

# Modelling the Interaction of Intonation and Lexical Tone in Vietnamese

Kieu-Phuong Ha and Martine Grice

IfL Phonetik, University of Cologne, Germany

kha@smail.uni-koeln.de, martine.grice@uni-koeln.de

## Abstract

Analysis of telephone conversations within the framework of Conversation Analysis reveals that Vietnamese makes pragmatic use of intonational tones. Pitch contours on one-word utterances in backchannels, requests for information, turn exits and repair initiations are compared. Target words bear one of two lexical tones: the low-level tone *thanh huyền* or the high-level tone *thanh ngang*. Results shed light on the interaction between these tones and those used for signalling communicative functions (intonation). An autosegmental analysis of this interaction is provided.

**Index Terms:** prosody, tone-intonation interaction, Vietnamese, pragmatics

## 1. Introduction

Over several decades research on Vietnamese intonation has concentrated largely on different sentence types or on syntactic phrasing and, recently, on emphasis and focus. For sentence types, different pitch contours were identified intuitively [18, 17], whereby it was supposed that only the phonetic characteristics of lexical tones were modified, leaving their phonological contrasts intact [cf. 3]. Furthermore, acoustic studies show that Vietnamese makes use of variations in F0 range as markers of emphasis and focus [9, 7].

Since the speech materials used in the above studies were read aloud without real interactional contexts, it is unclear to what extent these observations can be transferred to spontaneous speech. Recent work has identified a number of functions of Vietnamese prosody used at the utterance level [4, 5]. Adopting the framework of Conversation Analysis [13, 14, 6], these investigations have shown that pitch contours, in the sense of intonation, can be used as interactively relevant cues in Vietnamese conversation. This study aims to address the question of how lexical tones interact with intonation and to formalise this interaction within autosegmental phonology.

Section 2 provides an overview of four different functions of prosody in Vietnamese and illustrates the pitch contours used to express them on words with two different lexical tones; section 3 suggests a formalisation of the tone-intonation interaction and provides evidence for the analysis. Conclusions will be drawn in section 4.

The results presented here are based on a corpus of 18 telephone conversations with 28 speakers of the Northern Vietnamese dialect, 15 female and 13 male, aged between 11 and 63, recorded in Hanoi and Berlin. The corpus consists of approximately 57 minutes of speech in total.

## 2. Functions of prosody

While investigating discourse markers and one-word utterances in Northern Vietnamese everyday telephone conversations, we found different pitches on the same discourse marker/utterance depending on the interactional context or activity in conversation [4, 5]. For convenience, the

investigated words and their lexical tones are provided in the following table, along with their typical citation forms:

Table 1. *Tonal properties of investigated words*

| Word          | Lexical tone<br>(Tone name)    | Stylised<br>contour in<br>citation form | Gloss |
|---------------|--------------------------------|---|-------|
| ò / ì [ɔ]/[ɯ] | low-level<br>( <i>huyền</i> )  | -----                                   | yes   |
| gì [zi]       | low-level<br>( <i>huyền</i> )  | -----                                   | what  |
| vâng [vɤŋ]    | high-level<br>( <i>ngang</i> ) | -----                                   | yes   |
| ai [aɪ]       | high-level<br>( <i>ngang</i> ) | -----                                   | who   |

### 2.1. Backchannels (BC)

In backchannels, utterances spoken by the hearer to signal her/his attention to the current speaker's talk without taking the floor, a falling or low level pitch contour is frequently used on the words *ò* and its variation *ì* (both meaning 'yes' and bearing the lexical low-level tone). This realisation is consistent with the lexical tone of these two words. What is striking is that the word *vâng* (also meaning 'yes', often used when addressing elders) which has a lexical *high-level* tone can also be produced with a *falling* or *low level pitch* when it is used as a backchannel. That is, *vâng* does not display the lexical component of its high-level tone when produced as a backchannel, indicating that, in this context, the intonationally motivated pitch contour may be able to override the lexical tone. Figure 1 provides pitch contours on *vâng* in backchannels (with an English translation of speaker OD):

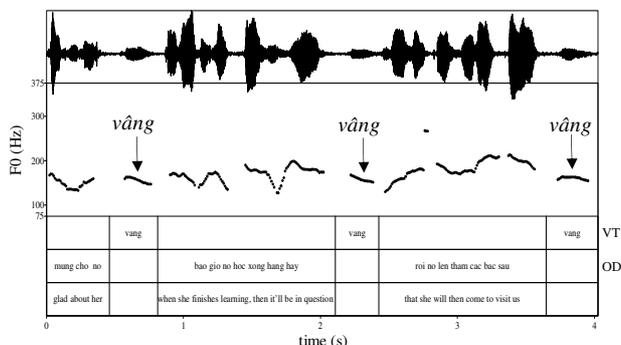


Figure 1: *Three instances of falling/low level pitch on vâng ('yes', lexical high-level) spoken by speaker VT (female) in backchannels*

## 2.2. Requests for information during telephone openings (Req)

In Schegloff's work on American English [15], standard openings in telephone calls involve a summons/answer sequence, an identification sequence, a greeting sequence and an exchange of how-are-you sequences. Somewhere at the beginning, the caller must give the reason for the call, which is also the first topic of the conversation.

In the Vietnamese corpus, *ờ*, *ừ* (lexical low-level tone) and *vâng* (lexical high-level tone) also occur during the opening of some telephone calls. In this position, a *rising pitch contour* can be used on these words – regardless of their lexical tones – by speakers to *acknowledge information* (e.g. asked by callers) and at the same time to *request the identification and/or the reason for the call*. (Note that this was only the case for female speakers).

## 2.3. Turn exit technique (T-Ex)

The *rising pitch contour* in requests for information typical of the female speakers during telephone openings is used by both female and male speakers during the body of telephone calls. This is usually followed by a turn change. This pitch contour can thus function as what is referred to in Conversation Analysis as a turn exit technique [13]. That means Vietnamese speakers can use this pitch contour on acknowledgements to signal that they are at the end of the turn and that the other speakers have the right/obligation to take the next turn. Example (1) is an excerpt from a telephone call made by two male speakers, HV and CT. They are friends, aged 56 and 60. CT explains that his son cannot go to school at the age of 6 years and 8 months. In the transcript ‘[’ stands for overlapped speech, ‘(.)’ for micro pause, ‘( )’ for incomprehensible words, ‘:::’ for lengthening, L% and H% for falling and rising final pitch movements and ‘→’ for investigated line. An English translation is provided in italics.

- (1) CT nếu mà::[: đến a::: sinh từ tháng chín [đỗ về (.) trước  
*If till uh if he were born before September,*  
 HV [đủ ( ) tròn tuổi [ờ  
*Reach the age of 7 yes*  
 CT là được sau [thì là thôi  
*he could go to school, born after that he cannot*  
 HV [mmL% mm L% mm L%  
*mm mm mm*  
 CT sau là phải sang năm  
*Born after September, he has to wait till next year.*  
 HV mm L% mm L%  
*Mm mm.*  
 → CT ờ H%  
*Yes.*  
 HV thế là cũng ( ) chậm mất một năm nhờ  
*That means he'll go to school almost a year late,*  
*right?*  
 CT ((laughs) ờ L%)  
*Yes.*

During CT's explanation, HV signals that he is listening and acknowledges the information provided by CT with a low fall. HV takes the floor after the acknowledgement produced by CT with a rise in pitch (Figure 2). Note that the rise is on *ờ* which has a lexical low-level tone.

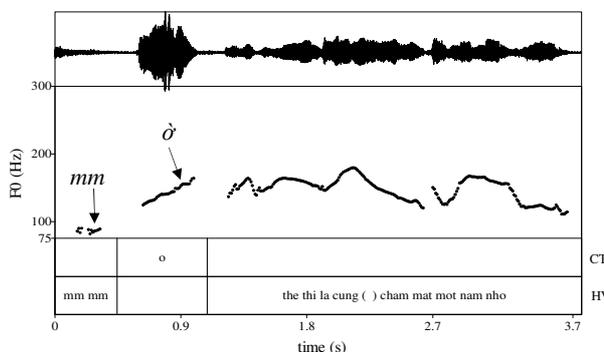


Figure 2: Rising pitch contour on *ờ* ('yes', lexical low-level) spoken by speaker CT (male) as turn exit technique (and low level pitch on *mm* as backchannel spoken by speaker HV)

The requests for information during telephone openings and the turn exit techniques discussed here have a similar function. It is therefore not surprising that the typical pitch contour is the same. Table 2 provides the pitch contours used on *ờ*, *ừ* and *vâng* in backchannels, requests for information during telephone openings and turn exits. Note that not all investigated words are found in all interactional contexts.

## 2.4. Repair initiations after mishearing (RI)

In spontaneous speech, problems in the production, hearing or understanding of utterances cannot be avoided. Thus, there must be a mechanism for interlocutors to manage their talk, so that they "are working with similar understanding of what one another is saying and meaning", and in order to maintain "mutual orientation to common topics and fields of reference in talk-in-interaction" [6: 64-65, cf. 16]. This mechanism is referred to in Conversation Analysis as "repair", with three phases: (i) trouble source turn, (ii) initiation of repair and (iii) performance of repair ([14], see example (2)). In the work on prosodic cues of self-repair by Nakatani and Hirschberg [10], a repair is divided into three intervals: the 'reparandum interval', the 'disfluency interval' and the 'repair interval'. While [10] dealt with *self-initiated* self-repairs, i.e. the repair is both initiated and performed by the speaker of the trouble source turn, this study investigates so-called *other-initiated* self-repairs, i.e. the repair is performed by the speaker of the trouble source turn but initiated by the other interlocutor.

- (2) F: This is nice, did you make this?  
 (i) K: No. Samu made that.  
 (ii) F: Who?  
 (iii) K: Samu.

[14: 367-68]

Table 2. Pitch contours on *ờ*, *ừ* and *vâng* in different contexts

| Word                 | Stylised contour in citation form | Backchannels (BC) | Requests for info (Req) (female) | Turn exit technique (T-Ex) |
|----------------------|-----------------------------------|-------------------|----------------------------------|----------------------------|
| <i>ờ / ừ</i> [ʌ]/[u] | -----                             | -----             | -----                            | -----                      |
| <i>vâng</i> [vʌŋ]    | -----                             | -----             | -----                            | -----                      |

Sequences like (2) in the Vietnamese corpus show that this language uses tonal means to initiate repairs. One-word utterances like *gì* and *ai* (with different lexical tones, as listed in Table 1) are produced with a *high rising pitch* when speakers have problems hearing an utterance and thus initiate a repair from conversational partners. In some cases, repair initiations are accompanied by an additional loudness or a fast tempo. The typical pitch contours used in this context are provided in schematic form in the following table:

Table 3. *Pitch contours used in repair initiations (RI)*

| Word           | Stylised contour in citation form | RI in fast speech | RI in slow, careful speech |
|----------------|-----------------------------------|-------------------|----------------------------|
| <i>gì</i> [zì] | -----                             | -----             | -----                      |
| <i>ai</i> [aɪ] | -----                             | -----             | -----                      |

The extent to which the lexical tone is discernible in repair initiations depends on the speech rate or style. The falling component of *gì* disappears in fast speech, but is retained to some extent in slow, careful speech. It is important to note that in repair initiations, no matter whether the lexical tones remain or not and whether in fast speech or slow, careful speech, there must be a (high) rising pitch at the end of the phrase.

### 3. Interaction of tone and intonation

Vietnamese possesses a great number of *sentence particles*, for example to form questions [17]. However, speakers can also form questions without particles. In this case, a high pitch at the end of the phrase functions as a marker for question modality [3]. This is analysed as an intonation phrase *final boundary tone*. Evidence for a final high boundary tone has also been found in the spontaneous speech reported above: in requests for information during telephone openings (2.2) and in repair initiations (2.4), both of which are questions, and in turn exits (2.3). The analysis of backchannels provides evidence for low boundary tones, supported by our findings on the semantically empty word *mm* discussed in section 3.3. Before proceeding with the formalisation of tone and intonation in conversation, the tone system of Vietnamese is summarised, based on results of previous research.

#### 3.1. Vietnamese tone

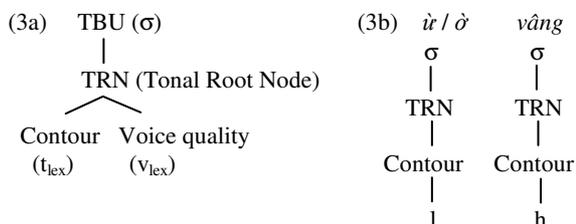
Vietnamese tones are characterised not only by pitch (F0), but also by voice quality [1, 11, 12, 8]. The six lexical tones are distributed across two tone registers which are also claimed to be related to voice quality (creaky voice and breathiness) [12]. Since the analysis of register in Vietnamese is highly controversial (see [2] also for discussion from perceptual perspective), we restrict this study to the question of how lexical tones interact with intonational tones without considering the voice quality. Table 4 provides the properties of the two lexical tones investigated (l = low, h = high):

Table 4. *Characterisation of investigated lexical tones*

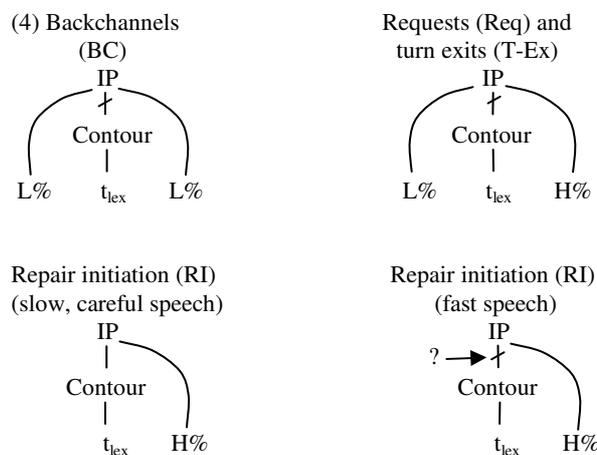
| Words        | Lexical tones               | Phonological analysis (Contour) |
|--------------|-----------------------------|---------------------------------|
| <i>ừ / ừ</i> | low-level ( <i>huyền</i> )  | l                               |
| <i>gì</i>    | low-level ( <i>huyền</i> )  | l                               |
| <i>vâng</i>  | high-level ( <i>ngang</i> ) | h                               |
| <i>ai</i>    | high-level ( <i>ngang</i> ) | h                               |

#### 3.2. Interaction of tone and intonation

The current study uses an autosegmental model adapted from [12] exhibiting a sisterhood relationship between contour and voice quality features of lexical tones. (3a) illustrates the structure of the tonal representation associated with the Tone Bearing Unit (TBU) which in Vietnamese is the syllable. (3b) shows an analysis of sample words with two lexical tones *huyền* and *ngang* (voice quality not considered).



(4) represents the formalisation of the interaction between tone and intonation in the investigated utterances:



Backchannels are analysed as having a low boundary tone at the beginning and at the end of the intonation phrase (IP). In *ừ* and *ừ*, the pitch contour is consistent with the lexical tone (low-level). However, this is not the case for *vâng*, which has a lexical high-level tone. Thus, regardless of the lexical tone, backchannels are typically produced low level or falling.

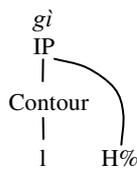
Similarly, in requests for information and turn exits, the initial low and final high boundary tones are reflected in a rising contour which is not affected by the lexical tone (see Table 2).

By contrast, repair initiations in slow, careful speech are produced with a final (high) rise, preceded by the lexical tone. In fast speech the lexical tones appear to be lost, possibly reflecting delinking of the lexical tones. Compared to requests for information and turn exits, the boundary H% in repair initiations accounts on its own for the high rise, presumably leaving more room for the realisation of the lexical tones in slow, careful speech.

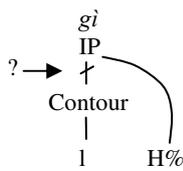
(5) represents for each communicative function an example of the interaction between tone and intonation:



RI in slow, careful speech



RI in fast speech



### 3.3. Evidence from the semantically empty word *mm*

In the corpus, *mm*, a semantically empty word, can be found in backchannels (acknowledgments), although less frequently than *ờ* and *ừ*, and in repair initiations. If the word *mm* has no lexical tone, we should expect to find evidence of the intonation alone. This means that intonational contours on *mm* should also look like the ones in backchannels and repair initiations on the words *ờ*, *ừ*, *vâng*, *gi*, and *ai*.

As a backchannel, *mm* is produced low falling or low level (see *mm* in Figure 2, spoken by speaker HV (male)). Figure 3 shows the pitch contour produced on *mm* in a repair initiation by speaker HO (female). The repair is initiated with a rising pitch which is similar to the pitch in repair initiations shown in Table 3. These findings serve to validate the formalisation of the two contexts in (4).

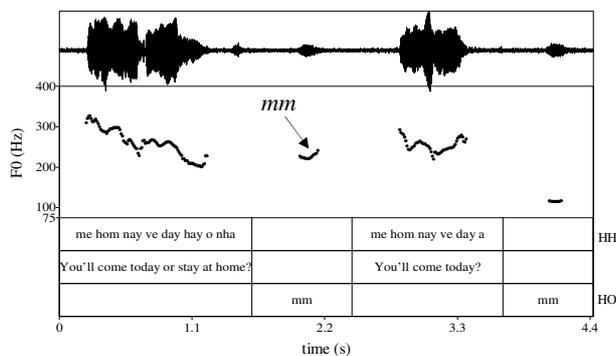


Figure 3: Pitch contour on *mm* produced by speaker HO (female) in a repair initiation

## 4. Discussion and conclusion

Results show that Vietnamese makes use of intonational tones to convey different contexts in talk-in-interaction. The typical form-function relations investigated are as follows: In backchannels, low boundary tones (L%) are used at the beginning and at the end of the phrase, whereas lexical tones appear to be lost. In requests for information during telephone openings and in turn exits, a rising contour (initial L% and final H%) is produced, again not affected by the lexical tones. In functional terms we are dealing with one-syllable utterances which bear initial and final boundary tones, possibly making it difficult for lexical tones to be realised. Speech rate does not appear to influence the realisation of these tones.

Repairs, by contrast, are realised with only one final boundary tone. In fast speech the lexical tones also appear to be absent, while in slow, careful speech they are present in the first portion of the signal. However, there is insufficient data to ascertain whether lexical tones are phonologically delinked in fast speech or whether we are dealing with a gradual process of (e.g.) coarticulation.

We have provided evidence (a) for input to the tonal phonology of Vietnamese not only from lexical but also from postlexical sources, and (b) for an analysis in which lexical tones can be masked or even overridden by edge tones. However, our study deals with only two lexical tones *thanh*

*huyền* and *thanh ngang* in function words and has looked exclusively at one-syllable phrases. Future work needs to include the other lexical tones on function words as well as content words, and also needs to consider longer phrases and different speech rates.

## 5. Acknowledgements

We would like to thank Marc Brunelle, Stefan Baumann and Christine Röhr for stimulating discussion and feedback, and the A.R.T.E.S. Graduate School, University of Cologne, for financial support of this research project.

## 6. References

- [1] Alves, M. (1997): "Tonal Features and the Development of Vietnamese Tones". *Working Papers in Linguistics*: Department of University of Hawaii at Manoa, Vol 27, 1-13.
- [2] Brunelle, M. (2009): "Tone Perception in Northern and Southern Vietnamese". In *Journal of Phonetics* 37, 79-96.
- [3] Do, T. D. & Tran, T. H. & Boulakia, G. (1998): "Intonation in Vietnamese". In Hirst, D. & Di Cristo, A. (eds.): *Intonation systems. A survey of twenty languages*. Cambridge University Press, 395-416.
- [4] Ha, K. P. (submitted): "Prosody of Vietnamese from an Interactional Perspective: Examples of 'Yes' in Everyday Telephone Calls". *Journal of Southeast Asian Linguistics*.
- [5] Ha, K. P. (2009). "Prosodic Means in Repair Initiation as an Activity in Vietnamese Conversation". Oral presentation at: *Workshop Linguistics of Vietnamese*, University of Stuttgart, July 6-7, 2009.
- [6] Hutchby, I. & Wooffitt, R. (2008): *Conversation Analysis*. 2nd edition, Polity Press, UK.
- [7] Jannedy, S. (2007): "Prosodic Focus in Vietnamese". In Ishihara, S. & Jannedy, S. & Schwarz, A. (eds.): *Interdisciplinary Studies on Information Structure* 08, 209-30.
- [8] Michaud, A. (2004): "Final Consonants and Glottalization: New Perspectives from Hanoi Vietnamese". In *Phonetica* 61, 119-46.
- [9] Michaud, A. & Vu, N. T. (2004): "Glottalized and Nonglottalized Tones under Emphasis: Open Quotient Curves Remain Stable, F0 Curve is Modified". *Proceedings of the international conference Speech Prosody 2004*. Nara, Japan, 745-48.
- [10] Nakatani, C. & Hirschberg, J. (1994): "A corpus-based study of repair cues in spontaneous speech". In *Journal of the Acoustical Society of America* 95, 1603-16.
- [11] Nguyen, V. L. & Edmondson, J. A. (1998): "Tones and voice quality in modern northern Vietnamese: Instrumental case studies". In *Mon-Khmer Studies* 28, 1-18.
- [12] Pham, A. H. (2003): *Vietnamese Tone. A New Analysis*. New York, Routledge.
- [13] Sacks, H. & Schegloff, E. A. & Jefferson, G. (1974): "A simplest systematics for the organization of turn-taking for conversation". In *Language* 50, 696-735.
- [14] Schegloff, E. A. & Jefferson, G. & Sacks, H. (1977): "The preference for self-correction in the organization of repair in conversation". In *Language* 53, 361-82.
- [15] Schegloff, E. A. (1986): "The routine as achievement". In *Human Studies* 9, 111-52.
- [16] Schegloff, E. A. (1992): "Repair after next turn: the last structurally provided defense of intersubjectivity in conversation". In *American Journal of Sociology* 97, 1295-345.
- [17] Thompson, L. C. (1965): *A Vietnamese Reference Grammar*. In *Mon-Khmer Studies XIII-XIV*. University of Washington. Reprint 1987: *A Vietnamese Reference Grammar*. Seattle: University of Hawaii.
- [18] Tran, H. M. (1967): "Tones and Intonation in South Vietnamese". *Series A - Occasional Papers #9, Papers in Southeast Asian Linguistics No. 1*. Nguyen D. L., Tran, H. M. & D. Dellinger (eds.). Canberra, Linguistics Circle of Canberra.