relations to lexical tones: that it is essentially intonational, and to be described along a continuum of degrees instead of in terms of lexical tonal categories. All in all, Cantonese sentence-final particles constitute a borderline case, neither clearly toneless (as in Mandarin), nor clearly aligning with the language’s main tone system.

By contrast, in Vietnamese, there is a consensus that final particles possess a lexical tone, which is phonetically affected (but not phonologically erased) by intonation. All of the six tones found in smooth syllables (syllables without a final stop) are found on sentence-final particles. From a phonetic/phonological point of view, Vietnamese sentence-final particles are also much closer to the language's other morphemes than they are in Cantonese.

Several studies have been conducted about the interaction between lexical tone and intonation of final particles in Vietnamese (Ha and Grice, 2010; Brunelle, Ha, and Grice, 2012; Mac et al., 2015). The intonational change in the tone of final particles occurs through a variation in the amplitude of the lexical tone curve (expansion, or conversely compression, of the lexical tone), and through a shift in register (up or down) (Brunelle, Ha, and Grice, 2012, p. 4). These phenomena can be described within the framework of a superpositional model of prosody, in which tone and intonation are conceived as superimposed: local modulations (tones) are grafted onto a modulation that spans larger domains (intonational group, utterance).

The fact that the Vietnamese tone system incorporates phonation-type characteristics makes it easier to investigate whether sentence-final particles have a lexical tone. For tones that comprise medial or final glottalization, the systematic presence of these precise phonation-type characteristics provides strong evidence that the hypothesized (lexical) tone is present. Two examples are the particle *a*, which consistently has final glottal constriction, and the particle *đã*, which consistently has medial glottalization.<sup>7</sup>

Looking back on the above reflections, it seems that the research question *whether sentence-final particles have a lexical tone or not* leads to enter into language-specific issues that can be thorny (as in the case of Cantonese), without bringing out a clear general picture. A suggestion here is to stand back somewhat from these synchronic questions, by reformulating the research issue as: *To what extent do sentence-final particles retain a consistent lexical tone through time?* The latter question arguably helps arrive at insights into a range of languages of Southeast Asia.

#### 5.5.1.2 Are the tones of sentence-final particles diachronically stable?

In some cases, where the question whether function words such as sentence-final particles have lexical tones of their own is hard to answer, it can be useful to look into their etymology. In cases where one is lucky to find good hypotheses on etymology, it is important to look into the regularity (or lack of such) of historical developments to the present-day form.

The etymology of many of the Vietnamese sentence-final particles is straightforward. Here are two examples. The sentence-final imperative *đi* /*đi*/ is derived from the verb

<sup>7</sup>This is not to say that glottal constriction may not serve intonational functions. In German, for instance, medial glottal constriction conveys negation in vocalizations that span the phonetic range between [ʔaʔa], [ʔmʔm] and [ʔäʔä] (Kohler, 1998, p. 269). But this is of a different nature from the Vietnamese facts, where there is no iconic relationship between the function of sentence-final particles and their phonation type, whereas there is a precise coincidence between these phonation types and patterns lexically associated with specific lexical tone classes.

‘to go’, with which it is still homophonous. The morpheme *đã* /**ɗa**/, which in nonfinal position means ‘already; past’, acquires a modal/attitudinal meaning when used as a sentence-final particle: ‘Let me sleep!’ can be expressed as *Hãy để tôi ngủ đi* (EXHORTATIVE + to let + 1SG + to sleep + IMPERATIVE), with final imperative sentence-final particle, or as the stronger *Để tôi ngủ đã*, where the sentence-final particle *đã* conveys an indication that the speaker’s decision to go to sleep is not open to discussion anymore. The evolution is semantically unsurprising: the past, unlike the future, is the domain of fact and certainty, which explains how the past morpheme *đã* could give birth to a sentence-final particle that expresses a type of assertion precluding contradiction. Here, the origin of the grammaticalized morpheme is clear, and the lexical tone remains unchanged. But this case should not blind the observer to different configurations.

In various languages of the Austroasiatic, Tai-Kadai and Sino-Tibetan families, there are some function words that behave differently from content words in prosodic terms (and also in terms of vowels and consonants, but that topic will not be discussed here). These function words can be sentence-final particles, topic markers, adpositions and the like. Those function words that can be related to content words in which they diachronically originate have sometimes undergone tonal developments that do not follow the same rules as content words. This provides evidence that the prosodic properties of function words can go their own sweet way, cutting off ties to the lexical words in which they originate: sometimes losing their tone altogether, sometimes acquiring another tone. Not to mention sentence-final particles of expressive origin: the wealth of sentence-particles whose segmental composition is simply the vowel /a/ strongly suggests that those are of expressive origin.

It is against this general background that the Muong facts set out below are to be understood. It will be argued here<sup>8</sup> that the situation is one where there is less clear association of lexical tones to sentence-final particles than there is in Vietnamese.

### 5.5.2 *Tone and intonation on sentence-final particles in Muong*

Table 5.1 shows particles observed in interrogative sentences in Muong. A detailed inventory of sentence-final particles in the language (aiming at comprehensiveness) remains to be established, as part of a reference grammar of Muong.

The first question raised about these sentence-final particles is whether the notion of *boundary tones* is relevant to describing their intonation.

### 5.5.3 *Boundary tones or tone coalescence?*

In many languages, interrogation is characterized by a rising melody or the raising of a final portion of the utterance, and the suppression of the declination line (Pike, 1948; Di Cristo, 1998). Muong is no exception to this statistical trend, which is not a

---

<sup>8</sup>The argument in this section is identical in its essentials with that published in Michaud, M.-C. Nguyễn, and Scholvin (2021).

Table 5.1: Ten final particles involved in interrogative statements in Muong.

Num.	Particle in IPA	Modulation	Function	Vietnamese equivalent
1	/cǎŋ/	high-ascending, sometimes with an initial descent	polar interrogative particle	<i>không</i>
2	/wə/	high-descending, ends in breathy voice	polar interrogative particle concerning a situation in progress	<i>à, há</i>
3	/cuə/	high-ascending, sometimes with a final descent	polar interrogative particle: has the process started or not yet?	<i>chưa</i>
4	/ci/	high-ascending, sometimes with a final descent	interrogative particle 'which'	<i>gì</i>
5	/nə/	high-descending then ascending	interrogative particle 'where'	<i>đâu</i>
6	/cɤ no/	/cɤ/ : high-flat, /no/ : high-descending then rising	interrogative particle 'when'	<i>bao giờ</i>
7	/ʔǎj/	high-descending then ascending	interrogative particle 'who'	<i>ai</i>
8	/hə/	low-flat, stops on a glottal constriction (glottic closure)	interrogative particle in confirmation questions: 'isn't it'.	<i>nhì</i>
9	/re/	high-ascending	final deictic particle of an interrogative statement; expresses curiosity, surprise, emphasis	<i>thế, đấy</i>
10	/kuə/	low-flat, stops on a glottal constriction (glottic closure)	particle indicating agreement with the interlocutor	<i>phải, đúng</i>

universal (Rialland, 2007): it can be seen from Table 5.1 that most of its interrogative markers bear an ascending contour.

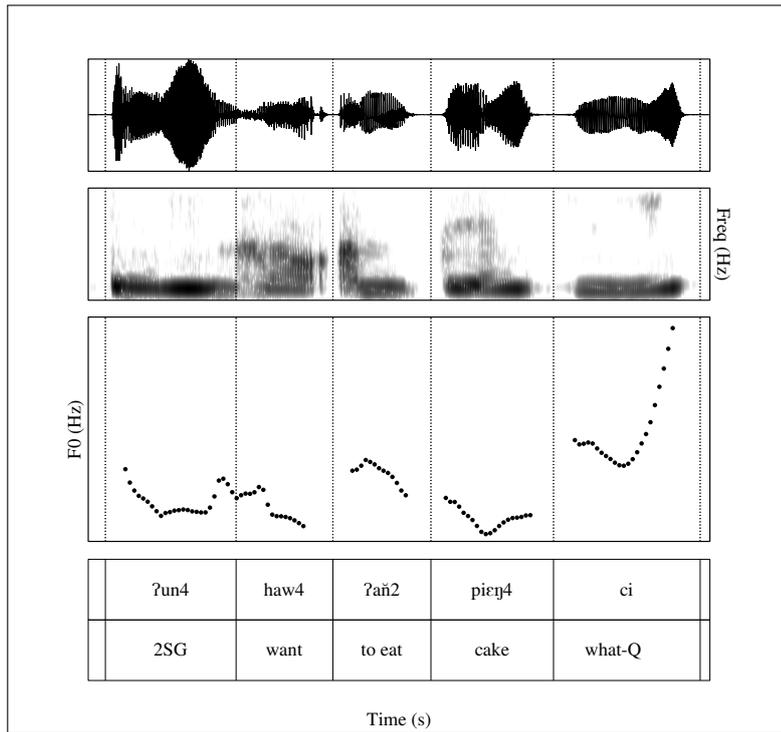
Statement 2 presents an example containing [ci], an interrogative marker that enters the composition of questions about ‘what, which, which’. The tone diacritics on the final particles provide a stylization of their pitch curve: [4] for the ascending curve of the syllables [ci4] and [re4] in Example 2 and 3, [1] for the descending curve of the syllable [ci1] in Example 3. This solution is chosen for the sake of clarity, in order to provide an indication of the prosodic realization of the particles without assigning a lexical tone to each one (which would involve an element of arbitrariness: see the warnings in Section 2.1.2). Figure 5.10 shows the phonetic realization of these two statements.

- (2) /ʔun<sup>4</sup> haw<sup>4</sup> ʔän<sup>2</sup> pieŋ<sup>4</sup> ci4/  
 2SG want to-eat cake which-INTERROG  
 ‘Which cake do you want to eat?’
- (3) /näj<sup>2</sup> kɔ<sup>4</sup> pieŋ<sup>4</sup> ci1 re1/  
 today to-have cake which-INTERROG PART.DEM  
 ‘Which cake do you have today?’

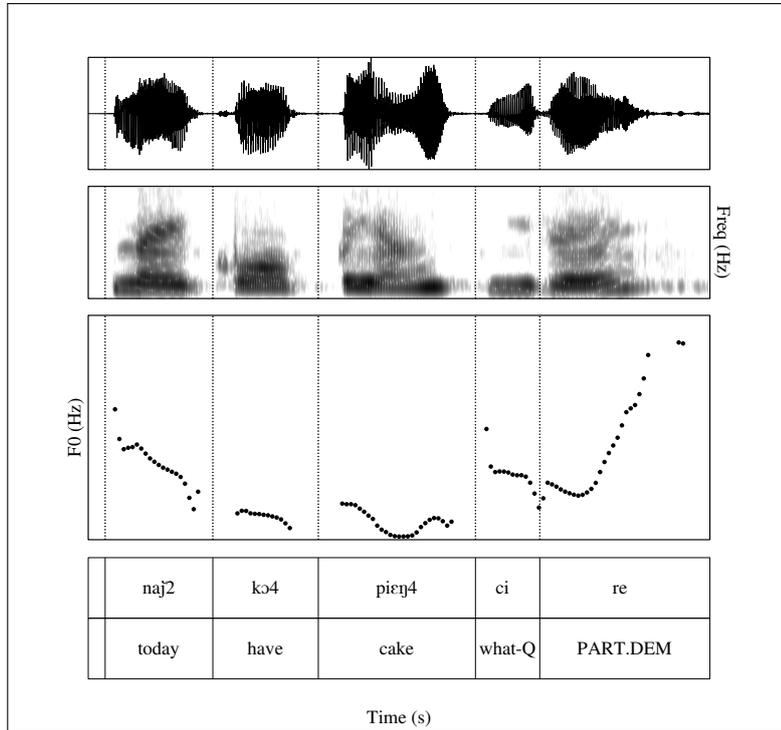
In Example 2, the interrogative particle /ci/ shows a modulation: descending, then clearly ascending at the end. If we had to assign a lexical tone to this word, it would definitely be **Tone 3**, an ascending tone – even if we have to note that the rise observed in this example is of a magnitude which clearly exceeds that of an ordinary **Tone 3**. However, when we compare Example 2 with Example 3, it is clear that it would be mistaken to attribute the final rising contour of the statement to a rising lexical tone which the morpheme /ci/ would carry. Indeed, the addition of a second particle, /re/, which is placed after the interrogative particle /ci/, has the effect of shifting the rising tone towards this second particle. The word /ci/, on the other hand, is lowered, and its duration is clearly less than in Example 2. The same phenomenon is observed for four other particles in Table 5.1: /cǎŋ/, /cuə/, /nɔ/, /ʔǎj/.

The phenomenon seems to us to lend itself to two rather divergent interpretations. One would be to consider that the final particles have no tone of their own, and that entities are freely expressed which we will describe, according to our theoretical preferences, as intonational morphemes (in the spirit of Delattre’s or Rossi’s reflections, mentioned in Section 2.1.1) or as boundary tones (Lieberman, 1975; Karlsson, House, and Svantesson, 2012): intonational effects which are concentrated at the edge of the intonational domain (in particular, at the beginning and at the end of the utterance). From this perspective, in Example 2, an interrogative intoneme (or boundary tone) is found at the end of the utterance, where it manifests itself phonetically as a strong rise. In other words, it is the interrogative intonation that takes precedence here. The differences between Example 2 and Example 3 in the realization of the interrogative morpheme /ci/ (strongly rising in Example 2, falling in Example 3) would then be explained as a direct consequence of the fact that the same intonational event is

5.5 Intonation in Kim Thuong Muong: the view from sentence-final particles



(a) Example 2



(b) Example 3

Figure 5.10: Acoustic signal, spectrogramme and fundamental frequency plot of the examples 2 and 3.

realized on a different sequence of syllables: two in Example 3, as opposed to only one in Example 2.

A second way of looking at the facts would be to imagine that a tonal reassociation phenomenon takes place. The emphatic/deictic particle /**re**/ would be present underlyingly in Example 2, but reduced to the point where it leaves only its tone, which would be reassociated with the previous morpheme: the interrogative particle /**ci**/. According to this hypothesis, the descending-ascending contour on the morpheme /**ci**/ ‘what’ in Example 2 would result from the combination (a kind of amalgam) between the two morphemes /**ci**/ and /**re**/. Thus, Example 3 would manifest the full form of which Example 2 is a reduced version. In Example 3, the interrogative morpheme /**ci**/ would manifest its lexical tone, which would be a descending tone. The striking similarity between the terminal contours of the two statements, which suggests that they are phonologically identical, would reflect the identity of the underlying lexical tone sequences.

Such processes of tonal reassociation are not as common in East Asian languages as they are in the Sub-Saharan domain, but they are nevertheless attested there (Hyman, 2007; Michaud and Xueguang, 2007; Jacques, 2011; Michaud, 2017). Tonal reassociation during syllable reduction can be considered to belong to the field of tonal sandhi, understood in a broad sense (Chen, 2000, p. 25). This type of speech reduction is familiar to us from the example of Vietnamese. When the tempo of a conversation accelerates, it is common for final particles to “go away” and only their melody remains. For example, in Example 4, the final particle can be reduced, leaving only an ascending final contour, where we recognize the mark of its lexical tone, B<sub>1</sub> (rising).

- (4) /**kaj**<sup>B<sub>1</sub></sup> **zi**<sup>A<sub>2</sub></sup>                      **dʔj**<sup>B<sub>1</sub></sup>  
 objet    what-INTERROG PART.DEM  
 ‘What is that?’

The reduction of deictic particles becomes clear when comparing the North Vietnamese and South Vietnamese dialects. North Vietnamese uses disyllabic pronouns consisting of the pronoun proper followed by the deictic *áy*: *cô áy* /**koA1 ʔʔjB1**/ ‘she’, *anh áy* /**ʔʔjA1 ʔʔjB1**/ ‘he’, *bà áy* /**ʔʔjA2 ʔʔjB1U**/ ‘she (informal)’, and so on. South Vietnamese, on the other hand, tends to reduce these forms by retaining only the pronouns, but with a change in tone (shift to C<sub>2</sub> tone). It seems reasonable to imagine that this evolution has been done because of the phonetic proximity between the pitch curve of the strongly coarticulated disyllable and that of the lexical C<sub>2</sub> tone, which, in Southern Vietnamese, has a low-ascending contour.

These observations are proposed as hypotheses: it remains to be verified by psycholinguistic means to what extent the tone of monosyllabic pronouns is well identified with the lexical tone C<sub>2</sub> ‘*thanh hỏi*’. At least the analysis seems plausible, from the point of view of the language system concerned, and also typologically. The fact that the reduced syllables are deictics is consistent with typological expectations about syllabic reduction.

To summarize, both hypotheses outlined above have their strengths. The second

one seems to us more satisfactory insofar as it brings the intonation facts, which are notoriously difficult to analyze, back to the field of mechanisms that are familiar to us: coarticulation and syllabic reduction. However, the first hypothesis is not without utility in the perspective of a dynamic synchronicity (André Martinet, 1990) sensitive to the changes of which the current situation bears the seeds. Indeed, the final intonation of Examples 2 and 3 could break the ties with its segmental origins (assuming that these are confirmed). Rather than being drawn into the tonal system and entering a pre-existing lexical category (such as South Vietnamese pronouns, which are in the C2 tonal category), it could be perpetuated as an intoneme, or boundary tone, associated with interrogation.

One point that seems important for this discussion is that the state described in Table 5.1 has variants that seem quite distinct. Looking at the descending-ascending curve in Example 2, we are inclined to the second hypothesis, while the simple final rise that constitutes the other variant of the interrogative contours makes us more inclined to opt for the first hypothesis. This type of hesitation is typical of “on the edge” situations, which inform about the permeabilities that can exist between tonal and intonational systems. It is particularly important not to be locked into a model that excludes *a priori* the existence of such interactions. Just as in biology, well-established certainties have delayed the recognition of the existence of certain mechanisms of information transfer between cells (Prochiantz and Joliot, 2003), so a phonological model that would compartmentalize tones and intonation in a watertight manner (or, conversely, represent them in an entirely homogeneous manner) would impede in-depth observation of the type of phenomena exemplified by Muong intonation.

The following paragraph is meant to be another touch on the subject. It deals with one of the most delicate points of intonation modeling: the question of intonation tones.

#### 5.5.4 An intonational tone in Muong?

In the majority of cases, the interrogative intonation in Muong is realized by a strong rise on the last words of the sentence, especially on the final particles. Nevertheless, there is also another case (less frequent, but clearly and systematically attested) in which the question ends on a final particle realized with a descending  $f_0$ .<sup>9</sup> We refer again to Table 5.1 to see that the second particle, /wə/, and the eighth, /hɔ/, belong to this second type. We will focus here on the particle /hɔ/, used to ask for confirmation or agreement. This particle is used either at the end of a sentence or alone in response to a question, as illustrated in Example 5.

- (5) /ha<sup>2</sup> nəj<sup>2</sup> rǎŋ<sup>4</sup> hɔ-ʔ/  
 today sunny right-INTERROG.PTCL  
 ‘It’s sunny today, isn’t it?’

<sup>9</sup>The observations in this paragraph were reported in a journal article (in French): Michaud, M.-C. Nguyễn, and Scholvin 2021.

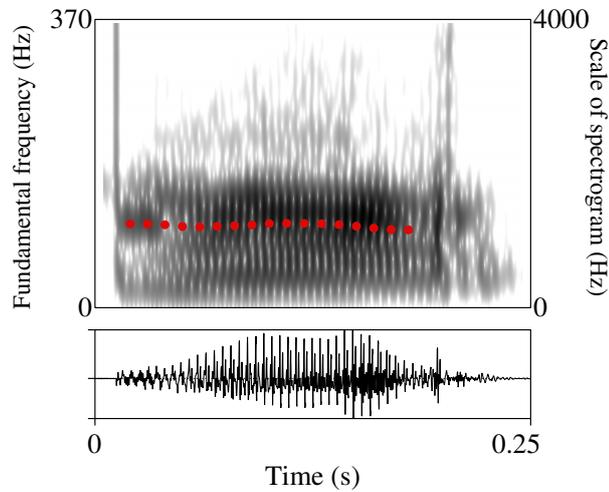


Figure 5.11: The final particle /hɔ̌/, taken from a dialogue prepared before recording.

- (6) /kuəɰʔ/  
 right-PTCL  
 Response: ‘right, absolutely!’

The final particle /hɔ̌/ in Example 5 attracts attention in that it carries a pitch contour which is distinct from all the lexical tonal categories in the system. To the ear, its prosodic realization has a striking proximity to the glottalized Vietnamese B2 tone (the *nặng* tone of the Vietnamese orthography). They share a strong final glottal constriction. However, the Vietnamese B2 tone is, in its canonical realization, a descending tone, due to the fact that it starts from fundamental frequency values that are not low to begin with (Brunelle, D. D. Nguyễn, and K. H. Nguyễn, 2010), while the  $f_0$  contour of the /hɔ̌/ particle is plain (flat), as illustrated in Figure 5.11.

If we look for the tone closest to (or rather, least distant from) this contour among the lexical tones of Muong, we must look for tones 1 to 5, which can appear on a syllable without a final occlusion, such as /hɔ̌/. Among these five tones, glottalization is only present in **Tone 4**. However, because of the descending-ascending modulation and the passage in creaky voice (phonation mechanism zero) that characterize it, **Tone 4** presents a very different appearance from the “tone” of the final particle /hɔ̌/. Moreover, this “tone” only appears in two particles, both represented in Examples 5 and 6, one in the question (the particle /hɔ̌/, just mentioned), the other in the answer: the particle /kuəɰʔ/, which, used alone, marks assent. It is therefore hardly possible to speak of an eighth tone of the system. In intuitive terms, it seems that a tonal niche has been created: a specific communicative function (marking strong assent) has availed itself of a phonetic-phonological pattern introduced through North Vietnamese (of which all Muong speakers are speakers to varying degrees): that of Vietnamese tone B2. It

would seem that Muong speakers are no less sensitive than others (native speakers of European languages, e.g. French or German) to the auditory impression produced by the strong glottal constriction present in two of the tones of North Vietnamese, and that they have borrowed this pattern – that of a Vietnamese tone – to make an intoneme out of it (an “intonational tone”). In terms of communication, this intonational tone is all the more univocal as its use is restricted. Without venturing to predict its future fortune, in a context of gradual replacement of Muong by Vietnamese, it is conceivable that this intonational tone will come to be used on an increased number of morphemes, while retaining (at least initially) the semantics currently associated with it.

This delicate pattern appeared as an apt final flourish for the present chapter, as it highlights the fact that glottalization is not only a salient and interesting phenomenon in itself: it also opens up a potential for interaction with other characteristics of linguistic systems. This final paragraph thus illustrates how glottalization, woven together with intonation, links up with the broader topic of creativity – phenomena that arise in language contact (in this instance, Muong and Northern Vietnamese), and in speech communication generally.



## Chapter 6

---

### Conclusion and perspectives

... he had a private instinct that a proof once established is better left so.

---

Mark Twain, *The Man That Corrupted Hadleyburg*, 1899, cited in Garellek, Gordon, et al. (2020)

In the present study, experimental phonetic means were deployed to gain a better understanding of phonetic dimensions that are known to be important cues to linguistic tone in an East/Southeast Asian context: fundamental frequency, which has pitch as its perceptual counterpart, and glottal open quotient, a parameter that relates to phonation types. A special focus of this thesis consisted in taking up the challenge of studying in detail the phonation of a glottalized tone.

On a methodological note, the study confirms the usefulness and feasibility of the use of electroglottography as part of the experimental setup in phonetic studies of tone in a fieldwork setting. Challenges such as the overall more noisy signals (lower signal-to-noise ratio) for female speakers did not prove insuperable in creating, in the field, a data set that holds up to professional standards in experimental phonetics.

An essential part of this final chapter is to provide a brief and clear recapitulation of the answers to the research questions formulated in the first chapter.

- Concerning the first question: “What is the role of the creaky voice in Kim Thuong Muong’s tone system?” It can confidently be concluded that although creak is cross-linguistically associated with lower pitch, this does not have the power to ‘predict’ glottalization phenomena in Kim Thuong Muong in full phonetic detail. The phasing of glottalization inside rhymes carrying **Tone 4** does not result from low-level phonetic phenomena. The association of creaky voice to Kim Thuong Muong **Tone 4** can safely be interpreted as due to a phonological specification.
- Concerning the second question: “To what extent is there phonetic variation in the glottalized tone?” It has been demonstrated in this thesis that glottalization in Kim Thuong Muong **Tone 4** has creaky voice as its canonical realization and glottal constriction as a variant. The sub-classification of types of creaky voice found in the implementation of this tone was initiated in this study by qualitative observations, with the help of a simple script, CreakDet. Bringing out

the conditioning of the variants constitutes a program which I intend to carry out in future work.

Looking back to take stock of the results, I am most aware of all that remains to be done. In particular: conducting state-of-the-art statistical analysis of the extracted parameters. The parameters extracted from the electroglottographic signal ( $f_{0\_dEGG}$  and  $O_{q\_dEGG}$ ) were only used here for descriptive statistics and visual display: in-depth statistical analysis remains to be carried out, teasing out the contribution of various parameters including (i) lexical tones, (ii) experimental conditions and (iii) speakers, of course, but also (iv) segmental contexts (intrinsic properties of vowels and co-intrinsic properties of initial consonants), which well deserve attention, not least due to their well-documented importance in tonogenetic processes. Thanks to the help of Solange Rossato (Laboratoire d'informatique de Grenoble, UMR 5712), some basic descriptive statistics are already in progress and will be made public in the upcoming work after this study.

Two obvious directions in which to supplement the present work would be to conduct perception studies,<sup>1</sup> and to relate audio and electroglottographic data to physiological observations and modeling.<sup>2</sup> These research directions constitute a captivating research program for the mid term.

There has nonetheless been some progress. The documentation of the Kim Thuong Muong tone system now reaches a stage where fundamental tonemic issues can be considered solved. Phonological and phonetic questions pertaining to tonation can be explored based on a reasonably rich and diversified corpus that contains vocabulary, narratives, conversations and songs in addition to phonetic/phonological experiments. Beyond the results reported in the present volume, this set of data holds potential for further studies in experimental phonetics, and may also prove a useful resource for other fields of linguistic research.

---

<sup>1</sup>Work on the perception of tonal contrasts in Southeast Asia has developed over recent decades, partly thanks to digital technology facilitating signal processing and perception tests. Among specialists in this field, I am aware of the work of Burnham and Francis (1997), Svantesson and House (2006), Abramson and Tingsabath (1999) and Abramson, Nye, and Luangthongkum (2007), Zsiga and Nitisaraj (2007), Brunelle and Jannedy (2007), Brunelle (2009b), Brunelle and Finkeldey (2011), Brunelle (2012), and Brunelle and Kirby (2016), Kirby (2010) and Kirby (2014), and Gruber (2011).

<sup>2</sup>Obtaining images of the larynx during phonation, through laryngography, is a desirable complement to electroglottography, already used in pioneering research (Esling, 1984; Brunelle, Hung, and Duong, 2010).

# Appendices



## Appendix A

---

### Appendix 1: Challenges and strategies of phonetic experimental study in an unwritten language in the context of a bilingual community in the field

#### A.1 Data collection software vs. hands-on monitoring of recording sessions by the investigator

“The second half of the 20th century was the dawn of information technology; and we now live in the digital age” (Niebuhr and Michaud, 2015, p. 1). Research centres in phonetics are at the vanguard of technological progress, and data collection procedures now routinely rely on specialized software. Thus, SpeechRecorder, a platform-independent audio recording software, is customized to the requirements of speech recordings. Its features make it extremely appealing to phoneticians who need to record a dataset under carefully controlled experimental conditions. It allows for a range of prompts: text, image, audio or even video prompts can be used to elicit speech in many different ways. XML-formatted recording scripts allow for a flexible organization of recording sessions, including options for automatic recordings and randomized prompt selection. Configurable speaker and experimenter screens offer an uncluttered and appealing graphical interface, as shown in Figure A.1.

No wonder that this piece of software, carefully designed by programmers working hand in hand with phoneticians, is a great success. The tool, which dates back to the beginning of this century (Draxler and Jansch, 2004), reached version 6 in 2020. Clearly, the way forward is in adopting such tools, which greatly facilitate data acquisition.

There is, however, one aspect in which the use of data acquisition software does not necessarily constitute an off-the-shelf solution: it assumes familiarity with digital tools and also (implicitly) with experimental protocols. Digital literacy and knowledge of data collection procedures are unevenly distributed among the population, and they are also unevenly distributed across the world map. Full success with SpeechRecorder and similar software can confidently be expected from most people who participate in data collection campaigns at phonetics institutes. While I am not aware of any study of the general demographic statistics of participants in data collection campaigns at phonetics laboratories worldwide, it seems likely that a fair share received college-level education.

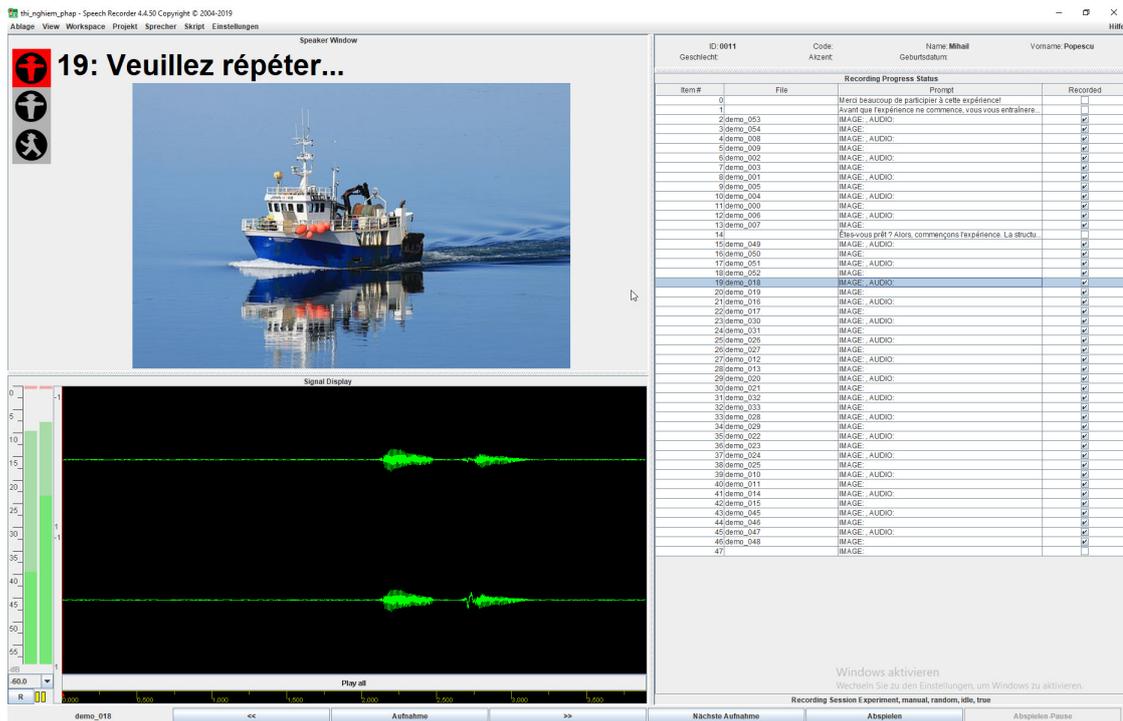


Figure A.1: Screenshot of the SpeechRecorder software from Vera Scholvin’s doctoral study on “Prosody in French-Vietnamese Language Contact”, with permission.

The picture of village life in Vietnam looks widely different, raising important issues in terms of experimental design. It looks as if SpeechRecorder provided coverage for any language, since the tool offers “Unicode text to display most of the world’s languages”<sup>1</sup> and audio prompts can be used for languages that are unwritten. But interacting with a computer did not seem the right way to go for my data collection campaign, given the lack of prior familiarity of Muong consultants with computer interfaces. The consultants’ understanding of the task, and their interpretation of the investigator’s purposes and intentions, are of paramount importance to the data collection process, and deserve much attention. Exploratory work led to adoption of a custom, ‘handcrafted’ protocol, which is somewhat at variance with procedures that are currently standard in many laboratory settings, but which I believe is true in spirit to the essentials of scientific research.

### A.1.1 Learning from an earlier pilot study: difficulties of operating at a phonemic level

To arrive at perfect symmetry of a corpus, which is highly desirable for the systematic study of individual parameters and their interplay, it is tempting to bypass the com-

<sup>1</sup><https://www.bas.uni-muenchen.de/Bas/software/speechrecorder/>

plexities of the lexicon of the target language: to study phonological factors on their own. The use of pseudo-words (sometimes called logatoms) allows the investigator to focus on phoneme-level phenomena, building higher-level phonological units from the inventory of phonemes in the target language. A list of items is generated by exploring phonotactic combinations in a systematic way. This device is widely used in studies of Vietnamese (Earle, 1975; Gsell, 1980; Michaud, 2004b) as well as in other languages (for English: Dollaghan, Biber, and Campbell 1993; Michael S. Vitevitch et al. 1997b; Goswami, Gombert, and Barrera 1998; Hay, Drager, and Thomas 2013a).

However, the use of forms not directly related to a meaning (that of a lexical item) raises difficulties in an unwritten language because there is no way to illustrate phonemic combinations to the consultants without building some sort of transcription system, adapting the writing system of another language, with the risk of phonological interferences. In the case of Muong, the obvious choice is the writing system with which the speakers are familiar through schooling, namely Vietnamese writing. This involves a number of potential pitfalls. The Muong people share with other communities of speakers of unwritten languages the characteristic that they (overall) speak their language with very little awareness of its phonemic system. On the other hand, the schooling that they receive in Vietnamese includes grapho-phonemic drills in Vietnamese spelling – a writing system which encapsulates various peculiarities that date back to pioneer 17th-century work by European linguists with Romance backgrounds (André-Georges Haudricourt, 2010). Therefore, code-switching between the sound systems of Muong and Vietnamese, and perhaps blending both in various proportions, are extremely difficult to avoid when using for Muong a Vietnamese-based phonemic transcription system.

Here as elsewhere in this study, the approach taken is an empirical one: if non-words allow for a successful investigation, then I am happy to use them. But a pilot study reported in my M.A. thesis (M.-C. Nguyễn, 2016, p. 22) with syllables some of which do not constitute real morphemes (words) led to the clear conclusion that this method was not applicable given the fieldwork settings. This pilot study seems worth recapitulating here *in extenso* in order to shed full light on this important topic.

The experiment using non-words (i.e., designed as a phonemic exercise) was referred to in the M.A. thesis as “Experiment 2”. It was designed as a supplement to “Experiment 1”, which was based on real words: recording two minimal sets built on the basis of the word list available at the time. In order to gain insights into intrinsic and co-intrinsic effects on fundamental frequency (coarticulatory effects of rhymes and initials on  $f_0$ ), it seemed worth trying out whether a reading experiment could yield interesting results despite the obvious difficulties associated with conducting a reading task in an unwritten language.

In term of vowels, I selected maximally different vowels, therefore choosing /i/, /a/ and /u/: a high front vowel, /i/; a low vowel, /a/; and a high back (rounded) vowel, /u/. As for consonants, unvoiced stops appeared as a convenient set for ease of segmentation, so I selected three unvoiced plosives: /p/, /t/ and /k/ (note that Muong, unlike Vietnamese, has /p/ in vocabulary of Vietic stock). It appeared interesting to have two series at similar places of articulation, allowing for a comparison of nasals

and stops, so I added the three corresponding nasals, /**m**/, /**n**/ and /**ŋ**/. (The two glottalized consonants /**ʔ**/ and /**ɗ**/ were excluded because they are expected to have strong coarticulatory effects on phonation type – a central topic in the study of tone. Aspirated /**tʰ**/ is alone in its series and was therefore left out too.) In a further effort to collect a set of data that would shed light on the full tone system of Kim Thuong Muong, I included, along with smooth syllables (syllables ending in a vowel or a nasal consonant, and carrying one of the tones from Tone 1 to Tone 5), some stopped syllables: syllables ending with a stop, and carrying tones 6 and 7 (the two tones of stopped syllables). In Kim Thuong Muong, there are four possibilities for a final consonant in a stopped syllable, namely /-**p**/, /-**t**/, /-**c**/ and /-**k**/. A quick check in the word list showed that, in the initial list available at the time, which contained 620 words, /-**k**/ was the most frequent final consonant (occurring in 44 words), the second most frequent being /-**t**/ (20 tokens), the third /-**p**/ (13 tokens), and the fourth and last /-**c**/ (6 tokens). Therefore, in terms of lexical frequency, /-**k**/ comes first. On the other hand, /-**k**/ has a strong identity in terms of coarticulation; in Vietnamese and Muong, this is reflected in entrenched patterns of coarticulation within the rhyme. In Hanoi Vietnamese, the rounding of vowels /**ɔ u o**/ moves onto final velar consonants; the phenomenon only affects velars. In other words, among the six final consonants /-**p**, -**t**, -**k**, -**m**, -**n**, -**ŋ**/ of Hanoi Vietnamese, the velar consonants /**k**/ and /**ŋ**/ have special patterns of coarticulation. This is not only the case after rounded vowels. Haudricourt in his study of 1952 showed that Vietnamese has a phenomenon of shortening of vowels in front of velar consonants, and of palatalization of the final after a front vowel, such as the realization of /**ɛk**/ as [**ɛc**], and /**ik**/ as [**ic**] (André-Georges Haudricourt, 1952). My own observations suggest that /-**k**/ in Kim Thuong Muong likewise tends to shorten a preceding vowel, so /-**k**/ was not selected for this experiment despite its status as the most commonly occurring final consonant.

In terms of formant movement, the final consonant that is likely to have the simplest coarticulatory influences is /**p**/: it lowers (slightly) the formants, especially at the transition from the vowel to the consonant, while /**t**/ has a stronger effect on the second formant (which tends towards a target of roughly 1,800 Hz) which is especially salient after back vowels, for example in the sequence /**ut**/, where a dramatic rise in the second formant can be found. In addition, /-**p**/, as a labial consonant, is the only consonant that does not conflict directly with the tongue position of the vowels. For those reasons, I chose /-**p**/ as a final consonant, hoping for data that would be easier to interpret.

To sum up, the final set of phonemes for creating target words consisted of: six initial consonants /**p**-, **t**-, **k**-, **m**-, **n**-, **ŋ**-, three nuclear vowels /**i**, **a**, **u**/, and final /-**p**/ for stopped syllables. Figure A.2 shows the corresponding cards.

The corpus of Experiment 2 was made up of the combinations of the above phonemes with 5 + 2 tones: 5 tones on smooth syllables and 2 tones on stopped syllables. Hence, we had  $[(6 \times 3) \times 5 + (6 \times 3 \times 1) \times 2] = 126$  syllables in total (90 tokens of smooth syllables, and 36 tokens of stopped syllables). Each syllable was said twice, hence the total number of tokens was  $126 \times 2 = 252$ .

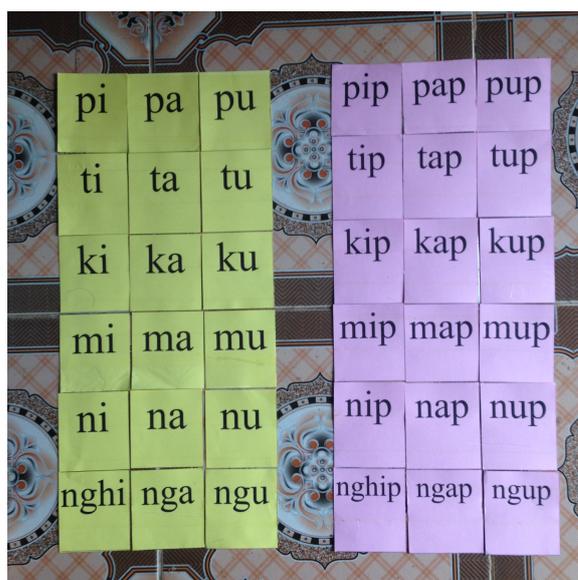


Figure A.2: Experiment 2: Cards for target words. Yellow cards for smooth syllables and pink cards for stopped syllables.

An important issue when preparing the stimuli was the order in which tones are recorded. Explanations about tones are not the easiest part when using written stimuli, because the proposed system has to be learned entirely: to avoid interferences with Vietnamese, the orthographic representations used for Vietnamese tones are not used for Kim Thuong Muong; the tone symbols chosen are numbers, a convention which does not exist in many writing systems, and which is not used in Vietnamese, the system with which the speakers are already familiar. Given the difficulty of cueing a tone, it seemed simpler at first to divide the data into seven sets, arranging them by tone: eliciting all the syllables that have Tone 1, then all those that have Tone 2, and so on. This is easier for the consultants, as they keep tone constant while the syllable's initial and rhyme change. However, a drawback of this procedure is that the consultants' attention is drawn to the segments (consonants and vowels), which change from one syllable to the next, and are thus in focus, whereas tone is constant from one syllable to the next within each of the 7 tonal sets. Tone can therefore tend to become less and less clearly articulated, because the ordering of items makes it a less contrastive phonological component of the syllable. Since the focus of the research was on tone, what was wanted was 'canonical' distinction: the fullest realization of the tonal templates. So it seemed advisable to change tone systematically from one token to the next. This solution was implemented in the final set of recordings. Tones were cued by displaying a pane with four images for each tone illustrating words that bear this tone: see Figure A.3.

An obvious drawback is that the consultant's experience of reading and writing is intimately linked with their practice of the Vietnamese language. It is possible to give



Figure A.3: Experiment 2: Cards used in an attempt to cue tones.

consultants some training in a writing system adapted from Vietnamese to Muong, and to give them the instruction to speak in Muong and not in Vietnamese. But despite my best efforts, under this setup, ‘classroom’ behavior surfaced at various points: getting back into the groove of the sound system of Vietnamese. As mentioned above, all consultants had received training in Vietnamese writing through primary schooling, whereas none had received training in Muong writing: neither at school nor later.

The obstacle was found to be insuperable. The sound systems of Vietnamese and Muong simply could not be efficiently kept apart. The failure of the non-words experiment in my master study led to the choice made here to use only attested words. Later in chapters of result and discussion, this topic will be taken up again, contrasting the present study’s complete results with my previous work. To preview the results, they confirm that there is a gaping difference between the “read” materials of “Experiment 2” and those elicited later by means of more ecologically valid methods: those classically employed in fieldwork on unwritten languages, to which we return in the next section.

Trying to summarize in a few words what is specific about phonetic data collection in an immersion fieldwork context, I would highlight two points. First, for the sake of data homogeneity, one has to exert constant attention to ensure that the consultants are comfortable and focused. Second, it is advisable to cast the net wide and record more types of speech data than is strictly necessary for the specific research goal at hand, as it is likely to be difficult (costly in funding, time and effort) to gather additional data from the same speakers later as the necessity for it arises. Thus, in this study, the focus is on the characteristics, interactions, and variety in the realizations of a phonologically glottalized lexical tone. The more spontaneous and natural the speech, the more challenging it is for the experimenter to tease out the influence of various factors, so my first focus was on sets of words that minimally contrast for tone (recording words in isolation and in a carrier sentence). But the ultimate goal is to understand how acts of communication work in the real-world diversity of contexts, so I also recorded narratives, conversations and other data which, in phoneticians’ typology, fall into the wide category of “spontaneous speech”. Even though exploitation of the latter part of the data set has barely begun, it provided a useful point of reference from the start, and I consult it regularly over the months and years. Thinking early and often about possible hypotheses for later work allowed me to look beyond carefully controlled materials, to more complex data, like someone studying human gait who processes laboratory data on walking but likes to observe the complex patterns of various styles of dancing, and to think of connections. Researchers working on the world’s top 5% best-documented and most widely spoken languages (as a rule of thumb: about 300 languages, including, as a matter of course, most national languages) can do this easily, by serendipitous observation of data that are not part of their experimental data set: language is everywhere. But when working on a newly-documented language variety, as is the case here, one has to record data by oneself.

The discussion which follows bears on the reasons for choosing specific types of materials, the selection strategies, their advantages and drawbacks, and ways to overcome

some limitations, as well as the method that was found to be applicable for Kim Thuong Muong. It is hoped that some of these reflections also hold for other unwritten languages, striking a hopeful note for phonetic experiments in the field.

## **A.2 Strategy for finding (near-) minimal sets in an unwritten language**

On the basis of the experiment involving real words designed in 2016, the strategy going forward was clear: extending the same experiment by finding additional minimal sets, and selecting pictures (photos) as stimuli for the data collection task. To be more specific, the process of finding (near-) minimal sets and pairs can be described in terms of the four following steps.

**Step 1: Vocabulary collection.** This is an indispensable prerequisite for finding (near-) minimal sets. Thanks to [the EFEO-CNRS-SOAS word list](#) which was designed for conducting in-depth lexical investigation when doing fieldwork on languages of Southeast Asia, we enjoy the benefits of a good tool to collect a good amount of basic vocabulary of Kim Thuong Muong. Due to time constraints in 2016 (when carrying out fieldwork as part of the master's degree), only 600 items from this word list were recorded. This amount of vocabulary was already sufficient to find a couple of tonal minimal sets. One of the goals of the current study was to extend the list of minimal sets and to add minimal pairs consisting of checked syllables. The more vocabulary we have, the more possibilities there are to find such sets (not to mention the obvious usefulness of a well-groomed word list, with the ultimate aim of publishing a full-fledged dictionary). Therefore, during this doctoral project, I availed myself of the additional field trips and favourable conditions for scientific work to elicit the full list: about 2,900 words. An enriched basis was thus available for tasks such as the search for (near-) minimal sets and pairs.

**Step 2: Transcription in IPA.** It is highly recommended to make the transcription in International Phonetic Alphabet immediately during the recording sessions. Thus we can easily check with the consultant(s) and make sure that every phonemic transcription is correct. If, for any reason, it cannot be done during the recording, it should at least be done during the field trip, rather than taking untranscribed recordings back to the laboratory. It certainly makes for a more reliable workflow, and it saves time and effort because verifications are much easier during fieldwork than later. In fieldwork we are always surrounded by locals who can help us verify. For instance, during the recording session of the vocabulary collection, I ran into a difficulty in perceiving the differences between two level tones (**Tone 1** and **Tone 5**), in contexts where they are co-articulated with other tones: in particular, in compound words. Therefore, I usually had to double-check with the consultant by asking her to repeat the word several times slower, saying the two syllables of the compound word separately.

The word list that I used for this step is available online in Open Office format (.ods) and MS-Excel format. I worked in the Excel file, for no better reason than because that was the format I had used from the start, in the framework of the [DO-RE-MI-FA](#)

project which aimed to digitize the audio recording collections of Michel Ferlus, many of which rely on this word list.

The search for tonal minimal sets was much facilitated by the work previously carried out on the language's phonological system. The essentials of the system of phonemes and tones had been worked out since my 2016 study. The background tasks of continuing vocabulary collection and double-checking transcriptions for accuracy and consistency went smoothly, and did not bring up any significant changes to the analyses produced at the time.

**Step 3: The separation of phonemic components.** After recording the entire word list and inputting them into the Excel file, the next step is to separate and classify each minimum phonemic component of each word into columns, in terms of: initial consonant, glide, vowel, final consonant, tone. This step serves at least three purposes.

- Firstly, we can easily examine the phonemic system from various angles, such as: listing the vowels (mono-vowels or diphthongs), the consonants, the position(s) they can occur in, and their possible combinations.
- Secondly, we can consider systematically (by the function of sorting and filtering on Excel) any uncertain contrasts and double-check to make sure whether they are distinct phonemes, allophonic variations, or just free variations. In this study, an outstanding puzzle was whether there exists a distinction between the alveolar trill /r/ and the alveolar tap/ flap /r/, or between the alveolar fricative /z/ and the palatal fricative /j/. By gathering all the puzzling words and applying the method just mentioned, we have identified them as free variations, not distinct morphemes.
- Thirdly, using the same Excel function of sorting and filtering, one may be lucky to find some minimal sets already nice and complete. Obviously, that depends on the amount of collected vocabulary. Otherwise, it is at least possible to find tonal minimal pairs or incomplete minimal sets. This is the material for the last step.

**Step 4: Supplementing incomplete sets by searching for missing words.** A complete minimal set for this study contains 5 words that distinguish 5 tones (on a smooth syllable). As a starting-point towards this goal, we searched and picked incomplete sets that have at least three target words (out of the desired total of five). It would be possible to start from a minimal pair (2/5) or even from a single word (1/5) and search for new items to supplement it, but that is highly impractical. Forcibly searching for lexical items containing a desired combination of segments and tones is tedious. It threatens to feel like a never-ending and disorienting work for consultants.

First, I prepared the incomplete sets in advance: listing them in order to know which tones are already available and which tones are missing in the set. The ideal setup for this step is that there are at least two native consultants who can help at the same time, supporting each other in the task of examining and supplementing the minimal sets. In this study, I took advantage of the fact that I was staying in a local family. I asked for help from a family member. A small financial reward was associated with participation in the task, which helped clarifying that the task was to be taken seriously. In addition, I asked other members of the family to come around to give extra help.

As an example of incomplete set, here is one of the four minimal sets in Table 3.1, in which we had already found three target words from the vocabulary list, namely:

1. /**la**j<sup>5</sup>/ “tongue”
2. /**la**j<sup>3</sup>/ “to return”
3. /**la**j<sup>4</sup>/ “to drive”

We first asked the consultant to confirm these three words, in order to verify if they really belong to the set, and also to clarify to the consultant what was the intended segmental make-up of the syllable for which we needed to find other tones. After that, we tried to pronounce the missing word(s) by combining the syllable with each missing tone, asking whether or not there exists such a word. She was able to find items with the two desired remaining tones:

1. /**la**j<sup>1</sup>/, the first syllable of the compound word /**la**j<sup>1</sup> tɔ̃m<sup>5</sup>/ which is “a bamboo fence to block lake drains to prevent fish from escaping through the gutter” (as shown in Figure A.4c)
2. /**la**j<sup>2</sup>/, which means “to carry stuff or people on motorcycle or bicycle (a two-wheeled vehicle)”.

In cases where she could not find the words with the missing tones because they don't exist in Kim Thuong Muong, we switched to the solution of the near-minimal set. The rhyme and the tone are kept while the first consonant is changed with other consonants that have the same articulation mode and a place of articulation close to that of the reference consonant. The sets from the 9<sup>th</sup> to the 12<sup>th</sup> in Table 3.1 have been found this way.

### A.3 Using illustrative photos to stimulate the target word: an appropriate method in unwritten language

Concerning the method of production, reading aloud the speech materials appears to be one of the most frequent choices because it is easy to implement and ensures consistency and accuracy of the data. However, Muong has no written form, hence reading was not an option. Indeed, in the previous study, we tried to borrow the Vietnamese alphabet for elicitation, as speakers of Muong are also proficient in this national language, and the failure of that experiment (despite various efforts) showed that it was clearly not the way to go. Vietnamese is so similar to Muong (maybe somewhat like German and Dutch, *mutatis mutandis*) that there is, in our experience, a high risk of interference between languages, creating experimental bias. To overcome this challenges, elicitation by means of photos appeared as the best choice.

A strength of this method is that it increases the naturalness of the speech data, at least more than the reading method. Indeed, an inherent characteristic of reading a text is that people tend to speak more formally and articulate more carefully than when they are involved in a free conversation, which is good for phonological exploration but may eliminate some interesting phonetic variations. This method also allows for a relative control of tempo, avoiding the changes in speech tempo (i.e. speeding up or slowing down) that frequently occur in spontaneous speech (for a review, see Niebuhr

and Michaud, 2015). On the other hand, a drawback of this method is that the speakers might be distracted by a memory task. As mentioned in the section of the word lists (3.1.1), some words are more abstract than others, for which a direct illustration cannot be found. Therefore, these words must be explained in detail to the speaker at the training stage and required to memorize their implication. This may compete with the requirements of good phonetic realization: it is expected that consultants will be able to focus on a clear phonetic realization of the tone. However, if some items are more difficult to remember than others, a sudden recall after a moment of hesitation (which may be awkward or stressful for the speakers) may be accompanied by a different phonetic realization. This is not to say, of course, that the task is beyond the speakers' abilities. Indeed, this experiment was carried out without great challenges with all speakers at different ages, levels of education, occupations, etc. Still, it seems that they would be more confident (and thus more consistent in their realization of the tones) if they were guided carefully through the list, item by item. And during the recording, whenever they encounter problems with hesitation or mispronunciation, they are asked, and given the freedom to re-produce them. This is clarified at the beginning to reduce participants' anxiety about making mistakes.

The selection of photos required some care so as to cue the intended monosyllable unambiguously. The choice was made in consultation with Muong speakers; this also offered an opportunity to familiarize them with the experimental procedure. Illustrative photos are selected based on certain criteria. Ideally, they were taken locally, so that participants can easily acquaint themselves with the subjects or activities, especially in case they are local concepts. Figure A.4 provides examples of photos taken in the field (by myself). However, when the number of illustrative photos is relatively numerous (here is 66 items), the actual photography is not the optimal way because they take a lot of time to search for objects in surrounding or setting the scene that is not always available. For this reason, in most cases, we searched and chose photos from the Internet,<sup>2</sup> as long as they meet the criteria of being as direct and clear as possible. In other words, if the photo illustrates a noun, the thing should be the main subject on a plain background (white or black). If the photo illustrates a verb (an action), perhaps the context is important, then we try to avoid the presence of people, especially the faces of people to prevent it from stealing the attention of the speaker. If it was unavoidable, then the people and the surroundings should be similar to the local ones, especially avoid using images of foreign people and sceneries. Otherwise, minimalist sketch is also an option to avoid distracting scenes as much as possible. For example, the illustrative photos for target words /kɔ̃<sup>3</sup>/ “to speak” in minimal sets N<sup>o</sup>6 and /ŋa<sup>5</sup>/ “to fall” in minimal sets N<sup>o</sup>9 are using this method (as reproduced in Figure A.5).

---

<sup>2</sup>The photos were downloaded between 2018 and 2019 from websites few of which provided information on author, persons appearing on the photo, or copyright; efforts to identify the original sources were unsuccessful, and hence no credits are provided in Figure A.6. Readers who can contribute such pieces of information are invited to contact me.



(a) Minimal set N°3: /pa<sup>1</sup>/



(b) Minimal set N°5: /taj<sup>3</sup>/



(c) Minimal set N°4: /laj<sup>1</sup>/



(d) Minimal set N°7: /kien<sup>5</sup>/

Figure A.4: Four examples of illustrative photos that were taken in the field in 2018 to serve the experiment: the best way to pick photos to illustrate target words, especially local concepts. You can find the translations of these words in the Table 3.1.



(a) /kɔ̃³/ “to speak”, minimal sets N°6



(b) /ŋa⁵/ “to fall”, minimal sets N°9

Figure A.5: Two examples of illustrative photos that were picked by the method of minimalist sketch in order to avoid distracting scenes as much as possible.

The case of adjectives will be more difficult because they are more abstract than other types of lexical words in nature. They are used to describe the characteristics of a thing or activity, which means that there is no other way to illustrate them directly but to borrow the image of the object possessing those properties. Being aware of this disadvantage, we have kept the use of adjectives in our word list to a minimum. Only four of the sixty-six target syllables could be recognized as adjectives, which is:

1. Minimal sets N°2: /rɔ̃¹/ in ʔr⁵ rɔ̃¹ “idle”
2. Minimal sets N°2: /rɔ̃²/ “to be satiated”
3. Minimal sets N°10: /ka¹/ “big”
4. Minimal sets N°11: /kaj⁴/ “female”

We can easily notice that the illustrations of these four adjectives all bear a somewhat indirect relationship to the intended word, reflecting the difficulty of finding a context, action, or state that is exactly spot-on: so relevant that it allows for immediate identification, and so easy for speakers to remember that the experiment flows smooth as silk. The photos A.6a and A.6b do not directly illustrate /rɔ̃¹/ “idle” and /rɔ̃²/ “to be satiated”. We must employ a specific example of “idle”, which is a moto-taxi driver resting on his motorcycle because there are no customers, and the context implying “to be satiated” is someone with a conspicuously full stomach in front of empty dishes. In the photos A.6c and A.6d, in order to evoke the words “big” and “female”, subjects exemplifying these characteristics are placed in a comparative opposition. Some might find fault with the use of foreign persons and contexts in illustrative photos for /ʔr⁵ rɔ̃²/ “to be satiated” (A.6b) and /ka¹/ “big” (A.6c), as also in two other cases: /paj⁵ (tʰäj¹)/ “arm span” and /ʔr¹ (kien³)/ “beside”. But to me it is clear that such a criticism would miss the key point: participants understand and accept the principle to prioritize the selection of an image that is closest to the content that is being illustrated, rather than other principles such as homogeneity in terms of the cultural contexts from which the



(a) Minimal sets N<sup>o</sup>2: /rɔ<sup>1</sup>/ “idle”



(b) Minimal sets N<sup>o</sup>2: /rɔ<sup>2</sup>/ “to be sated”



(c) Minimal sets N<sup>o</sup>10: /ka<sup>1</sup>/ “big”



(d) Minimal sets N<sup>o</sup>11: /kaj<sup>4</sup>/ “female”

Figure A.6: The illustrative photos for four adjective

pictures are extracted, none of which was an object of puzzlement for participants.

Once all the photos had been selected, they were arranged into slideshows. This method appeared to be more optimal than using cardboard (carton cards) as had been done in my previous (M.A.) study, especially because the number of target words was relatively large. One risk to take into account, if the slideshow is shown on a laptop computer, is that the computer’s fan will start making a whirring noise, which will compromise the quality of the recording. But apart from this potential difficulty, the digital method is relatively convenient for organizing and controlling the photos in the desired sequence: it makes it easy to replace photos with others that are more appropriate according to the participants’ comments, and is particularly advantageous

for experimenters to handle data collection by themselves if there is no assistant to help showing the photos (an assistance that I was fortunate to have for my M.A.). It saves a lot of preparation time and are more eco-friendly because there is no need to print, cut and arrange the photo cards in order. Once the photos have been placed in the slides, they are already well-organized, with exactly the same order and quantity, no risk of messing up or losing them.

A further advantage is that the display of minimal sets will be much easier and more efficient to visualize in a slideshow. As can be seen in the demo slideshow showing illustrative photos for the elicitation of the first minimal sets (Figure. 3.5, the minimal set is started by an empty slide with only the number of its order of appearance, as a signal announcing the beginning of a particular set. The next slide shows five photos together to help the speaker recognize the expected minimal set. After that, the photos are presented one by one to draw attention to each particular element when the speaker has to utter them within the carrier sentence.



## Appendix B

---

# Some practical details about the recording settings

## B.1 Equipment and experimental setup in the field

### B.1.1 *Recorders and microphones*

I have occasionally heard the comment that technical detail about recordings in publications in linguistics is irrelevant, as devices now possess fewer specificities. Two digital (solid-state) recorders admittedly differ less than a Digital Audio Tape recorder and a compact tape recorder, for instance. But microphones remain widely different from one another, and so do electroglottographs, with consequences on the recorded data that do not seem negligible. Let us therefore go into some technical detail, with apologies to readers for whom some pieces of information seem superfluous (such as indicating the brand names of the equipment that was used). For want of having carried out comparisons between different equipments, it is not possible to tell whether (and to what extent) different results would be obtained with different equipment; information on the topic of recorders and microphones are offered with a view to completeness and explicitness.

For recording, this study used the Roland 4-channel recorder<sup>1</sup> (as Figure B.1a), which allows for the simultaneous recording of up to four channels. In the case of the present study, one of the channels was used for the electroglottographic signal, leaving three channels available for recording acoustic signals. This is a high quality recording solution that linguists can bring to their fieldwork locations. In order to use this recorder, the accompanying equipment needed includes:

- (Figure B.1b): packs of alkaline batteries Duracell AA 1.5V. Each four alkaline batteries will provide approximately one hour of operation.
- (Figure B.1c): two head-mounted microphones Sennheiser HSP4,<sup>2</sup> with MZA 900 P4 phantom power, hereafter referred to for short as *head mic*.
- (Figure B.1d): a microphone AKG C535EB,<sup>3</sup> hereafter referred to for short as *table microphone*.

---

<sup>1</sup>All the information about Roland 4-channel recorder can be found here: <https://proav.roland.com/global/products/r-44/>.

<sup>2</sup>Information of Sennheiser HSP4 microphone is available online: <http://fr-fr.sennheiser.com/microphones-serre-tete-hsp-4>

<sup>3</sup>Information of AKG C535EB microphone is available online: <http://www.ake.com/pro/p/c535eb>

- (Figure B.1e): a pro-audio XLR Male - XLR Female cable for connecting mic outputs to the recorder.
- (Figure B.1f): an adapter converting a 6.3mm stereo jack socket plug to a 3.5mm stereo jack for connecting the recorder to a headphone in order to listen the recording directly during the session.

In fact, the Roland recorder can operate with 2 types of power: AC adapter and AA batteries. However, since the electric network is not grounded in the village, if using the AC adapter for power, there will be conductive metal in contact with the speaker's skin during he/she wears the head-worn microphone. If the voltage differential between the two threads is not 220V vs. zero, but, say, 320V vs. 100V, there is a 100-V difference with the ground, which results in current down into the earth through the consultant's body. An internet search about electric shocks due to non-grounded microphones yields a harvest of horror stories about strong electric shock, with real threats to the consultants' peace of mind during recording and even to their health. This is a serious issue in the case of audio recording. The same problem was also reported by Alexis Michaud during his fieldwork in remote areas of Yunnan. A good solution would have consisted in asking an electrician to provide grounding for the entire household, thus improving the family's security, in addition to solving the issue encountered in fieldwork. But no electrician was available locally to perform the operation. Given this situation, it appeared safest to rely on batteries instead, conducting all recordings on alkaline batteries. With four alkaline batteries each, the Roland recorder can operate in about an hour. Therefore, to record at least 20 people for the experiment and also for the vocabulary collection task, I estimated and brought 80 batteries to the field. That is a huge number of batteries. The fact is that the batteries after the recording can still be used for other devices for a while. I gave them to local people who want to recharge their clocks, lamps, toys, etc. However, I later regretted realizing that they will be discarded in the environment since there is no local system for the treatment of electronic waste yet. This is to say, the batteries may be a safety solution to help overcome electrical problems when using a recording device, but its potential problems of damaging the local environment must also be taken into account consciously.

### **B.1.2 *Electroglottographic device***

Electroglottography, an important exploratory technique for the present study, is presented in Section 2.4 below. In the present paragraph, I just say a few words about practical issues concerning the choice of an electroglottographic device, and its use as a part of the experimental setup that I took to the field.

There are several electroglottographic devices on the market. The one made available to me was a Glottal Enterprise EG2-PCX<sup>4</sup> (hereafter EGG equipment), shown in Figure B.2.

---

<sup>4</sup>Information about the Glottal enterprises EG2-PCX electroglottograph is available from the manufacturer's web page: <http://www.glottal.com/Electroglottographs.html>

B.1 Equipment and experimental setup in the field



(a) Solid-state digital recorder



(b) Alkaline batteries



(c) Head-mounted microphone



(d) Standard microphone



(e) XLR to XLR cable



(f) Jack to minijack adapter

Figure B.1: Roland 4-channel recorder and its accompanying devices.

Beside the main device – the electroglottograph proper: a 2-channel electroglottograph –, the manufacturer also provides supporting components (Figure B.2) as mentioned on their website ([here](#)). Among them, what we really needed, and brought to the field, includes:

- (G) 35mm dual channel electrodes
- (E) Power supply
- (I) Tube of Electrogel
- EG2-PCX2 Manual

This equipment is connected to the Roland recorder via an XLR (male) - minijack cable as shown in Figure B.2b, so that the output, i.e. the EGG signal, can be recorded simultaneously with the acoustic signals. The EGG signal cannot be checked (by eye) directly during recording, but it can be listened to through a channel on the Roland recorder where it is set by the user. In my case, for the current study, I placed it in the third channel.

In addition, there are features in the equipment that allow us to control and know the quality of the signal in a rough way. The first one is the larynx position indicator as in Figure B.3c. This indicator is useful when the researcher attaches two electrodes to the speaker's neck. This signal can be consulted to know when the electrodes are placed in the right position as the LEDs is quite stable in green. Otherwise, if the LEDs move too far to the left or right into the yellow zone or worse into the red zone, it means that the electrodes are placed too high or too low, and still need to be adjusted.

The second feature - calibrateable LED signal strength indicator (as Figure B.3d), can be used to verify that the waveform is strong enough to reliably indicate vocal fold contact area variation. In general, men's signals are stable and good at 8 to 10 points. The women's signal is usually not as good, usually at 5 to 8 points, and many only at 2 to 3 points. The result of such weak signal is mostly difficult or impossible to analyze. These are the cases that were archived but not processed as indicated in the Table 3.7 as "No analysis" because of "Weak EGG signal".

Unlike the Roland recorder, this EGG equipment is equipped with six rechargeable batteries inside, so it can be recharged by the power supply before the recording session. This is convenient because it will not encounter the problems of electrical shock, machine noise, electronic waste or recording interruption due to battery change, etc. However, one experience to note is that when transporting this equipment through customs in the case of moving from abroad (as in my case, from France to Vietnam), be sure to keep in mind that if the equipment is put in carry-on luggage to keep it secure, we should always bring the user manual with it, as customs officers may want to check the battery information to see if it's safe to bring it on board. I ran into this problem in 2019, they were about to delay my flight and prevent me from taking this equipment on board because I could not provide the battery information. It took me a while to show them the website and explain the function of the equipment.



(a) (A) 2-Channel Electroglottograph w/microphone preamplifier with (G) 35mm dual channel electrodes and other components



(b) XLR (male) - minijack cable to connect with recorder

Figure B.2: The Glottal enterprises EG2-PCX and its accompanying components.

## Analog Line-Level Output for Use With Any Data Acquisition System

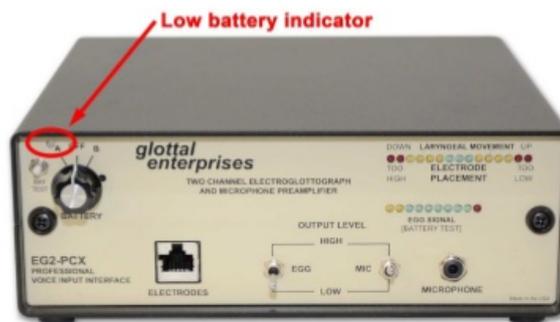


- User can easily connect whatever data acquisition system they prefer to use such as Dataq, National Instruments or one of the Glottal Enterprises analysis systems (Phasecomp, Waveview or Aeroview)

Glottal Enterprises Inc. Electroglossograph Features Feb 2015

(a) Position of the minijack connector for the output.

## Low Battery Indicator Light



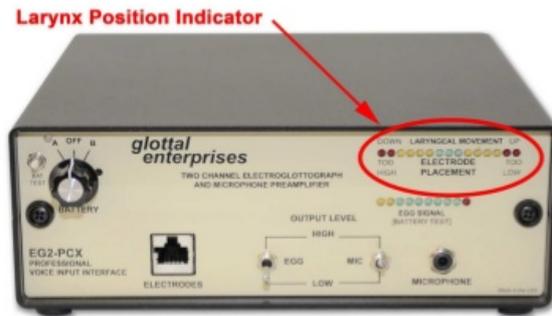
- There are two rechargeable batteries. Can easily switch to second battery when the first one is out.
- Each battery should run for 12 hours.
- The EGG CANNOT run when plugged in. It was designed this way to avoid any noise from the line power.
- The EGG is portable and can be taken into places where there is no power available.

Glottal Enterprises Inc. Electroglossograph Features Feb 2015

(b) Battery indicator light: (i) green for full charge, (ii) red for almost empty battery.

Figure B.3: Some features to note regarding the use of EGG equipment Glottal enterprises EG2-PCX. Reproduced from Hao (2015) (here).

### Two-Channel Configuration Allowing Electrode Placement Feedback

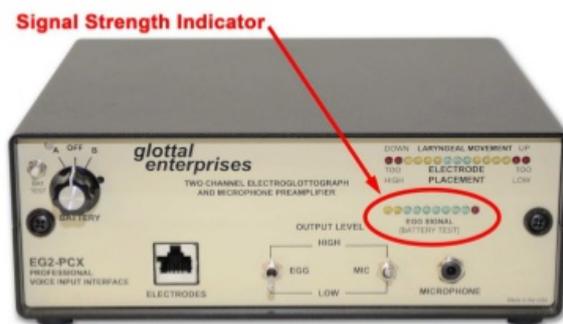


- Allows user to know that they are using the optimum placement of the electrodes.
- You are assured that electrodes will be in the same place for consistency when comparing data from one session to another.
- Saves the user time as they will know during the recording process if improper electrode placement will result in an unacceptable recording.

Glottal Enterprises Inc. Electroglottograph Features Feb 2015

(c) Larynx position indicator

### Signal Strength Indicator on Unit



- Signal Strength Indicator allows you to separate between a good signal and those that are too low for accurate waveform representation.

Glottal Enterprises Inc. Electroglottograph Features Feb 2015

(d) Signal strength Indicator

Figure B.3: Some features to note regarding the use of EGG equipment Glottal enterprises EG2-PCX. Reproduced from Hao (2015) (here).

### B.1.3 Recording environment

All equipment was transported from France to the field work site by plane and then by taxi. They were safely transported in a solid and waterproof polyester suitcase. Due to the frequent high humidity in the area, except during recording, the equipment was always packed carefully in the suitcase and avoided being placed on the floor.

The recording sessions were conducted directly in my Muong teacher's house (Figure B.4), where I stayed during the field trips. The Muong people have the custom of living on the upper floor to avoid feral animals. In the past, in the original architecture of traditional houses (i.e. *nhà sàn*), the first floor was used as a stable for buffaloes, cows or poultry. Today, they rebuilt it with bricks to make an extra room where their children can sleep and to store grain. Recording in this room has both advantages and disadvantages. The advantage is that with the wooden ceiling, and the rice bags filled in the house, the echoes of the cement walls are greatly reduced. Also, to minimize noise, the air vents are temporarily sealed with styrofoam. These conditions together with the high quality of the recorder allow the recordings to be relatively good from a technical point of view, despite the fact that the recordings were made on a farm, not in a recording studio.

On the other hand, the disadvantage also come from the issue of wooden ceiling. The fact is that, in long sessions of recording, it was unavoidable that family members walked in the upper room and made noise as the wooden floor is not solid enough. To remedy this, the recordings were usually carried out when family members were working in the fields, and they were asked to try walking lightly. Another problem was related to the hot and humid weather in the area. There was no other way, but a fan was used throughout the recording, despite the fact that it made noise. More importantly, this ensured that the speaker did not experience discomfort during the performance and did not sweat too much, which could have affected the recording of the EGG signal when the two electrodes were attached relatively tightly to the speaker's neck.

### B.1.4 Recording set-up

Before each recording session, the recorder and EGG equipment were prepared in the following sequence of steps:

1. Place the EGG equipment (fully charged) (B.2a) on the table
2. Place the Roland recorder (B.1a) on the EGG equipment. Make sure a memory card (SDHC) is inserted inside.
3. Insert 4 batteries (B.1b) in the Roland recorder.
4. Plug the male connector of (i) three XLR-XLR cables (B.1e) into the first, second, and fourth female connectors on the Roland recorder, and (ii) an XLR-minijack (B.2b) into the third position.
5. Plug the two head mics into the phantom power supplies (B.1c, then into the females of the XLR cables which were connected to the first and fourth channels of the recorder.



Figure B.4: The room on the ground floor of my Muong teacher's house was used as a "field recording studio" during all my field trips from 2016 to 2019.

6. Plug the table mic (B.1d) into the female end of the XLR cable which was connected to the second channel of the recorder.
7. Plug the minijack of the XLR-minijack cable (B.2b), which was connected to the third channel of the recorder, into the EGG analog signal output (B.3a) in the back panel.
8. Plug the dual-electrodes (G) in to the EGG equipment (A) (B.2a) in the front panel.

At recording, the position of these equipments was as shown on Figure B.5. The two main equipments, the Roland recorder and Glottal enterprises EG2-PCX, were placed on the table. The dual-channel electrode collar was attached to the neck of the speaker to measure the translaryngeal electrical resistance at two adjacent locations; these two signals are combined as the main EGG output, which thus occupies only one channel (the third one) in the resulting WAV file. Electrode placement was conducted as recommended by Nathalie Henrich, locating the last ring of the trachea and placing the electrodes above that point of reference. The input to the Roland steady-state recorder includes four channels:

- (i) the 1<sup>st</sup> channel – audio signal: the first head mic, attached so that the microphone proper was above the corner of the mouth of the speaker;
- (ii) the 2<sup>nd</sup> channel – audio signal: the table mic was put in front of the speaker and on the table (the distance between table mic and speaker was about 30-40cm);
- (iii) the 3<sup>rd</sup> channel – EGG signal: the dual-channel electrode collar was attached to the neck of the speaker;
- (iv) the 4<sup>th</sup> channel – audio signal: the second head mic (in a similar position with first one) was worn by the investigator-instructor (to keep track of the dialogue).<sup>5</sup>

## **B.2 List of files: main experiment and additional materials**

This appendix provides a list of all the audio files collected during field trips in the course of the present study. It includes 224 audio files (the total recording time is around 27 hours and a half) falling into four categories: experiments about tone (including the files of the main experiment for the present thesis), vocabulary elicitation, narratives, and dialogues. My thesis only uses a fraction of this database. The rest is intended for use in further work, and for gradual improvement. There are plans for hosting these data in the [Pangloss](#) Collection, which already hosts Muong materials collected by Michel Ferlus from 1983 to 1996. Some of my data on the experiment of minimal sets and pairs are already available from the [Pangloss](#) Collection<sup>6</sup>.

All files are in .WAV format.

For reasons of page width, the following table only provides the following pieces of basic information:

---

<sup>5</sup>The second microphone was also used when eliciting conversations – not studied here –, in which case it was worn by the second consultant.

<sup>6</sup><https://pangloss.cnrs.fr/corpus/Muong>

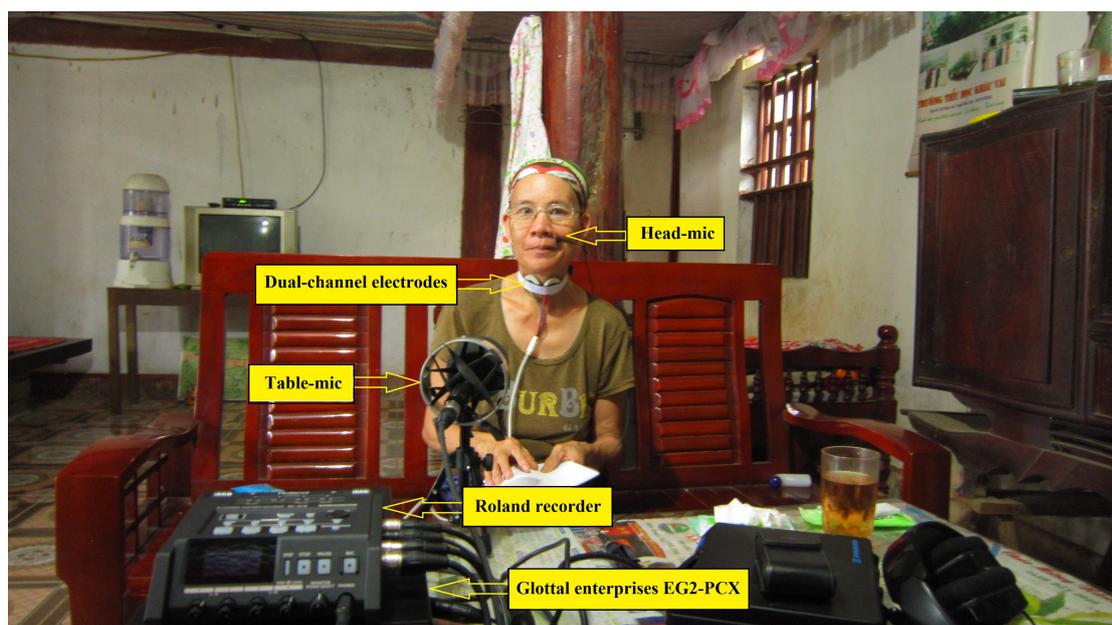


Figure B.5: The position of equipments on the recording (an illustration from the recording of speaker F5 in August 2015).

1. Name of audio files. The file name already contains four pieces of information: (i) the language code: the identifier of each document begins with <crdo-> (the former name of the platform hosting the Pangloss Collection) followed by the ISO code of the language in the Ethnologue catalogue of languages (the code for Mường is MTQ), followed by a short dialect identifier in capital letters, i.e., KTM; (ii) the code of the speaker who recorded these audio files: “F” for Female and “M” for Male, followed by a number (from 1 to 21) referring to the order in which the consultants participated in the research (detailed information about the consultants is provided in Table 3.3); (iii) the basic content of the recording and (iv) an indication about successive repetitions of recordings with similar contents: “1” for the first run and “2” for the second run (remember that some experimental routines were run twice). Underscores are used as separators between pieces of information.
2. Date of recording.
3. Duration of the audio file.

N°	Name of WAV file	Date	Duration
1	crdo-MTQ_KTM_F1F3_BAMBOO-SHOOTS_1	October 2, 2018	0:02:24
2	crdo-MTQ_KTM_F1_BAMBOO-SHOOTS_2	October 2, 2018	0:03:23
3	crdo-MTQ_KTM_F1F3_BAMBOO-SHOOTS_3	October 2, 2018	0:01:35
4	crdo-MTQ_KTM_F1_WEDDING_1	October 2, 2018	0:09:43
5	crdo-MTQ_KTM_F1_MOSS_1	October 2, 2018	0:02:11

Appendix B Some practical details about the recording settings

N°	Name of WAV file	Date	Duration
6	crdo-MTQ_KTM_F1_COM-LAM	October 2, 2018	0:02:13
7	crdo-MTQ_KTM_F1_CHOI-HOI_1	October 2, 2018	0:03:14
8	crdo-MTQ_KTM_F1_CHOI-HOI_2	October 2, 2018	0:03:47
9	crdo-MTQ_KTM_F1_CULTIVATE_1	October 2, 2018	0:07:00
10	crdo-MTQ_KTM_F1_BANH-CHUNG_1	October 2, 2018	0:03:23
11	crdo-MTQ_KTM_F1_BANH-CHUNG_2	October 2, 2018	0:04:00
12	crdo-MTQ_KTM_F1_FUNERAL-OF-UNCLE-TU_1	October 2, 2018	0:08:33
13	crdo-MTQ_KTM_F1_FUNERAL-OF-UNCLE-TU_2	October 2, 2018	0:08:15
14	crdo-MTQ_KTM_F6_WORKING-IN-FARM	October 16, 2018	0:04:36
15	crdo-MTQ_KTM_F6_MEDICINAL-PLANTS	October 16, 2018	0:01:56
16	crdo-MTQ_KTM_F6_FOLK-HEALING	October 16, 2018	0:01:31
17	crdo-MTQ_KTM_F6_DEMONIC-POSSESSION	October 16, 2018	0:02:07
18	crdo-MTQ_KTM_F6_FOLK-GAMES	October 16, 2018	0:02:11
19	crdo-MTQ_KTM_F6_FAMILY	October 16, 2018	0:04:00
20	crdo-MTQ_KTM_F6_WEDDING	October 16, 2018	0:03:53
21	crdo-MTQ_KTM_F6_ACCIDENTS	October 16, 2018	0:01:51
22	crdo-MTQ_KTM_F6_FOLK-TALE_1	October 16, 2018	0:02:43
23	crdo-MTQ_KTM_F6_FOLK-TALE_2	October 16, 2018	0:03:15
24	crdo-MTQ_KTM_F6_FOLK-TALE_3	October 16, 2018	0:03:23
25	crdo-MTQ_KTM_F6_FOLK-TALE_4	October 16, 2018	0:02:54
26	crdo-MTQ_KTM_F6_FISH	October 16, 2018	0:01:43
27	crdo-MTQ_KTM_F6_SLAUGHTER	October 16, 2018	0:02:07
28	crdo-MTQ_KTM_F6_BASSINET	October 16, 2018	0:01:39
29	crdo-MTQ_KTM_F6_NHA-SAN	October 16, 2018	0:03:58
30	crdo-MTQ_KTM_F1_MINIMALSET_1	October 23, 2018	0:10:55
31	crdo-MTQ_KTM_F1_DOG_1	October 23, 2018	0:02:36
32	crdo-MTQ_KTM_F1_BANH-CHUNG_3	October 23, 2018	0:03:13
33	crdo-MTQ_KTM_F3_MINIMALSET	October 25, 2018	0:16:53
34	crdo-MTQ_KTM_F3_BANH-CHUNG_1	October 25, 2018	0:03:26
35	crdo-MTQ_KTM_F3_BANH-CHUNG_2	October 25, 2018	0:04:34
36	crdo-MTQ_KTM_F3_A-WORKING-DAY	October 25, 2018	0:01:38
37	crdo-MTQ_KTM_F3_CATCHING-FISH.wav	October 25, 2018	0:02:13
38	crdo-MTQ_KTM_F3_FAMILY	October 25, 2018	0:02:50
39	crdo-MTQ_KTM_F1_MINIMALSET_2	October 26, 2018	0:11:20
40	crdo-MTQ_KTM_F1_BANH-CHUNG_4	October 26, 2018	0:02:57
41	crdo-MTQ_KTM_F1_DOG_2	October 26, 2018	0:03:14
42	crdo-MTQ_KTM_F1_FUNERAL	October 26, 2018	0:06:35
43	crdo-MTQ_KTM_F1_SLAUGHTER	October 26, 2018	0:02:43
44	crdo-MTQ_KTM_F1_MOSS_2	October 26, 2018	0:02:49
45	crdo-MTQ_KTM_F1_FAMILY	October 26, 2018	0:03:05
46	crdo-MTQ_KTM_F1_WEDDING_2	October 26, 2018	0:06:43
47	crdo-MTQ_KTM_F1_A-WORKING-YEAR	October 26, 2018	0:03:42

B.2 List of files: main experiment and additional materials

N°	Name of WAV file	Date	Duration
48	crdo-MTQ_KTM_F1_MAKING-CAKE	October 26, 2018	0:04:50
49	crdo-MTQ_KTM_F1_CULTIVATE_2	October 26, 2018	0:07:19
50	crdo-MTQ_KTM_F1_BAMBOO-SHOOTS_4	October 26, 2018	0:02:05
51	crdo-MTQ_KTM_F1_FOLK-GAMES	October 26, 2018	0:02:49
52	crdo-MTQ_KTM_F1_HIV	October 26, 2018	0:04:50
53	crdo-MTQ_KTM_M1_MINIMALSET	October 26, 2018	0:14:23
54	crdo-MTQ_KTM_M1F1_BANH-CHUNG	October 26, 2018	0:03:15
55	crdo-MTQ_KTM_M1_LABOR	October 26, 2018	0:01:50
56	crdo-MTQ_KTM_M1_FISH	October 26, 2018	0:01:32
57	crdo-MTQ_KTM_M1_CULTIVATE	October 26, 2018	0:02:31
58	crdo-MTQ_KTM_M1_BAMBOO-SHOOTS	October 26, 2018	0:04:33
59	crdo-MTQ_KTM_M1_TRAVELLING	October 26, 2018	0:06:52
60	crdo-MTQ_KTM_M1_CHOI-HO	October 26, 2018	0:02:50
61	crdo-MTQ_KTM_M1_FOLK-GAMES	October 26, 2018	0:04:03
62	crdo-MTQ_KTM_M1_FUNERAL	October 26, 2018	0:11:58
63	crdo-MTQ_KTM_M7_MINIMALSET	October 27, 2018	0:16:33
64	crdo-MTQ_KTM_M7F1_BANH-CHUNG	October 27, 2018	0:04:36
65	crdo-MTQ_KTM_M7M1_CHILDHOOD	October 27, 2018	0:10:26
66	crdo-MTQ_KTM_M7F1_WORKING-IN-HANOI	October 27, 2018	0:06:07
67	crdo-MTQ_KTM_M1M7_APPENDICITIS	October 27, 2018	0:06:24
68	crdo-MTQ_KTM_M7M1_FATHER	October 27, 2018	0:14:24
69	crdo-MTQ_KTM_M1M7_ALCOHOL	October 27, 2018	0:06:47
70	crdo-MTQ_KTM_M1M7_FRIEND	October 27, 2018	0:08:10
71	crdo-MTQ_KTM_M1M7_POLITICS	October 27, 2018	0:04:14
72	crdo-MTQ_KTM_M1M7_CUSTOMS	October 27, 2018	0:09:47
73	crdo-MTQ_KTM_M8_MINIMALSET	October 28, 2018	0:21:52
74	crdo-MTQ_KTM_M8_BANH-CHUNG	October 28, 2018	0:05:14
75	crdo-MTQ_KTM_M8_LIFE	October 28, 2018	0:08:27
76	crdo-MTQ_KTM_F7_MINIMALSET	October 28, 2018	0:29:51
77	crdo-MTQ_KTM_F7_BANH-CHUNG	October 28, 2018	0:05:39
78	crdo-MTQ_KTM_F7_FAMILY	October 28, 2018	0:04:46
79	crdo-MTQ_KTM_F7_TEACHING	October 28, 2018	0:03:53
80	crdo-MTQ_KTM_F7_CHILDHOOD	October 28, 2018	0:06:59
81	crdo-MTQ_KTM_M5_MINIMALSET	October 29, 2018	0:17:32
82	crdo-MTQ_KTM_M5_BANH-CHUNG	October 29, 2018	0:05:34
83	crdo-MTQ_KTM_M5_WORKS-IN-FARM	October 29, 2018	0:07:16
84	crdo-MTQ_KTM_M5_ANCIENT-LIFE	October 29, 2018	0:11:36
85	crdo-MTQ_KTM_M5_LIFE-IN-WAR	October 29, 2018	0:03:49
86	crdo-MTQ_KTM_M5_ALCOHOL	October 29, 2018	0:04:40
87	crdo-MTQ_KTM_M5_FAMILY	October 29, 2018	0:07:38
88	crdo-MTQ_KTM_F8_MINIMALSET	October 29, 2018	0:19:12
89	crdo-MTQ_KTM_F8F1_BANH-CHUNG	October 29, 2018	0:04:57

*Appendix B Some practical details about the recording settings*

N°	Name of WAV file	Date	Duration
90	crdo-MTQ_KTM_F8F1_WELL	October 29, 2018	0:02:54
91	crdo-MTQ_KTM_F8F1_A-WORKING-DAY	October 29, 2018	0:01:31
92	crdo-MTQ_KTM_M9_MINIMALSET	October 29, 2018	0:13:53
93	crdo-MTQ_KTM_M9_BANH-CHUNG	October 29, 2018	0:06:03
94	crdo-MTQ_KTM_M9_FUNERAL	October 29, 2018	0:03:53
95	crdo-MTQ_KTM_M9_WEDDING	October 29, 2018	0:07:44
96	crdo-MTQ_KTM_M9_ANT-SONG	October 29, 2018	0:00:51
97	crdo-MTQ_KTM_M9_FOLK-GAMES	October 29, 2018	0:03:34
98	crdo-MTQ_KTM_M9_STUDYING	October 29, 2018	0:03:13
99	crdo-MTQ_KTM_M9_GIVING-BIRTH	October 29, 2018	0:01:53
100	crdo-MTQ_KTM_M9_HERD	October 29, 2018	0:01:37
101	crdo-MTQ_KTM_F9_MINIMALSET	October 30, 2018	0:16:39
102	crdo-MTQ_KTM_F9_BANH-CHUNG	October 30, 2018	0:05:38
103	crdo-MTQ_KTM_F9_SON	October 30, 2018	0:04:48
104	crdo-MTQ_KTM_M10_MINIMALSET	October 30, 2018	0:17:03
105	crdo-MTQ_KTM_M10_BANH-CHUNG_1	October 30, 2018	0:03:21
106	crdo-MTQ_KTM_M10_BANH-CHUNG_2	October 30, 2018	0:07:41
107	crdo-MTQ_KTM_M10_WORKING	October 30, 2018	0:06:05
108	crdo-MTQ_KTM_M10_FAMILY_1	October 30, 2018	0:02:05
109	crdo-MTQ_KTM_M10_MUONG-HOA-BINH	October 30, 2018	0:04:23
110	crdo-MTQ_KTM_M10_WORKING-DIFFICULTIES	October 30, 2018	0:02:13
111	crdo-MTQ_KTM_M10_TET2018	October 30, 2018	0:02:34
112	crdo-MTQ_KTM_M10_FAMILY_2	October 30, 2018	0:01:27
113	crdo-MTQ_KTM_M11_MINIMALSET	October 30, 2018	0:13:30
114	crdo-MTQ_KTM_M11_BANH-CHUNG	October 30, 2018	0:03:35
115	crdo-MTQ_KTM_M11_A-WORKING-DAY	October 30, 2018	0:00:52
116	crdo-MTQ_KTM_M11_FAMILY	October 30, 2018	0:03:25
117	crdo-MTQ_KTM_M11_CULTIVATE	October 30, 2018	0:01:41
118	crdo-MTQ_KTM_M11_TET-HOLIDAY	October 30, 2018	0:01:47
119	crdo-MTQ_KTM_M11_CHILDHOOD	October 30, 2018	0:01:09
120	crdo-MTQ_KTM_F10_MINIMALSET	October 30, 2018	0:15:14
121	crdo-MTQ_KTM_F10_BANH-CHUNG	October 30, 2018	0:03:14
122	crdo-MTQ_KTM_F11_MINIMALSET	October 30, 2018	0:13:41
123	crdo-MTQ_KTM_F11_BANH-CHUNG	October 30, 2018	0:04:51
124	crdo-MTQ_KTM_F11_A-WORKING-DAY	October 30, 2018	0:01:05
125	crdo-MTQ_KTM_F11_A-WORKING-YEAR	October 30, 2018	0:02:08
126	crdo-MTQ_KTM_F11_FAMILY	October 30, 2018	0:01:50
127	crdo-MTQ_KTM_M12_MINIMALSET	October 31, 2018	0:22:17
128	crdo-MTQ_KTM_M12_BANH-CHUNG	October 31, 2018	0:08:02
129	crdo-MTQ_KTM_M12_FOLK-TALE	October 31, 2018	0:00:56
130	crdo-MTQ_KTM_M12_ARMY	October 31, 2018	0:04:51
131	crdo-MTQ_KTM_M12_FAMILY	October 31, 2018	0:03:36

B.2 List of files: main experiment and additional materials

N°	Name of WAV file	Date	Duration
132	crdo-MTQ_KTM_M12_LIFE	October 31, 2018	0:05:15
133	crdo-MTQ_KTM_F12_MINIMALSET	October 31, 2018	0:16:35
134	crdo-MTQ_KTM_F12_BANH-CHUNG_1	October 31, 2018	0:05:30
135	crdo-MTQ_KTM_F12_BANH-CHUNG_2	October 31, 2018	0:05:25
136	crdo-MTQ_KTM_F12_A-WORKING-DAY	October 31, 2018	0:01:49
137	crdo-MTQ_KTM_F12_CULTIVATE	October 31, 2018	0:05:03
138	crdo-MTQ_KTM_F12_COM-LAM	October 31, 2018	0:01:52
139	crdo-MTQ_KTM_F12_COOKING-FISH	October 31, 2018	0:00:45
140	crdo-MTQ_KTM_F12_FAMILY	October 31, 2018	0:04:56
141	crdo-MTQ_KTM_F12_TET-HOLIDAY	October 31, 2018	0:05:00
142	crdo-MTQ_KTM_F13_MINIMALSET	November 1, 2018	0:19:37
143	crdo-MTQ_KTM_F13_BANH-CHUNG_1	November 1, 2018	0:00:58
144	crdo-MTQ_KTM_F13_BANH-CHUNG_2	November 1, 2018	0:04:35
145	crdo-MTQ_KTM_F13_LABOR-EXPORT	November 1, 2018	0:12:31
146	crdo-MTQ_KTM_F13_A-WORKING-YEAR	November 1, 2018	0:03:10
147	crdo-MTQ_KTM_F13_FAMILY	November 1, 2018	0:07:30
148	crdo-MTQ_KTM_F13_ANT-SONG	November 1, 2018	0:00:12
149	crdo-MTQ_KTM_F13_LULLABY_1	November 1, 2018	0:00:39
150	crdo-MTQ_KTM_M13_MINIMALSET	November 1, 2018	0:20:31
151	crdo-MTQ_KTM_M13_BANH-CHUNG	November 1, 2018	0:05:34
152	crdo-MTQ_KTM_M13_FAMILY	November 1, 2018	0:01:55
153	crdo-MTQ_KTM_M13_CULTIVATE	November 1, 2018	0:01:37
154	crdo-MTQ_KTM_M13_ALCOHOL	November 1, 2018	0:02:49
155	crdo-MTQ_KTM_M13_WORKS-IN-FARM	November 1, 2018	0:01:56
156	crdo-MTQ_KTM_M13_FOLK-GAMES	November 1, 2018	0:02:38
157	crdo-MTQ_KTM_M13_TET-HOLIDAY	November 1, 2018	0:03:01
158	crdo-MTQ_KTM_M13_WEDDING	November 1, 2018	0:02:03
159	crdo-MTQ_KTM_M13_MEDICINAL-PLANTS	November 1, 2018	0:03:14
160	crdo-MTQ_KTM_F14_MINIMALSET	November 2, 2018	0:16:31
161	crdo-MTQ_KTM_F14_BANH-CHUNG	November 2, 2018	0:06:54
162	crdo-MTQ_KTM_F14_FOLK-TALE_1	November 2, 2018	0:02:48
163	crdo-MTQ_KTM_F14_FOLK-TALE_2	November 2, 2018	0:01:43
164	crdo-MTQ_KTM_F14_FOLK-TALE_3	November 2, 2018	0:02:11
165	crdo-MTQ_KTM_F14_FOLK-TALE_4	November 2, 2018	0:01:06
166	crdo-MTQ_KTM_F14_ANT-SONG	November 2, 2018	0:01:53
167	crdo-MTQ_KTM_F14_LULLABY	November 2, 2018	0:02:35
168	crdo-MTQ_KTM_F15_LULLABY_1	November 2, 2018	0:01:59
169	crdo-MTQ_KTM_F15_FOLK-SONG	November 2, 2018	0:01:20
170	crdo-MTQ_KTM_F15_ANT-SONG	November 2, 2018	0:00:14
171	crdo-MTQ_KTM_F15_LULLABY_2	November 2, 2018	0:01:18
172	crdo-MTQ_KTM_F13_FOLK-GAMES	November 2, 2018	0:06:34
173	crdo-MTQ_KTM_F13_LULLABY_2	November 2, 2018	0:01:22

Appendix B Some practical details about the recording settings

N°	Name of WAV file	Date	Duration
174	crdo-MTQ_KTM_F14_FOLK-TALE_5	November 2, 2018	0:02:55
175	crdo-MTQ_KTM_F13_FOLK-TALE_VIE	November 2, 2018	0:04:20
176	crdo-MTQ_KTM_F13F14_FOLKSONG_1	November 2, 2018	0:07:21
177	crdo-MTQ_KTM_F13F14_FOLKSONG_2	November 2, 2018	0:06:23
178	crdo-MTQ_KTM_F14F13_FOLKSONG_3	November 2, 2018	0:09:40
179	crdo-MTQ_KTM_F14F13_FOLKSONG_4	November 2, 2018	0:16:04
180	crdo-MTQ_KTM_M14_MINIMALSET	November 2, 2018	0:19:05
181	crdo-MTQ_KTM_M14_BANH-CHUNG	November 2, 2018	0:03:21
182	crdo-MTQ_KTM_M14_A-WORKING-DAY	November 2, 2018	0:00:49
183	crdo-MTQ_KTM_M14_FAMILY	November 2, 2018	0:02:33
184	crdo-MTQ_KTM_M14_TET-HOLIDAY	November 2, 2018	0:01:10
185	crdo-MTQ_KTM_M14_LIVING-IN-FARM	November 2, 2018	0:00:47
186	crdo-MTQ_KTM_M12_MINIMALSET_2	August 15, 2019	0:23:28
187	crdo-MTQ_KTM_M12_BANH-CHUNG_2	August 15, 2019	0:04:10
188	crdo-MTQ_KTM_F16_MINIMALSET	August 16, 2019	0:19:53
189	crdo-MTQ_KTM_F16_A-WORKING-YEAR	August 16, 2019	0:01:13
190	crdo-MTQ_KTM_F16_BANH-CHUNG	August 16, 2019	0:02:28
191	crdo-MTQ_KTM_F16_LULLABY	August 16, 2019	0:01:03
192	crdo-MTQ_KTM_F16_PRIVATE-LIFE	August 16, 2019	0:03:44
193	crdo-MTQ_KTM_F16_TET-HOLIDAY	August 16, 2019	0:01:28
194	crdo-MTQ_KTM_F17_MINIMALSET	August 16, 2019	0:28:12
195	crdo-MTQ_KTM_F17_BANH-CHUNG_1	August 16, 2019	0:02:01
196	crdo-MTQ_KTM_F17_BANH-CHUNG_2	August 16, 2019	0:02:00
197	crdo-MTQ_KTM_F17_ANT-SONG	August 16, 2019	0:00:40
198	crdo-MTQ_KTM_F17_PRIVATE-LIFE	August 16, 2019	0:02:00
199	crdo-MTQ_KTM_F17_TET-HOLIDAY	August 16, 2019	0:01:13
200	crdo-MTQ_KTM_F17_WORKING	August 16, 2019	0:01:03
201	crdo-MTQ_KTM_F18_MINIMALSET	August 19, 2019	0:14:57
202	crdo-MTQ_KTM_F18_BANH-CHUNG	August 19, 2019	0:02:24
203	crdo-MTQ_KTM_F18_WORKING	August 19, 2019	0:01:53
204	crdo-MTQ_KTM_F19_MINIMALSET	August 19, 2019	0:13:24
205	crdo-MTQ_KTM_F19_BANH-CHUNG	August 19, 2019	0:02:26
206	crdo-MTQ_KTM_F19_A-WORKING-DAY	August 19, 2019	0:01:09
207	crdo-MTQ_KTM_F20_MINIMALSET	August 20, 2019	0:14:18
208	crdo-MTQ_KTM_F20_BANH-CHUNG	August 20, 2019	0:03:21
209	crdo-MTQ_KTM_F20_A-WORKING-DAY	August 20, 2019	0:01:09
210	crdo-MTQ_KTM_F20_DAUGHTERS	August 20, 2019	0:00:51
211	crdo-MTQ_KTM_F20_MUSIC-ACTIVITY	August 20, 2019	0:01:29
212	crdo-MTQ_KTM_F20_MOSS	August 20, 2019	0:01:05
213	crdo-MTQ_KTM_F1_WORDLIST	August 20, 2019	1:41:54
214	crdo-MTQ_KTM_F3_WORDLIST_P1	August 21, 2019	0:52:51
215	crdo-MTQ_KTM_F3_WORDLIST_P2	August 21, 2019	0:59:05

*B.2 List of files: main experiment and additional materials*

<b>N°</b>	<b>Name of WAV file</b>	<b>Date</b>	<b>Duration</b>
216	crdo-MTQ_KTM_F3_WORDLIST_P3	August 22, 2019	0:51:04
217	crdo-MTQ_KTM_F3_WORDLIST_P4	August 22, 2019	1:02:05
218	crdo-MTQ_KTM_F3_WORDLIST_P5	August 22, 2019	0:24:56
219	crdo-MTQ_KTM_F3_WORDLIST_P6	August 22, 2019	1:02:05
220	crdo-MTQ_KTM_F3_WORDLIST_P7	August 22, 2019	0:37:32
221	crdo-MTQ_KTM_F21_MINIMALSET	August 22, 2019	0:15:39
222	crdo-MTQ_KTM_F21_BANH-CHUNG	August 22, 2019	0:01:50
223	crdo-MTQ_KTM_F21_A-WORKING-DAY	August 22, 2019	0:01:09
224	crdo-MTQ_KTM_F21_RICE-MILLING	August 22, 2019	0:01:08



## Appendix C

---

### Data analysis and visualization

#### C.1 Labeling the tone system: numeric labels and association with colors

This section aims to draw attention to the way the object of this study – the tone system of Kim Thuong Muong – is labeled and displayed: the assignment of arbitrary numeric labels, and the association of tones with colors in a principled way.

For ease of reference, each tone within the system is assigned a number, adopting a continuous numbering: from 1 to 5 for the five tones in smooth syllables, and 6 and 7 for the two tones in checked syllables. Such a way of labeling tone systems is common in synchronic studies of various tonal languages, such as Mandarin, a language for which the synchronic labeling of tones from 1 to 4 is adopted in phonetic-phonological studies (Howie, 1974; Jongman et al., 2006), sidestepping the diachronic complexities of the traditional categories of Chinese philology (Jacques, 2006). The ordering of tones when assigning them a number (or when drawing up a list of the tones when they have names of their own, as in Vietnamese and Middle Chinese) can, in principle, have a logic of its own, such as: simplest tone first, as in Vietnamese and Mandarin, lowest tone first, as in Joseph Rock's numbering of Naxi tones (Rock, 1963), or highest tone first, as in Fu Maoji's numbering of the same Naxi tones (Fù, 1981, p. 2). Frequency in the lexicon could also be a criterion (as, again, in Naxi, where both Joseph Rock and Fu Maoji list the Rising tone, a rare tone, at the end of the list, as the fourth tone). For Muong, the labeling by numbers has no such motivation: numbers were assigned to tones in the order in which they were discovered, in early stages of fieldwork, and the numbering was kept thereafter for the sake of consistency. Thus, the numbering does not follow any systematic language-internal logic. No clear and easy mnemonic associations between numbers and any property of the tones stands out *a posteriori* either.

A drawback of the numeric labels is that their lack of phonetic/phonological explicitness can make it difficult to remember them. In an effort to address this problem and create a degree of mnemonic association of the labels with the tones' phonological and phonetic characteristics, each tone is assigned a particular color. This is intended to create ties between the tones and the means used to represent them, with a view to making it easier for the reader to recognize the intended tones when they are mentioned by their numeric labels. Here we will clarify which colors are used to represent which tones,

and explain why we selected those colors. The developments that follow inevitably require previewing some of the results to be set out later in this dissertation, as the choice of colors is linked to phonetic/phonological properties of the tone.

The selected colors are based on two criteria: (i) giving an impression that is compatible with the intended tone, and (ii) setting a good contrast for optimal observation. On the first criterion, in order to help the reader get familiar with the tone system, we attempt to apply the method of tone→color perception. Hereafter, every time each tone is mentioned, it will be highlighted with its associated color<sup>1</sup>, as in Table C.1.

### C.1.o.1 Tone and color perception

The study of the synesthesia of sound and colors is a fascinating topic. The first time I was exposed to this strand of research was in the spectrogram reading course of professor Jacqueline Vaissière (at École doctorale 622, Sorbonne Nouvelle). Figure C.1 is the piece of that course where she explains why and how each vowel can be represented by a certain color, according to her perception.

Indeed, there are many interesting findings about the relationship between sounds and colors. For instance, several studies on vowel-color mapping find that lighter colors (yellow, green) tend to be associated with more front vowels (higher F<sub>2</sub>; e.g., /i/, /e/), darker colors (e.g., red, brown, blue) with back vowels (/o/, /u/) (Wrembel, 2009; Marks, 2014). These findings echo psychophonetic intuitions by Fónagy about front vowels being closer to the lips, and hence to the light of day, whereas back vowels are deeper inside the vocal tract, as it were, and hence steeped in darkness (Fónagy, 1983).

Now coming to pitch, a robust pattern seems to be the association of brighter colors with higher pitch, and darker colors with lower pitch (Marks, 2014). A cross-linguistic overview is provided by Anikin and Johansson (2019): see in particular their Table 1, pp. 3-4. Here, following the idea of sound-color synesthesia, we try to associate each tone with an appropriate color, used systematically in the figures showing results for f<sub>o</sub> ΔEGG and O<sub>q</sub> ΔEGG. Specifically: **Tone** ■ **1** is associated with green color because I get a sense of *nature* when listening this tone. The flat contour presumably gives a feeling of stability, lightness, and tranquility. It is somewhat like a soft breath: if it were wind, it would be like a breeze. The other level tone of the system, **Tone** ■ **5**, is the highest in pitch, giving a sensation of something bright, clear, cheerful and soaring. The color of a vivid clean sky was assigned for this tone: cyan. Blue is the color I see when listening to **Tone** ■ **2** because its sharp fall makes me perceive something forceful, strong, and quick like waves in a choppy sea, or the image of dolphins rushing away and diving quickly into the water. In contrast, the movement of **Tone** ■ **3** is somewhat lighter and less intense. It reminds me of the rising movement of the sun at dawn. Not that the rise is slow and sluggish: it is also fast, but somehow calm, gentle and serene. Therefore, a light orange is used as the color for representing this tone. Last but not least, the glottalized tone, labeled as **Tone** ■ **4**, is associated with red, the color

---

<sup>1</sup>The highlighted colors here are not exactly those used in the figures because we also have to take into account the need for contrast with the black letters and tracings, for highest readability.

**PERCEPTION**  
 Perceptivement, /i/ est une voyelle aérienne. Je la vois jaune clair, plus brillante pour les focales que pour les autres, plus tranchante, plus claire. Rimbaud la voyait rouge.

**Question 6 : Et vous, de quelle couleur est /i/ ?**

Figure 7: la façon dont je vois les voyelles du français

Figure C.1: Color perception of French vowels. Reproduced (with permission) from the spectrogram reading course of Pr. Jacqueline Vaissière in 2018.

of lava stone. We think it is an excellent match for this tone, since the glottalization has creaky voice as its most frequent phonetic realization, bringing the perception of something solid, rough, and at the same time pretty light, like volcanic stones (full of bubbles of air – reminiscent of the impression that, in creaky voice, bubbles are escaping between the adducted vocal folds).

Additionally, there are also two tones in checked syllables, which are classified as a sub-system within the tone system, distinct from the sub-type of 5 tones in smooth syllables. These two tones were included in the experiment for the sake of completeness, but they are not an important object of study in the present dissertation. Therefore, they are both backgrounded, using shades of grey to set them apart as a separate sub-type. **Tone 6** is represented by brighter grey in comparison with the shade used for **Tone 7**, reflecting in visual terms their distinguishing characteristic in term of pitch level: **Tone 6** is higher than **Tone 7**.

The explanation in Table C.1 is a brief summary that will be useful to help remember the basic distinguishing impressions associated with the various tones, which may be all the more helpful as there is no accompanying sound.

It needs to be emphasized that the color perception of the tones described here is based solely on my own impressions. Importantly, no claim is made that these impressions are a worthwhile path for accessing an underlying logic of the system and connect it to scientific (phonetic) observations. Thus, my impression that **Tone 1** is *like soft breath* is clearly at variance with measurements of open quotient: if the tone were technically breathy,  $O_{q \text{ dEGG}}$  would be higher than for other tones, but (to preview results from Chapter 4)  $O_{q \text{ dEGG}}$  measurements for **Tone 1** do not at all stand out within the five-tone system: technically, the tone is not phonetically breathy. The impressionistic description and scientific study are on two distinct levels. Others are likely to have different perceptions and opinions. The aim in assigning colors to tones is to provide a consistent visualization of tones, and thus help the reader and myself to quickly and effectively establish a relationship between the three dimensions: (i) the labeled tone, (ii) its representative number, and (iii) salient acoustic characteristics of the tone.

Table C.1: List of representative colors for tones accompanied by a brief explanation of reasons for the choice of colors

Tone	Short label	color	Impressionistic associations
<b>Tone 1</b>	mid-low level tone	green	<b>perception of nature:</b> light, calm, peaceful, elegant, slightly breathy (like the wind)
<b>Tone 2</b>	falling tone	blue	<b>perception of sea:</b> heavy, sharp, quick move (like the jump of dolphin into water)
<b>Tone 3</b>	rising tone	orange	<b>perception of sun rising:</b> moving up in peace, stability, and optimism
<b>Tone 4</b>	glottalized tone	red	<b>perception of lava stone:</b> solid, rough but light
<b>Tone 5</b>	high level tone	cyan	<b>perception of sky:</b> bright, clear, cheerful, something flies
<b>Tone 6</b>	high checked tone	light grey	<i>high tone of checked syllables</i>
<b>Tone 7</b>	low checked tone	dark grey	<i>low tone of checked syllables</i>

### C.1.0.2 Color contrast for good readability and visibility

A second criterion for the choice of colors has to do with visual clarity. We need to appraise to what extent the colors which by themselves appear suitable for the tones are in good contrast with one another, in order to facilitate and optimize visual inspection of the materials that display experimental results. In other words, colors are intended to bring out the tonal space and the nuances within this space. Achieving

good color contrast simply involves choosing colors that differ clearly from one another. Crisp contrast adds visual clarity and appeal. On the other hand, contrast should be applied in a balanced way, especially in a complex combination: too much contrast can do more harm than good (somewhat like excessive use of spice in cooking), by creating a confusing or visually jarring appearance. The colors chosen in the previous step were therefore subjected to inspection from the point of view of their combinatorial properties as colors for display on the same graphs, fine-tuning them for better contrasts according to basic principles based on hue and lightness (tint and shade).

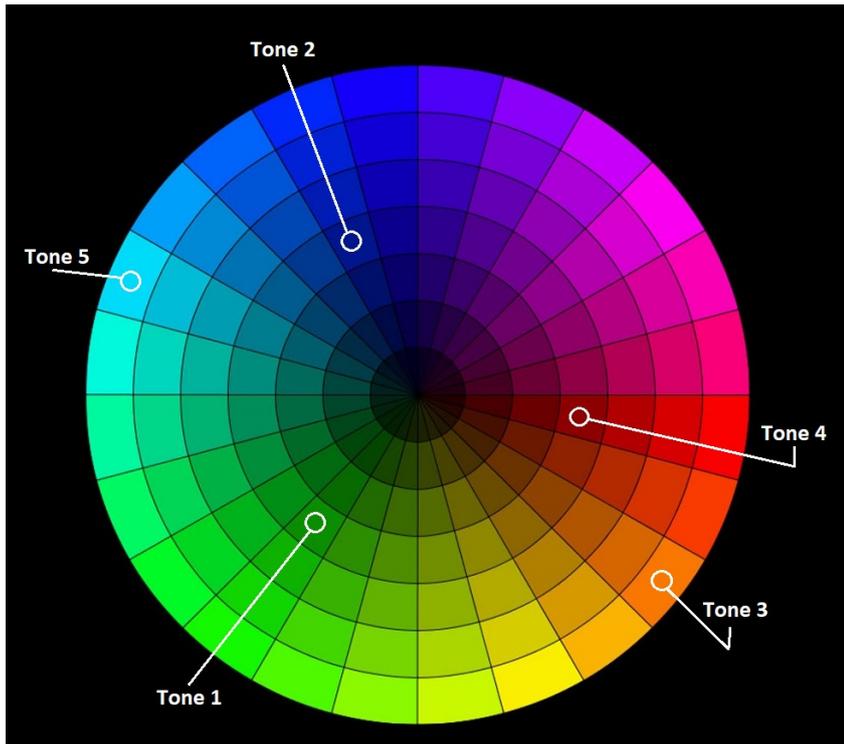


Figure C.2: The distribution of representative colors of 5 tones on the color wheel.

A color wheel is a visual representation of the relationships of color hues. Two colors that sit opposite each other on the color wheel are called complementary colors. These colors create highest contrast to each other at maximum saturation. However, this stark contrast can make them visually jarring. Besides, we can also have analogous combinations in which 2 to 5 (ideally 2 to 3) colors are adjacent to each other on the color wheel. In comparison with complementary combination, the contrast between analogous colors is admittedly lower, but it tends to create a calming, likeable impression (as I learnt from a [“color combination cheat sheet”](#)). A color wheel is a good way to help us think about how different hues relate to one another. It can serve as a practical way to determine whether a pair (or more) of colors will relate to one another in a harmonious way. Figure C.2 shows the position in the color wheel of

the five representative colors associated with five tones as used in all the figures in the present work. It can be seen that there are no pairs of colors in highest contrast, or in other words, in purely complementary relationship. However, the three colors *blue* (for **Tone 2**), *green* (for **Tone 1**) and *red* (for **Tone 4**) almost form an equilateral triangle in the colors wheel, which means that the contrast between those colors is fairly strong.

Special mention needs to be made of the tinge of red chosen for the glottalized tone labeled as **Tone 4**. This tone is central to the present study; the choice of dark red as a color for this tone not only meets the first criterion (color sensations in response to this tone), but it is also suitable to set out the tone as a “star” within the system. According to color psychology, red is the most intense color, which can provoke the strongest emotions, and capture attention.

The above considerations constitute the main considerations in the choice of color combination for three of the five tones of smooth syllables. The other two get cyan (for **Tone 5**) and orange (for **Tone 3**). These have an analogous relationship with blue and red, respectively. Therefore, in order to enhance the contrast between these colors, we select the lightest colors in brightness scale for cyan and orange so that they contrast clearly with the darker colors previously selected.

A last step for this consideration is to test in practice how well the colors render together, and whether they allow for good legibility of figures that show the two main parameters discussed in the present work:  $f_{o\text{ dEGG}}$  and  $O_{q\text{ dEGG}}$ . Regarding those figures in Chapter 4, it seems fair to say that the colors match reasonably well, even at places where the tones overlap. These considerations seemed to vindicate the choices made here.

One more note for this section is that there are differences between highlight colors used in the thesis (in  $\LaTeX$  format) and colors in figures produced by MATLAB. The reason is that we must take into account the color contrast for better readability and visibility since the highlighted colors within the text are surrounded by black letters whereas in the figures the colors display the  $f_{o\text{ dEGG}}$  or  $O_{q\text{ dEGG}}$  lines against a white background. The solution is that we attempt to select relatively similar colors but in different tinges and shades in the  $\LaTeX$  palette compared to those chosen in MATLAB so as to maintain an effective association of colors to the label names of tones. Especially, the three dark colors of green, blue, and red selected for **Tone 1**, **Tone 2**, and **Tone 4** in plotting figures are adjusted to much lighter colors in the text. We considered and selected the exact colors by consulting the list of ready-to-use RGB triplets codes from  $\LaTeX$  ([here](#)) and MATLAB ([here](#)).

Needless to say, these quick remarks leave many topics unaddressed, such as the issue of data access for color-blind readers. I apologize for this shortcoming, which readers with different sensitivities to color can remedy by relying on the different shapes of the marker symbols used for different tones.

## Appendix D

---

### Additional graphs produced as a result of this study

This appendix provides an additional set of figures that were plotted to visualize the results of this study but were not presented in the Results and Discussion chapters.

**D.1 The Kim Thuong Muong tone system by fundamental frequency in semitone**

**D.2 The glottalized tone in Kim Thuong Muong: the distribution of fundamental frequency and open quotient values are presented in box plot**

**D.3 The correlation between fundamental frequency and open quotient by a scatter plot**

Appendix D Additional graphs produced as a result of this study

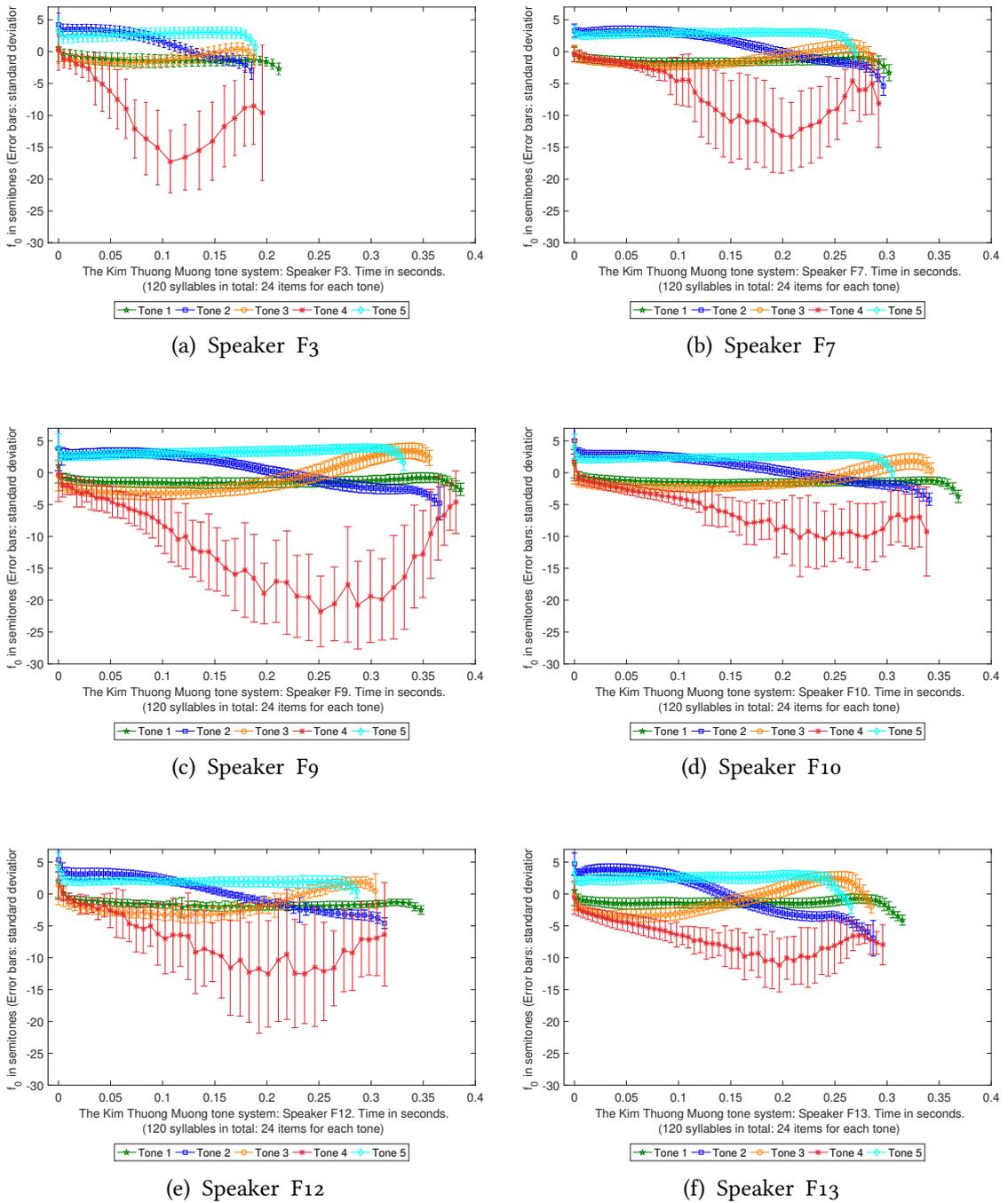
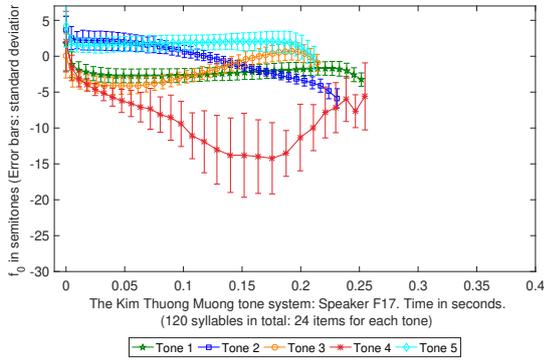
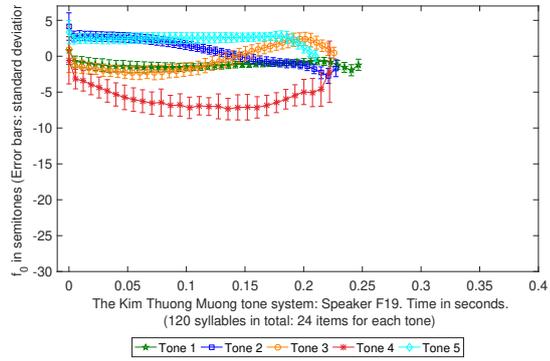


Figure D.1: Kim Thuong Muong's tone system presented in semitones: speaker by speaker. Same data as in Figure 4.1.

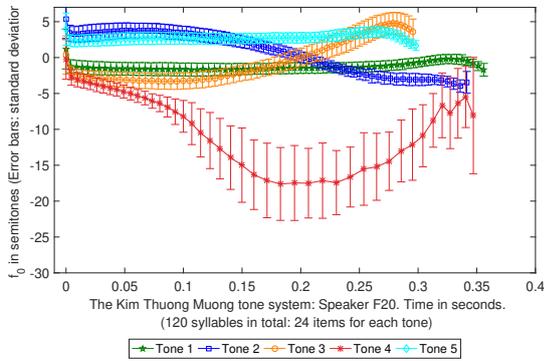
D.3 The correlation between fundamental frequency and open quotient by a scatter plot



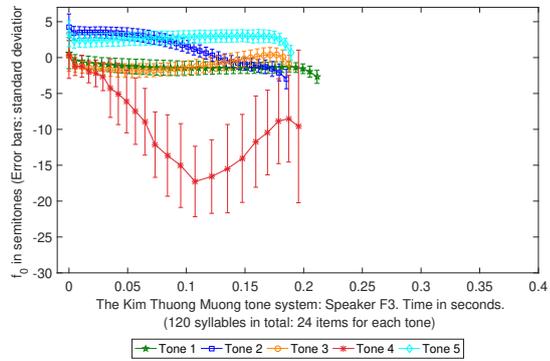
(g) Speaker F17



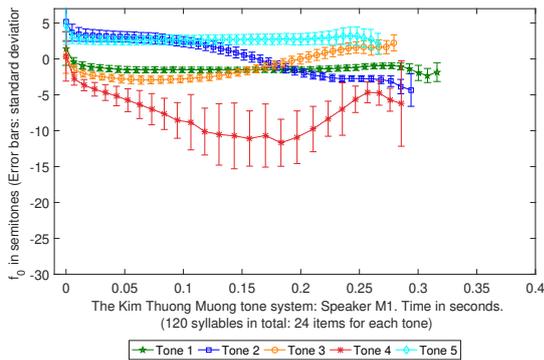
(h) Speaker F19



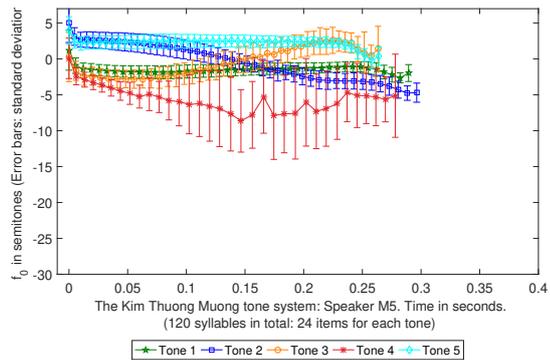
(i) Speaker F20



(j) Speaker F21



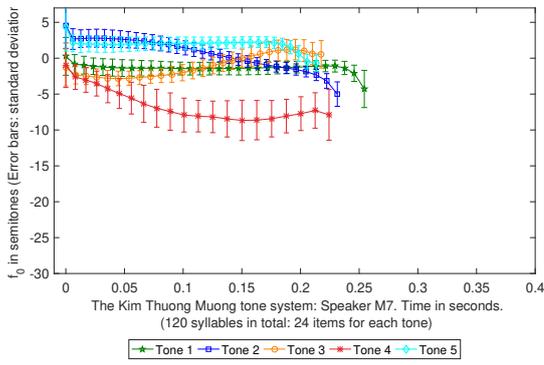
(k) Speaker M1



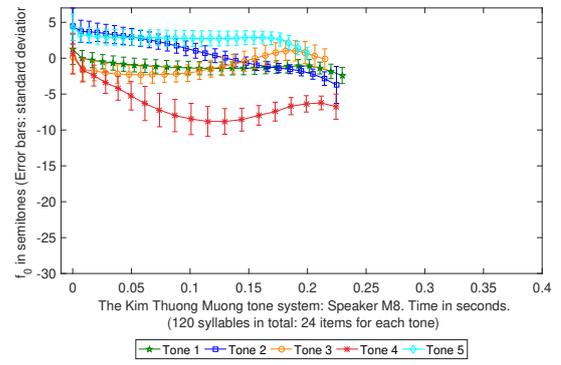
(l) Speaker M5

Figure D.1: Kim Thuong Muong's tone system:  $f_0$  <sub>dEGG</sub> in semitones, speaker by speaker.

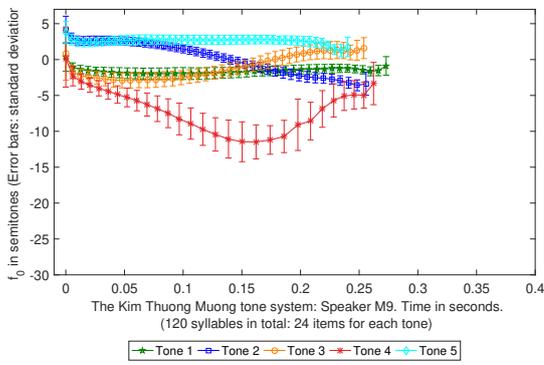
Appendix D Additional graphs produced as a result of this study



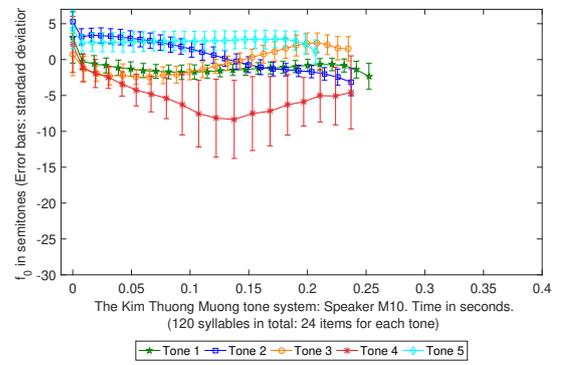
(m) Speaker M7



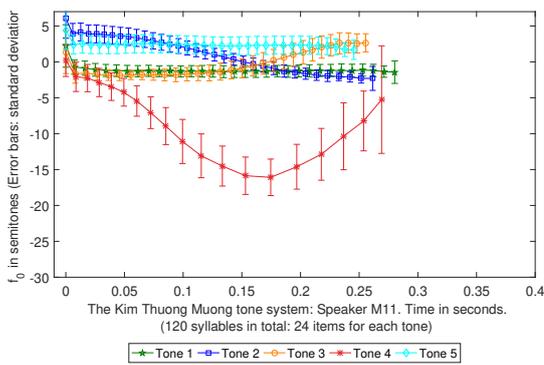
(n) Speaker M8



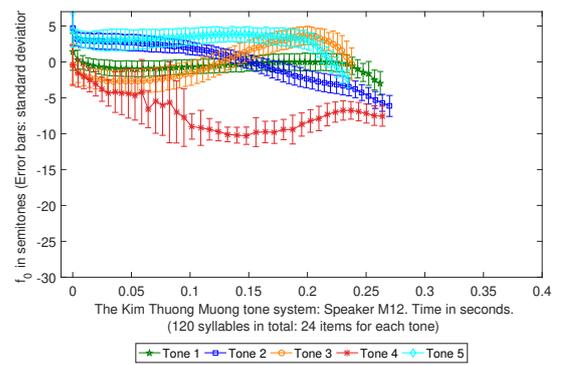
(o) Speaker M9



(p) Speaker M10



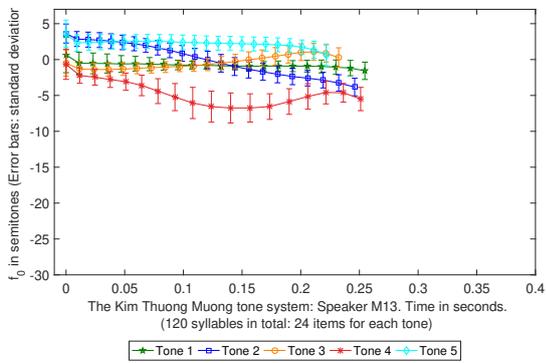
(q) Speaker M11



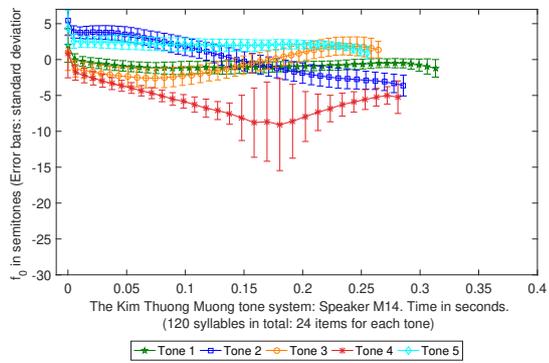
(r) Speaker M12

Fig D.1: Kim Thuong Muong's tone system is presented by  $f_0$  d<sub>EGG</sub> in semitones: speaker by speaker.

D.3 The correlation between fundamental frequency and open quotient by a scatter plot



(s) Speaker M13



(t) Speaker M14

Fig D.1: Kim Thuong Muong's tone system is presented by  $f_0$  dEGG in semitones: speaker by speaker.

The KTM Tone 4 by 20 speakers (F: woman, M: man): distribution of  $f_0$  and  $O_q$  values (24 syllables per speaker)

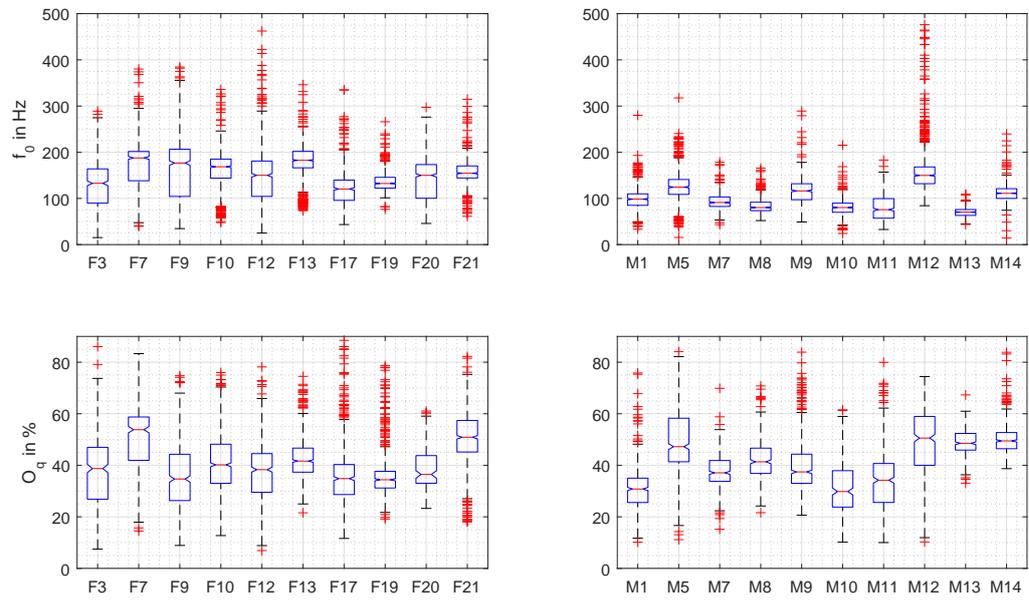
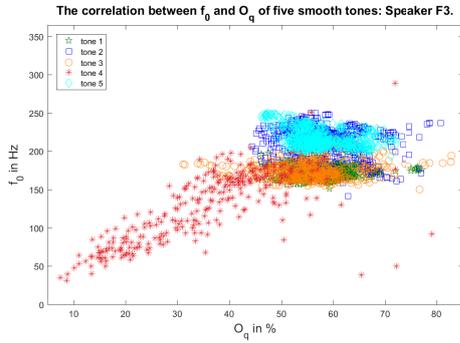
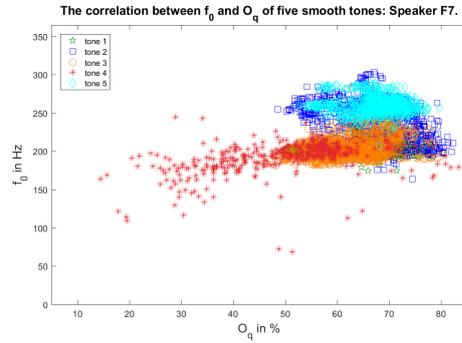


Figure D.2: The glottalized tone in Kim Thuong Muong: the distribution of  $f_0$  <sub>dEGG</sub> and  $O_q$  <sub>dEGG</sub> values are presented in box plot.

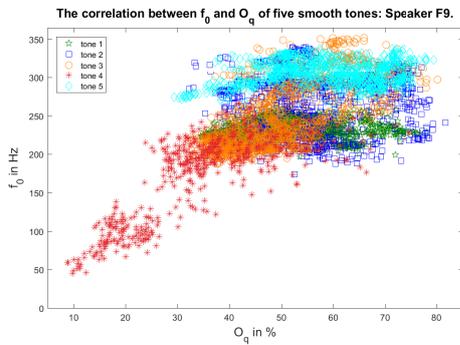
D.3 The correlation between fundamental frequency and open quotient by a scatter plot



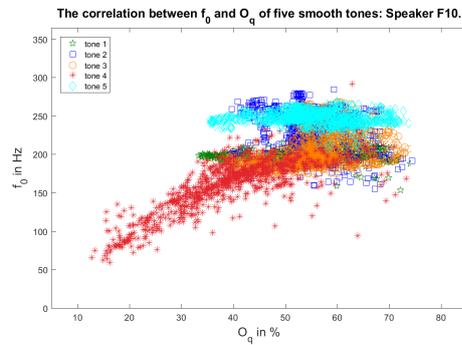
(a) Speaker F3



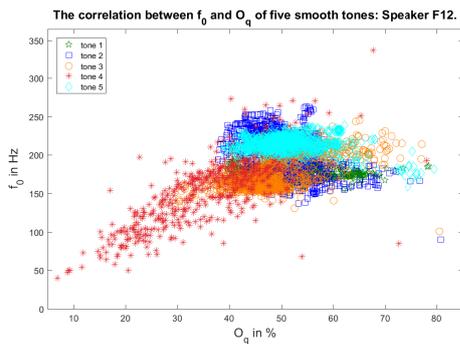
(b) Speaker F7



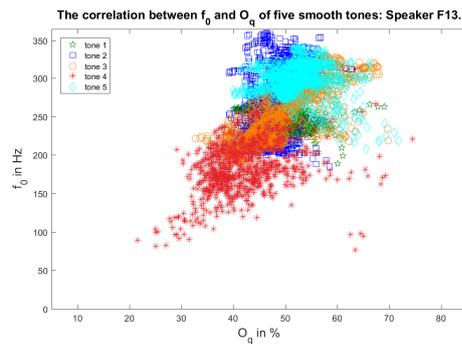
(c) Speaker F9



(d) Speaker F10



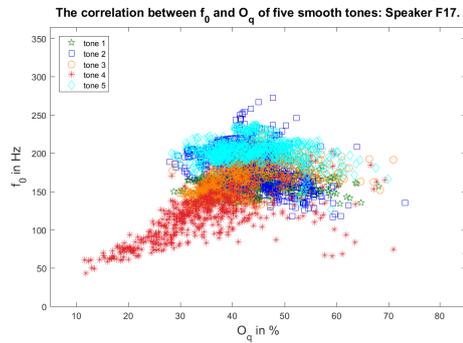
(e) Speaker F12



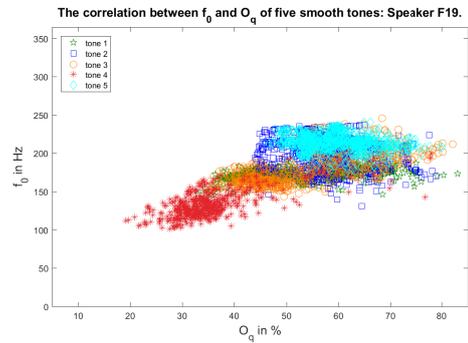
(f) Speaker F13

Figure D.3: The correlation between  $f_0$  <sub>dEgg</sub> and  $O_q$  <sub>dEgg</sub>.

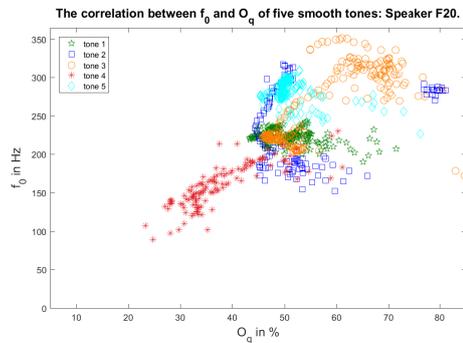
Appendix D Additional graphs produced as a result of this study



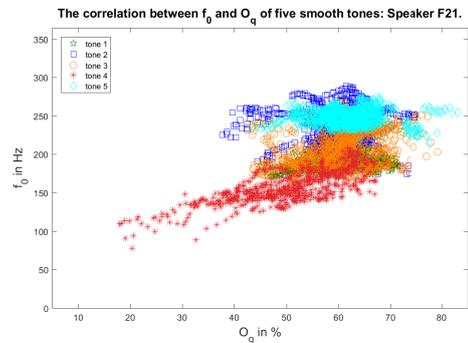
(g) Speaker F17



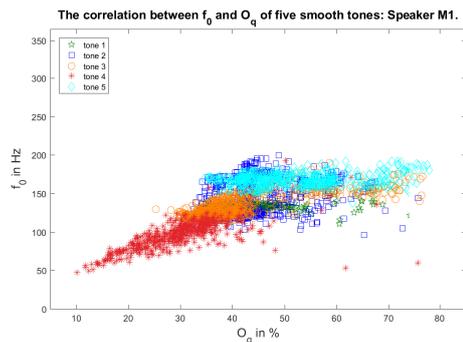
(h) Speaker F19



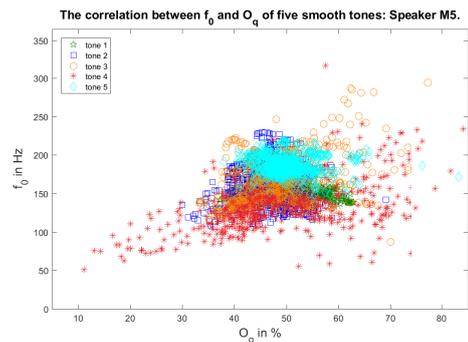
(i) Speaker F20



(j) Speaker F21



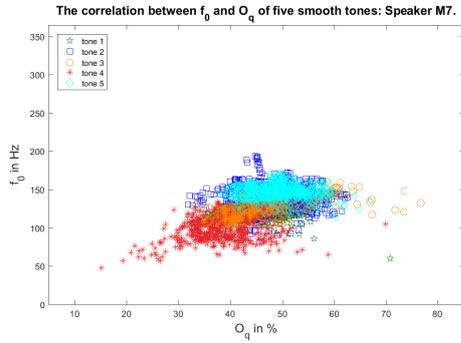
(k) Speaker M1



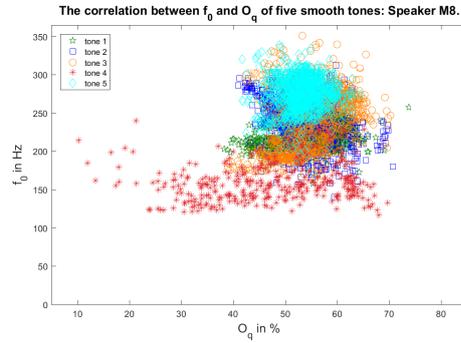
(l) Speaker M5

Fig D.3: The correlation between  $f_0$  <sub>dEGG</sub> and  $O_q$  <sub>dEGG</sub>.

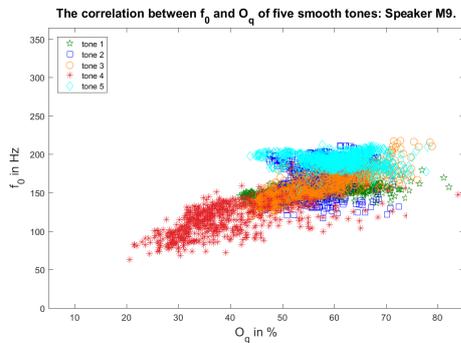
D.3 The correlation between fundamental frequency and open quotient by a scatter plot



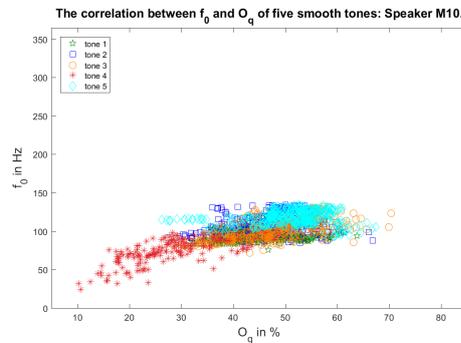
(m) Speaker M7



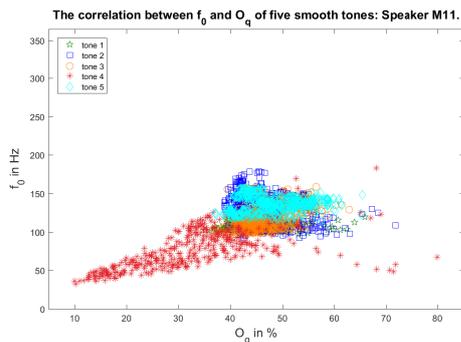
(n) Speaker M8



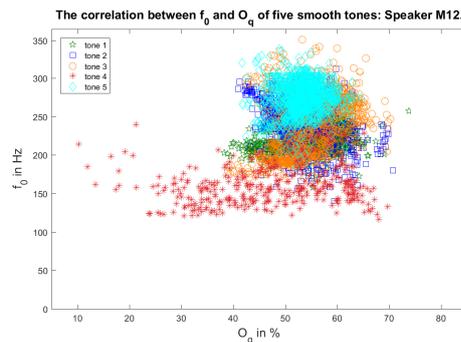
(o) Speaker M9



(p) Speaker M10

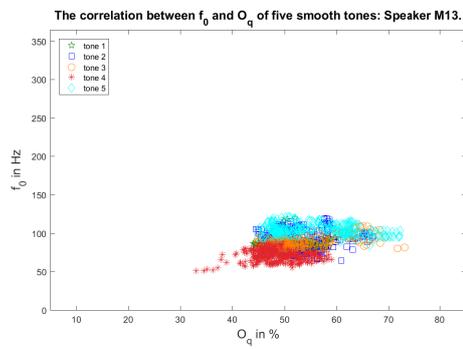


(q) Speaker M11

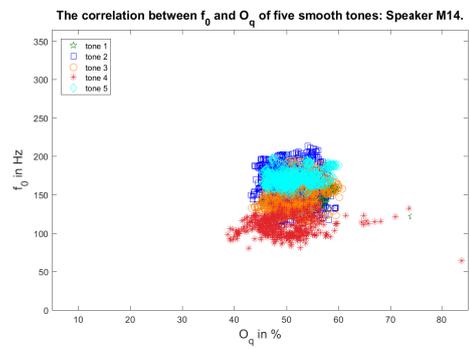


(r) Speaker M12

Fig D.3: The correlation between  $f_{0 \text{ dEGG}}$  and  $O_{q \text{ dEGG}}$ .



(s) Speaker M13



(t) Speaker M14

Fig D.3: The correlation between  $f_{0 \text{ dEGG}}$  and  $O_{q \text{ dEGG}}$ .

## Bibliography

- Abramson, Arthur S., Patrick W Nye, and Therapan Luangthongkum (2007). "Voice register in Khmu: Experiments in production and perception." In: *Phonetica* 64.2-3, pp. 80-104.
- Abramson, Arthur S. and Kalaya Tingsabadh (1999). "Thai final stops: cross-language perception." In: *Phonetica* 56.3-4, pp. 111-122.
- Adams, Oliver, Trevor Cohn, et al. (2017). "Phonemic transcription of low-resource tonal languages." In: *Proceedings of ALTA 2017 (Australasian Language Technology Association Workshop)*. Brisbane, pp. 53-60. URL: <https://halshs.archives-ouvertes.fr/halshs-01656683>.
- Adams, Oliver, Benjamin Galliot, et al. (2021). "User-friendly automatic transcription of low-resource languages: plugging ESPnet into Elpis." In: *Proceedings of ComputEL-4: Fourth Workshop on the Use of Computational Methods in the Study of Endangered Languages*. Hawai'i. URL: <https://halshs.archives-ouvertes.fr/halshs-03030529>.
- Anderson, V. (1950). *Training the speaking voice*. Oxford University Press.
- Andruski, Jean E (2006). "Tone clarity in mixed pitch/phonation-type tones." In: *Journal of Phonetics* 34.3, pp. 388-404.
- Anikin, Andrey and Niklas Johansson (2019). "Implicit associations between individual properties of color and sound." In: *Attention, Perception, & Psychophysics* 81.3, pp. 764-777.
- Barker, Milton E. and Muriel A. Barker (1976). *Mường-Vietnamese-English dictionary*. Huntington Beach, CA: Summer Institute of Linguistics, 537 pp.
- Barker, Miriam A. (1993). "Bibliography of Mường and other Vietic language groups, with notes." In: *The Mon-Khmer Studies Journal* 23, pp. 197-243. URL: <http://purl.org/sealang/barker1993bibliography.pdf>.
- Barlow, Jessica A. and Judith A. Gierut (2002). "Minimal pair approaches to phonological remediation." In: *Seminars in speech and language*. Vol. 23. 01, pp. 57-68.
- Batliner, A. et al. (1993). "MÜSLI: A Classification Scheme For Laryngealizations." In: *Proc. ESCA Workshop on Prosody*, pp. 176-179.
- Bell, Alexander Melville (1867). *Visible speech: The science of universal alphabets*. N. Trubner & Co, London, New York.
- Blomgren, M. et al. (1998). "Acoustic, aerodynamic, physiologic, and perceptual properties of modal and vocal fry registers." In: *Journal of the Acoustical Society of America* 103.5, pp. 2649-2658.
- Bowler, Ned W. (1964). "A fundamental frequency analysis of harsh vocal quality." In: *Communications Monographs* 31.2, pp. 128-134.

## Bibliography

- Brunelle, Marc (2009a). “Northern and Southern Vietnamese tone coarticulation: A comparative case study.” In: *Journal of Southeast Asian Linguistics* 1, pp. 49–62.
- (2009b). “Tone perception in Northern and Southern Vietnamese.” In: *Journal of Phonetics* 37, pp. 79–96.
- (2012). “Dialect experience and perceptual integrality in phonological registers: Fundamental frequency, voice quality and the first formant in Cham.” In: *Journal of the Acoustical Society of America* 131.4, pp. 3088–3102.
- (2015). “Effects of lexical frequency and lexical category on the duration of Vietnamese syllables.” In: *Proceedings of ICPhS XVIII*. Glasgow.
- Brunelle, Marc and Joshua Finkeldey (2011). “Tone perception in Sgaw Karen.” In: *Proceedings of the 17th International Congress of Phonetic Sciences (ICPhS 17)*, pp. 372–375.
- Brunelle, Marc, Kiều Phương Hạ, and Martine Grice (2012). “Intonation in Northern Vietnamese.” In: *The Linguistic Review* 29.1, pp. 3–36.
- Brunelle, Marc, Nguyen Khac Hung, and Nguyen Duy Duong (2010). “A laryngographic and laryngoscopic study of Northern Vietnamese tones.” In: *Phonetica* 67.3, pp. 147–169.
- Brunelle, Marc and Stefanie Jannedy (2007). “Social effects on the perception of Vietnamese tones.” In: *International Congress of Phonetic Sciences*. Saarbrücken, pp. 1461–1464.
- Brunelle, Marc and James Kirby (2016). “Tone and phonation in Southeast Asian languages.” In: *Language and Linguistics Compass* 10.4, pp. 191–207.
- Brunelle, Marc, Duy Duong Nguyễn, and Khac Hùng Nguyễn (2010). “A laryngographic and laryngoscopic study of Northern Vietnamese tones.” In: *Phonetica* 67.3, pp. 147–169. URL: <http://www.karger.com/Article/Abstract/321053>.
- Brunelle, Marc, Thành Tấn Tạ, et al. (2020). “Transphonologization of voicing in Chru: Studies in production and perception.” In: *Laboratory Phonology: Journal of the Association for Laboratory Phonology* 11.1, p. 15. DOI: [10.5334/labphon.278](https://doi.org/10.5334/labphon.278).
- Bùi, Thiện (2010). *Dân ca Mường: Phần tiếng Mường*. NXB Văn hóa Dân tộc.
- Burnham, Denis and Elizabeth Francis (1997). “The role of linguistic experience in the perception of Thai tones.” In: *Southeast Asian linguistic studies in honour of Vichin Panupong*, pp. 29–47.
- Catford, John Cunnison (1977). *Fundamental problems in phonetics*. Edinburgh: Edinburgh University Press.
- Chan, Marjorie K.M. (1998). “Sentence particles je and jek in Cantonese and their distribution across gender and sentence types.” In: *Engendering Communication: Proceedings of the Fifth Berkeley Women and Language Conference*. Ed. by Suzanne Wertheim, Ashlee C. Bailey, and Monica Corston-Oliver. Berkeley: UC Berkeley, pp. 117–128.
- Chao, Yuen-ren (1930). “A system of tone letters.” In: *Le Maître phonétique* 45, pp. 24–27.
- (1934). “The non-uniqueness of phonemic solutions of phonetic systems.” In: *Bulletin of the Institute of History and Philology, Academia Sinica* 4.4. Reprinted in:

- Readings in linguistics, ed. by Martin Joos, 38-54. New York: American Council of Learned Societies, 1958, pp. 363-397.
- Chen, Matthew Y. (2000). *Tone sandhi: Patterns across Chinese dialects*. Cambridge Studies in Linguistics 92. Cambridge: Cambridge University Press.
- Childers, Donald G. and Ashok K. Krishnamurthy (1984). "A critical review of electroglottography." In: *Critical reviews in biomedical engineering* 12.2, pp. 131-161.
- Childers, Donald G. and C.K. Lee (1991). "Vocal quality factors: Analysis, synthesis and perception." In: *Journal of the Acoustical Society of America* 90.5, pp. 2394-2410.
- Christophersen, Paul (1952). "The glottal stop in English." In: URL: <http://www.tandfonline.com/doi/pdf/10.1080/00138385208596879>.
- Clements, Nick, Alexis Michaud, and Cédric Patin (2011). "Do we need tone features?" In: *Tones and Features*. Ed. by Elizabeth Hume, John Goldsmith, and W. Leo Wetzels. Berlin: De Gruyter Mouton, pp. 3-24.
- Coleman, Robert F. (1963). "Decay characteristics of vocal fry." In: *Folia Phoniatica et Logopaedica* 15.4, pp. 256-263.
- Colton, Raymond H. and Edward G. Conture (1990). "Problems and pitfalls of electroglottography." In: *Journal of Voice* 4.1, pp. 10-24.
- Creissels, Denis (1994). *Aperçu sur les structures phonologiques des langues négro-africaines*. Grenoble: ELLUG.
- Cruttenden, Alan (1986). *Intonation*. Cambridge Textbooks in Linguistics. Cambridge: Cambridge University Press.
- Crystal, David (1975). *The English tone of voice: Essays in intonation, prosody and paralanguage*. Hodder Arnold.
- (2011). *Dictionary of linguistics and phonetics*. Vol. 30. John Wiley & Sons.
- Cuisinier, Jeanne (1948). *Les Mu'ò'ng: Géographie humaine et sociologie*. Vol. 45. Institut d'ethnologie.
- Dallaston, Katherine and Gerard Docherty (2019). "Estimating the prevalence of creaky voice: A fundamental frequency-based approach." In: *Proceedings of the 19th International Congress of Phonetic Sciences. Australasian Speech Science and Technology Association Inc*, pp. 532-536.
- Đao, Dich Muc and Anh-Thu T Nguyen (2018). "Acoustic Correlates of Statement and Question Intonation in Southern Vietnamese." In: *Journal of the Southeast Asian Linguistics Society* 11.2, pp. 19-41.
- Davidson, Lisa (2021). "The versatility of creaky phonation: Segmental, prosodic, and sociolinguistic uses in the world's languages." In: *WIREs Cognitive Science* 12.3, e1547. DOI: <https://doi.org/10.1002/wcs.1547>. URL: <https://wires.onlinelibrary.wiley.com/doi/abs/10.1002/wcs.1547>.
- De Boer, Bart (2011). "First formant difference for /i/ and /u/: A cross-linguistic study and an explanation." In: *Journal of Phonetics* 39.1, pp. 110-114.
- Degottex, Gilles et al. (2014). "COVAREP: a collaborative voice analysis repository for speech technologies." In: *Acoustics, Speech and Signal Processing (ICASSP), 2014*

## Bibliography

- IEEE International Conference on*. DOI 10.1109/ICASSP.2014.6853739. IEEE, pp. 960–964. URL: <https://github.com/covarep/covarep>.
- Di Cristo, Albert (1998). “Intonation in French.” In: *Intonation systems: a survey of twenty languages*. Ed. by Daniel Hirst and Albert Di Cristo. Cambridge: Cambridge University Press, pp. 195–218.
- DiCanio, Christian T (2009). “The phonetics of register in Takhian Thong Chong.” In: *Journal of the International Phonetic Association*, pp. 162–188.
- Diffloth, Gérard (1989). “Proto-Austroasiatic creaky voice.” In: *Mon-Khmer Studies* 15, pp. 139–154.
- Dilley, L. and S. Shattuck-Hufnagel (1996). “Glottalization of word-initial vowels as a function of prosodic structure.” In: *Journal of Phonetics* 24, pp. 423–444.
- Ding, Picus Sizhi (2013). “From intonation to tone: the case of utterance-final particles “aa” and “wo” in Cantonese [article in Chinese].” In: *Modern Linguistics* 1.2, pp. 36–41.
- Do, The Dung, Thien Hường Tran, and Georges Boulakia (1998). “Intonation in Vietnamese.” In: *Intonation systems: a survey of twenty languages*. Ed. by Daniel Hirst and Albert Di Cristo. Cambridge: Cambridge University Press, pp. 395–416.
- Dobrovolsky, Michael and Francis Katamba (1996). “Phonology: the function and patterning of sounds.” In: *Contemporary Linguistics: An Introduction*. Essex: Addison Wesley Longman Limited.
- Dollaghan, Christine, Maureen Biber, and Thomas Campbell (1993). “Constituent syllable effects in a nonsense-word repetition task.” In: *Journal of Speech, Language, and Hearing Research* 36.5, pp. 1051–1054. ISSN: 1092-4388, 1558-9102. DOI: [10.1044/jshr.3605.1051](https://doi.org/10.1044/jshr.3605.1051). URL: <http://pubs.asha.org/doi/10.1044/jshr.3605.1051>.
- Downing, Laura J. and Annie Rialland (2016). “Introduction.” In: *Intonation in African tone languages*. Ed. by Laura J. Downing and Annie Rialland. Phonology and Phonetics 24. Berlin: De Gruyter, pp. 1–16.
- Draxler, Christoph and Klaus Jänsch (2004). “SpeechRecorder-a Universal Platform Independent Multi-Channel Audio Recording Software.” In: *LREC*. Citeseer.
- Drugman, Thomas, John Kane, and Christer Gobl (2014). “Data-driven detection and analysis of the patterns of creaky voice.” In: *Computer Speech & Language* 28.5, pp. 1233–1253.
- Earle, M.A. (1975). *An acoustic phonetic study of Northern Vietnamese tones*. Tech. rep. SCRL Monograph. Santa Barbara: Speech Communications Research Laboratory, p. 211.
- Enfield, Nicholas J and Bernard Comrie (2015). “Mainland Southeast Asian languages: State of the art and new directions.” In: *Languages of Mainland Southeast Asia: The state of the art*.
- Erickson, Donna (2005). “Expressive speech: Production, perception and application to speech synthesis.” In: *Acoustical science and technology* 26.4, pp. 317–325.
- Esling, John (1984). “Laryngographic study of phonation type and laryngeal configuration.” In: *Journal of the International Phonetic Association* 14, pp. 56–73.
- (1996). “Pharyngeal consonants and the aryepiglottic sphincter.” In: *Journal of the International Phonetic Association* 26.02, pp. 65–88.

- (1999). "The IPA Categories "Pharyngeal" and "Epiglottal" Laryngoscopic Observations of Pharyngeal Articulations and Larynx Height." In: *Language and Speech* 42.4, pp. 349-372.
- Esling, John, Katherine E Fraser, and Jimmy G Harris (2005). "Glottal stop, glottalized resonants, and pharyngeals: A reinterpretation with evidence from a laryngoscopic study of Nuuchahnulth (Nootka)." In: *Journal of Phonetics* 33.4, pp. 383-410.
- Esling, John, Scott R. Moisik, et al. (2019). *Voice quality: the laryngeal articulator model*. Vol. 162. Cambridge University Press.
- Fabre, Philippe (1957). "Un procédé électrique percutané d'inscription de l'accolement glottique au cours de la phonation: glottographie de haute fréquence." In: *Bulletin de l'Académie Nationale de Médecine* 141, pp. 66-69.
- (1958). "Etude comparée des glottogrammes et des phonogrammes de la voix humaine." In: *Annuaire Oto-rhino Laryngologie* 75, pp. 767-775.
- (1959). "La glottographie électrique en haute fréquence: Particularités de l'appareillage." In: *Comptes rendus des séances de la Société de biologie et de ses filiales* 153.8-9, pp. 1361-1364.
- (1961). "Glottographie respiratoire, appareillage et premiers résultats." In: *Comptes rendus hebdomadaires des séances* 252.9, p. 1386.
- Ferlus, Michel (1979). "Formation des registres et mutations consonantiques dans les langues mon-khmer." In: *Mon-Khmer Studies* 8, pp. 1-76.
- (1982). "Spirantisation des obstruantes médiales et formation du système consonantique du vietnamien." In: *Cahiers de linguistique - Asie Orientale* 11.1, pp. 83-106.
- (1995). "Particularités du dialecte vietnamien de Cao Lao Hạ (Quảng Bình, Vietnam)." In: *Dixièmes Journées de Linguistique d'Asie Orientale*. Paris. URL: <http://halshs.archives-ouvertes.fr/halshs-00922735>.
- (1996). "Langues et peuples viet-muong." In: *Mon-Khmer Studies* 26, pp. 7-28.
- (1998). "Les systèmes de tons dans les langues viet-muong." In: *Diachronica* 15.1, pp. 1-27.
- (1999). "Les disharmonies tonales en viet-muong et leurs implications historiques [Irregular tonal correspondences within Vietic and their historical implications]." In: *Cahiers de Linguistique - Asie Orientale* 28.1, pp. 83-99.
- (2001). "Les hypercorrections dans le thổ de Làng Lỗ (Nghê An, Vietnam) ou les pièges du comparatisme." In: *Quinzièmes Journées de Linguistique de l'Asie Orientale*. Ecole des Hautes Etudes en Sciences Sociales, Paris. URL: <http://halshs.archives-ouvertes.fr/halshs-00922722>.
- (2004). "The origin of tones in Viet-Muong." In: *Papers from the Eleventh Annual Meeting of the Southeast Asian Linguistics Society 2001*. Ed. by Somsong Burusphat. Tempe, Arizona: Arizona State University Programme for Southeast Asian Studies Monograph Series Press, pp. 297-313.
- (2009). "What were the four divisions of Middle Chinese?" In: *Diachronica* 26.2, pp. 184-213.
- (2013). "Arem, a Vietic language (an overview)." In: September 4-5, 2013. Canberra: The Australian National University.

Bibliography

- Ferlus, Michel (2014). "Arem, a Vietic language." In: *Mon-Khmer Studies* 43, pp. 1–15. URL: <http://www.mksjournal.org/mksj43.pdf#page=5>.
- Fernández-Llamazares, Álvaro et al. (2021). "Scientists' Warning to Humanity on Threats to Indigenous and Local Knowledge Systems." In: *Journal of Ethnobiology* 41.2, pp. 144–169. DOI: [10.2993/0278-0771-41.2.144](https://doi.org/10.2993/0278-0771-41.2.144). URL: <https://doi.org/10.2993/0278-0771-41.2.144>.
- Fónagy, Ivan (1983). *La vive voix: essais de psycho-phonétique*. Ed. by Louis-Jean Calvet. "Langages et Sociétés". Paris: Payot.
- Fougeron, Cécile (1999). "Prosodically conditioned articulatory variations: A review." In: *UCLA Working Papers in Phonetics* 97, pp. 1–68.
- Fù, Mào jì (1981). *Nàxīyǔ túhuà wénzì "bái biānfú qǔ jīng jì" yánjiū* (A study of a Naxi pictographic manuscript, "White Bat's Search for Sacred Books", vol.1). Computational Analyses of Asian and African Languages: Monograph Series 6 6. Tokyo: CAAAL.
- Gandour, Jack (1974). "Consonant types and tones in Siamese." In: *Journal of Phonetics* 2, pp. 37–50.
- Gao, Jiayin (2015). "Interdependence between tones, segments and phonation types in Shanghai Chinese." Ph.D. Université Paris 3-Sorbonne Nouvelle.
- (2016). "Sociolinguistic motivations in sound change: On-going loss of low tone breathy voice in Shanghai Chinese." In: *Papers in Historical Phonology* 1, pp. 166–186.
- Garellek, Marc (2015). "Perception of glottalization and phrase-final creak." In: *The Journal of the Acoustical Society of America* 137.2, pp. 822–831.
- Garellek, Marc, Matthew Gordon, et al. (2020). "Toward open data policies in phonetics: What we can gain and how we can avoid pitfalls." In: *Journal of Speech Science* 9.1. ISSN: 2236-9740. URL: <https://halshs.archives-ouvertes.fr/halshs-02894375>.
- Garellek, Marc, Patricia Keating, et al. (2013). "Voice quality and tone identification in White Hmong." In: *The Journal of the Acoustical Society of America* 133.2, pp. 1078–1089.
- Gerratt, Bruce R. and Jody Kreiman (2001). "Toward a taxonomy of nonmodal phonation." In: *Journal of Phonetics* 29.4, pp. 365–381.
- Gobl, Christer (2010). "Voice Source Variation and Its Communicative Functions." In: *The handbook of phonetic sciences*, pp. 378–423.
- Gordon, Matthew and Peter Ladefoged (2001). "Phonation types: a cross-linguistic overview." In: *Journal of Phonetics* 29, pp. 383–406.
- Goswami, Usha, Jean Emile Gombert, and Lucia Fraca de Barrera (Jan. 1998). "Children's orthographic representations and linguistic transparency: Nonsense word reading in English, French, and Spanish." In: *Applied Psycholinguistics* 19.1. Publisher: Cambridge University Press, pp. 19–52. ISSN: 1469-1817, 0142-7164. DOI: [10.1017/S0142716400010560](https://doi.org/10.1017/S0142716400010560).
- Gruber, James Frederick (2011). "An articulatory, acoustic, and auditory study of Burmese tone." In: URL: <https://repository.library.georgetown.edu/handle/10822/558130>.

- Gsell, René (1980). "Remarques sur la structure de l'espace tonal en vietnamien du sud (parler de Saïgon)." In: *Cahier d'études vietnamiennes, Département de Langues et Civilisations de l'Asie Orientale de l'Université Paris 7* 4.
- Gussenhoven, Carlos (2002). "Phonology of intonation." In: *Glott International* 6.9/10, pp. 271–284.
- Ha, Kieu-Phuong and Martine Grice (2010). "Modelling the interaction of intonation and lexical tone in Vietnamese." In: Chicago.
- Hampala, Vít et al. (2016). "Relationship between the electroglottographic signal and vocal fold contact area." In: *Journal of Voice* 30.2, pp. 161–171.
- Han, Miekko S. and Kong-On Kim (1974). "Phonetic variation of Vietnamese tones in disyllabic utterances." In: *Journal of Phonetics* 2, pp. 223–232.
- Hanson, Helen M (2009). "Effects of obstruent consonants on fundamental frequency at vowel onset in English." In: *The Journal of the Acoustical Society of America* 125.1, pp. 425–441.
- Hao, Zhang (2015). *Electroglottograph features*. URL: <https://www.slideshare.net/HaoZhang12/031215-eg2-pcx2electroglottographfeatures>.
- Harris, Jimmy G (1999). "States of the glottis for voiceless plosives." In: *Proceedings of the 14th international congress of phonetic sciences*. Vol. 3, pp. 2041–2044.
- (2001). "States of the glottis of Thai voiceless stops and affricates." In: *Essays in Tai linguistics*. Ed. by Kalaya Tingsabadh and Arthur S. Abramson. Bangkok: Chulalongkorn University Press, pp. 3–11.
- Haudricourt, André-Georges (1952). "Les voyelles brèves du vietnamien." In: *Bulletin de la Société de Linguistique de Paris* 48.1. Translated into English by Alexis Michaud, pp. 90–93. URL: <https://hal.archives-ouvertes.fr/halshs-01631474/>.
- (1953). "La place du vietnamien dans les langues austroasiatiques." In: *Bulletin de la Société de Linguistique de Paris* 49.1. Translated into English by Alexis Michaud, pp. 122–128. URL: <https://halshs.archives-ouvertes.fr/halshs-01631477/>.
- (1954a). "Comment reconstruire le chinois archaïque." In: *Word* 10.2-3. Translated into English by Guillaume Jacques, pp. 351–364. URL: <https://halshs.archives-ouvertes.fr/halshs-01631479/>.
- (1954b). "De l'origine des tons en vietnamien." In: *Journal Asiatique* 242. Translated into English by Marc Brunelle, pp. 69–82. URL: <https://halshs.archives-ouvertes.fr/halshs-01678018/>.
- (1961). "Bipartition et tripartition des systèmes de tons dans quelques langues d'Extrême-Orient." In: *Bulletin de la Société de Linguistique de Paris* 56.1. Translated into English by Christopher Court, pp. 163–80. URL: <http://sealang.net/sala/archives/pdf4/haudricourt1972two.pdf>.
- (2010). "The origin of the peculiarities of the Vietnamese alphabet." In: *Mon-Khmer Studies* 39, pp. 89–104.
- Hay, Jennifer, Katie Drager, and Brynmor Thomas (2013a). "Using nonsense words to investigate vowel merger." In: *English Language and Linguistics* 17.2. Publisher: Cambridge University Press, p. 241.

## Bibliography

- Hay, Jennifer, Katie Drager, and Brynmor Thomas (2013b). "Using nonsense words to investigate vowel merger1." In: *English Language & Linguistics* 17.2, pp. 241–269.
- Hayes, La Vaughn H. (1992). "Vietic and Việt-Muông: a new subgrouping in Mon-Khmer." In: *Mon-Khmer Studies* 21, pp. 211–228.
- Hedelin, P. and D. Huber (1990). "Pitch period determination of aperiodic speech signals." In: *IEEE*, pp. 361–364.
- Heffner, R-MS (1950). *General phonetics*. University of Wisconsin Press, Madison.
- Henderson, Eugénie JA (1952). "The main features of Cambodian pronunciation." In: *Bulletin of the School of Oriental and African studies, University of London*, pp. 149–174.
- Henrich, Nathalie (2001). "Etude de la source glottique en voix parlée et chantée : modélisation et estimation, mesures acoustiques et électroglottographiques, perception." Ph. D. PhD thesis. Paris: Université Paris 6, Acoustique.
- Henrich, Nathalie et al. (2004). "On the use of the derivative of electroglottographic signals for characterization of nonpathological phonation." In: *Journal of the Acoustical Society of America* 115.3, pp. 1321–1332.
- Herbst, Christian T. (2020). "Electroglottography – An Update." In: *Journal of Voice* 34.4, pp. 503–526. ISSN: 0892-1997. DOI: <https://doi.org/10.1016/j.jvoice.2018.12.014>. URL: <https://www.sciencedirect.com/science/article/pii/S0892199718304612>.
- Hoang, Cao Cường (1985). "Bước đầu nhận xét về đặc điểm ngữ điệu tiếng Việt (trên cú liệu thực nghiệm)." In: *Tạp chí Ngôn ngữ (3) [Vietnamese linguistic journal]* 3, pp. 40–49.
- Hockett, Charles F. (1963). "The problem of universals in language." In: *Universals of language*. Ed. by Joseph H. Greenberg. Vol. 2. MIT press Cambridge, Mass., pp. 1–29.
- Holder, William (1669). *Elements of speech: An essay of inquiry into the natural production of letters*. London:T.N. for J. Martyn Printer to the R. Society.
- Hollien, Harry (1963). "An investigation of vocal fry." In: *National Institute of Health research grants*, pp. 5–8.
- Hollien, Harry and John F. Michel (1968). "Vocal fry as a phonational register." In: *Journal of Speech, Language, and Hearing Research* 11.3, pp. 600–604. URL: <http://dx.doi.org/10.1044/jshr.1103.600>.
- Hollien, Harry, P. Moore, et al. (1966). "On the nature of vocal fry." In: *Journal of Speech and Hearing Research* 9, pp. 245–247.
- Hollien, Harry and Ronald W. Wendahl (Mar. 1968). "Perceptual study of vocal fry." In: *The Journal of the Acoustical Society of America* 43.3, pp. 506–509. ISSN: 0001-4966.
- Hombert, Jean-Marie (1978). "Consonant types, vowel quality and tone." In: *Tone : a Linguistic Survey*. Ed. by Victoria A. Fromkin. New York: Academic Press, pp. 77–111.
- Howie, J.M. (1974). "On the domain of tone in Mandarin." In: *Phonetica* 30, pp. 129–148.
- Hyman, Larry M. (2007). "Kuki-Thaadow: an African tone system in Southeast Asia." In: *Annual Report of UC Berkeley Phonology Lab*, pp. 1–19.
- Jacques, Guillaume (2006). "Introduction à la phonologie historique du chinois." In:

- (2011). “Tonal alternations in the Pumi verbal system.” In: *Language and Linguistics* 12.2, pp. 359–392.
- Jones, Daniel (1956). *An outline of English phonetics*. (8th ed), E.P. Dutton & Co, New York.
- Jongman, Allard et al. (2006). *Perception and production of Mandarin Chinese tones*. na.
- Jun, Sun-Ah (2005). “Prosodic typology.” In: *Prosodic typology: the phonology of intonation and phrasing*. Ed. by Sun-Ah Jun. Oxford: Oxford University Press, pp. 430–458.
- Karlsson, Anastasia, David House, and Jan-Olof Svantesson (2012). “Intonation adapts to lexical tone: the case of Kammu.” In: *Phonetica* 69.1-2, pp. 28–47.
- Keating, Patricia, Christina Esposito, et al. (2010). “Phonation contrasts across languages.” In: *UCLA Working Papers in Phonetics* 108, pp. 188–202.
- Keating, Patricia, Marc Garellek, and Jody Kreiman (2015). “Acoustic properties of different kinds of creaky voice.” In: *Proceedings of the 18th International Congress of Phonetic Sciences, Glasgow*.
- Kingston, John (2009). “Segmental influences on Fo: Automatic or controlled?” In: *Volume 2 Experimental Studies in Word and Sentence Prosody*. De Gruyter Mouton, pp. 171–210.
- Kirby, James (2010). “Dialect experience in Vietnamese tone perception.” In: *Journal of the Acoustical Society of America* 127.6, pp. 3749–3757.
- (2011). “Vietnamese (Hanoi Vietnamese).” In: *Journal of the International Phonetic Association* 41.3, pp. 381–392.
- (2014). “Incipient tonogenesis in Phnom Penh Khmer: Acoustic and perceptual studies.” In: *Journal of Phonetics* 43, pp. 69–85.
- (2018). “Onset pitch perturbations and the cross-linguistic implementation of voicing: Evidence from tonal and non-tonal languages.” In: *Journal of Phonetics* 71, pp. 326–354.
- Kirk, Paul L, Jenny Ladefoged, and Peter Ladefoged (1993). “Quantifying acoustic properties of modal, breathy and creaky vowels in Jalapa Mazatec.” In: *American Indian linguistics and ethnography in honor of Laurence C. Thompson*, pp. 435–450.
- Klatt, Dennis and Laura Klatt (1990). “Analysis, synthesis, and perception of voice quality variations among female and male talkers.” In: *Journal of the Acoustical Society of America* 87, pp. 820–857.
- Kochanski, Greg P. and Chilin Shih (2003). “A Quasi-glottogram signal for voicing and power estimation.” In: *Journal of the Acoustical Society of America* 114.4, pp. 2206–2216.
- Kohler, Klaus J. (1994). “Glottal stops and glottalization in German.” In: *Phonetica* 51, pp. 38–51.
- (1998). “The development of sound systems in human language.” In: *Approaches to the evolution of language*. Ed. by J. Hurford, C. Knight, and Michael Studdert-Kennedy. Cambridge: Cambridge University Press, pp. 265–278.

Bibliography

- Kohler, Klaus J. and Oliver Niebuhr (2011). "On the role of articulatory prosodies in German message decoding." In: *Phonetica* 68, pp. 1–31.
- Kuang, Jianjing (2013a). "Phonation in Tonal Contrasts." Ph. D. thesis. PhD thesis. Los Angeles: University of California.
- (2013b). "The tonal space of contrastive five level tones." In: *Phonetica* 70.1-2, pp. 1–23. DOI: [10.1159/000353853](https://doi.org/10.1159/000353853).
- (Aug. 2017). "Creaky voice as a function of tonal categories and prosodic boundaries." In: ISCA, pp. 3216–3220.
- Kwok, Helen (1984). *Sentence particles in Cantonese*. Hong Kong: Centre of Asian Studies.
- Labov, William (1994). *Principles of linguistic change. Internal factors*. Language in Society 20. Oxford: Basil Blackwell.
- (2001). *Principles of linguistic change. Social factors*. Language in Society 29. Oxford: Basil Blackwell.
- (2010). *Principles of linguistic change. Volume 3: Cognitive and cultural factors*. Language in Society 39. Oxford, UK & Malden, USA: Wiley-Blackwell.
- Ladd, Robert (1981). "On intonational universals." In: *Advances in Psychology*. Vol. 7. Elsevier, pp. 389–397.
- (2008). *Intonational phonology*. Cambridge University Press. ISBN: 1-139-47399-9.
- Ladefoged, Peter (1975). *A Course in Phonetics*. New York: Harcourt Brace College.
- (May 1981). "The relative nature of voice quality." In: *Journal of the Acoustical Society of America* 69.S1, S67–S67. ISSN: 0001-4966. DOI: [10.1121/1.386168](https://doi.org/10.1121/1.386168). URL: <http://asa.scitation.org/doi/10.1121/1.386168>.
- (1983). "The linguistic use of different phonation types." In: *Vocal fold physiology: Contemporary research and clinical issues*, pp. 351–360.
- Ladefoged, Peter and Ian Maddieson (1996). *The Sounds of the World's Languages*. Ed. by M. Kenstowicz et al. Phonological Theory. Oxford, U.K. & Cambridge, Massachusetts: Blackwell.
- Laver, John (1980). *The phonetic description of voice quality*. Cambridge: Cambridge University Press.
- (1994). *Principles of Phonetics*. Cambridge: Cambridge University Press.
- Lee, Wai-Sum (2003). "A phonetic study of the neutral tone in Beijing Mandarin." In: *Proceedings of the 15th International Congress of Phonetic Sciences (ICPHS 2003)*.
- Lee, Wai-Sum and Eric Zee (2017). *Tone and intonation*. Ed. by Rint Sybesma. Leiden/Boston.
- Lehiste, Ilse (1975). "The phonetic structure of paragraphs." In: *Structure and Process in Speech Perception*. Ed. by A. Cohen and S.G. Noteboom. Berlin: Springer, pp. 195–206.
- Leung, Wai-Mun (2009). "A study of the Cantonese hearsay particle wo from a tonal perspective." In: *International Journal of Linguistics* 1.1, pp. 1–14.
- Lewis, M Paul, Gary F Simons, and Charles D Fennig (2009). *Ethnologue: Languages of the world*. Vol. 9. SIL international Dallas, TX.

- Li, Charles N. and Sandra A. Thompson (1981). *Mandarin Chinese: a functional reference grammar*. Berkeley: University of California Press.
- Li, Wu Wing (2009). "Sentence-final particles in Hong Kong Cantonese: Are they tonal or intonational?" In: *Proceedings of InterSpeech 2009*. Brighton, UK, pp. 2291–2294.
- Lieberman, Mark (1975). "The intonational system of English." Ph. D. PhD thesis. Cambridge, Massachusetts: MIT. Distributed by Indiana University Linguistics Club,
- Lindblom, Björn (1990). "Explaining phonetic variation: a sketch of the H&H theory." In: *Speech production and speech modelling*. Ed. by William J. Hardcastle and Alain Marchal. Dordrecht: Kluwer, pp. 403–439.
- Luksaneeyanawin, Sudaporn (1983). "Intonation in Thai." PhD thesis. University of Edinburgh,
- Mac, Dang-Khoa et al. (2015). "Influences of speaker attitudes on glottalized tones: a study of two Vietnamese sentence-final particles." In: *Proceedings of ICPHS XVIII (18th International Congress of Phonetic Sciences)*. Areas: (i) Tone; (ii) SpeechProsody; (iii) Phonation and Voice Quality. Glasgow.
- Malécot, André (1975). "The glottal stop in French." In: *Phonetica* 31.1, pp. 51–63.
- Marks, Lawrence E. (2014). *The unity of the senses: Interrelations among the modalities*. Google-Books-ID: qjqoBQAAQBAJ. Academic Press. ISBN: 978-1-4832-6033-4.
- Martinet, André (1956). *La Description phonologique avec application au parler franco-provençal d'Hauteville (Savoie)*. Genève: Droz.
- (1981). "Fricatives and spirants." In: *Suniti Kumar Chatterji commemoration volume*. Ed. by Bhakti Prasad Mallik. Burdwan, West Bengal, India: Burdwan University Press, pp. 145–151.
- (1990). "La synchronie dynamique." In: *La linguistique* 26.2, pp. 13–23.
- Maspero, Henri (1912). "Etude sur la phonétique historique de la langue annamite: les initiales." In: *Bulletin de l'Ecole Française d'Extrême-Orient* 12, pp. 1–127.
- Mazaudon, Martine and Alexis Michaud (2008). "Tonal contrasts and initial consonants: a case study of Tamang, a 'missing link' in tonogenesis." In: *Phonetica* 65.4, pp. 231–256.
- McGlone, Robert E. (1967). "Air flow during vocal fry phonation." In: *Journal of Speech, Language, and Hearing Research* 10.2, pp. 299–304.
- Michailovsky, Boyd (1988). *La langue hayu*. Ed. by Sylvain Auroux. Sciences du langage. Paris: CNRS Editions.
- Michailovsky, Boyd et al. (2014). "Documenting and researching endangered languages: the Pangloss Collection." In: *Language Documentation and Conservation* 8, pp. 119–135. ISSN: 1934-5275. URL: <http://hdl.handle.net/10125/4621>.
- Michaud, Alexis (2004a). "A measurement from electroglottography: DECFA, and its application in prosody." In: *Speech Prosody 2004*. Ed. by Bernard Bel and Isabelle Marlien. Nara, Japan, pp. 633–636.
- (2004b). "Final consonants and glottalization: new perspectives from Hanoi Vietnamese." In: *Phonetica* 61.2-3, pp. 119–146.
- (2012). "Monosyllabicization: patterns of evolution in Asian languages." In: *Monosyllables: from phonology to typology*. Ed. by Nicole Nau, Thomas Stolz, and Cornelia

## Bibliography

- Stroh. Berlin: Akademie Verlag, pp. 115–130. URL: <http://halshs.archives-ouvertes.fr/halshs-00436432/>.
- Michaud, Alexis (2017). *Tone in Yongning Na: lexical tones and morphotonology*. Studies in Diversity Linguistics 13. Berlin: Language Science Press. ISBN: 978-3-946234-86-9. URL: <http://langsci-press.org/catalog/book/109>.
- Michaud, Alexis, Michel Ferlus, and Minh-Châu Nguyễn (2015). “Strata of standardization: the Phong Nha dialect of Vietnamese (Quảng Bình Province) in historical perspective.” In: *Linguistics of the Tibeto-Burman Area* 38.1, pp. 124–162. DOI: [10.1075/ltba.38.1.04mic](https://doi.org/10.1075/ltba.38.1.04mic).
- Michaud, Alexis and Barbara Kühnert (2006). “A pilot study on the Fo curve of syllable-initial sonorants, comparing nasals, lenis stops and fortis stops.” In: *13e Colloque de Villetaneuse sur l’anglais oral (ALOES 2006)*. Villetaneuse, France. URL: <https://halshs.archives-ouvertes.fr/halshs-01631243>.
- Michaud, Alexis and Tuan Vu-Ngoc (2004). “Glottalized and nonglottalized tones under emphasis: open quotient curves remain stable, Fo curve is modified.” In: *Speech Prosody 2004*. Ed. by Bernard Bel and Isabelle Marlien. Nara, Japan, pp. 745–748.
- Michaud, Alexis, Tuấn Vu-Ngoc, et al. (2006). “Nasal release, nasal finals and tonal contrasts in Hanoi Vietnamese: an aerodynamic experiment.” In: *Mon-Khmer Studies* 36, pp. 121–137.
- Michaud, Alexis, Minh-Châu Nguyễn, and Vera Scholvin (2021). “L’intonation dans les langues tonales : des réflexions générales et deux études de cas.” In: *Études de linguistique appliquée (ELA)* 199.1. URL: <https://halshs.archives-ouvertes.fr/halshs-03189736>.
- Michaud, Alexis and Bonny Sands (Aug. 2020). “Tonogenesis.” In: *Oxford Research Encyclopedia of Linguistics*. Oxford University Press. ISBN: 978-0-19-938465-5. DOI: [10.1093/acrefore/9780199384655.013.748](https://doi.org/10.1093/acrefore/9780199384655.013.748). URL: <https://oxfordre.com/linguistics/view/10.1093/acrefore/9780199384655.001.0001/acrefore-9780199384655-e-748>.
- Michaud, Alexis and Jacqueline Vaissière (2015). “Tone and intonation: Introductory notes and practical recommendations.” In: *KALIPHO - Kieler Arbeiten zur Linguistik und Phonetik* 3, pp. 43–80.
- Michaud, Alexis, Jacqueline Vaissière, and Minh-Châu Nguyễn (2015). “Phonetic insights into a simple level-tone system: ‘careful’ vs. ‘impatient’ realizations of Naxi High, Mid and Low tones.” In: *ICPhS XVIII (18th International Congress of Phonetic Sciences)*.
- Michaud, Alexis and He Xueguang (2007). “Reassociated tones and coalescent syllables in Naxi (Tibeto-Burman).” In: *Journal of the International Phonetic Association* 37.3, pp. 237–255.
- Michel, John F. (1964). “Vocal fry and harshness.” PhD Thesis. University of Florida.
- (1968). “Fundamental frequency investigation of vocal fry and harshness.” In: *Journal of Speech, Language, and Hearing Research* 11.3, pp. 590–594. URL: [+%20http://dx.doi.org/10.1044/jshr.1103.590](https://dx.doi.org/10.1044/jshr.1103.590).

- Michel, John F. and Harry Hollien (1968). "Perceptual Differentiation of Vocal Fry and Harshness." In: *Journal of Speech, Language, and Hearing Research* 11.2, pp. 439–443.
- Moisik, Scott R. and John Esling (2014). "Modeling the biomechanical influence of epilaryngeal stricture on the vocal folds: A low-dimensional model of vocal–ventricular fold coupling." In: *Journal of Speech, Language, and Hearing Research* 57.2. Publisher: ASHA, S687–S704.
- Moisik, Scott R., John Esling, et al. (2015). "Multimodal imaging of glottal stop and creaky voice: Evaluating the role of epilaryngeal constriction." In: *The Scottish Consortium for ICPhS*.
- Moore, Paul and Hans Von Leden (1958). "Dynamic variations of the vibratory pattern in the normal larynx." In: *Folia Phoniatica et Logopaedica* 10.4, pp. 205–238. ISSN: 1021-7762, 1421-9972. DOI: [10.1159/000262819](https://doi.org/10.1159/000262819).
- Moser, Henry M. (June 1942). "Symposium on unique cases of speech disorders; presentation of a case." In: *Journal of Speech Disorders* 7.2, pp. 173–174. ISSN: 0885-9426.
- Nguyen, Thi Thanh Hoa and Georges Boulakia (1999). "Another look at Vietnamese intonation." In: *International Congress of Phonetic Sciences*. Vol. 3. San Francisco, pp. 2399–2402.
- Nguyen, Thi-Lan et al. (2013). "The interplay of intonation and complex lexical tones: how speaker attitudes affect the realization of glottalization on Vietnamese sentence-final particles." In: *Proceedings of Interspeech 2013*. Lyon.
- Nguyen, Van-Tai (2005). *Ngữ âm tiếng Mường qua các phương ngữ [The phonetics of the Mường language across its various dialects]*. Hanoi: NXB Từ điển Bách khoa.
- Nguyễn, Minh-Châu (2016). "The tone system of Kim Thượng Mường: an experimental study of fundamental frequency, duration, and phonation types." MA thesis. Hanoi: VNU-USSH - Vietnam National University - Department of Linguistics. URL: <https://dumas.ccsd.cnrs.fr/dumas-01405496/>.
- Nguyễn, Minh-Châu et al. (2019). "A glottalized tone in Muong (Vietic): a pilot study based on audio and electroglottographic recordings." In: *ICPhS XIX (19th International Congress of Phonetic Sciences)*.
- Niebuhr, Oliver (2013). "The acoustic complexity of intonation." In: *Nordic Prosody XI*. Ed. by Eva Liina Asu and Pärtel Lippus. Frankfurt: Peter Lang, pp. 15–29.
- Niebuhr, Oliver and Alexis Michaud (2015). "Speech data acquisition: the underestimated challenge." In: *KALIPHO - Kieler Arbeiten zur Linguistik und Phonetik* 3, pp. 1–42. URL: <https://halshs.archives-ouvertes.fr/halshs-01026295>.
- Noël-Armfield, George (1931). *General phonetics, for missionaries and students of language*. W. Heffer & sons Ltd, Cambridge.
- Nolan, Francis (2003). "Intonational equivalence: an experimental evaluation of pitch scales." In: *Proceedings of the 15th International Congress of Phonetic Sciences*. Barcelona, pp. 771–774.

## Bibliography

- Nolan, Francis (2006). "Intonation." In: *The Handbook of English Linguistics*. Ed. by Bas Aarts and April McMahon. Malden, MA, USA: Blackwell Publishing, pp. 433–457. DOI: [10.1002/9780470753002.ch19](https://doi.org/10.1002/9780470753002.ch19).
- O'Connor, J.D. and G.F. Arnold (1973). *Intonation of colloquial English*. Vol. 2nd edition. London: Longman.
- Ohala, John (1983). "Cross-language use of pitch: an ethological view." In: *Phonetica* 40, pp. 1–18.
- Orlikoff, Robert F. (1998). "Scrambled EGG: the uses and abuses of electroglottography." In: *Phonoscope* 1.1, pp. 37–53.
- Pain, Frédéric et al. (2014). *EFEO-CNRS-SOAS word list for linguistic fieldwork in Southeast Asia*. Tech. rep. Hanoi: International Research Institute MICA. URL: <https://halshs.archives-ouvertes.fr/halshs-01068533/>.
- Phan, John (2012). "Mường is not a subgroup: phonological evidence for a paraphyletic taxon in the Viet-Muong sub-family." In: *Mon-Khmer Studies* 40, pp. 1–18.
- Pierrehumbert, Janet (1980). "The Phonology and Phonetics of English Intonation." Ph. D. Thesis. PhD thesis. Cambridge, Massachusetts: Massachusetts Institute of Technology (distributed by the Indiana University Linguistics Club),
- Pike, Kenneth L. (1947). *Phonemics: a technique for reducing languages to writing*. University of Michigan publications. Linguistics, 3. Ann Arbor: University of Michigan Press.
- (1948). *Tone languages. A technique for determining the number and type of pitch contrasts in a language, with studies in tonemic substitution and fusion*. Ann Arbor: University of Michigan Press.
- Pillot-Loiseau, Claire et al. (Nov. 2019). "The evolution of creaky voice use in read speech by native-French and native-English speakers in tandem: a pilot study." In: *Anglophonia* 27. ISSN: 1278-3331, 2427-0466. DOI: [10.4000/anglophonia.2005](https://doi.org/10.4000/anglophonia.2005). URL: <http://journals.openedition.org/anglophonia/2005>.
- Pittayaporn, Pittayawat (2007). "Prosody of final particles in Thai: Interaction between lexical tones and intonation." In: Saarbrücken. URL: [https://linguistics.ucla.edu/people/jun/Workshop2007ICPhS/Pittayaporn\\_Thai.pdf](https://linguistics.ucla.edu/people/jun/Workshop2007ICPhS/Pittayaporn_Thai.pdf).
- Post, Mark (2006). "Compounding and the structure of the Tani lexicon." In: *Linguistics of the Tibeto-Burman Area* 29.1, pp. 41–60.
- Prochiantz, Alain and Alain Joliot (2003). "Can transcription factors function as cell–cell signalling molecules?" In: *Nature Reviews Molecular Cell Biology* 4.10. Publisher: Nature Publishing Group, pp. 814–819.
- Przedlacka, Joanna (2000). "Estuary English: glottaling in the Home Counties." In: *Oxford University Working Papers in Linguistics, Philology and Phonetics* 5, pp. 19–24.
- Recasens, Daniel and Meritxell Mira (2013). "Voicing assimilation in Catalan three-consonant clusters." In: *Journal of Phonetics* 41.3-4, pp. 264–280.
- Redi, Laura and S. Shattuck-Hufnagel (2001). "Variation in the realization of glottalization in normal speakers." In: *Journal of Phonetics* 29.4, pp. 407–429.
- Rhodes, Alexandre de (1651). *Dictionarium Annamiticum Lusitanum et Latinum*. Rome.

- Rialland, Annie (2007). "Question prosody: an African perspective." In: *Tones and tunes. Volume 1: Typological studies in word and sentence prosody*. Ed. by Tomas Riad and Carlos Gussenhoven. Phonology and Phonetics. Berlin/New York: Mouton de Gruyter, pp. 35–62.
- Rice, Keren (2014). "On beginning the study of the tone system of a Dene (Athabaskan) language: Looking back." In: *Language Documentation and Conservation* 8, pp. 690–706.
- Rock, Joseph (1963). *A Na-Khi – English encyclopedic dictionary*. Serie Orientale Roma, no. 28. Roma: Istituto Italiano per il Medio ed Estremo Oriente.
- Roettger, Timo B., Bodo Winter, and Harald Baayen (Mar. 2019). "Emergent data analysis in phonetic sciences: Towards pluralism and reproducibility." In: *Journal of Phonetics* 73, pp. 1–7. ISSN: 0095-4470. DOI: [10.1016/j.wocn.2018.12.001](https://doi.org/10.1016/j.wocn.2018.12.001). URL: <http://www.sciencedirect.com/science/article/pii/S0095447018300810>.
- Rossi, Mario (Jan. 2001). "L'intonation." In: *Modèles linguistiques* XXII-1.43, pp. 103–137. ISSN: 0249-6267, 2274-0511. DOI: [10.4000/ml.1463](https://doi.org/10.4000/ml.1463). URL: <http://journals.openedition.org/ml/1463>.
- Rothenberg, Martin (1992). "A multichannel electroglottograph." In: *Journal of Voice* 6.1, pp. 36–43.
- Roubeau, Bernard, Nathalie Henrich, and Michèle Castellengo (2009). "Laryngeal vibratory mechanisms: the notion of vocal register revisited." In: *Journal of Voice* 23.4, pp. 425–38.
- Seitz, Philip Franz D (1986). "Relationships between tones and segments in Vietnamese." PhD thesis. Graduate School of Arts and Sciences, University of Pennsylvania.
- Seyfeddinipur, Mandana and Felix Rau (2020). "Keeping it real: Video data in language documentation and language archiving." In: *Language Documentation & Conservation* 14. ISBN: 1934-5275 Publisher: University of Hawaii Press, pp. 503–519.
- Sherman, Dorothy and Eugene Linke (Dec. 1952). "The influence of certain vowel types on degree of harsh voice quality." In: *Journal of Speech and Hearing Disorders* 17.4, pp. 401–408.
- Silverman, D. (1997). *Phrasing and recoverability*. London: Routledge.
- Silverman, D. et al. (1995). "Phonetic structures in Jalapa Mazatec." In: *Anthropological Linguistics* 37, pp. 70–88.
- Sprigg, Richard Keith (1966). "The glottal stop and glottal constriction in Lepcha, and borrowing from Tibetan." In: *Bulletin of Tibetology* 3.1, pp. 5–14.
- Steien, Guri Bordal and Kofi Yakpo (2020). "Romancing with tone: On the outcomes of prosodic contact." In: *Language* 96.1. ISBN: 1535-0665 Publisher: Linguistic Society of America, pp. 1–41.
- Svantesson, Jan-Olof and David House (2006). "Tone production, tone perception and Kammu tonogenesis." In: *Phonology* 23, pp. 309–333.
- Sweet, Henry (1877). *A handbook of phonetics, including a popular exposition of the principles of spelling reform*. Vol. 2. Clarendon Press, Oxford.
- Sybesma, Rint and Boya Li (2007). "The dissection and structural mapping of Cantonese sentence final particles." In: *Lingua* 117.10, pp. 1739–1783.

## Bibliography

- Tạ, Tấn Thành (Apr. 2021). "The tone system of Ruc and tonogenesis in Vietic languages." en. In: *Science & Technology Development Journal - Social Sciences & Humanities* 5.1. Number: 1, pp. 955-976. ISSN: 2588-1043. DOI: [10.32508/stdjssh.v5i1.568](https://doi.org/10.32508/stdjssh.v5i1.568).
- Ternes, Elmar (2006). *The phonemic analysis of Scottish Gaelic, based on the dialect of Applecross, Ross-shire*. Vol. Third revised edition, with an additional chapter. Dublin: School of Celtic Studies of the Dublin Institute for Advanced Studies.
- Thieberger, Nicholas and Michel Jacobson (2010). "Sharing data in small and endangered languages." In: *Language documentation: practice and values*. Ed. by L.A. Grenoble and L. Furbee. Amsterdam & Philadelphia: John Benjamins, pp. 147-158.
- Timcke R., Leden H., and Moore P. (Apr. 1959). "Laryngeal vibrations: Measurements of the glottic wave: part ii. physiologic variations." In: *A.M.A. Archives of Otolaryngology* 69.4, pp. 438-444. ISSN: 0096-6894. DOI: [10.1001/archotol.1959.00730030448011](https://doi.org/10.1001/archotol.1959.00730030448011).
- Tran, Thi Hue (2010). "Tinh thái giảm nhẹ trong diễn ngôn tiếng Việt [Attenuation in Vietnamese discourse]." MA thesis. Ho Chi Minh City: Ho Chi Minh City Normal University.
- Từ, Trần (1988). "Người Mường ở Hòa Bình cũ." In: *Người Mường với văn hóa cổ truyền Mường Bì*. Sở Văn hóa Thông tin Hà Sơn Bình.
- Vaissière, Jacqueline (1983). "Language-independent prosodic features." In: *Prosody: Models and Measurements*. Ed. by Anne Cutler and Robert Ladd. Berlin: Springer Verlag, pp. 53-66.
- Vaissière, Jacqueline, Kiyoshi Honda, et al. (2010). "Multisensor platform for speech physiology research in a phonetics laboratory." In: *Journal of the Phonetic Society of Japan* 14.2, pp. 65-77. URL: [https://www.jstage.jst.go.jp/article/onseikenkyu/14/2/14\\_KJ00007408569/\\_pdf](https://www.jstage.jst.go.jp/article/onseikenkyu/14/2/14_KJ00007408569/_pdf).
- Vaissière, Jacqueline and Alexis Michaud (2006). "Prosodic constituents in French: a data-driven approach." In: *Prosody and syntax: Cross-linguistic perspectives*. Ed. by Ivan Fónagy, Yuji Kawaguchi, and Tsunekazu Moriguchi. Usage-based linguistic informatics. Amsterdam: John Benjamins, pp. 47-64.
- Van Riper, C. and John V. Irwin (1958). *Voice and articulation*. Prentice-hall, Englewood Cliffs. N.J.
- Vasile, Aurelia et al. (2020). "Le Digital Object Identifier, une impérieuse nécessité ? L'exemple de l'attribution de DOI à la Collection Pangloss, archive ouverte de langues en danger." In: *I2D - Information, données & documents* 2, pp. 156-175. URL: <https://halshs.archives-ouvertes.fr/halshs-02870206>.
- Vitevitch, Michael S et al. (1997a). "Phonotactics and syllable stress: Implications for the processing of spoken nonsense words." In: *Language and speech* 40.1, pp. 47-62.
- (Jan. 1997b). "Phonotactics and Syllable Stress: Implications for the Processing of Spoken Nonsense Words." In: *Language and Speech* 40.1. Publisher: SAGE Publications Ltd, pp. 47-62. ISSN: 0023-8309. DOI: [10.1177/002383099704000103](https://doi.org/10.1177/002383099704000103). URL: <https://doi.org/10.1177/002383099704000103>.
- Vydrina, Alexandra (2017). "A corpus-based description of Kakabe, a Western Mande language: prosody in grammar." PhD thesis. Sorbonne Paris Cité. URL: <https://tel.archives-ouvertes.fr/tel-01801759/>.

- Weinreich, Uriel (1957). "On the description of phonic interference." In: *Word* 13.1, pp. 1–11.
- (2011). *Languages in contact: French, German and Romansh in twentieth-century Switzerland*. John Benjamins Publishing.
- Weinreich, Uriel, William Labov, and Marvin R Herzog (1968). *Empirical foundations for a theory of language change*. Austin: University of Texas Press.
- Wendahl, R. W., G. P. Moore, and Harry Hollien (1963). "Comments on vocal fry." In: *Folia Phoniatica et Logopaedica* 15.4, pp. 251–255. ISSN: 1021-7762, 1421-9972.
- Whalen, Dough H. and Andrea G. Levitt (1995). "The universality of intrinsic Fo of vowels." In: *Journal of Phonetics* 23, pp. 349–366.
- Wrembel, Magdalena (2009). "On hearing colours—Cross-modal associations in vowel perception in a non-synaesthetic population." In: *Poznan Studies in Contemporary Linguistics* 45.4. Publisher: De Gruyter Mouton, pp. 595–612.
- Yu, Kristine and Hiu Wai Lam (2014). "The role of creaky voice in Cantonese tonal perception." In: *Journal of the Acoustical Society of America* 136.3, pp. 1320–1333.
- Zerbian, Sabine (2010). "Developments in the study of intonational typology." In: *Language and Linguistics Compass* 3.1, pp. 1–16.
- Zhang, Zhenrui and Fang Hu (2020). "Neutral tone in Changde Mandarin." In: *Proceedings of InterSpeech 2011*.
- Zsiga, Elizabeth and Rattima Nitisaroj (2007). "Tone features, tone perception, and peak alignment in Thai." In: *Language and Speech* 50.3, pp. 343–383.





## Glottalization, tonal contrasts and intonation: an experimental study of the Kim Thuong dialect of Muong

**Abstract:** All languages in the Vietic subbranch of Austroasiatic have at least one glottalized tone. This thesis zooms in on one of these languages: Muong (in Vietnamese orthography: *Mường*, endonym: /**mon**<sup>3</sup>/), spoken in Kim Thuong (Phu Tho, Vietnam). Twenty speakers recorded twelve tonal minimal sets of the five tones of smooth syllables, plus three tonal minimal pairs of the two tones of checked syllables, under two conditions: in isolation and in a carrier sentence. Acoustic and electroglottographic recordings allow for estimating fundamental frequency, glottal open quotient and duration. These parameters are compared across tones, experimental conditions and speakers, in order to contribute to a better understanding of glottalization as a feature of linguistic tones. First, allotones of the phonologically glottalized tone in Muong (Tone 4) are classified on a phonetic basis, confirming the consistent presence of *creak*. It is tempting to contrast it with the *glottally constricted* tones of Northern Vietnamese (with which Muong is in sustained language contact). However, the phonological discussion emphasizes that analysis of Tone 4 as a prototypical “creaky tone” would be a pitfall. Tone 4 behaves in key respects like the other tones in the system: it is not defined solely by phonation type. Moreover, the range of phonetic (allotonic) variation of Tone 4 includes cases of glottal constriction. Use of a phonetic nomenclature for types of glottalization serves as a basis for describing the interaction of glottalization with intonation.

**Keywords:** glottalization, creaky voice, phonation types, tone systems, experimental phonology, phonetic fieldwork, electroglottography, Vietic languages, Muong language

## Glottalisation, oppositions tonales et intonation : étude expérimentale du dialecte muong de Kim Thuong

**Résumé :** Toutes les langues de la branche viétique de la famille austroasiatique possèdent au moins un ton glottalisé. La présente thèse se concentre sur l'une de ces langues : le muong (en orthographe vietnamienne : *Mường*, endonyme : /**mon**<sup>3</sup>/), parlé à Kim Thuong (Phu Tho, Vietnam). Vingt locuteurs ont enregistré douze ensembles minimaux des cinq tons des syllabes sans occlusives finales, et trois paires minimales des deux tons des syllabes à occlusives finales, dans deux conditions : à l'isolée et dans une phrase-cadre. Les signaux acoustiques et électroglottographiques recueillis permettent d'estimer fréquence fondamentale, quotient ouvert et durée. Ces paramètres sont comparés entre tons, entre conditions expérimentales et entre locuteurs, afin de parvenir à une meilleure compréhension de la glottalisation en tant que caractéristique d'un ton lexical. Tout d'abord, les allotones du ton phonologiquement glottalisé en muong (le ton 4) sont classés sur des bases phonétiques. Il est tentant d'opposer ce ton, caractérisé par la présence régulière d'une voix craquée, avec les tons B2 et C2 du vietnamien du nord (avec lequel le muong est en contact linguistique soutenu), caractérisés par une constriction glottale. Cependant, une analyse phonologique du ton 4 comme prototype de « ton en voix craquée » masquerait la complexité des faits : le ton 4 fait partie d'un système au sein duquel il n'est pas défini exclusivement par un type de phonation. En outre, la plage de variation allotonique du ton 4 comprend des cas de constriction glottale. Une nomenclature phonétique des types de glottalisation sert de base à la description du ton 4 et de son interaction avec l'intonation.

**Mots-clés :** langues viétiques, langue muong, systèmes de tons, types de voix, glottalisation, voix craquée, phonétique expérimentale, électroglottographie