

A PHONETIC STUDY OF VIETNAMESE TONES: ACOUSTIC AND ELECTROGLOTTOGRAPHIC MEASUREMENTS

Vu Ngoc Tuan, Christophe d'Alessandro & Sophie Rosset

LIMSI-CNRS, BP 133, F91403, Orsay France

ABSTRACT

Vietnamese is a tone language using 6 different tones. The phonology of Vietnamese makes use of both melodic tones (pitch movements and registers) and voice quality tones. In this paper we present a phonetic description of the 6 tones of Vietnamese. Acoustic and electroglottographic data are provided. Two speakers (one female, one male) participated in the recordings. Phonetic realization of phonological tones using these acoustic parameters is discussed. The results indicate that several phonetic features are able to make possible the phonological contrasts. The first feature is the mean melodic register (high-low), represented mainly by the fundamental frequency. The second feature is the melodic movement (fall, rise, and fall-rise) represented mainly by the fundamental frequency. The third feature is voice quality, represented mainly by the voice open quotient and energy. Finally, a fourth feature seems to be the tone duration. However, all these features are not equally effective for all the tones, and some features seem correlated.

1. INTRODUCTION

One of the most striking features of the phonology of Vietnamese is its tonal system. In most tone languages, tones are melodic tones, mainly correlated with fundamental frequency (F0) [5, 6], for instance a “rising tone” is realized in the acoustic domain by a rising pitch movement. The situation is more intricate for Vietnamese. The tonal distinctions are based on melodic movements (rises, falls, fall-rises), on melodic registers (high and low) and on voice source distinctions (clear voice vs creaky voice). These “voice type tones” are produced by a glottal constriction, a change in laryngeal tenseness, which results in a binary opposition between “clear” voice and “creaky” voice [2]. Creaky voice seems to indicate a strong glottal constriction, sometimes referred as “voix étranglée” (strangled voice) in the French literature on Vietnamese.

The Vietnamese language makes use of 6 tones. Four tones are purely melodic and register tones. The main acoustic correlates of these tones should be melodic movements. Two tones are voice type tones : their main acoustic correlates seem rely on glottal tenseness in addition to F0. Although several studies have been devoted to the phonology of Vietnamese (particularly tonogenesis [9, 11]), few phonetic-acoustic data are available, and new work seemed needed.

The aim of this study is to investigate the acoustic correlates of Vietnamese tones, and particularly of voice type tones. For studying these tones, specific investigation of the laryngeal activity is needed. Therefore both acoustic and electroglottographic measures are presented. The correlations between acoustic parameters (voice quality, F0, energy, durations) and the 6 tones are searched for. On the one hand, this is an important problem for the phonetic

description of the Vietnamese language. On the other hand, the composite nature of Vietnamese tones raise interesting questions in the field of prosodic and voice quality studies. This is a unique opportunity to study the interaction between voice quality changes and pitch movements in tonal perception. However, the perception aspect will not be developed in the framework of this paper, although it is one of the main motivation of this research.

The paper is organized as follows. In Section II, the speech corpus and the measurement methodology are described. In Section III acoustic analyses of the corpus are presented. Section IV is a discussion of the results obtained and a conclusion.

2. CORPUS AND MEASUREMENTS

2.1. Acoustic parameters

This study is in the framework of the acoustic theory of speech production [3]. Speech is produced by a voice source filtered by the vocal tract. In this framework, the tonal aspects of speech are mainly controlled by the voice source. The acoustics of the voice source can be described by the glottal flow signal, which is generally represented with the help of several acoustic parameters [7]:

- F0** the fundamental frequency, or frequency of voicing, inverse of TO, the fundamental period. This parameter is the main acoustic correlate of melodic tones. It can be measured on both the acoustic and electroglottographic (EGG) signals.
- A** the amplitude of voicing. This parameter represents the amplitude (and therefore the energy) of the voice source. This parameter may vary in case of laryngeal constriction, like in voice type tones. This parameter can be measured on the acoustic signal.
- Oq** the open quotient. This parameter represents the ratio between the open phase of the glottis and the fundamental period. Therefore, it is an indication of the voice tenseness (pressed vs relaxed voice). Oq is likely to vary during the production of laryngeal constrictions, like in voice type tones. This parameter can be measured on the EGG signal.
- Sq** the speed quotient. This parameter represents the speed of vocal folds closure, i.e. the ratio between the closing phase and the opening phase. This parameter is an indication of the symmetry of the glottal pulse. This parameter is difficult to measure. It is generally correlated to the source spectral tilt.
- Qa** the shape of the glottal flow signal at glottal closure, or the spectral tilt. This parameter is an indication of vocal effort. It is rather difficult to measure directly. It may vary much with vocal effort, rather than glottal constriction.

Vq the degree of voicing. This parameter measures the ratio between the periodic and the aperiodic component of the voice source. This ratio is 1 for unvoiced source, and zero for perfectly periodic sounds. For voiced speech sound, some degree of aperiodicity is always present, and large for breathy vowels.

Duration the relative durations of the same segments may help to contrast the different tones.

In this work, we measured 4 parameters : F0, A, Oq and durations. It can be guessed that F0 is the main acoustic correlate of melodic tones, and the Oq and A are the main acoustic parameters of the voice type tones. Vocal effort parameters (represented mainly by Sq, Qa and of course A) seem less significant for both melodic and voice type tones. F0 and A can be measured directly on the acoustic signal, but reliable measures of Oq require the use of EGG.

2.2. Speech corpus

The speech corpus contains three types of Vietnamese isolated words and nonsense words: isolated vowels, CV syllables and CVC syllables. There are 12 vowels in Vietnamese, 9 normal duration vowels (i, ê, e, u, ô, o, ơ, ư, a) and 3 short vowels, and 22 consonants (b, c, ch, d, đ, ph, g(h), gi, h, kh, l, m, n, nh, ng(h), r, s, t, tr, v and x), 21 of which can be found as syllable initial consonant. The short vowels are a followed by nh or ch, a followed by u or y and â (another vowels which does not belong to the set of normal duration vowels). The tonal system of Vietnamese makes use of 6 tones distributed in a low and a high register, that are derived (according to [4]) from three inflexions of ancient Chinese as follows: ngang-huyền, sắc-nặng and hỏi-ngã. The tones of Vietnamese are:

ngang high-neutral melodic tone.

huyền low-falling melodic tone.

sắc high-rising melodic tone.

nặng low-fall-creaky melodic and voice type tone.

ngã high-neutral-creaky-rise melodic and voice type tone.

hỏi low-fall-neutral-rise melodic tone.

2.3. Measurement methodology

Simultaneous recording of the acoustic and EGG signals have been conducted. Two native speakers of Vietnamese (one male, the first author of this paper, and one female) participated in the recordings. Both speakers are native speakers of the Vietnamese language spoken in the north of Vietnam, a dialect of Vietnamese which makes use of voice type tones (this is not the case for the Vietnamese spoken in the south of Vietnam). The acoustic signal has been recorded in a soundproof booth using a Bruel & Kjaer measurement microphone 30cm in front of the mouth, on the first track of a Digital Audio Tape (DAT) recorder. The EGG signal has been recorded simultaneously on the second track of the DAT. The parameters are measured as follows:

F0 is obtained as the inverse of the fundamental period T0. The fundamental period is the duration between the large negative peaks in the derivative of the EGG signal (DEGG) [10].

A short term energy of the acoustic signal is used as an indirect measure of the amplitude of voicing.

Oq is obtained using the DEGG signal. It is the ratio of the duration between the small positive peak and the large negative peak of the DEGG signal and T0 (the duration between two large negative peaks) [10].

Duration is measured directly on the oscillograms.

3. RESULTS

In this section, the acoustic and EGG measurements obtained for each tone are analyzed in terms of fundamental frequency, open quotient, duration and energy. Figures 1-4 display the data obtained for the 6 tones (ngang, huyền, sắc, nặng, ngã, hỏi). In the corpus there are isolated words and vowels. The data for all the 9 normal duration vowels following a consonant /k/ that were present in the corpus are plotted in Figures 2 (female speaker) and 4 (male speaker). The data for one vowel /a/ are plotted in Figures 1 (female speaker) and 3 (male speaker).

In Figures 1 and 3, the F0 contour (Hz) is displayed at the top of the panels, the Oq contour (%) is displayed at the bottom of the panels, and the energy contour (dB) is displayed in between.

Figures 2 and 4 display the F0 contours (Hz) at the top of the panels, and the Oq contours (%) at the bottom of the panels.

ngang The first panels (for Figures 1-4) display the data obtained for the high neutral tone ngang. For most examples, this neutral melodic tone is a falling melodic movement rather than a flat F0 tone. However, F0 is in the high register of the speaker. One can notice a final rising movement for open quotient. This is typical of normal phonation, and indicates that voice is relaxing at the end of the utterance. Durations are average.

huyền The second panels display the data obtained for the low-falling melodic tone huyền. A falling melodic movement is clearly noticeable, and the melody is in the low register. There is a slight final rising movement for open quotient, but data are rather widespread. It seems that voice is relaxing at the end of the utterance, but less than for the neutral tone ngang. Durations are average.

sắc The third panels display the data obtained for the high-rising melodic tone sắc. A rising melodic movement is clearly noticeable, and the melody is in the high register. There is a clear final rising movement for open quotient. It seems that voice is relaxing at the end of the utterance, like for ngang. Durations are average.

nặng The fourth panels display the data obtained for the low-fall-creaky melodic and voice type tone nặng. A large falling movement is noticeable for the female speaker, but on the contrary, there is a rise for the male speaker (or the pitch movement is almost flat). For both speakers, the melody is in the low register. There is a clear falling movement for open quotient. This is an indication of glottal constriction: voice is more tense at the end of the tone. Durations are shorter than average. This is probably another indication of laryngeal closure and voice tenseness.

ngã The fifth panels display the data obtained for the high-fall-creaky-rise melodic and voice type tone ngã. Large falling and rising movements are noticeable, and the melody ends in the high register. A falling and rising movement is also noticeable for Oq (a relaxed-constricted-relaxed voice source).

For the female speaker, there is also a drop in energy during the glottal constriction, resulting in an interruption of voicing. Durations are average.

hỏi The sixth panels display the data obtained for the low-fall-rise melodic tone **hỏi**. Large falling and rising movements are noticeable, and the melody stays in the low register. No movement of Oq is noticeable: the voice seems rather relaxed. Durations are longer than average.

4. DISCUSSION AND CONCLUSION

The results confirm that several phonetic features are used to make the distinction between the 6 phonological tones of Vietnamese.

The first feature is the mean melodic register (high-low) that is clearly contrasted. The phonological distinction high/low results generally in higher and lower F0 data in the phonetic domain. For F0 glissandos, pitch perception corresponds to a weighted time average of the frequencies present [1]. Then, the register distinction is probably based on the average F0, but more attention is paid to the end of the tone. Therefore, the register distinction is probably more contrasted viewed through a perceptual model, compared to raw acoustic data.

The second feature is the melodic movement (fall, rise, and fall-rise). It must be pointed out that a phonological movement may be realized by a different phonetic movement. For instance the neutral phonologic tone seems realized by a falling phonetic movement. Another example is the fall-creaky tone that is actually realized by rising melodic movements by the male speaker. This is of course possible as far as the contrasts between tones is maintained.

The third feature is voice quality. The main acoustic correlates of the clear voice/creaky voice phonological distinction are open quotient and energy. The female speaker made much more constriction than the male speaker (this is not the case in normal voice for languages like English [8]), resulting in a drop in energy and voicing. For both speakers, creaky voice correspond to a falling Oq movement (more tension in the voice, a smaller open quotient). As there is almost no data on perception of open quotient movements, it is difficult to know the perceptual contrast given by Oq movements. This is one of our research projects.

A fourth feature is duration contrasts. The fall-rise tones are longer than average, and the fall-creaky tones are shorter than average. This may be an effect of the tones tenseness.

All the features are not effective for all the tones. For instance, the final glottal constriction seem sufficient for indicating a **nặng** tone. Therefore, the melodic movement, and even the melodic register may show large individual variability.

Some features that seem highly correlated have gone unnoticed in the literature. For instance, duration variation seems systematic and important, but seem correlated with tenseness (**nặng**, shorter duration than average) or complex melodic movements (**hỏi**, fall-neutral-rise melodic movement). Along the same line, creaky voice seems correlated with F0 falls.

Other voice source parameters of interest have not been considered here: the degree of voicing, the spectral tilt. They are probably less important features for Vietnamese tonal distinction, but they could play a role and will be considered in our future work.

In this paper we presented a phonetic description of the 6 tones of Vietnamese. It is well known in the linguistic literature that this language uses both melodic and voice quality features in its tonal

system. The main acoustic correlates of the phonological tones are fundamental frequency, voice open quotient, and amplitude of voicing. Duration seems also to play a role, but it may be only a consequence of the glottal constriction. Acoustic and electroglottographic measurements of these parameters are provided here, for the first time to the author's knowledge.

This work will be extended in several directions. Collection of acoustic and EGG data is still in progress, because it seems important to process the data of more subjects. However, acoustic analysis alone may not be sufficient for studying the relative importance of the glottal parameters. Therefore, an analysis-by-synthesis paradigm is necessary for studying the effect of F0, Oq, energy and the other glottal parameter on tone perception. Perceptual experiments using controlled synthetic stimuli will be designed for measurements of thresholds for voice type tones perception by native and non-native speakers of Vietnamese. This will give some insight into the role played by voice quality for tonal perception. Finally, we plan to use the results obtained to text-to-speech synthesis in Vietnamese.

5. REFERENCES

- [1] d'Alessandro C., Rosset S. and Rossi J.P. (1998). "The pitch of short-duration fundamental frequency glissandos". *J. Acoust. Soc. Am.*, 104(4), pp 2339-2348.
- [2] Diffloth G. (1989), "Proto-Austroasiatic creaky voice", *Mon-Khmer Studies*, 15, pp 139-154.
- [3] Fant G. (1995), "The LF-model revisited. Transformations and frequency domain analysis", *STL-QPSR*, Vol. 2-3, 1995, pp 119-56.
- [4] Ferlus, M. (1982), "Spirantisation des obstruantes médiales et formation du système consonantique du Vietnamien", *Cahiers de Linguistique Asie Orientale* Vol. XI 1, pp 83-106.
- [5] Ferlus, M. (1996), "Langues et peuples viet-muong", *Mon-Khmer Studies* 26, pp 7-28. (Viet-Muong languages and nations).
- [6] Ferlus, M. (1998), "Les systèmes de tons dans les langues Viet-Muong", *Diachronica* XV:1, pp 1-27. (The systems of tones in Viet-Muong languages).
- [7] Hanson H.M., (1997) "Glottal characteristics of female speakers: Acoustic correlates", *J. Acoust. Soc. Am.*, 101, pp 466-481.
- [8] Hanson H.M. and Chuang E. S. (1999) "Glottal characteristics of male speakers: Acoustic correlates and comparison with female data", *J. Acoust. Soc. Am.*, 106(2), pp 1064-1077.
- [9] Haudricourt, H. G. (1954), "De l'origine des tons en Vietnamien", *Journal Asiatique* 242, pp 69-82. (On the origin of tones in Vietnamese).
- [10] Henrich N., Doval B., d'Alessandro C., Castellengo M. (2000), "Open quotient measurements on EGG, speech and singing signals" 4th workshop "Advances in Objective Laryngoscopy, Voice and Speech Research", Jena, Germany, pp. 93-101.
- [11] Thurgood G. (1999), "The origin of tones in Vietnamese: revising the model and the analysis", XXXII International Conference on Sino-Tibetan Languages and Linguistics, University of Illinois et Urbana-Champaign.

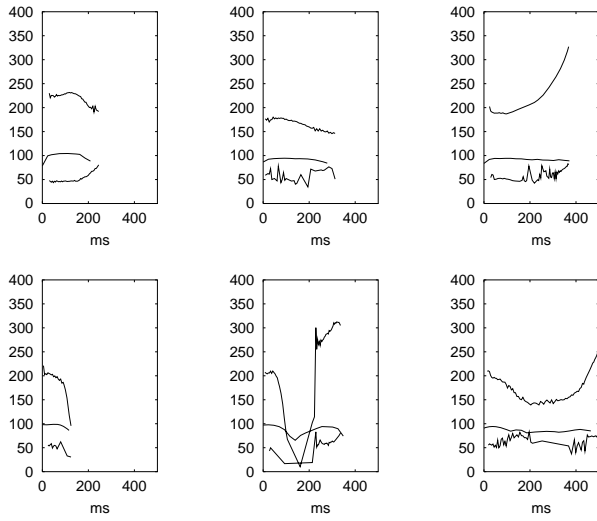


Fig. 1. Data for the tones ngang, huyen, sắc, nặng, ngã, hoi (from left to right and from top to bottom). Data for a vowel /a/. For each panel, from top to bottom: F0 (Hz), energy (dB), Oq (%). Female speaker.

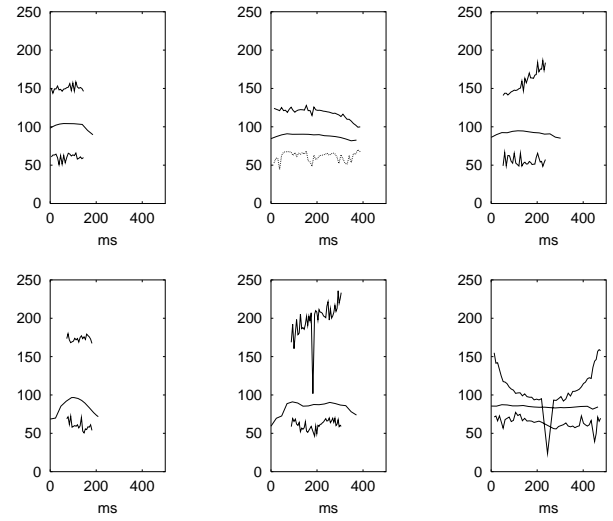


Fig. 3. Data for the tones ngang, huyen, sắc, nặng, ngã, hoi (from left to right and from top to bottom). Data for a vowel /a/. For each panel, from top to bottom: F0 (Hz), energy (dB), Oq (%). Male speaker.

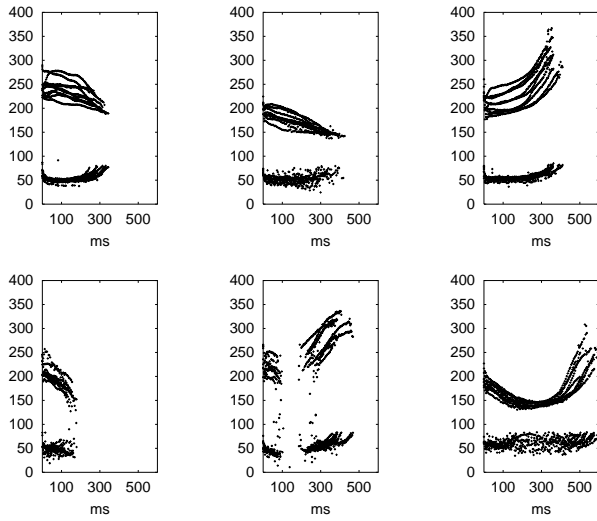


Fig. 2. Data for the tones ngang huyen, sắc, nặng, ngã, hoi, (from left to right and from top to bottom). Contours for all vowels following a consonant /k/. For each panel, from top to bottom: F0 (Hz), Oq (%). Female speaker.

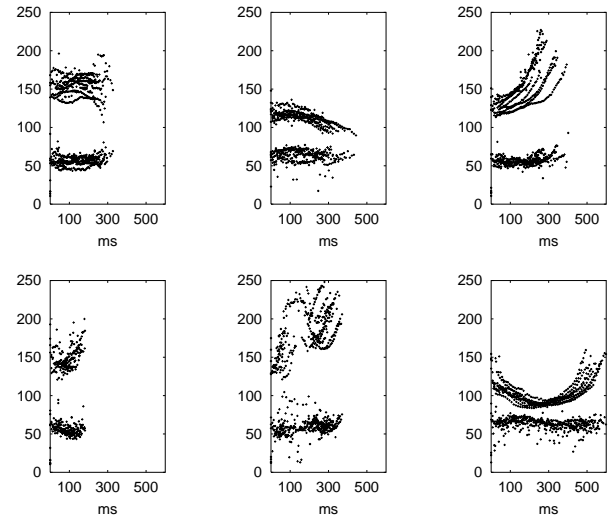


Fig. 4. Data for the tones ngang huyen, sắc, nặng, ngã, hoi, (from left to right and from top to bottom). Contours for all vowels following a consonant /k/. For each panel, from top to bottom: F0 (Hz), Oq (%). Male speaker.