The interplay of intonation and complex lexical tones: how speaker attitudes affect the realization of glottalization on Vietnamese sentence-final particles

Thi-Lan Nguyen, Alexis Michaud, Do-Dat Tran, Dang-Khoa Mac

HAL Id: halshs-00824076
https://shs.hal.science/halshs-00824076v2

Submitted on 12 Jun 2013

The interplay of intonation and complex lexical tones: how speaker attitudes affect the realization of glottalization on Vietnamese sentence-final particles

Thi-Lan Nguyen[1], *Alexis Michaud*[1 2]*, Do-Dat Tran*[1]*, Dang-Khoa Mac*[1]

[1] International Research Institute MICA, HUST – CNRS/UMI-2954 – Grenoble INP,
Hanoi University of Science and Technology
[2] Langues et Civilisations à Tradition Orale (LACITO), CNRS – Univ. Paris 3 Sorbonne Nouvelle
{Thi-Lan.Nguyen, Alexis.Michaud, Do-Dat.Tran, Dang-Khoa.Mac}@mica.edu.vn

Abstract

A salient aspect of the tone system of Hanoi Vietnamese is its use of phonation-type characteristics. This pilot study investigates intonational variation in the realization of two tones: tone 3 *(ngã)*, a rising tone with a strong glottalization in its first half, and tone 6 *(nặng)*, which starts on a middle pitch and usually falls dramatically because of a strong glottalization in its second half. This study focuses on how speaker attitude affects the realization of glottalization on two sentence-final particles (SFPs) carrying tones 3 and 6: *đã* [IPA: ɗa3], conveying tense-aspect-modality information, and *ạ* [IPA: a6], conveying politeness. Audio and electroglottographic recordings from 4 male speakers suggest that glottalization is phased earlier for *surprise* than for *declaration*. *Irritation* also tends to be reflected in earlier glottalization, but with an added glottal constriction at the very end. A methodological challenge is that phonetic realizations of tones 3 and 6 span a wide range of states of the glottis. A procedure is proposed for detecting the complex-repetitive patterns found in cases of lapse into creaky phonation (vocal fry). This helps quantify glottalization phenomena, with a view to arriving at a model that can be used in speech processing.

Index Terms: glottalization, creaky voice, phonation types, tone, intonation, Vietnamese, sentence-final particles, speaker attitude, electroglottography

1. Introduction

1.1. Vietnamese sentence-final particles: expressive morphemes carrying lexical tone

Sentence-final particles (hereafter SFPs) are expressive words indicating sentence mode, evidential nuances, and affective coloring – functions that tend to be fulfilled by intonation in a language such as English [1]. The relationship between intonation and SFPs is not simply one of functional equivalence, however, since SFPs also carry intonational information: sentence-level intonational phenomena are known to cluster on SFPs. One and the same SFP can take on different nuances depending on intonation.

There are about ten SFPs in Mandarin, thirty in Cantonese [2], and about the same number in Vietnamese [3]; SFPs are ubiquitous in colloquial speech. Unlike in Mandarin, where SFPs belong among toneless syllables [4], [5], Vietnamese SFPs undisputedly carry lexical tone.

A previous study brought out the influence of the SFP's function – the nuances that it conveys – on its phonetic realization [6]. The present pilot study extends the scope of analysis beyond F0: the research issue is *how speaker attitude affects the realization of glottalization* for tones 3 and 6.

1.2. Tones 3 and 6 (*ngã, nặng*): the role of glottalization

A salient aspect of the Northern Vietnamese tone system is its use of phonation-type characteristics [7]–[10]. In particular, tones 3 and 6 are glottalized. Tone 3 (also referred to by its orthographic label, *ngã*, its etymological label, C2, or the English descriptor 'broken tone') is a rising tone with a strong glottalization in its first half. Tone 6 (a.k.a. *nặng*, B2, 'drop tone') starts on a middle pitch and usually falls dramatically because of a strong glottalization in its second half. Vietnamese tones illustrate the decisive importance of the synthesis of phonation types for new-generation speech processing: for instance, the Fujisaki model, which focuses exclusively on F0, would require substantial additions before it can handle such phenomena [11].

1.3. Intonational variability of tones

Another challenge for natural language processing, as well as for linguistic modeling, comes from the fact that Vietnamese has salient intonational phenomena [12]. The surface realization of tones depends greatly on intonation: phrasing, prominence, and the expression of attitudes and emotions. It has been reported that the F0 curve of tone 6 differs widely between a 'neutral' context and an 'emphatic/impatient' context, whereas the degree of glottal constriction appears to be little affected [13]. It appeared worthwhile to investigate how speaker attitude affects the realization of glottalization, a phonetic dimension which is cross-linguistically known to convey "paralinguistic" information [14], [15].

2. Method

There already exists a corpus designed for the study of social attitudes in Vietnamese [16], but it does not contain SFPs. We therefore decided to record new data. Speech data acquisition is an underestimated challenge [17], especially when attempting to capture such elusive aspects of speech as attitudinal information. Special attention was therefore paid to the elaboration of materials and recording procedures.

2.1. Data collection procedure

2.1.1. Materials: target sentences and contexts

Due to the great amount of carry-over tonal coarticulation in Vietnamese [18], the tones of SFPs are strongly influenced by those of preceding syllables [19]. We therefore devised sentence (1) using only syllables carrying tone 1 (*ngang*, A1), which is phonetically the simplest: a level, non-low tone.

(1) *Lam   lên   công ty*
     proper_name to_go/up   workplace/company
     'Lam goes to work' / 'I'm going to work'

This sentence was then associated with SFPs *ạ* [IPA: a6], conveying politeness, and *đã* [IPA: ɗa3], conveying tense-aspect-modality information. This yields (2) *Lam lên công ty ạ* and (3) *Lam lên công ty đã*. Finally, sentences (1-3) were placed inside dialogues, which were precisely contextualized. The attitudes under study are (i) POLITENESS, associated lexically to the SFP *ạ*, and (ii) DECLARATION, IRRITATION, and SURPRISE, elicited by context. The general context is as follows: Lam, Minh and An are three friends who have just moved into a shared flat; today is Saturday, a day when they neither go to class nor to their workplace; but Lam is suddenly requested to go to work for urgent business.

Table 2 presents the speech materials. Labels for the intended attitudes follow the terminology proposed by [17], which distinguishes 16 attitudes, and which treats sentence mode (declarative, interrogative or imperative) as part of speaker attitude.

Table 2. Intended attitudes. 'Sentence'=contextualized sentence; DECL=DECLARATION, IRRI=IRRITATION, SURP=SURPRISE, POLI=POLITENESS, AUTH=AUTHORITY.

sentence	SFP	dialogue	DECL	IRRI	SURP	POLI	AUTH
α	*ạ*	3			yes	yes	
β	*ạ*	4		yes	yes	yes	
γ	*ạ*	5	yes			yes	
δ	*đã*	1	yes				
ε	*đã*	1	yes	yes			
ζ	*đã*	2	yes				yes
η	*đã*	2	yes	yes			yes

The data are not fully symmetrical because of the semantics of SFPs. The SFP *ạ*, conveying politeness, is fully compatible with interrogative as well as declarative sentence mode, whereas *đã* is not compatible with interrogation.

2.1.2. Elicitation method and speakers

Two different approaches to data collection were used. *The first* aimed at maximal ecological validity, eliciting the intended attitude through contextualization, from two speakers who were unaware of the purpose of the study. (Their speaker codes, assigned as part of a larger database, are M4 and M5, respectively.) *The second* aimed at maximal clarity in contrasting different attitudes: two speech scientists (M6 and M7) who were aware of the purpose of the study deliberately expressed the intended attitude as identified by the labels in Table 2. M4 and M5 are aged 24; they have university education in software engineering. M6 and M7 are aged 26 and 31, respectively. They were born in Hanoi, and are permanent residents there, apart from a total of 2 years in France for M7. All four can speak some English, and M7 is also fluent in French.

2.1.3. Recording conditions

The recordings were conducted at the MICA Institute's sound-treated booth. M4 and M5 received information about electroglottography (EGG) and its full innocuity [21], [22]. Then they were given time to familiarize themselves with the scripts of the dialogues. Questions were answered through discussion of the context. After this, the two speakers read the dialogues three times, then swapped roles and read another three times. They were instructed to read 'like actors' – an indirect way to elicit a vivid, expressive dialogue.

The two colleagues who participated as speakers (M6 and M7) managed the task themselves; the first author served as dialogue partner when requested.

The electroglottographic signal from an EG2-PC (for one of the speakers) and the audio (from one microphone for each speaker) were recorded as three synchronized WAV files (44,100 Hz, 24-bit). These recordings are available online as part of the MICA Institute's AuCo project; long-term archiving and online availability are guaranteed through the Pangloss Collection [20]. See:

http://lacito.vjf.cnrs.fr/archivage/languages/Vietnamese_en.htm

2.1.4. Data analysis

The analysis method used is based on the derivative of the EGG signal [23], which allows for the measurement of cycle length (and hence F0), glottal open quotient (Oq) [24], and a parameter called DECPA: Derivative-Electroglottographic Closure Peak Amplitude [25]. Semi-automatic measurements were conducted using PEAKDET (available online [26]). The PRAAT and MATLAB scripts devised for the present study are available from the first author's web page.

3.   Results

3.1. The expression of SURPRISE

Fig. 1 illustrates differences in the realization of SFP /a6/ in association with a *declarative/neutral* attitude (Fig. 1a) and an attitude of *surprise* (Fig. 1b). (All figures are grouped on the third page of this paper.) In Fig. 1a, glottalization is final; this is the canonical realization of tone 6 [27], [28]. In Fig. 1b, only mild medial glottalization is observed: it is evidenced by a decreased Oq and amplitude of the audio signal. It is followed by a final portion with rising F0 and Oq, which lengthens the syllable considerably. Measurements averaged over six tokens bring out the same difference in the phasing of glottalization: see Fig. 2. This results in strikingly different allophones for tone 6: earlier phasing results either in a mild glottal constriction followed by modal voicing up until the end of the syllable (as in Fig. 1 and 2) or in the production of a 'voiced tail' after a complete glottal constriction (as in some of the data from the other three speakers, not shown here for reasons of space).

3.2. The expression of IRRITATION

Fig. 3 illustrates differences in the realization of SFP /ɗa3/ in association with a DECLARATIVE/NEUTRAL attitude (Fig. 3a) and an attitude of IRRITATION (Fig. 3b). In Fig. 3a, glottalization is medial: it is preceded and followed by modal-voice portions of comparable duration. In Fig. 3b, there is an early glottalization (evidenced by a dip in Oq and in the amplitude of the audio signal), and a glottalized voicing offset (a final glottal constriction). The syllable in Fig. 3b is strongly glottalized in more ways than one: the syllable is glottalized at two places (medial *and* final); and Oq is lower throughout the syllable. There is also a salient difference in F0, visible from Fig. 3 and from the average curves in Fig. 4 (same speaker).
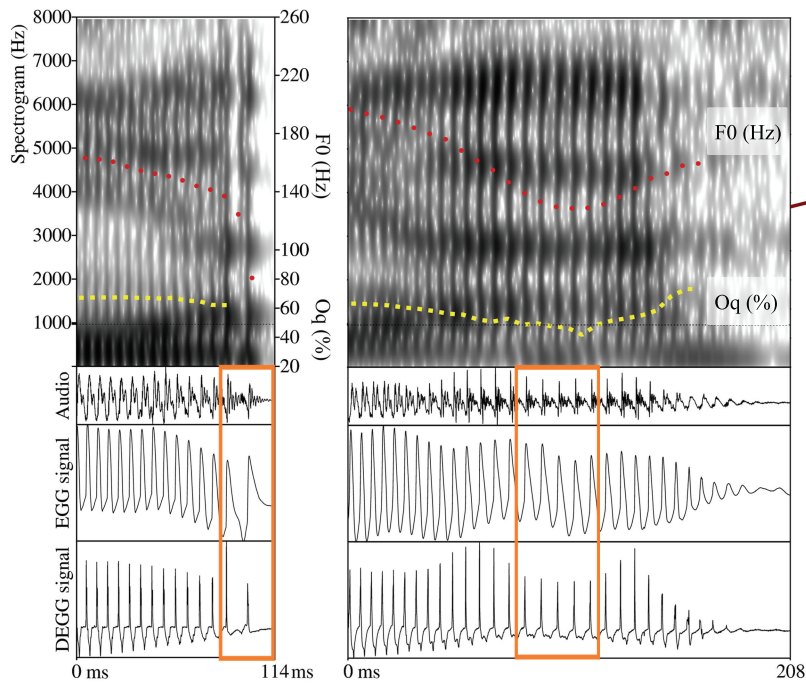
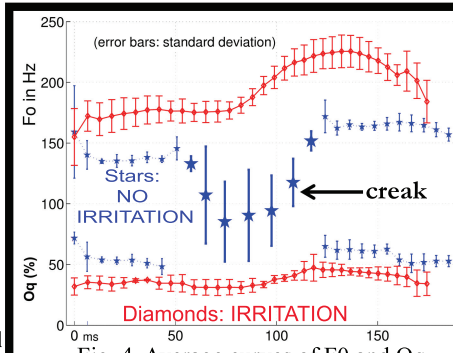Fig. 2. Average curves of F0 and Oq over 6 tokens of /a6/, speaker M7

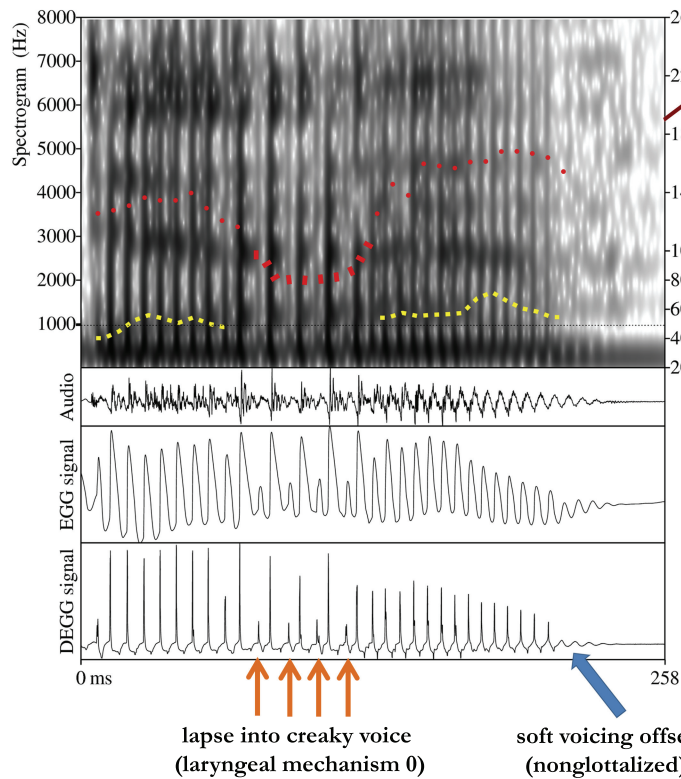Fig. 4. Average curves of F0 and Oq over 6 tokens of /ɖa3/, speaker M6

Fig. 1a: **declarative/neutral**

Figure 1b: attitude: **surprise**. Speaker M7.

Fig. 1. Two realizations of glottalization on SFP /a6/. 1a: **final**; 1b: **mild, medial**

lapse into creaky voice
(laryngeal mechanism 0)

soft voicing offset
(nonglottalized)

mild glottalization

glottalized voicing offset
(final glottal constriction)
of intonational origin
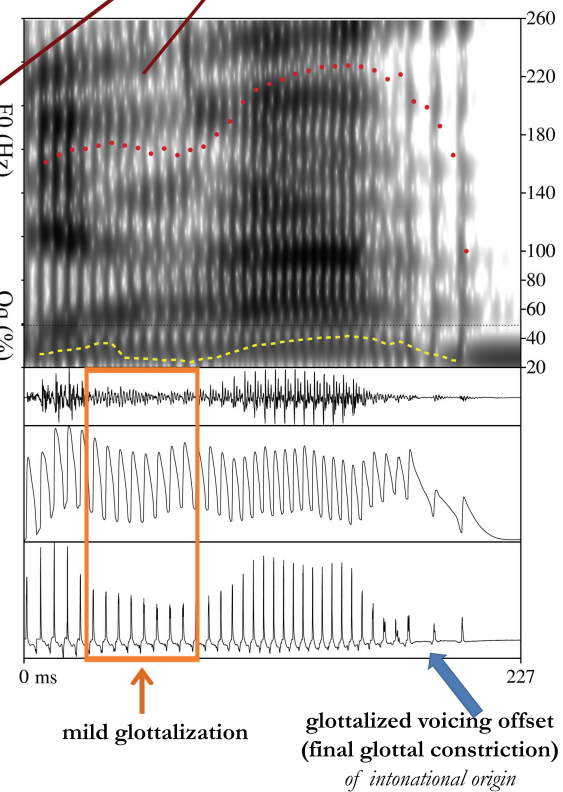
Fig. 3a: **declarative/neutral**

Fig. 3b: attitude: **irritation**. Speaker M6.

Fig. 3. Two realizations of glottalization on SFP /ɖa3/. 3a: **medial creak**; 3b: **early glottalization+final constriction**

To venture a generalization from these observations, IRRITATION, like SURPRISE, is reflected in earlier glottalization than DECLARATIVE/NEUTRAL attitude, but in the case of IRRITATION, earlier glottalization may be one of several consequences of an overall *stronger* glottalization, whereas for SURPRISE, the earlier phasing of glottalization appears to serve a widely different purpose: it can be interpreted as clearing the way for the realization of an additional (nonglottalized) voiced portion at the end of the syllable, which functions as a clear attitudinal cue.

At the present stage, only data from two speakers (M6 and M7) are presented. The data for the other two speakers (M4 and M5) were also fully analyzed; no figures for these speakers are provided for want of space. We plan to include these data in statistical treatment (after speaker normalization), and also to extend the analysis to sentences β (combining IRRITATION and SURPRISE) and ζ+η (to study the added attitude of AUTHORITY: see Table 2 above). At the present stage, let us simply mention the general observation that the data for speakers M4 and M5 show the same trends as for speakers M6 and M7, though with less extreme differences between attitudes. This relates to the differences in the collection procedure mentioned in §2.1.2: speakers M6 and M7 expressed the intended attitudes forcibly, with a stronger emotional tinge than was obtained through contextualized dialogues.

4. Discussion

4.1. The interplay of attitudes and lexical tones

The observation of stronger glottalization for IRRITATION is consistent with the cross-language, psychophonetic tendency for negative emotions to be accompanied by constriction in the various "valves of the throat" [29]: see [14]. Conversely, for SURPRISE, the presence of a voiced interval with modal/soft phonation and high F0 at the end (as a 'voiced appendix' to tone 6; see Fig. 1b, contrasting with Fig. 1a, and the averaged values in Fig. 2) can be interpreted as a means of resolving the headlong conflict between (i) strong glottal constriction, as lexically specified for tone 6, and (ii) higher F0, which tends to be associated with SURPRISE cross-linguistically.

To attempt a generalization over the small set of data examined here: later phasing of glottalization is associated with assertion (*declarative* attitude, in the typology used here), and earlier phasing is associated with a suspension of assertion.

4.2. The issue of the quantification of creaky voice

Qualitatively, the glottalization in Fig. 3a consists in a lapse into creaky voice (laryngeal mechanism zero [30]). This is not specific to tone 3: the canonical, hyperarticulated realization of tone 6 involves a strong final glottal constriction [7], [8], [13], [27], as in the example shown in Fig. 1a, but the recorded data show that a less intense glottalization (simply a brief lapse into creaky voice) is relatively common in non-hyperarticulated contexts, or in association with non-assertive attitudes.

The complex-repetitive waveforms frequently encountered in vocal fry/creaky voice [31] raise an issue for the measurement of F0 and Oq. Phonation mechanisms O and I (a.k.a. "creaky voice" and "chest voice"; see [30]) are therefore dealt with separately in our analysis of the data.

First, the presence of mechanism 0 (creak) is detected through a measurement of irregularities in the amplitude of the positive peak found on the dEGG signal at the glottis-closure-instant. This parameter is called DECPA, for Derivative-Electroglottographic Closure Peak Amplitude [25], [32] (another label that has been proposed for this parameter is PIC, for Peak Increase in Contact [33]). Sudden variation in DECPA in glottalization is even more salient than the differences in glottal cycle length. DECPA is one of the parameters measured by the PEAKDET script [26].

On this basis, syllable rhymes are divided temporally into successive portions: for instance, in Fig. 3a, mechanism I, then 0, and finally I again. For portions in mechanism 0, we base the measurement of cycle length on the entire complex-repetitive patterns, rather than on each isolated pulse as detectable on the EGG signal and its derivative (dEGG). Said differently, minor closing peaks within complex-repetitive patterns are left aside when calculating the inverse of cycle duration. The values thus obtained are called "F0" for convenience, even though the terms "cycle" and "fundamental frequency" are not fully appropriate since the signal is not quasi-periodic. We refrain from proposing an estimation of Oq for creaky voice, since the notion of open quotient only applies to glottal cycles that simply comprise a closed phase and an open phase – a situation which differs widely from that found in creaky voice.

Averaged representations can then be created for sets of syllable rhymes all of which comprise a creaky-voice portion, as in Figure 4, where the portion of medial creak is represented with larger-sized stars, and with no corresponding Oq estimation. (The procedure for calculating these averaged representations is set out in [27].)

5. Conclusions

A recent study of Northern Vietnamese intonation emphasizes "global F0 and intensity and local sentence-final F0" as the predominant acoustic correlates of intonation structure [34]. The present pilot study proposes a complementary perspective, inspired by (i) cross-linguistic findings about the great diversity of the acoustic correlates of intonation [35] and (ii) observations about the importance of phonation types in Northern Vietnamese. The results suggest that there is a considerable range of intonational (allotonic) variation in the realization of glottalization, and that it contributes to expressing speaker attitude – and is not unlikely to convey other types of prosodic information as well.

Perspectives for future work include:

- Conducting verifications through statistical treatment of the production data, and perception tests.
- For phonetic/phonological modeling: investigating to which extent there exists an analogy between the phasing of glottalization and the phasing of F0 curves ("tonal alignment" [36]).
- For automatic speech processing: offering fine-grained control of glottalization phenomena in speech synthesis, as a means to imitate more closely the intonational mechanisms of natural languages.

6. Acknowledgments

7. References

[1] M. K. M. Chan, "Sentence-final particles in Cantonese: a gender-linked survey and study," in *Proc. of NACCL 11*. Cambridge (Mass.), 1999, pp. 87–101.

[2] H. Kwok, *Sentence particles in Cantonese*. Hong Kong: Centre of Asian Studies, 1984.

[3] Trần Thị Huệ, "Tình thái giảm nhẹ trong diễn ngôn tiếng Việt [Attenuation in Vietnamese discourse]," M.A. thesis, Ho Chi Minh City Normal University, 2010.

[4] C. N. Li and S. A. Thompson, *Mandarin Chinese: a functional reference grammar*. Berkeley: University of California Press, 1981.

[5] S. Niimi, Yan Qun, S. Horiguchi, and H. Hirose, "An electromyographic study on laryngeal adjustment in production of the Mandarin Chinese light tone," *Annual Bulletin of the Research Institute for Logopedics and Phoniatrics*, vol. 24, pp. 7–17, 1990.

[6] Thi-Lan Nguyen and Do-Dat Tran, "Tonal co-articulation on particles in Vietnamese," in *Proc. of IALP 2012*, Hanoi, 2012.

[7] Van Loi Nguyên and J. A. Edmondson, "Tones and voice quality in modern northern Vietnamese: instrumental case studies," *Mon-Khmer Studies*, vol. 28, pp. 1–18, 1998.

[8] M. Brunelle, Khắc Hùng Nguyễn, and Duy Dương Nguyễn, "A Laryngographic and Laryngoscopic Study of Northern Vietnamese Tones," *Phonetica*, vol. 67, no. 3, pp. 147–169, 2010.

[9] M. Brunelle, "Tone perception in Northern and Southern Vietnamese," *Journal of Phonetics*, vol. 37, pp. 79–96, 2009.

[10] J. Kirby, "Dialect experience in Vietnamese tone perception," *Journal of the Acoustical Society of America*, vol. 127, no. 6, pp. 3749–3757, 2010.

[11] H. Mixdorff, Hung-Bach Nguyen, H. Fujisaki, and Chi-Luong Mai, "Quantitative Analysis and Synthesis of Syllabic Tones in Vietnamese," in *Proc. of Eurospeech*, Geneva, 2003, pp. 177–180.

[12] Tran Do Dat and E. Castelli, "Register of Vietnamese tones in continuous speech," in *Proc. of SLTU 2008*, Hanoi, 2008.

[13] A. Michaud and T. Vu-Ngoc, "Glottalized and Nonglottalized Tones under Emphasis: Open Quotient Curves Remain Stable, F0 Curve is Modified," in *Proc. of Speech Prosody 2004*, Nara, Japan, 2004, pp. 745–748.

[14] I. Fónagy, *La vive voix: essais de psycho-phonétique*. Paris: Payot, 1983.

[15] C. Gobl and A. Ní Chasaide, "The role of voice quality in communicating emotion, mood and attitude," *Speech Communication*, vol. 40, pp. 189–212, 2003.

[16] Dang-Khoa Mac, V. Aubergé, A. Rilliard, and E. Castelli, "Audio-visual prosody of social attitudes in Vietnamese: building and evaluating a tones balanced corpus," in *Proc. of Interspeech 2009*, Brighton, 2009, pp. 2263–2266.

[17] O. Niebuhr and A. Michaud, "Speech data acquisition: the underestimated challenge," submitted.

[18] M. Brunelle, "Northern and Southern Vietnamese tone coarticulation: A comparative case study," *Journal of Southeast Asian Linguistics*, vol. 1, pp. 49– 62, 2009.

[19] Thi-Lan Nguyen and Do-Dat Tran, "Tonal co-articulation on function words: a pilot study of sentence-final particles in Hanoi Vietnamese," *International Journal of Asian Language Processing*, submitted.

[20] "The Pangloss Collection." [Online]. Available: http://lacito.vjf.cnrs.fr/archivage/presentation_en.htm.

[21] P. Fabre, "Un procédé électrique percutané d'inscription de l'accolement glottique au cours de la phonation: glottographie de haute fréquence," *Bulletin de l'Académie Nationale de Médecine*, vol. 141, pp. 66–69, 1957.

[22] R. J. Baken, "Electroglottography," *Journal of Voice*, vol. 6, no. 2, pp. 98–110, 1992.

[23] N. Henrich, C. d'Alessandro, M. Castellengo, and B. Doval, "On the use of the derivative of electroglottographic signals for characterization of nonpathological phonation," *J. of the Acoustical Society of America*, vol. 115, no. 3, pp. 1321–1332, 2004.

[24] Vu-Ngoc Tuân, C. d'Alessandro, and A. Michaud, "Using open quotient for the characterization of Vietnamese glottalized tones," in *Proc. of Interspeech 2005*, Lisboa, 2005, pp. 2885–2889.

[25] A. Michaud, "A Measurement from Electroglottography: DECPA, and its Application in Prosody," in *Proc. of Speech Prosody 2004*, Nara, Japan, 2004, pp. 633–636.

[26] "Open-Source Software for Analysing the Electroglottographic Signal." [Online]. Available: http://voiceresearch.free.fr/egg/.

[27] A. Michaud, "Final consonants and glottalization: new perspectives from Hanoi Vietnamese," *Phonetica*, vol. 61, no. 2–3, pp. 119–146, 2004.

[28] L. C. Thompson, *A Vietnamese Reference Grammar*. Washington: University of Washington Press, 1965.

[29] J. A. Edmondson and J. Esling, "The valves of the throat and their functioning in tone, vocal register and stress: laryngoscopic case studies," *Phonology*, vol. 23, no. 2, pp. 157–191, 2006.

[30] B. Roubeau, N. Henrich, and M. Castellengo, "Laryngeal vibratory mechanisms: the notion of vocal register revisited," *Journal of Voice*, vol. 23, no. 4, pp. 425–38, 2009.

[31] M. Blomgren, Chen Yang, L. N. Manwa, and H. R. Gilbert, "Acoustic, aerodynamic, physiologic, and perceptual properties of modal and vocal fry registers," *Journal of the Acoustical Society of America*, vol. 103, no. 5, pp. 2649–2658, 1998.

[32] C. Esposito, "An acoustic and electroglottographic study of White Hmong tone and phonation," *Journal of Phonetics*, vol. 40, pp. 466–476, 2012.

[33] P. Keating, C. Esposito, M. Garellek, S. ud D. Khan, and J. Kuang, "Phonation contrasts across languages," *UCLA Working Papers in Phonetics*, vol. 108, pp. 188–202, 2010.

[34] M. Brunelle, K. P. Hạ, and M. Grice, "Intonation in Northern Vietnamese," *The Linguistic Review*, vol. 29, no. 1, pp. 3–36, 2012.

[35] O. Niebuhr, "On the acoustic complexity of intonation," in *Nordic Prosody XI*, Frankfurt: Peter Lang, 2013, pp. 1–14.

[36] D. House, "Perception of tone with particular reference to temporal alignment," in *Proc. of the symposium "Cross-linguistic studies of tonal phenomena"*, S. Kaji, Ed. Tokyo: RILCAA, 2003, pp. 203–216.