



Prosody of Interrogative and Affirmative Sentences in Vietnamese Language: Analysis and Perceptive Results

Vũ Minh Quang - Trần Đỗ Đạt - Eric Castelli

International research center MICA
IP Hanoi – CNRS/UMI-2954 – INP Grenoble
1-Dai Co Viet - Hanoi, Vietnam

{Minh-Quang.Vu ; Do-Dat.Tran ; Eric.Castelli}@mica.edu.vn

Abstract

This paper presents a new study on the prosody of Vietnamese language. Sentence pairs containing one interrogative sentence and one affirmative sentence, which have the same tones and the same number of syllables to avoid the effects of lexical tones and of co-articulation, are recorded in order to analyze their prosody evolution. Comparisons allow us to characterize differences between interrogative and affirmative sentences at sentence prosody level. Our work is completed by a perceptual study on re-synthesized sound where all syllables of the sentence are replaced by the vowel /a/ to hide lexical meaning, while the prosody of the sentence is kept unchanged. Our goal is to see if sentence prosody carries any information about sentence nature characteristics, and then whether it enables listeners to classify sentence type (in this case interrogative and affirmative), despite the complex form of this prosody in tonal languages. The obtained results show that information on sentence type is present at the end of the second half of the last syllable and that about 70% of sentences are properly classified for female synthesis voice.

Index terms: prosody, intonation, analysis, perceptive

1. Introduction

For the non-tonal Western languages (French or English) it was validated that sentence prosody carries extra linguistic information, such as the emotions or the state of the speaker, or the sentence type (affirmative, interrogative or exclamative [1, 2]). To automatically evaluate the type of sentences for detection or classification purposes, it is possible to analyze the sentence signal directly by using the prosodic characteristics of the sentence, without any need for lexical information from, for example, an Automatic Speech Recognition (ASR) engine. In this case, the measured and analyzed parameters take into account the evolution of the intonation during sentence statement: register of F0, increase of F0 at the end of the sentence or other parameters derived from the values of F0, for example [3, 4]. However, in the case of tonal languages (like Mandarin or Vietnamese), the melody contour of the intonation is complex. It is composed of macro-variations corresponding to the intonation of the sentence and of micro-variations corresponding to the lexical tone applied to each syllable of mono (or bi) syllabic words. This is why the direct application to tonal languages of analysis methods developed for non-tonal languages is very likely to fail, because tonal micro-variations

tend to scramble the extra-lexical information coded on the sentence prosody. In the case of Vietnamese language, in order to differentiate the interrogative sentences from other sentence types, the use of specific words called "interrogative classifiers" (không, gì, chưa, for example) is practically systematic. Therefore, the main goal of our study here can be summarized as follows: is there in Vietnamese, a tonal language whose prosody is complex, any extra-linguistic information characterizing the sentence type conveyed by prosody and used during acts of dialogue? The answer will on one hand enable us to go further into our knowledge of the language, and on the other hand (in case of positive answer), will allow us to consider the realization of an automatic classification of sentence type that is independent of ASR system.

2. Prosodic analysis

2.1. Methodology and corpus preparation

Up to now, very few studies have deeply analyzed the phonology of the Vietnamese language. We can cite some recent works relating to the lexical tone [5, 6, 7] and to the prosody of the sentence [8, 9]. After analyzing sentences of "read" and "spontaneous" corpora, Lê T. X. [8] and Nguyễn Thị T. H. & Boulakia [9] noted that there is a difference in height of F0 between sentence types. By evaluating their register level, [8 & 9] presented that the assertive sentence is marked with a low register whereas the interrogative and the injunctive sentence have a high register. Moreover, [9] made the report that, on the intonation contour level, a descending slope does not always correspond to a declarative sentence. On the duration level, the interrogative statements have a faster rhythm than the assertive and injunctive ones, although the difference in duration between the two last is not significant [9]. As for the intensity level, it is generally stronger in the interrogative sentence, and the intensity of the final syllables is often more significant than the other syllables of the sentence [9]. Based on these reports, we wish to further determine the prosodic differences between interrogative sentences and affirmative sentences. For this purpose, we built a specific corpus made up of pairs of interrogative/affirmative sentences. The two sentences of one pair have the same tonal context and the same number of syllables. The choice of identical tones enables us to eliminate the influence from the syllable tones on the general intonation of the sentence, and also to control the micro-variations of the intonation. Furthermore, to also eliminate all the phenomena of co-articulation, which could interfere with our prosodic

analysis, sentences had the same word structure, or we used words with little pronunciation difference. All these sentences were integrated in significant dialogues; so that their pronunciation is the most natural possible (we recorded the totality of the dialogues, and then extracted chosen sentences for analysis). Each dialogue is repeated five times by six native speakers (3 men and 3 women) from Hanoi (the North region, considered to be official pronunciation of the Vietnamese language). It is noted that, in Vietnamese, for the construction of interrogative sentences, besides using practically and systematically the "interrogative classifier" words, speakers can add at the end of the sentence certain words which are normally optional. However, this addition possibility, which does not make change to sentence's meaning, depends strongly on the habitude, on the way the person speaks, on the context in which the dialogue occurs, on the expression of respect and/or courtesy towards the interlocutor, etc. Because these optional final words can carry any of the six tones of Vietnamese, the final portion of the intonation contour can be modified by the contour of the tone of these final words. This is why, for each selected interrogative sentence, we decided to incorporate into the corpus a certain number of variants which have final words with different tones, in order to study as much as possible the forms of sentence intonation contours. The Table 1 presents the 14 selected sentence pairs investigated. A complete sentence is formed by the root part (underlined text) followed by one of words in ending part (words in bracket "[]" and separated by "'"). We noted that one root can combine with one of many terminals while the sentence's meaning is still unchanged.

Table 1: 14 pairs of affirmative (A) and interrogative (I) sentences in corpus.

Affirm 1 to 5	<u>Hôm nay là ngày</u> + [ba mươi ba mươi rồi ba mươi vậy ba mươi đây ba mươi bảy*] <i>Today is thirty (today is thirty seven*)</i>
Interro 1 to 5	<u>Hôm nay là ngày</u> + [bao nhiêu bao nhiêu rồi bao nhiêu vậy bao nhiêu thế bao nhiêu hả] ? <i>What is the date today?</i>
Affirm 6 to 9	<u>Tên anh ta là</u> + [Trì Trì rồi Kỳ Cậy* Kỳ Thế**] <i>His name is + [Trì Trì Kỳ Cậy* Kỳ Thế**]</i>
Interro 6 to 9	<u>Tên anh ta là</u> [gì rồi vậy thế] ? <i>What is his name?</i>
Affirm 10 to 12	<u>Anh ăn cơm</u> + [không vậy Không Thế **] <i>He eats rice only (He eats the rice Không Thế**)</i>
Interro 10 to 12	<u>Anh ăn cơm</u> + [không không vậy không thế] ? <i>Do you eat rice?</i>
Affirm 13	<u>Em ăn bánh Chè</u> <i>I eat Chè cake.</i>
Interro 13	<u>Em ăn bánh nhé?</u> <i>You will eat cake?</i>
Affirm 14	<u>Ba giờ chỉ gặp anh Nghĩa</u> <i>I meet Mr Nghĩa at three o'clock</i>
Interro 14	<u>Bao giờ chỉ gặp anh Nghĩa?</u> <i>When do you meet Mr Nghĩa ?</i>

2.2. Analysis results

For each signal, we have analyzed the contour of the fundamental frequency F0 using the software Praat (example presented in Figure 1).

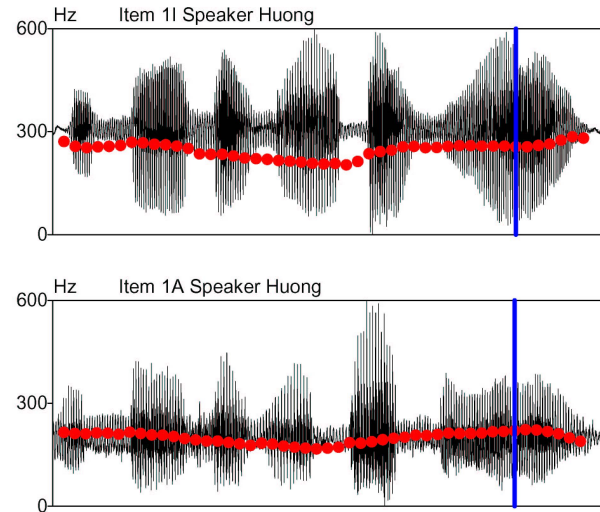


Figure 1: Two sentences (1I and 1A of Table 1) having the same number of syllables and the same tone. F0 contour is in red; on top: interrogative sentence, on bottom: affirmative sentence

By studying each pair of sentences presented in Table 1, we found that the main part of differences in intonation is at the end of the sentence (zone located on Figure 1 after the vertical bar): the contour of the last syllable or of its second half tends to increase for the interrogative sentences. A statistical study presented in Table 2 confirms this tendency: 85% of interrogative (I) sentences have an F0 contour increasing at the end of sentence. We recover here a well-known tendency for the non-tonal languages like French.

Table 2: F0 contours of the last half of the last syllable: Count (and percent) of contour rising/falling

Sentence type \ F0 contour	Rising contour	Falling contour
Interrogative	357(85%)	63(15%)
Affirmative	190(45%)	230(55%)

However, there's special case of 3 affirmative sentences (A) with the contour of the last half of the last syllable also increased for 29 of 30 recordings (even 30 of 30). These sentences are: "Em ăn bánh Chè"; "Tên anh ta là Kỳ Thế"; "Anh ăn cơm Không Thế". All of them contain terminal word which is a proper noun of ton5 (rising tone). Due to this fact, speakers have tendency to pronounce them very clearly to avoid misunderstanding which makes sentence intonation contour on this region very closed to syllable contour. The first sentence contains even two words of rising tone at the end. Both of these facts influence strongly the global intonation contour of the sentence as illustrated in the following Figure 2.

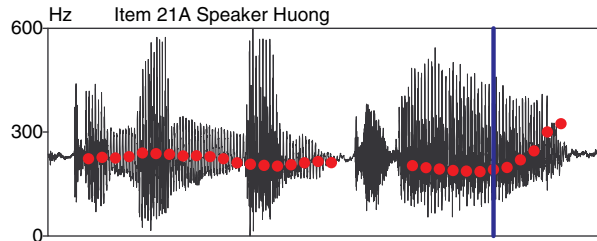


Figure 2: The sentence "Em ăn bánh Ché" with both two final words of rising tone

If we look closer to the form of sentence final part in case of different final tons, we can remark that: in case of I sentences, the contour of sentence final part increases regardless of the tone of the final syllable, or, increases in all cases of final tons (from 1 to 6). But it's not the same in case of A sentences where the slope of this region varies with final tons: it increases in case of ton3 (ngã-broken) or ton5 (sắc-rising), but decreases with remaining tons as illustrated on the following Table 3:

Table 3: Contours of the last-half of the last syllable in interrogative sentence with terminal words of different tones: count (and percentage %) contours rising/falling.

	ton1	ton2	ton3	ton4	ton5	ton6
Interrogative Rising	36 (60)	69 (77)	30 (100)	24 (80)	116 (97)	82 (91)
Interrogative Falling	24 (40)	21 (23)	0 (0)	6 (20)	4 (3)	8 (9)
Affirmative Rising	1 (2)	9 (10)	30 (100)	13 (43)	119 (99)	18 (20)
Affirmative Falling	59 (98)	81 (90)	0 (0)	17 (57)	1 (1)	72 (80)

[8 and 9] suggested that the sentences of the type I are marked with a higher register. For this point, statistical study on our corpus shows that sentences of type I have an average value of F0 higher than that of sentences of type A (Table 4). However, while the difference is significant for female speakers, that of male speakers are weak and smaller than the values of the corresponding standard deviation. Thus, unlike work of [8 and 9], the effect of register is not very significant in our corpus. For the duration, we find the same tendency noted in [9]: the duration of the interrogative sentences is on average smaller 10% than that of the affirmative ones (12% in [9]).

Table 4: F0 average (and standard deviation) in Hz of I and A sentences of six speakers. (M = Male, F = Female).

Speaker	Diệp F	Hương F	Lan F	Thành M	Khoa M	Phương M
Interrogative	261 (22)	259 (22)	253 (12)	143 (10)	144 (11)	124 (8)
Affirmative	226 (19)	217 (18)	245 (14)	136 (7)	127 (10)	119 (6)

3. Perception of I/A sentences

3.1. Methodology and corpus preparation

We wish to verify whether the differences detected in our analysis are actually perceived as a factor for listeners to

classify interrogative sentences and affirmative sentences, or, in other words, that prosody of the sentence, in spite of its complexity due to the presence of tones, carries information which allows listeners to make this classification. We used sentence pairs in the corpus described above. For each sentence, after extraction of the prosodic contour, we used that contour to synthesize a pseudo sentence in which all syllables are replaced by the vowel /a/. Because the lexical signification of the words does not exist anymore, we thus eliminate the possibility for the listener to recognize a question only by the presence of an "interrogative" word. We have reproduced as accurately as possible, not only the intonation contour, but also the duration of the voiced/unvoiced segments, along with the intensity contour. Then, listeners heard these pseudo-sentences and were asked to determine whether the perceived synthetic sentences were interrogative or affirmative.

3.2. Synthesis for perception test

The extraction of this prosodic information is carried out by the software Praat with an analyzing windows of 20 ms for F0 and 5ms for the intensity. For the synthesis, we extracted two periods from signal of a vowel /a/ in one sentence uttered by the same speaker of our preceding study. A male voice and a female voice were synthesized.

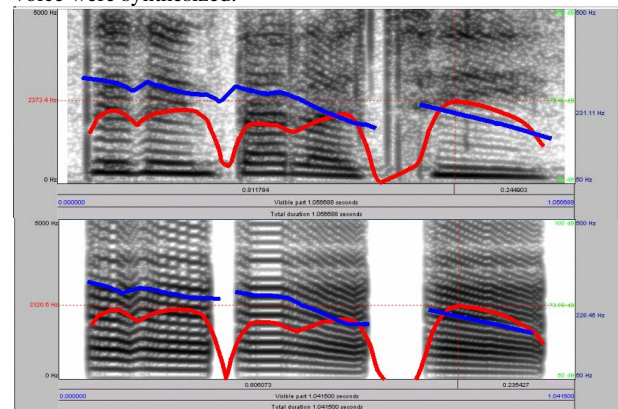


Figure 3: Spectrogram, F0 contour (blue) and energy contour (red) of source signal (a) and corresponding synthesis signal (b). Sentence 6a of Table 1

The TD-PSOLA algorithm was used to concatenate these extracts of signal, and to control the pitch (F0) of synthesized sentences, and zones corresponding to an unvoiced signal (consonants) in these sentences were replaced by silence. We thus obtained a corpus made up of 13 synthetic "interrogative" pseudo-sentences and of 13 synthetic "affirmative" pseudo-sentences, without any semantic information. Six listeners (3 men and 3 women) participated in our perception test. They had to choose between two answers "I" or "A". Each listener did the test 5 times for female voice and 5 times for male voice, and for each time, the order of the sentences proposed to him was random.

3.3. Perception results

The perception test results are presented in Table 5. For female synthesis voice, the global correct recognition rate on the whole of I and A is approximately 70%. We can see that the interrogative sentences were better recognized (approximately 74 % of good answers) than affirmative sentences (only 63%).



For male synthesis voice, the global recognition rate is only 60% while other rate for I/A sentences are also weaker than those of female voice.

Table 5: Correct recognition rate (in percent) of sentence types

Rates \ Synthesis voice	Female	Male
Global Correct Recognition rate	69%	58%
Correct Recognition rate of I sentences	74%	61%
Correct Recognition rate of A sentences	63%	55%

We can see that with female synthesis voice, listeners can do the recognition and classification better than with male voice. Based on analyzed results, we found that the F0 variations of this male speaker are less significant than that of female speaker.

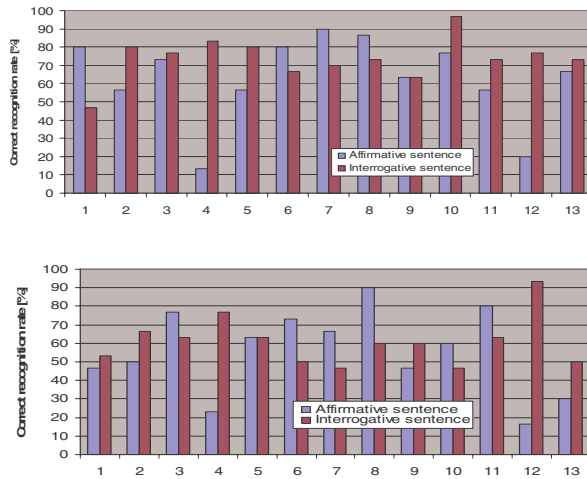


Figure 4: Correct recognition rate of 13 pairs of sentences I/A.

Synthesis of: on top female voice, on bottom male voice.

Figure 4 details the results for the 13 pairs of I/A sentences. With female voice, for 10 of these 13 pairs, the interrogative sentence is well recognized with a rate higher than 70%, and with the 10th pair, this rate even reaches 95%. However for pairs 4 and 12, the affirmative sentence is very poorly recognized (12% and 20% respectively).

3.4. Discussion

While trying to correlate these perceptive results with those of our analysis on the intonation contour production, we found that listeners had a tendency to consider a sentence as being interrogative if it presents a rising intonation at the end, and to consider the sentence as being affirmative in the opposite case. This assumption seems to be valid to explain the case of pairs 4 and 12 where recognition rate of I-type sentences is much higher than that of A-type sentences. For these two pairs, all the sentences present a last syllable having tone 5 (rising tone), which makes the final part of the intonation contour of the sentence raise, both for interrogations and assertions. The fact that the global correct recognition rate of sentences A and I is approximately 70% (and some of them have correct recognition rate higher than 90%) shows that the prosodic parameters of Vietnamese sentence transport extra-linguistic information which can allow listeners to discriminate sentence types. Beside other factors of intensity and duration, as for the non-tonal languages, this information is coded by the fact that the

intonation goes up or not at the end of the sentence. However, this information can be scrambled by the modulation of prosodic contour by the lexical tone: listeners can badly classify assertions in case that produced sentences present a final syllable with the rising tone. Questions can be badly classified if their final syllable carries a falling tone. The use of interrogative words to eliminate ambiguities is thus necessary and logical.

4. Conclusions

In the production level, our study help us to characterize the prosody of simple sentences of Vietnamese language (dialogue), by eliminating the influence from tone: the differences between interrogative and affirmative sentences are characterized primarily by a difference in F0 contour (increasing or decreasing) at the end of the sentence (second half of the last syllable), and by a modification of speaking rate. However, for our study, the change of register seems weaker than for [8 and 9]. At the perceptive level, we showed that, as for the non-tonal languages, the prosody of the sentence transports extra-linguistic information of the type of the sentence, but they are not always discriminative, due to the presence of the lexical tone.

5. References

- [1] Rossi M. "L'intonation, le système du français : description et modélisation" Editions Ophrys, 1999, ISBN : 2-7080-0912-5.
- [2] Hirst, D.J. & Di Cristo, A. (Eds.) "Intonation Systems. A Survey of 20 Languages" Cambridge: Cambridge University Press.
- [3] Shriberg, E., Bates, R., Taylor, P., Stolcke, A., Jurafsky, D., Ries, K., Coccaro, N., Martin, R., Meteer, M. & Van Ess-Dykema, C. "Can Prosody Aid the Automatic Classification of Dialog Acts in Conversational Speech?" *Language and Speech* 41, pp. 439-487, 1998.
- [4] Vu M.Q., Castelli E., Boucher A. & Besacier L. "Classification de parole en Question et Non-Question par arbre de décision" *SFC 05, 12^{èmes} Rencontres de la Société Francophone de Classification* - Montréal, 2005
- [5] Nguyen Q.C., Pham Thi N.Y. & Castelli E. "Shape vector characterization of Vietnamese tones and application to automatic recognition" *ASRU 2001 Madonna di Campiglio*, cdrom
- [6] Pham Thi N. Y., Castelli E. & Nguyen Q.C. "Gabarits des tons vietnamiens" *JEP 2002 Nancy*, pp 23-26, juin 2002
- [7] Michaud A. & Vu N.T. "Glottalised and non glottalised tones under emphasis: open quotient curves remain stable, F0 curve is modified" *Speech Prosody, Nara, Japan*. 745-748, 2004
- [8] Lê Thị X., "Etude contrastive de l'intonation expressive en français et en vietnamien". *Thèse en linguistique : Paris, Université Paris 7*, 1989
- [9] Nguyễn Thị T.H. & Boulakia, G. "Another look at vietnamese intonation" *14th International Congress of Phonetic Sciences, San Francisco, California*, pp. 2399-2402, 1999
- [10] Yuan, J.; Shih, C.; Kochanski, G. P., Comparison of Declarative and Interrogative Intonation in Chinese. *Proceedings of Speech Prosody 2002, Aix-en-Provence, France*, pp. 711-714.