Prosody in Vietnamese: 
Intonational Form and Function of Short Utterances in Conversation

Kieu-Phuong Ha

A-PL 002 / SEAMLES 001

This is the first book illustrating the use of intonation in Northern Vietnamese based on a hybrid approach combining Conversation Analysis and Laboratory Phonology. The results show that the role of intonation in (Northern) Vietnamese has been greatly underestimated in previous investigations. The interaction between intonation and lexical tone is analysed in the framework of autosegmental phonology. New evidence suggests that the melody at the edges of utterances in Northern Vietnamese can be analysed as a combination of the lexical tone of the final syllable and a boundary or intonational tone used to express communicative functions. The study also has implications for the analysis of intercultural communication.

URL: http://hdl.handle.net/1885/9543
# Table of contents

## 1. Introduction

- 1.1 Prosody research
- 2.1.1 Interactional approach
- 2.1.2 Experimental approach
- 2.1.3 Intonation in tone languages
- 2.1.4 Modelling tone-intonation interaction

## 2. Theoretical issues

- 2.2 Vietnamese
  - 2.2.1 Lexical tones
  - 2.2.2 Phạm’s model (2003b)
  - 2.2.3 Proposal of a neutral tone model
  - 2.2.4 Intonation

## 3. Telephone corpus description

- 3.1 Materials and methods
- 3.2 Utterances investigated
  - 3.2.1 Backchannels
  - 3.2.2 Repair initiations
- 3.3 Transcription/annotation

## 4. Prosody of backchannels

- 4.1 Background
- 4.2 Case-by-case analysis
  - 4.2.1 Falling/low level pitch contour on backchannels carrying a lexical low-falling tone
  - 4.2.2 Falling/low level pitch contour on the backchannel vâng (lexically high-level)
- 4.3 The non-lexical item mm
- 4.4 Quantitative analysis
5. **Other interactional contexts** .................................................................39
   5.1 Turn-yielding .......................................................................................39
   5.1.1 Background ...................................................................................39
   5.1.2 Case-by-case analysis ......................................................................40
   5.1.3 Quantitative analysis .......................................................................42
   5.2 Request for information during telephone openings .......................43
   5.2.1 Background ...................................................................................43
   5.2.2 Case-by-case analysis ......................................................................44
   5.2.2.1 Final high/rising pitch on acknowledgment words carrying a lexical low-level/falling tone ........................................................44
   5.2.2.2 Final rising pitch contour on the acknowledgment token vâng carrying a lexical high-level tone .....................................................................45
   5.2.3 Deviant cases ...................................................................................46
   5.3 Confirmation/agreement .................................................................49
   5.3.1 Case-by-case analysis ......................................................................49
   5.3.2 Quantitative analysis .......................................................................50
   5.4 Interim conclusions ...........................................................................50

6. **Prosody of one-word repair initiations** ...............................................52
   6.1 Background .......................................................................................52
   6.2 Case-by-case analysis .........................................................................53
   6.3 Quantitative analysis ...........................................................................60
   6.4 Longer repair initiation utterances ...................................................60
   6.5 The non-lexical item mm ...................................................................61
   6.6 Repair initiations vs. backchannels ..................................................63
   6.7 Interim conclusions ...........................................................................64

7. **Interaction between lexical tone and intonation** ................................ 67
   7.1 Data and methods .............................................................................67
   7.2 Backchannels and other interactional contexts ..................................68
   7.2.1 Backchannels ................................................................................68
   7.2.2 Other interactional contexts ...........................................................71
   7.3 Repair initiations ...............................................................................74
   7.4 Autosegmental representation of tone-intonation interaction ..............80
   7.5 Interim conclusions and discussion ...................................................83
8. Perception of paralinguistic meanings in backchannels........................................ 85
   8.1 Introduction............................................................................................................. 85
   8.2 Stimuli preparation ............................................................................................ 87
   8.3 Experiment I: A forced-choice task ..................................................................... 88
      8.3.1 Design and methods .................................................................................... 88
      8.3.2 Results .......................................................................................................... 89
   8.4 Experiment II: A semantic scaling task ............................................................... 90
      8.4.1 Design and methods .................................................................................... 90
      8.4.2 Results .......................................................................................................... 90
   8.5 Interim conclusions and discussion ..................................................................... 93

9. Summary and conclusions........................................................................................... 95

References.......................................................................................................................... 97

Appendix A: ....................................................................................................................... 104

Appendix B: ....................................................................................................................... 105
**Abbreviations used in glosses**

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>(CÓ)...KHÔNG</td>
<td>yes/no question frame, for example, <em>anh (có) khỏe không?</em> (word-for-word translation: you (are) healthy or not) ‘How are you?’</td>
</tr>
<tr>
<td>CONT</td>
<td>continuous form (like <em>-ing</em> in English)</td>
</tr>
<tr>
<td>ĐƯỢC</td>
<td>means ‘to obtain’, ‘to reach’</td>
</tr>
<tr>
<td>CLASS</td>
<td>classifier</td>
</tr>
<tr>
<td>IMP-PART</td>
<td>imperative particle</td>
</tr>
<tr>
<td>NEG</td>
<td>negation</td>
</tr>
<tr>
<td>PART</td>
<td>particle</td>
</tr>
<tr>
<td>PERS-SG</td>
<td>person, singular</td>
</tr>
<tr>
<td>POL-PART</td>
<td>politeness particle</td>
</tr>
<tr>
<td>Q-PART</td>
<td>question particle</td>
</tr>
<tr>
<td>SENT-PART</td>
<td>sentence particle</td>
</tr>
<tr>
<td>TOP-MARKER</td>
<td>topic marker</td>
</tr>
<tr>
<td>VOC-PART</td>
<td>vocative particle</td>
</tr>
<tr>
<td>Ý (mà)</td>
<td>sentence particle referring to the recognition of persons, places or matters by the speaker. Ý can also be used to elicit acknowledgement from the hearer.</td>
</tr>
</tbody>
</table>
This book would not have been possible without the help of many people. First of all, I would like to thank my supervisor, Martine Grice, for her great support from the very beginning of my doctoral studies at the University of Cologne. Her enthusiasm for research and phonetics has inspired me a great deal. She provided endless encouragement for me during the process of writing as well as in other difficult situations. The second person I would like to say thank you to is Marc Brunelle, my external supervisor from the University of Ottawa. Marc accompanied me from my first steps in looking for the research questions in the project. I have learned a lot from his innumerable comments on my papers, presentations and dissertation. Thank you to both of you, Martine and Marc!

I sincerely appreciate the a.r.t.e.s. research school of the University of Cologne for its financial and practical support during my PhD project from 2008 to 2011 including travel scholarships for my field work in Vietnam. Without this support this project would not have been possible.

I would not have accomplished this project without the support of the Cologne Phonetics Team: Martine Grice, Stefan Baumann, Anne Hermes, Jessica Di Napoli, Doris Mücke, Elina Savino, Sandra Vella, Johannes Becker, Nils Jagdfeld, Christine Röhr, Theo Klinker, Tingting Brengelmann, Martina Krüger, Henrik Niemann, Simon Ritter, Timo Röttger, Christine Riek and Silvia Dahmen. They helped me in designing and conducting experiments, gave me comments on the chapters in my thesis as well as constructive criticism on my practice talks before conferences. I thank you all! It is a pleasure for me to work with you as a team!

I thank Fritz Serzisko for attending the committee during my defense, Nikolaus Himmelmann for his comments on the drafts of some of my chapters, Bernd Pompino-Marschall for his supervision during my study in Berlin, and James Kirby for his comments on chapters in my thesis and all his conversations about the PhD.

Many thanks to Paul Sidwell, the managing editor of Asia-Pacific Linguistics, for his quick support in getting the thesis ready for publication.

A special thanks to Matthias Meissner for his encouragement both in work and life.

Finally, I want to express gratitude to my family, especially to my parents who have been with me in every step of my life. This book is dedicated to them.

Kieu-Phuong Ha
Cologne, November 2012
1 Introduction

Vietnamese, a Mon-Khmer language within the Austroasiatic family, is a tone language. Its complex tone system comprises six lexical tones in the Northern dialect and five in the Southern variety. The lexical tones exploit the entire pitch range and are characterised not only by means of pitch, but also by means of voice quality (Alves 1997, Nguyễn and Edmondson 1998, Phạm 2003a, 2003b, Michaud 2004). In contrast to some African tone languages such as Bantu languages, no tone spreading or downstep is found in Vietnamese. There is no attested fully-fledged tone sandhi although there is a degree of tonal coarticulation (Brunelle 2009b). As many other East Asian tone languages such as Mandarin Chinese, Cantonese and Thai, Vietnamese possesses a large number of particles for expressing grammatical functions. It has often been assumed that a complex tone system in combination with the extensive use of particles leaves little room for intonation to be used for expressing communicative functions at the utterance level.

Nonetheless, a number of studies based on descriptions of auditory observations as well as experimental data have shown that Vietnamese does modulate pitch to convey non-lexical information. In fact, it uses intonation to express a variety of different functions: different overall pitch contours were found to express different dialogue acts, such as statement vs. question (Thompson 1965, Trần 1967, Đỗ et al. 1998, Nguyễn and Boulakia 1999). Pitch contours have also been shown to play a role in the organisation of speech, for example, speech, where the falling contour signals finality and the rising contour non-finality (Trần 1967, Thompson 1965). Acoustic studies have found that variation in fundamental frequency (F0) – the physical property of the speech signal determined by the vibration rate of the vocal folds – can be used for pragmatic purposes: for example, a rising F0 and a higher F0 register are used as markers for emphasis (Michaud and Vũ 2004), and strong F0 variation is used as a marker of focus (Jannedy 2007). However, all of these studies are based on speech materials that were elicited by means of sentences or short paragraphs read aloud out of context and, more importantly, without interaction with an interlocutor.

The present study contributes to the growing body of research on Vietnamese intonation by using spontaneous speech and analysing speaker interactions. Our everyday conversation involves not only an exchange of messages but also a cooperation in which we convey to each other what is being heard, processed or understood. Besides words and their organisation, speech melody can be used extensively to signal whether we are listening and all is going well, or whether the speaker should backtrack and clarify. In this respect, several questions relating to tone languages like Vietnamese arise.

Research questions

Which functions can prosody, in particular intonation, convey in Northern Vietnamese? To what extent does it manifest itself, given that every single syllable/word in the language already has a lexical pitch? The current study aims to investigate the intonational form of
utterances in Northern Vietnamese discourse, focusing on their function as a conveyer of speaker-hearer interactions.

This aim is inspired by the Conversation Analysis framework (Sacks et al. 1974, Schegloff et al. 1977, inter alia), the main findings of which include the turn-taking system, the sequential organisation of talk and the activity of repair. According to this framework, talk-in-interaction is a structurally organised activity in social life.

**Objectives of research**

To investigate whether speakers of Vietnamese use pitch movements to manage the discourse and to convey activities in talk-in-interaction, a corpus of Northern Vietnamese telephone conversations was recorded and analysed. The study focuses predominantly on the prosody used in two strategies that interlocutors use to manage the discourse: the first is “backchannel-behaviour” which signals a listener’s attention to or interest in the talk of the speaker. In conversation it is important for interlocutors to exchange such Go on! I’m listening! signals appropriately, so as to function as an interested or cooperative listener. Backchannels are commonly short utterances (e.g. yes, right or exactly) or even utterances with non-lexical items (e.g. uh huh or mhm) with a distinctive prosody and which can be accompanied by facial expressions or gaze. This phenomenon reflects a strategy in conversation that we generally use unconsciously and to which we often pay little attention. Besides factors such as how often and in which position in talk we should produce backchannels, speech melody is largely responsible for our interpretation of such utterances, and consequently of another speaker’s attitude. Across cultures and languages the melody is not always the same. Thus, in intercultural communication or communication between native and non-native speakers of a language, conversational strategies which are used unconsciously often appear to be phenomena in speech where different linguistic patterns from one’s native language are transferred into one’s second or foreign language(s). Therefore, looking at speaker-hearer-interaction can help us to explain potential misunderstandings that do not appear to have to do with personalities or attitudes but instead with cultural or language backgrounds.

The second strategy is the activity of repair. In conversation it is unavoidable that we sometimes have trouble hearing or understanding what the other speaker has said. Repair is a universal system which provides us with a mechanism for signalling that the speaker should backtrack and, for instance, clarify. This system helps us manage our talk in such a way that we achieve mutual understandings, instead of talking about different topics or events which can consequently lead to misunderstandings or even failure in communication. Standard German, for example, uses a high overall pitch movement on the wh-question word was ‘what’, when the speaker has hearing problems, or an even higher overall pitch movement and a higher overall loudness on was, when what was previously uttered is contrary to his or her expectation (Selting 1996). With respect to the universal characteristics of repair, the current study will seek to find out whether Vietnamese utilises prosodic means in this conversational strategy.

From the perspective of speaker-hearer-interaction, these strategies can be categorised as merely hearer signals (Go on! I’m listening! or Stop! What did you say?), while from the perspective of the turn-taking system in Conversation Analysis, the backchannels can be grouped as signals contributing to the speech organisation (who is now the one who is talking) and repairs as signals exhibiting conversational activities (after a problem or misunderstanding, a hearer initiates a repetition of the prior utterance by the speaker). The current study will analyse these two strategies in line with the framework of Conversation Analysis. As will be shown later, speakers of Northern Vietnamese do use pitch variation on acknowledgment tokens to manage their interactive organisation: i.e. the low falling
pitch contour is used to signal attention (backchannels), whereas the rising pitch contour on exactly the same tokens is used to request the reason for the call, to yield the turn or to agree with the interlocutor. As for repair initiations, speakers use a rising pitch contour on all of the investigated tokens.

The analysis of the prosody in backchannels and repairs will focus on short utterances, their most common manifestation. In connected speech, backchannels and repairs are immediate measures used by a listener/addressee to renounce a speaking turn or to interrupt the ongoing speech flow of a speaker, respectively. In general, the speaker can continue the talk without disruption after receiving the listener’s signal that s/he is not taking the floor (backchannels). If the listener has lost the common ground, s/he has to re-establish it to keep the conversation going (repair initiations). Due to the rapid and immediate nature of this process, backchannels and repairs are often spoken with a fast tempo. These utterances thus tend to force lexical information into the background leaving room for intonational meanings (see Đỗ et al. 1998). In this particular environment, it is usually the intonation that enables conversational partners to manage their talk and (thus) maintain mutual understanding.

The investigated utterances comprise three categories of words in Northern Vietnamese:

(a) discourse particles (ờ, ừ and vâng, all can be translated as ‘yes’; dạ ‘politeness particle’; ơi ‘vocative particle’),
(b) function words (wh-question words gì ‘what’; ai ‘who’; sao ‘why/how’; hà ‘yes/no question word’),
(c) non-lexical items (mm).

The phonological properties of discourse particles and function words, which are commonly categorised as non-lexical words, have been reported to be significantly different from those of lexical words (Nespor and Vogel 1986). They are not as resistant to change during phonological processes as lexical words. For example, function words in Tokyo Japanese can lose their high tone accent when preceded by another accented word in the same phrase. This does not hold for lexical words (Selkirk 1996). Non-lexical items such as mm which have no lexical meanings cannot, by definition, have a lexical tone and can be seen as bearers of intonation alone. In conversation they will provide evidence that we are dealing with specific intonation contours. The words investigated in the current thesis have lexical tones and thus serve as initial objects of research to study the interaction of lexical tone and intonation.

In languages with complex tone systems, one often expects that speakers make use of overall pitch effects – changes in the overall pitch or overall pitch range – to convey communicative functions at the utterance level. Lexical pitch should be intact so that the propositional content of the message is transmitted. The present study will show that the lexical tones in Northern Vietnamese can be modulated by intonational tones used to convey utterance-level meanings. The latter are analysed in Autosegmental-Metrical Phonology as post-lexical boundary tones which are local in nature and therefore can affect the adjacent lexical tone. As a result, the thesis argues that the melody at the edges of utterances in Vietnamese can be analysed as a combination of the lexical tone of the final syllable and a boundary tone used for pragmatic purposes.

It is common knowledge that intonation can not only fulfil linguistic functions, it can also be used to express paralinguistic meanings, including emotional state, speaker’s attitudes or affect. The fact that linguistic and paralinguistic signals can interact often leads to the assumption with regard to interpersonal interaction that it is usually not important what is said but how it is said. Often we do not interpret intonation in terms of how it contributes to the flow of our talk, rather we see it in terms of speakers’ personalities or
intentions – whether or not someone is polite or interested, for example (Tannen 1986). Backchannels are signals in conversation that easily lead to these interpretations, based on whether or not someone is seen as an interested or cooperative/polite listener. The study of Chen (2005) on language-specific perception of paralinguistic intonational meanings shows that speakers from different cultures or even from different speech communities can interpret aspects of pitch modulation in terms of interpersonal intentions such as politeness or friendliness in different ways. The current thesis addresses this aspect by investigating how native listeners of Northern Vietnamese perceive affective meanings conveyed by pitch variation in backchannels.

Outline of the thesis

In lines with the aims mentioned above, the thesis comprises the following chapters. The next chapter (Chapter 2) introduces theoretical frameworks which underlie the empirical analysis: i) Conversation Analysis, the methodology of which is adopted to analyse the Northern Vietnamese corpus, ii) prosody research in interaction focussing on functional analysis and natural spontaneous speech, and experimental studies of intonation focusing on laboratory data. In this chapter, the characteristics of Northern Vietnamese tones and an overview of intonation research in Vietnamese are provided.

Chapter 3 describes the speech materials and gives an overview of the target utterances. Qualitative as well as quantitative data on the phonetic form of the pitch contour on the acknowledgment tokensờ, ừ and vâng are presented in Chapters 4 and 5. Chapter 4 deals with backchannel context, whereas Chapter 5 focuses on turn-yielding context, requests for information during telephone openings, agreements and confirmations. Chapter 6 deals with the intonation of repair initiations. A phonological analysis of the interaction between intonation and lexical tone is provided in Chapter 7.

Chapter 8 presents two experiments testing the naturalness and the affective meanings of politeness and dominance perceived by native listeners of Northern Vietnamese when the final pitch on backchannel utterances is modified. Results will be discussed in terms of potential misunderstandings between Northern Vietnamese speakers and speakers of other languages as well as cultures.

Chapter 9 summarises the results of the study and addresses its contribution to theoretical and methodological issues within intonation research, in particular Vietnamese intonation.
2 Theoretical issues

In this chapter a brief introduction to the theoretical background for the empirical part of the study is provided. Methodologically, the investigation of intonational functions in Northern Vietnamese is based on the combination of both interactional/conversation-analytical and experimental approaches. It deals not only with naturally occurring data but also with scripted/manipulated speech materials. Furthermore, some issues of intonation in tone languages in general as well as the modelling of the interaction between lexical tone and intonation are addressed.

The last part of this chapter presents approaches to Vietnamese tonal phonology (the lexical tones). A simplified model is proposed serving as the basis for the corpus analysis and modelling. Finally, an overview of the work to date on intonation in Vietnamese is given.

2.1 Prosody research

2.1.1 Interactional approach

Over the last two decades, prosody has been established as an independent object of research in studies on talk-in-interaction. Research on prosody in interaction is based on the methodologies of Conversation Analysis (hereafter CA) (Sacks et al. 1974, Hutchby and Wooffitt 2008). CA is a sociological framework investigating talk as an action in which interlocutors accomplish social activities (requests, invitations, greetings, etc.) and through which they manage their social lives.

Methodology in Conversation Analysis

In general, materials analysed in CA are based on audio or audiovisual recordings of naturally occurring interactions. This kind of data is the preferred object of research because of its naturalness. It is possible from such data to discern how interlocutors themselves understand what is going on in talk. The starting point for the current study was to record naturally occurring interactions and to look at how speakers produce their utterances in real interactions. This approach made it possible to analyse backchannel and repair strategies, both of which are difficult to elicit in read speech. Both backchannel and repair initiation were found to be expressed by intonational patterns which appear to be relevant in discourse management and conversational activities among interlocutors.

Analysing natural interactions in CA is reflected in the step of transcribing data which involves a special consideration of details including gaps, pauses, voice quality such as breathiness, and even laughter. Conversational data from different languages, but especially English, has been largely transcribed using the conventions developed by Jefferson (2005). For prosodic investigations within CA, a transcription system called GAT (GesprächsAnalytisches Transkriptionssystem) has been developed (Selting et al. 1998). This system is compatible with Jefferson’s, but includes additional conventions which are specifically for prosodic analyses. It has recently been revised as GAT 2 which
differentiates three levels of transcripts: the minimal transcript as a working transcript, the basic transcript which is appropriate for sequential analysis, and the fine transcript for prosodic/phonetic analyses (Selting et al. 2009).

Although GAT 2 has been developed exclusively for German (see Couper-Kuhlen and Barth-Weingarten 2011 for a version for English) and the system might pose difficulties for non-CA scholars, the present study makes use of its conventions as far as the research purposes require. The fine transcripts for prosodic properties will only be transcribed at the utterance level for Northern Vietnamese (see Chapter 3 for a transcript sample). The prosodic properties at the syllable level, i.e. the lexical tones, are described separately (see also Chapter 3).

After transcribing the data, CA researchers proceed with identifying the potential phenomenon and consequently build collections of the phenomenon identified. One of the most important analytic principles of CA is the demonstration of results in the data itself by showing the orientation of interlocutors towards the structure of a device applied in a specific context. This can be done by analysing the ‘sequential organisation’ of the talk using the “next turn proof procedure” – how an utterance is responded to in the next turn, and so-called “deviant cases” (see Ten Have 1999). The latter are cases that appear to deviate from the assumed normative pattern resulting from the first observations and descriptions. Analysis of these cases demonstrates the relevance of the identified or normative pattern, in the way that different reactions of interlocutors/ recipients can be observed when this pattern is not used.

As will be shown in Chapter 5, the analysis of deviant cases can exhibit the underlying structure of such practices as requests for information during telephone openings, and consequently the orientation of interlocutors to this structure. However, this method is not always applicable due to their random occurrence in natural settings (see Bergmann 2001) and researchers should find other ways to validate the results. One way is to conduct interviews with interlocutors which can provide “insight into members’ common sense reasoning about prosody” (Couper-Kuhlen and Selting 1996: 25). Another is to use controlled recordings that involve features deviating from normative patterns to investigate how speakers respond to them (Couper-Kuhlen and Selting 1996). The present study makes use of the latter, controlled recordings, to investigate the relevance of pitch variation occurring in the same word in a given context. These materials are incorporated in the investigation of repair initiations (Chapter 6).

In principle, CA avoids treating phenomena as “statistical variables” and reinforces rather case-by-case analysis (Hutchby and Wooffitt 2008). This indicates that there is always a risk that the phenomena given are taken to be only illustrative. However, more recent studies on talk in interaction, in particular work at the intersection of CA and phonetics, have begun to combine both qualitative and quantitative data (Special Issue on the Phonetics of talk in interaction in Language and Speech 2012, guest editor Richard Ogden, see Ogden 2012 for an overview). In conversations on the telephone, in which visual cues are not available, backchannels and repairs tend to occur frequently to signal attention and to deal with problems understanding utterances, respectively. The analysis in the current study, thus, takes quantification into account attempting to make the data more representative.

**Contextualisation Theory**

The concepts of the interactional approach are inspired by Contextualisation Theory which goes back to Gumperz’s ethnographic studies of linguistic variability (see Gumperz 1982, 1992). Contextualisation Theory is grounded in sociolinguistics and investigates how linguistic and social knowledge affects practices in discourse and how this knowledge
enters into interpretation. Interlocutors’ language use can reflect relevant social information such as their social relationships/roles and frame the interpretation of utterances in terms of what is going on in interaction.

From the linguistic point of view, the concept of contextualisation refers to the process of inferencing and guiding the interpretation of words/utterances conveyed by contextualisation cues. In contrast to other theories of context, e.g. metapragmatics (Bateson et al. 1956) and linguistic pragmatics such as relevance theory (Sperber and Wilson 2004), the theory of contextualisation deals with naturally occurring data and focuses on particular classes of cues (Auer 1992). Studies on prosody as one type of these cues have shown that a variety of prosodic means have a function in conversation.

For English, Local (1992) identified the fall-rise pitch pattern produced with a drop in pitch range, acceleration and diminished loudness as a cue for parenthetical talk. According to Couper-Kuhlen (1992), different uses of pitch register in repetitions of a prior turn may distinguish two activities in conversation: while a repetition spoken in a relatively high pitch register can convey quoting, a repetition spoken in the same pitch range can convey mimicking. Turn-competitive incomings are found to be produced with relatively high pitch and loud volume (French and Local 1983). Activities such as complaints about a third party can be categorised as different activity types, depending on the use of prosody: sustained F0 above a speaker’s average F0, loudness, high F0 peaks and wide pitch span contextualise complaints seeking the affiliation of a conversational partner, whereas the reverse phonetic realisations are cues for complaints with a sequence closing function (Ogden 2010). Prosodic variation in responses to yes/no interrogatives was investigated by Raymond (2010) who found that speakers produce a slight rising intonation contour on these responses to project more talk or use closed forms (yep, nope, yeah with glottal stop) to convey turn completion.

For German, Auer (1996) investigated the role of prosody in conveying information structure in syntactic expansions (Ausklammerungen). While expansions contextualised as carrying rhematic information are spoken with at least one accent unit, the ones conveying (sub)thematic information are spoken with no further pitch accents. Günthner (1996) found that speakers of German use a complex combination of prosodic features, alongside lexicosemantic and sequential cues, to differentiate reproaches in why-questions from “real” why-questions. These patterns include a falling terminal pitch, a global increase of loudness, extreme rising-falling (or falling-rising) pitch movements, falling glides on lengthened syllables, narrow or verum focus, staccato accentuation and increase of tempo. Speech rate has also been investigated in relation to contextualisation in a study by Uhmann (1992) who proposed a two-dimensional model of speech rate. This model refers to the number of syllables per unit of time, on the one hand, and the number of accented syllables per unit of time, on the other. Her results show that a higher density of accented syllables within a stretch of syllables per unit of time indicates more relevance than a lower density thereof.

Studies on prosody in conversation have also been conducted for other languages such as Dutch (Plug 2005, Mazeland and Plug 2010), Finnish (Ogden and Routarinne 2005), and Japanese (Tanaka 2010). In general, the studies mentioned above are based on methodologies in Conversation Analysis and the analytical concepts of Contextualisation Theory (see Couper-Kuhlen and Selting 1996, Local and Walker 2005 for more details).

In a nutshell, the main methodological principles for prosody research in interaction are: i) the use of naturally occurring data, ii) data-driven analyses of functions/categories, iii) combination of sequential and phonetic-prosodic analyses, and iv) validation of analysis by identifying participants’ orientation to the functions (Selting 2010).
2.1.2 Experimental approach

*Laboratory Phonology and the form-meaning association*

Experimental studies within Laboratory Phonology aim to discover and explain aspects of speech melody in a controlled setting. To establish the form-meaning association, intonational parameters which may contribute to a certain meaning are modified, and listeners’ intuitions or judgments are consequently tested (Gussenhoven 2004). A number of experimental tasks have so far been designed: categorical perception, semantic differential, imitation and semantic scaling tasks.

The results serve to address the question of whether a given phonetic difference reflects a phonological contrast (discrete form-meaning association) or whether it should be analysed as a matter of different phonetic realisations of the same phonological category (gradient form-meaning association). The former association has often been claimed to involve linguistic meanings (Pierrehumbert and Hirschberg 1990, Ladd 1996, Gussenhoven 2004), whereas the latter association has often been claimed to involve degrees of paralinguistic meanings. However, the division between these two types of form-meaning association is not always clear, as gradient form-meaning associations can be grammaticalised and become discrete: for example, in a number of languages a high pitch phrase-finally which signals openness has been grammaticalised such that a high boundary tone (H%) is used to mark questions as opposed to a low boundary tone (L%) to mark statements (see Rietveld and Chen 2006).

Since the experiment in Chapter 8 of the current study tests the perception of naturalness, politeness and dominance in backchannels, this section provides the methodological background for this purpose and focuses on semantic scales, introducing the Equal Appearing Interval Scale (EAI). The reasons for choosing EAI are threefold. First, semantic scales can directly exhibit the relation between intonational meanings and phonetic manifestations. Second, the EAI scale allows for more rating space than other tasks such as forced-choice tasks, in which participants choose one from a set of typically two values. Third, the EAI scale has been proven to be sensitive and therefore particularly suitable for investigating the perception of pitch-related properties (Rietveld and Chen 2006).

The basic aim of using semantic scales is to investigate the perception of intonational meanings in a gradient way, specifically “how a certain intonational parameter is interpreted in different pragmatic or discoursal contexts by listeners with the same or different language backgrounds” (Rietveld and Chen 2006: 286). Different manipulated speech stimuli are judged by listeners by means of numbers or categories labelled on one or more semantic scales (see Figure 2-1). These numbers and categories correspond to parts along a continuum of a particular meaning: for example 1-7 whereby 1 = “not friendly” and 7 = “very friendly” (Equal Appearing Interval Scale) or 5-1 for indicating (dis)agreement with a certain statement about a semantic meaning (Likert Scale).
While the Equal Appearing Interval Scale allows a direct judgement, for example, of the degree of friendliness, the Likert Scale can be seen as an indirect way to judge the meanings, as it allows subjects to show to what extent they agree with a statement, such as “The speaker sounds friendly”.

The relationship between a meaning and a certain intonational property is reflected in the correspondence between these numbers and categories. In terms of statistical procedures, Rietveld and Chen (2006) recommend parametric methods for interval variables, such as mean judgments/mean normalised judgments (z-scores), and analysis of variance (ANOVA), even though values on a semantic scale are ordinal. The reason is that we do not know whether, for example, the difference between “not friendly” rated as 1 and 3 is the same as the difference between 5 and 7 on the same scale (Rietveld and Chen 2006).

The results presented in Chapter 8 support the claims by Chen (2005) that speakers of different languages or even of different speech communities can perceive a given phonetic contrast differently in terms of paralinguistic meanings. Results from the perception experiment with Northern Vietnamese listeners also provide evidence that the relation between intonational form and paralinguistic meanings (speakers’ attitudes or affect) is to a certain extent universal but there are nonetheless some language-specific aspects (Gussenhoven 2004, Chen 2005).

2.1.3 Intonation in tone languages

A number of intonational strategies have been reported in tone languages. This subsection addresses three of those before introducing the Autosegmental-Metrical framework used for describing intonation in these languages.

Sentence (final) particles

It is widely recognised that in a large number of tone languages the functions of (sentence-final) particles correspond to the functions of intonation in non-tonal languages. The extensive use of this word class appears to avoid the conflict between lexical tone and intonation. Particles can convey communicative functions at the grammatical level to form, for example, an interrogative sentence, or at the discourse level to convey aspects of speaker-hearer interaction. The richness of particles in different tone languages is reported in Yip (2002). Here we focus on East Asian languages. Mandarin has about 28 particles of which 26 have a tone (two are toneless), Cantonese has more than 35, and Vietnamese has more than 20. Particles in the two latter languages have so far been claimed to be all tonal.
Nevertheless, detailed investigations of their (inh erently) tonal status are still needed, especially for Vietnamese (see Yip 2002: 275).

Recent work on Thai, a language with both tonal and toneless particles, has shown that particles do not necessarily have to be an alternative to intonational tones in the language but can coexist with intonation. Pittayaporn (2007) found that in combination with intonation, toneless final particles can convey a range of discourse meanings. For example, the toneless final particle *na* can be used to convey ‘establishment’ of the content of an utterance. It can further convey who establishes this content, the hearer or the speaker, depending on whether it is produced with a high or low intonational tone.

**Phrase-final intonation**

Intonational tones, mainly phrase-final tones (Yip 2002), can convey grammatical and discourse meanings. These phrase-final tones have been widely analysed in the framework of Autosegmental-Metrical Phonology as edge or boundary tones, since they commonly occur at edges or phrase boundaries (see section 2.1.4 below). They are targets and can be realised in sequence with the lexical tone on the last syllable, or alternatively they can merge with it or neutralise it. Interactions resulting in sequencing and neutralisation of lexical tone will be discussed in Chapter 7.

As established in a number of East Asian tone languages, a high boundary tone can change a declarative into an interrogative when added to the lexical tone of the last syllable (Thai, Vietnamese – see also 2.2.4 for more details). In Thai, this boundary tone can interact with a lexical low tone on the last syllable, which then results in a mid-level tone (Luksaneeyanawin 1998: 382-383). A final high boundary tone can also convey other sentence types such as imperatives in Taiwanese (Peng 1997), or, together with a raised overall pitch register, echo questions in Mandarin (see also next section and Peng et al. 2005).

**Pitch register/range adjustment**

Pitch register/range adjustment, falling or rising trends in the top line or baseline of the pitch contour, or shifts in pitch range such as expansion or compression, can also convey meanings (Yip 2002). In African and Asian tone languages as well as pitch accent and intonation languages, declination, which involves a gradual decline in the pitch of high tones over the course of an utterance, is common in declaratives, whereas suspension of declination can be applied to convey interrogatives. The gradual lowering of high tones over the course of a phrase should be distinguished from the phonological process of downstep, which involves a floating low tone which triggers a step down in pitch on the following high tone. Downstep has so far been established in analyses of a number of unrelated African languages but not for East Asian tone languages. In these languages, it has been shown that an expansion of pitch range can mark emphasis or focus, whereas post-focal elements are found to be compressed (Xu 1999). Since the adjustment of pitch register/range commonly takes place over the entire utterance, the lexical tones remain, but are produced in a lower or higher pitch space/range: high tones becoming higher, low tones becoming lower.

### 2.1.4 Modelling tone-intonation interaction

**Autosegmental-Metrical Phonology**

The Autosegmental-Metrical model (hereafter AM) has been successfully adopted to describe intonation in a variety of languages, for example English (Gussenhoven 2004,
Ladd 2008), German (Grice and Baumann 2002) and Japanese (Pierrehumbert and Beckman 1988). Its main tenets are addressed here as background for the analysis of the interaction between intonation and lexical tone in Northern Vietnamese. In this model, the intonation contour of utterances is analysed as a string of tones which come from different sources depending on the domains in which they occur (Grice 2006, Pierrehumbert and Beckman 1988, Gussenhoven 2004, Ladd 2008).

Taking English as an example, the string of tones in the utterance like *Too many cooks spoil the broth* (see Figure 2-2 below) can be derived from the word level (also called ‘phonological word’) or from higher levels such as the intermediate phrase (ip) or intonation phrase (IP). The AM theory separates metrical and tonal features, separately investigating metrical structure as influencing the location of H* or L* pitch accents – tonal features associated with stressed syllables. Other tones are phrase tones or phrase accents (H-, L-) occurring at or around the edges of intermediate phrases and sometimes on metrically strong syllables, and boundary tones (H%, L%) occurring at the boundaries of intonation phrases.

![Figure 2-2: One way of analysing the contour on the sentence *Too many cooks spoil the broth* within Autosegmental-Metrical Phonology. From Grice (2006: 781).](image)

The model has been developed to capture the intonation of languages with different linguistic systems. For example, in Japanese pitch accents are associated with lexically accented syllables. The language has edge tones both at the utterance (IP) and accentual phrase levels (α). The tone bearing unit in the language is the mora. Analyses within the AM framework reflect so-called ‘secondary association’ of an accentual phrase final edge tone, a manifestation of tones characteristic for this language, as can be seen in Figure 2-3.

![Figure 2-3: Secondary association of the accentual phrase edge tone (L, emphasis: KPH) in the sentence *Ane-no akai se’etaa -wa do’ko desus -ka?* ‘Where is big sister’s red sweater?’ From Pierrehumbert and Beckman (1988: 128).](image)
The L edge tone in Japanese can be associated with both the higher prosodic domain(s), e.g. the accentual phrase (α), and with elements lower in the hierarchy, e.g. the mora. The latter association is secondary and occurs when the initial sonorant mora (on a in *akai*) in the following accentual phrase is unaccented. Otherwise, the L edge tone has only the primary association with the accentual phrase. For a detailed discussion on the associations of accentual phrase edge tones see Pierrehumbert and Beckman (1988).

The secondary association of edge tones has been proposed for other languages and varieties of those languages, such as Hungarian, Romanian, Greek, German and Swedish (Grice et al. 2000, Grice 2006). Edge tones may be associated not only with constituents such as mora, as in Japanese, but also with those higher in the hierarchy, namely with a word or foot. Although the place of the phrase accent/tone in intonational phonology is still controversial, the description of tones and their different sources within AM model appears to successfully account for a number of different languages and varieties.

**Major types of tone-intonation interaction in tone languages**

The interaction of intonation and lexical tone in East Asian tone languages has been modelled within two different approaches. The first can be referred to as a target-interpolation approach within the Autosegmental-Metrical framework. In this framework, lexical tonal targets alternate with phrase-level/post-lexical boundary tones which are local in nature and thus, tend to affect the pitch contour at the edges of prosodic constituents (for Thai see Pittayaporn 2007, for Naxi see Michaud 2006, for Cantonese see Wong et al. 2005, for Mandarin Chinese see e.g. Shih 1988, Peng et al. 2005, Fox et al. 2008). There is interpolation from one target to the next.

The second approach can be referred to as superpositional (for Thai see Luksaneeyanawin 1998, for Chinese see Shih 1988, Shen 1990, Fox et al. 2008) In this approach the melody of a sentence consists of local contours (lexical tones) overlaid or superimposed on global contours, falling or rising trends of the pitch contour, or shifts in pitch range (e.g. expansion or compression). Section 2.2.4 below will show that Vietnamese intonation has been investigated predominantly in terms of superposition, e.g. global F0. Less attention has been paid to target-interpolation models.

For the purpose of investigating universal and language-specific uses of intonation across languages, a common framework is needed. The current study aims to describe Northern Vietnamese intonation within a linguistic theory in line with the AM framework for several reasons. First, we emphasise the form and phonological structuring of speech melody in order to investigate sources for the input to the tonal phonology of Vietnamese.

Second, there has been evidence from different tone languages (see e.g. Yip 2002) suggesting that lexical tones and phrase-final intonation can be described by the same mechanisms, namely by distinctive pitch targets (see also Ladd 2008). For example, the high boundary tone at the right edge of the word conveying focus, as found in Naxi, can be associated with the last low-toned syllable which then results in a contour tone LH (Michaud 2006). The adding of this boundary tone to the last lexical tone of the utterance leads to multiple association, one of the characteristics of tone (Yip 2002: 69). Multiple tonal features (L and H) can surface on one host segment (in the last syllable) as illustrated in a) in Figure 2-4. Furthermore, the high boundary tone might be analysed as spreading left within the intonation phrase and delinking the lexical low tone of the preceding word as found in interrogatives in Thai. This corresponds to the characteristic of mobility of tone (Yip 2002). As illustrated in b) in Figure 2-4, the high boundary tone can be analysed to neutralise the preceding low tone (indicated by the heavy crossed line) which then results in a level tone (see Luksaneeyanawin 1998).
Third, there is evidence from articulation suggesting that intonation and lexical tone can be described by means of the same features. Speakers of tonal as well as non-tonal languages lower or raise their voices by using the same articulatory gestures and making adjustments to the vocal folds or by manipulating subglottal pressure of the air in the lungs. Intonation is also perceived and interpreted by native speakers of tone languages as lexical tone (Duanmu 2004). These similarities in form, along with similarities in structures, lead to an analysis of intonation and lexical tone using of the same phonological features. The current study uses the binary features of high and low, and analyses intonation in Northern Vietnamese in line with the AM framework in terms of sequences of tones.

Fourth, recent work on Northern Vietnamese intonation (Brunelle et al. 2012) found not only global effects but also local effects of F0 as conveyors of communicative functions. A low final boundary tone is found to convey declaratives and a high final boundary tone is found to convey unmarked interrogatives. Even though inter-speaker variation emerges, and the data analysed is based on read speech, results from the study show that it is possible to analyse intonation in Northern Vietnamese within the AM framework by means of local pitch targets.

2.2 Vietnamese

2.2.1 Lexical tones

The tone system in Northern Vietnamese consists of six lexical tones which occur on open syllables or syllables ending with a sonorant coda. These tones are characterised by pitch as well as voice quality (Alves 1997, Hoàng 1989, Nguyễn and Edmondson 1998, Phạm 2003a, 2003b, Michaud 2004, Brunelle 2009a, Kirby 2010). Figure 2-5 below illustrates the F0 contours of these tones produced by a male speaker. (The alphanumeric codes A1 to C2 are provided according to Michaud (2004) indicating the historical origin of tones; traditional tone names are given in blue).

Tone ngang A1 is high-level (or mid-level), tone sắc B1 is a rising tone which begins lower than ngang and then rises. Tone nặng C2 is rising-glottalised; it begins higher than sắc B1 and then rises abruptly after a period of glottalisation, commonly realised as a glottal stop in the middle of the tone. Tone huyền A2 is a low-falling tone (also often produced as low-level), and can start higher than sắc B1. It can sometimes be produced with breathy voice (Phạm 2003a). Tone hỏi C1 is a falling (-rising) tone which can be produced with breathiness and/or glottalisation (Nguyễn and Edmondson 1998, Michaud 2004). In Hanoi Vietnamese the rise of this tone is commonly absent.
In the telephone conversations, breathy voice is rare – and if present, it is difficult to distinguish from modal voice, whereas glottalisation is clearly discernible. The corpus analysis therefore concentrates mainly on glottalisation as another property of tone in addition to pitch. Table 2-1 summarises the properties of the six lexical tones in Northern Vietnamese and provides an abstract representation of these tones (see also Appendix A for an overview).

Table 2-1: Tonal properties of six Northern Vietnamese lexical tones and their abstract analysis based on auditory impression (h = high, l = low, ʔ = lottalisation, breathiness not considered). Both traditional names and alphanumerical codes are given.

<table>
<thead>
<tr>
<th>Tone name</th>
<th>ngang (A1)</th>
<th>huyền (A2)</th>
<th>sắc (B1)</th>
<th>nặng (B2)</th>
<th>hỏi (C1)</th>
<th>ngã (C2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pitch1</td>
<td>h</td>
<td>l</td>
<td>lh</td>
<td>(h)ʔ</td>
<td>hl(h)</td>
<td>lʔh</td>
</tr>
<tr>
<td>Citation form</td>
<td>high-level (or mid-level)</td>
<td>low-level (or falling)</td>
<td>rising</td>
<td>falling-glottalised</td>
<td>falling-(rising)</td>
<td>rising-glottalised</td>
</tr>
<tr>
<td>Stylised contour</td>
<td>··</td>
<td>···</td>
<td>···</td>
<td>··?</td>
<td>···?</td>
<td>··?</td>
</tr>
</tbody>
</table>

The high-level tone ngang A1 is analysed as high (h), the low-level/falling tone huyền A2 as low (l). The tone sắc B1 bears a low as well as high pitch, which results in a rise (lh). The falling-glottalised tone nặng B2 has a glottal stop after the falling contour consisting of a high and a low pitch target. However, the high pitch (h) is optional due to the glottalisation which restricts the realisation of pitch; therefore, the tone is often produced as low-glottalised ((h) lʔ) (Brunelle 2009a). The falling (-rising) tone hỏi C1 is described as involving a high-low-high pitch sequence, from which the final high can be

---

1 The labelling of pitch with lower case h and l follows Yip (1995), a summary of phonological issues dealing with tone in East Asian languages. In this study, upper case H and L denote the pitch register [±upper].
absent (h 1 (h)). The rising-glottalised tone ngã C2 is described as having a glottal stop in the middle of the rise (l̩ ʔ h).

There are also models which advocate an eight-tone paradigm in Northern Vietnamese. Two additional tones are claimed to occur on checked syllables, that is, those ending with oral stops. One of these models is proposed by Phạm (2003b) based on a new analysis of voice quality as the primary tonal feature. This model is reviewed briefly below.

### 2.2.2 Phạm’s model (2003b)

In her acoustic studies (Phạm 2001, 2003b), Phạm argues that it is not tonal height or shape that are distinctive, as has so far been proposed, but laryngeal features (breathiness and creakiness) which she refers to as “register”. Tonal shape or contour is claimed to be predictable from laryngeal features. By means of a wide range of evidence, the tone classification is proposed as involving two features: a) two registers ‘unmarked’ (characterised by modal voice) and ‘marked’ (characterised by breathiness or creakiness) and b) contours which group the tones in pairs (see Table 2-2 below).

An overview of the phonological representation of the tones is provided in Figure 2-6. In this representation pitch heights (low = l and high = h) are claimed to be derived from the phonetic implementation rules which are predictable from register. Tones with a register node (R) are lowered and have a final low pitch, whereas those without this node have a final high pitch (Phạm 2003b: 79). The relative markedness of the tones is also illustrated. Ngang (a modal high-level tone) is less marked than huyền which – in addition to a contour node (C) – also has a register node (R) due to its breathiness. Ngang (h) is less marked than sắc, since the contour of the latter is more complex (lh). Sắc (sac1) is in turn less marked than its counterpart nặng (nang1) because of the register node involving constricted glottis. In comparison to huyền, nặng possesses both a more complex contour node and a more complex register node ([spread] is analysed as less marked than [constricted]), and is therefore more marked. Finally, ngã is interpreted as the most complex tone, since it has the most complex contour node as well as the most complex register node.

Phạm’s tonal representation can account for a wide range of phonological processes in Vietnamese including reduplication, clitics, dialectal neutralisations and borrowings from non-tonal languages such as French or English. Her phonetic implementation rules, however, are not entirely transparent, i.e. they do not entirely match the phonetic realisations of tones as reported in the acoustic studies mentioned above, in particular the realisation of the final portion in tone ngã C2, emphasised in Figure 2-6. The final part of this tone is commonly produced with an abrupt rise (see also Figure 2-5), but is interpreted in Phạm’s model as having a falling or low pitch.
Table 2-2: Classification of tones based on their registers, i.e. unmarked vs. marked, and tonal shapes. The markedness increases from left to right and from top to bottom (adapted from Phạm 2003b: 40)

<table>
<thead>
<tr>
<th></th>
<th>EVEN</th>
<th>NON-EVEN</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Unmarked register</strong> (modal voice)</td>
<td>rise/fall</td>
<td>rise/fall (on checked syllables)</td>
</tr>
<tr>
<td>ngang A1</td>
<td>sắc B1 (rising)</td>
<td>sắc D1 (rising)</td>
</tr>
<tr>
<td>Marked register (breathiness/creakiness)</td>
<td>huyền A2</td>
<td>nặng B2 (falling)</td>
</tr>
</tbody>
</table>

Figure 2-6: Phonological representation of each tone in Vietnamese. Pitch heights \((l = \text{low}, \ h = \text{high})\) added according to the phonetic implementation rules proposed by Phạm (2003b) \((l, \text{emphasis}: \text{KPH})\). Sac1 and nang1 correspond to sắc B1 and nặng B2.

### 2.2.3 Proposal of a neutral tone model

For tonal representations in this study we simplify Phạm’s model (2003b) by merging the contour and register nodes. We do this for two reasons: a) to avoid introducing/maintaining a level of structure which has been contested in the recent literature (see Brunelle 2009a for detailed discussion), and b) to enable a sequencing of the tones and glottal stop/creakiness to be represented, in particular in the sequence \((h)\ l\ ʔ\) and \(l\ ʔ\ h\) (nặng and ngã, respectively). Figure 2-7 provides the tone structure proposed in this model, along with the illustration of a sample tone; the tone bearing unit (TBU) in Vietnamese is recognised to be the syllable.
The current study deals with tonal aspects at the lexical level, i.e. word tones, and the pragmatic/discourse level, i.e. the intonation used in discourse. We shall not carry out an analysis of tonal aspects at other levels, such as the morpho-phonological level (reduplications). Furthermore, we do not deal separately with the two additional tones on checked syllables, since they are generally regarded by native speakers to be variants of tones sắc B1 and nặng B2 (see Kirby 2011).

The neutral description of tone will be addressed again in Chapter 7, which analyses the interaction between lexical tone and intonation found in interactional contexts. In this analysis, the contours used in conversational contexts will be compared with those in citation form to find out how intonation influences the lexical tones of the words investigated.

### 2.2.4 Intonation

In addition to a complex tone system, Vietnamese also has a large number of particles for conveying grammatical functions. These include initial, final, phrasal, vocative, predicative and post-verbal particles (Thompson 1965). It has often been assumed that this complex lexical tone system coupled with an extensive use of particles would leave little room for intonation to express communicative functions. However, a number of studies based on auditory observations (Thompson 1965, Trần 1967, Nguyễn 1990, Đỗ 2009) as well as on experimental data (Đỗ et al. 1998, Nguyễn and Boulakia 1999, Vũ et al. 2006, Jannedy 2007) fail to confirm this assumption. In contrast, they have shown that pitch is in fact modulated for expressing discourse-related functions, sentence types, affect and information structure. These functions and their expression are summarised below.

Different functions in discourse are reported to be associated with different intonational patterns. For Northern Vietnamese, Thompson (1965) proposes a number of intonation contours which can occur on the final syllable of a phonological phrase: a) “Decreasing” intonation can be found at the end of a pause group or utterance. It conveys that the speaker has not finished and that he or she has been interrupted or has simply left
something unsaid. b) “Increasing” intonation accompanying final syllables conveys “exclamative statements and commands, and questions to which the answer is of relatively small interest to the speaker”. c) “Fading” intonation “includes commands, statements of fact which the speaker assumes the hearer will accept as true, and questions to which the speaker assumes the answer”. d) Finally, the “sustaining” intonation is used when the speaker is “unable or for some other reason fails to predict the result of his speech”. The latter includes “doubtful statements, emotionally coloured and indecisive imperatives and exhortations, and informational questions” (Thompson 1965: 107-109).

A survey of Southern Vietnamese (a further dialect of the language with a slightly simpler lexical tone system) describes three basic global intonation contours – falling, sustaining and increasing – within different sentence types (Trần 1967). The first intonation contour, falling, occurs in declarative statements, indicating that the speaker has completed his sentence. The second contour, sustaining, occurs in sentences which are non-final. Increasing intonation occurs mostly in interrogative and imperative exhortative sentences (see also Hoàng 1985). There is a general consensus that these global intonation contours can “modify the pitch characteristics of the tones, but not affect the tonemic contrast between them” (Trần 1967: 23).

Further phonetic characteristics of questions as compared to statements were found in Northern Vietnamese by Đỗ et al. (1998) and Nguyễn and Boulakia (1999) such as fast tempo, rising overall F0 and a final rise in F0. A recent study suggests that declaratives and interrogatives are distinguished from each other rather by a rise that is largely located at the end of the sentence-final question particle in interrogatives than an overall pitch register difference in the sense of local pitch range (Vũ et al. 2006).

Intonation has also been shown to express affect. Đỗ et al. (1998) found that “logical evidence” and anger are both marked with an extra-high register, a higher intensity and a rising F0 contour. Anger is found to have a faster tempo than “logical evidence”. More generally, Hoàng (1985) finds that F0 is higher in declaratives conveying a marked emotion than in other declaratives.

In Brunelle et al. (2012) we provide evidence for a great deal of inter-speaker variation in the realisation of intonation in different utterance types (question vs. statement) and affect (neutral vs. annoyed). We found that both global F0 and intensity as well as local pitch movements are used to differentiate between these communicative functions.

Recent experimental investigations have revealed that F0, intensity and duration are all markers of focus in Vietnamese. Strong F0 variations in subject, verb and object focus contexts were found in studies by Jannedy (2007). However, while subject, verb and object foci were easy to identify in perception tests, broad focus was not easily distinguished from narrow focus. Emphasis has been shown to be expressed by F0 rises and a higher F0 register (Michaud and Vũ 2004).

Generally speaking, most of the studies mentioned above refer to F0 in a gradual sense, such as in terms of global pitch range or register (Trần 1967; Hoàng 1985, Đỗ et al. 1998, Nguyễn and Boulakia 1999, Michaud and Vũ 2004, Vũ et al. 2006, Jannedy 2007, Brunelle et al. 2012). Only two refer explicitly to pitch in a discrete sense, in terms of intonational tones which are represented as tones at the end of intonation phrases (Vũ et al. 2006, Brunelle et al. 2012). The current investigation of short utterances based on naturally occurring data will show that Northern Vietnamese makes use of intonational tones and that these tones can influence the realisation of lexical tones. In fact, intonation can modify or even override the pitch specifications for individual lexical tones. The thesis makes a first attempt to describe the tone-intonation interaction in Autosegmental-Metrical Phonology.
3 Telephone corpus description

3.1 Materials and methods

The speech materials are based on 36 telephone calls (approximately 95 minutes of total speech time) made by 43 Northern Vietnamese participants (20 males and 23 females) in Hanoi. Their age varied between 11 and 63 years. Telephone calls were made to friends, family members or acquaintances in informal, natural settings. Speakers recorded their conversations after receiving instructions on how to use the technical equipment. The equipment consisted of a Digital Audio Tape recorder or an EDIROL digital recorder, and a microphone fixed to the landline telephone in speakerphone mode. Some of the conversations are multiparty, as in some cases more than two people were involved. Despite the fact that the voice of the speaker at the other end of the telephone line was not recorded directly, the overall acoustic quality of the acquired speech signal is sufficiently good for a reliable F0 contour analysis. Data recorded in telephone settings were chosen for the analysis, since in this type of communication only the audio channel is used. We can therefore be sure that there is no visual feedback which might obscure the analysis of the effect of the spoken signal. Table 3-1 below provides an overview of the corpus including details about identification codes for speakers, topics, duration of calls, remarks on conversational structure (particularly regarding the closing phases), as well as the relationship between the participants.

Laboratory data are also added to the corpus. Controlled dialogues were analysed to investigate the function of prosody in different activities in talk such as repair initiation vs. acknowledgment (see Chapter 6, section 6.6). Sentence lists were recorded to elicit citation forms of the investigated tokens. This allows an acoustic analysis of the interaction between the pitch contours of lexical tones and those of intonation (see Chapter 7).

Permission to record telephone conversations as well as laboratory data was obtained from participants orally. In order to safeguard privacy, the names of participants are not used here but are replaced throughout the analysis with labels of the type F or M (F = female, M = male) and a unique number, for example F2 or M3 (as shown in Table 3-1). Furthermore, proper names mentioned during conversations have been replaced by abbreviations.

The corpus analysis provides a qualitative as well as quantitative investigation of the phenomena in question. Adapting the methodology of Conversation Analysis (see Chapter 2, section 2.1.1), the qualitative investigation aims to identify the prosody of the phenomena, especially intonation, by looking at the sequential organisation of dialogues. In particular, recipient responses are taken into account. Results reveal that participants orient themselves to the patterns identified. As support for generalising statements as to how intonation contributes to interactional structures (backchannel versus request for information and turn-yielding, as will be shown below), analytic attention will be paid to deviant cases, i.e. cases that deviate from the assumed standard patterns. They reveal that in instances when the identified pitch contour is not used, conversational partners locate the utterances in another interactional context and therefore respond differently.
Table 3-1: Telephone corpus.

<table>
<thead>
<tr>
<th>Nr</th>
<th>Identification code</th>
<th>Topic of the conversation</th>
<th>Duration in minutes and (or) seconds</th>
<th>Conversational structure</th>
<th>Relationship of participants</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>M1-M2</td>
<td>Computer game 1</td>
<td>2m</td>
<td></td>
<td>Friends</td>
</tr>
<tr>
<td>2</td>
<td>M1-M2</td>
<td>Computer game 2</td>
<td>40s</td>
<td></td>
<td>Friends</td>
</tr>
<tr>
<td>3</td>
<td>M3-M1-F23-F1</td>
<td>Electricity</td>
<td>24s</td>
<td></td>
<td>Family members and acquaintance</td>
</tr>
<tr>
<td>4</td>
<td>F2-F3-F4</td>
<td>Friend talk</td>
<td>6m 19s</td>
<td></td>
<td>Friends</td>
</tr>
<tr>
<td>5</td>
<td>F2-F5-F6</td>
<td>Sister talk</td>
<td>7m 19s</td>
<td>No closing</td>
<td>Relatives</td>
</tr>
<tr>
<td>6</td>
<td>M4-M5</td>
<td>Entrance exam</td>
<td>4m 4s</td>
<td></td>
<td>Friends</td>
</tr>
<tr>
<td>7</td>
<td>M4-M6-M7</td>
<td>Study</td>
<td>2m</td>
<td></td>
<td>Friends</td>
</tr>
<tr>
<td>8</td>
<td>F2-F7</td>
<td>Family</td>
<td>5m 43s</td>
<td>No closing</td>
<td>Friends</td>
</tr>
<tr>
<td>9</td>
<td>M3-M8</td>
<td>Children</td>
<td>8m 52s</td>
<td></td>
<td>Friends</td>
</tr>
<tr>
<td>10</td>
<td>M3-M9</td>
<td>Exam notes</td>
<td>1m 21s</td>
<td>No closing</td>
<td>Friends</td>
</tr>
<tr>
<td>11</td>
<td>M3-F8</td>
<td>Bill 1</td>
<td>43s</td>
<td></td>
<td>Acquaintances</td>
</tr>
<tr>
<td>12</td>
<td>M3-F9</td>
<td>Bill 2</td>
<td>34s</td>
<td>No closing</td>
<td>Relatives</td>
</tr>
<tr>
<td>13</td>
<td>F10-F11</td>
<td>Dinner</td>
<td>1m 23s</td>
<td></td>
<td>Friends</td>
</tr>
<tr>
<td>14</td>
<td>F11-F12</td>
<td>Metro</td>
<td>3m 42s</td>
<td></td>
<td>Friends</td>
</tr>
<tr>
<td>15</td>
<td>F10-M10</td>
<td>Father-daughter</td>
<td>1m</td>
<td></td>
<td>Friends</td>
</tr>
<tr>
<td>16</td>
<td>M3-F13</td>
<td>Semester schedule</td>
<td>1m 17s</td>
<td></td>
<td>Acquaintances</td>
</tr>
<tr>
<td>17</td>
<td>M3-M9</td>
<td>Funeral</td>
<td>1m 12s</td>
<td></td>
<td>Friends</td>
</tr>
<tr>
<td>18</td>
<td>M3-M11</td>
<td>Not at home</td>
<td>17s</td>
<td></td>
<td>Acquaintances</td>
</tr>
<tr>
<td>19</td>
<td>M3-M12-F2</td>
<td>Address</td>
<td>2m 42s</td>
<td></td>
<td>Friends</td>
</tr>
<tr>
<td>20</td>
<td>M11-M1</td>
<td>Homework 1</td>
<td>17m</td>
<td>No closing</td>
<td>Friends</td>
</tr>
<tr>
<td>21</td>
<td>M11-M1</td>
<td>Homework 2</td>
<td>2m 7s</td>
<td></td>
<td>Friends</td>
</tr>
<tr>
<td>22</td>
<td>F14-F15</td>
<td>Colleague 1</td>
<td>16s</td>
<td></td>
<td>Acquaintances</td>
</tr>
<tr>
<td>23</td>
<td>M13-F16</td>
<td>Colleague 2</td>
<td>4m</td>
<td></td>
<td>Colleagues</td>
</tr>
<tr>
<td>24</td>
<td>M13-F14</td>
<td>Key</td>
<td>1m 36s</td>
<td></td>
<td>Friends</td>
</tr>
<tr>
<td>25</td>
<td>F14-F17</td>
<td>Lunch 1</td>
<td>23s</td>
<td>One-sided closing</td>
<td>Acquaintances</td>
</tr>
<tr>
<td>26</td>
<td>F14-F17</td>
<td>Lunch 2</td>
<td>14s</td>
<td></td>
<td>Acquaintances</td>
</tr>
<tr>
<td>27</td>
<td>F14-F18</td>
<td>Bill 3</td>
<td>2m 24s</td>
<td></td>
<td>Colleagues</td>
</tr>
<tr>
<td>28</td>
<td>M1-F19</td>
<td>Homework 3</td>
<td>1m 52s</td>
<td></td>
<td>Acquaintances</td>
</tr>
<tr>
<td>29</td>
<td>M1-M20</td>
<td>Chat 1</td>
<td>49s</td>
<td>One-sided closing</td>
<td>Friends</td>
</tr>
<tr>
<td>30</td>
<td>M1-M15</td>
<td>Chat 2</td>
<td>35s</td>
<td></td>
<td>Friends</td>
</tr>
<tr>
<td>31</td>
<td>F2-F20</td>
<td>Telephone number</td>
<td>2m 3s</td>
<td></td>
<td>Friends</td>
</tr>
<tr>
<td>32</td>
<td>M3-M16</td>
<td>Location</td>
<td>20s</td>
<td></td>
<td>Acquaintances</td>
</tr>
<tr>
<td>33</td>
<td>M3-M17</td>
<td>Meeting 1</td>
<td>2m 46s</td>
<td></td>
<td>Friends</td>
</tr>
<tr>
<td>34</td>
<td>M3-M18</td>
<td>Meeting 2</td>
<td>3m</td>
<td></td>
<td>Friends</td>
</tr>
<tr>
<td>35</td>
<td>M3-M19</td>
<td>Meeting 3</td>
<td>52s</td>
<td></td>
<td>Friends</td>
</tr>
<tr>
<td>36</td>
<td>F14-F22</td>
<td>Study abroad</td>
<td>2m 10s</td>
<td></td>
<td>Friends</td>
</tr>
</tbody>
</table>

The quantitative analysis provides statistical information of a descriptive nature about the frequency and distribution of the investigated tokens as well as their pitch contours. The phonetic realisation of one-word utterances, specifically in terms of the pitch contour, was identified auditorily and is illustrated using figures produced in Praat, a program for acoustic analysis (Boersma and Weenink 2010).
3.2 Utterances investigated
3.2.1 Backchannels

The utterances investigated in Chapter 4 comprise the discourse markers meaning ‘yes’ ơ [ɤ] and its variant ū [ɯ], vâng [vəŋ], and the non-lexical item mm as spoken in everyday telephone calls. The words ơ and ū are commonly used in conversations involving speakers of the same age (e.g. between friends) or same hierarchical position (e.g. colleagues). In contrast, vâng is commonly used as a polite response when addressing elders, people in higher positions, or people one meets for the first time. Table 3-2 provides tonal properties of the words, where mm bears no lexical tone and carries no lexical meaning. The words ơ, ū, and vâng carry two of the six lexical tones in Northern Vietnamese: ơ and ū have a lexical low-level tone, while vâng has a high-level tone.

Table 3-2: Tonal properties of investigated tokens.

<table>
<thead>
<tr>
<th>Token</th>
<th>Citation form</th>
<th>Stylised tone contour</th>
<th>Gloss</th>
<th>Denotation of lexical tone</th>
</tr>
</thead>
<tbody>
<tr>
<td>vâng [vəŋ]</td>
<td>high-level (or mid-level)</td>
<td>逐渐升高</td>
<td>yes (polite)</td>
<td>ngang (A1)</td>
</tr>
<tr>
<td>ơ [ɤ]</td>
<td>low-level (or falling)</td>
<td>逐渐升高</td>
<td>yes</td>
<td>huyền (A2)</td>
</tr>
<tr>
<td>ū [ɯ]</td>
<td>low-level (or falling)</td>
<td>逐渐升高</td>
<td>yes</td>
<td>huyền (A2)</td>
</tr>
<tr>
<td>mm</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

These utterances occur very frequently in the corpus, 1065 times in total, and in different pragmatic functions:

(a) BC: backchannels signalling the listener’s attention to the current talk and the intention of not taking the floor
(b) T_In: turn-initial acknowledgments when listeners are about to take the floor
(c) AN_yn: responses to signals produced by speakers actively eliciting affirmation or recognition of references made, e.g. to persons or places
(d) AN_imp: positive answers/responses to imperatives/requests which are clearly expressed by means of morphosyntactic structures
(e) Tel_clos: signals at telephone closings acknowledging the end of a conversation.

Table 3-3 provides an overview of these categories and their distribution in the whole corpus.

---

2 Examples for these signals are ý nhỉ ‘do you think so?’; thể hả ‘really?’, đúng không ‘right?’, à thể à ‘really?’; còn gì nữa ‘what else?’; hay sao ý, ... đấy or ... ý ‘you see?’.
3 For example, thể nhé ‘yes!’ or ‘ok!’.
Table 3-3: Distribution of annotated categories throughout the corpus.

<table>
<thead>
<tr>
<th>Category</th>
<th>BC</th>
<th>T_In</th>
<th>AN_yn</th>
<th>AN_imp</th>
<th>Tel_clos</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Token</td>
<td>516</td>
<td>213</td>
<td>212</td>
<td>49</td>
<td>75</td>
<td>1065</td>
</tr>
<tr>
<td>Percentage</td>
<td>48.4%</td>
<td>19.9%</td>
<td>19.8%</td>
<td>4.7%</td>
<td>7.1%</td>
<td>100%</td>
</tr>
</tbody>
</table>

The thesis focuses on utterances in backchannels only (n = 516). Results show that in the backchannel context, the utterances are commonly produced with a low or falling pitch. The analysis is then expanded addressing the question of what happens when speakers produce a rise instead of a fall on the investigated words. We found that the rising contour is used to yield the turn to the other interlocutor, to request information during telephone openings and to convey confirmation or agreement.

3.2.2 Repair initiations

The one-word utterances investigated in Chapter 6 are the yes/no question word hà, three wh-question words gì ‘what’, ai ‘who’, and sao ‘why/how’, the vocative particle ơi, the particle for politeness dạ ‘yes’, and the non-lexical item mm. These tokens carry four different lexical tones in Northern Vietnamese, the properties of which are provided in Table 3-4, along with their citation forms.

Table 3-4: Tonal properties of investigated words produced in repair initiations.

<table>
<thead>
<tr>
<th>Word</th>
<th>Citation form</th>
<th>Stylised tone contour</th>
<th>Gloss</th>
<th>Denotation of lexical tone</th>
</tr>
</thead>
<tbody>
<tr>
<td>ai [aɪ]</td>
<td>high-level (or mid-level)</td>
<td>– – – – –</td>
<td>who</td>
<td>ngang (A1)</td>
</tr>
<tr>
<td>sao [sao]</td>
<td>high-level (or mid-level)</td>
<td>– – – – –</td>
<td>why/how</td>
<td>ngang (A1)</td>
</tr>
<tr>
<td>ơi [ɤɪ]</td>
<td>high-level (or mid-level)</td>
<td>– – – – –</td>
<td>vocative particle</td>
<td>ngang (A1)</td>
</tr>
<tr>
<td>gì [zi]</td>
<td>low-level (or falling)</td>
<td>– – – – –</td>
<td>what</td>
<td>huyền (A2)</td>
</tr>
<tr>
<td>dạ [zd]</td>
<td>falling-glottalised</td>
<td>– ? –</td>
<td>yes (polite form)</td>
<td>nặng (B2)</td>
</tr>
<tr>
<td>hà [hɑ]</td>
<td>falling (-rising)</td>
<td>– – – –</td>
<td>yes/no quest. word</td>
<td>hỏi (C1)</td>
</tr>
<tr>
<td>mm</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note that utterances consisting of more than one word also occur in the corpus, such as cái gì as an alternative to gì ‘what’ or làm sao as alternative to sao ‘why/how’ (open class next turn repair initiations, see also Chapter 6, section 6.4). To guarantee consistency as far as the objects of study are concerned, the current study concentrates on one-word repair
initiations. Their multi-word alternatives are taken into account as evidence for the boundary effect of a high intonational tone in longer utterances.

3.3 Transcription/annotation

In the qualitative analysis, data were first transcribed orthographically. After establishing a collection of excerpts containing backchannels as well as other pragmatic categories, the data were transcribed in detail – as much as was relevant for the analysis – using the GAT 2 conventions (see Chapter 2, section 2.1.1, and Appendix B). Results are presented in the form of speech excerpts. They formally consist of a heading in bold, the name of the conversation, a context given in double brackets, a gloss of each constituent and a translation of the whole sentence. Following is an example illustrating a repair initiated by speaker M5 in line 4, a phenomenon analysed in Chapter 6:

Example (X): S6-Entrance exam ((M4 and M5 are friends, both aged 19))

1 M5 tôi chạy biết [nô thí (approx. 0.4s)]
   I  NEG know she take part in an exam
   I didn’t know about the exam (?).
2 M4 [nô dó  trường  gì]
   she pass university what
   In which university did she pass the exam?
3 (0.9s)
4 M5 <<all> gì?>
   What?
5 M4 (.nô dó  trường  gì)
   she pass university what
   In which university did she pass the exam?
6 M5 (0.6s) tôi cũng chạy biết nô dó  trường  gì  ighest
   I also NEG know she pass university what  ighest
   I don’t know in which university she passed the exam either.

For the analysis of the pitch contour on the investigated tokens, the intervals are annotated using Praat (Boersma and Weenink 2010) with the word in orthography, a gloss and its pragmatic category. Subsequently, the pitch contour was categorised as falling or non-falling. For the quantitative analysis, the frequency and distribution of the analysed tokens relating pragmatic function to pitch contour were automatically obtained using Praat scripts.
4 Prosody of backchannels

This chapter provides background information for backchannels and presents then a case-by-case analysis of the phonetic form of their pitch contour in the telephone corpus. In this qualitative analysis the low or falling pitch contour on the acknowledgment tokens ở, ừ and vâng is found to be the unmarked intonational pattern for backchannels signalling attention. Since the words ở and ừ also have a similar (low-level/falling) lexical pitch (huyền A2), evidence from the word vâng which has a lexical high-level tone (ngang A1) and from the non-lexical item mm is provided to ascertain the contribution of intonation to the pitch contour. This chapter also presents a quantitative analysis of the phenomenon and addresses some issues regarding cross-language comparisons.

4.1 Background

According to Duncan (1974: 166), words like ở, ừ, vâng and mm belong to a group of items which signal that attention is being paid and thus, constitute a type of backchannel behaviour. Also referred to in the literature as “continuers”, “acknowledgments”, “listener signals” or “recipiency tokens”, they play an important role in contributing to the flow of conversation (Rath 2001, Henne and Rehbock 2001). They are spoken by listeners to signal their attention to the current talk without taking the turn. In telephone calls, when speakers do not see each other, these signals make an indispensable contribution to the flow of the talk and to structuring the conversation. In fact, conversations can be disrupted when speakers do not backchannel. Backchannels in the corpus are very often elicited by a speaker using particles with a lexical rising tone such as nhé, nhớ, há, ý, or ā, and sometimes even with lengthening of the final syllable accompanied by a rise following the lexical tone when there is an absence of segmental material to carry the pitch.

In this section a number of recent studies on the prosody of backchannels are reviewed. They investigate a variety of languages based on different speech corpora ranging from task-oriented dialogues to spontaneous conversations. We begin with studies based on the latter speech materials. They are comparable to those analysed in the current study and will therefore be reported on in more detail than the others.

The primary function of backchannels is to signal attention. This includes the listener’s acknowledgment that he or she has understood, and/or that the listener is providing his/her agreement as to the continuation of the speaker’s talk (see Gardner 2003). In addition, these short items have also been found to signal “affiliation”, i.e. agreement/alignment, or “disaffiliation”, i.e. doubt, scepticism or hostility, of the listener towards the content of the current talk (Müller 1996). According to Müller’s study (1996) which is based on data from an Italian radio phone-in show, a type of institutional setting, these functional differences of backchannels appear to be distinguishable by prosody: affiliation can be signalled by matching the prosodic patterns to the prior turn of the speaker, whereas disaffiliation exhibits no “specific design to fit the particular speech object which they follow and acknowledge” (Müller 1996: 163). It is unclear, however, to what extent these results are language-specific.
A study of Australian, British and American English spontaneous speech corpora by Gardner (2003) indicated that, among others, both falling and falling-rising contours on *mm* tokens are used as acknowledgments. Details about the intonational patterns were reported specifically for the Australian corpus. The falling contour was found on a majority of monosyllabic “continuers” (199 out of a total of 290) signalling that the listener wants to hear more from the speaker and that s/he is not about to take the floor, whereas the falling-rising contour was found on only 44 monosyllabic tokens but on a majority of disyllabic *mm hm* tokens (28 out of a total of 32).

In a study by Ward (2004) on prosodic features in non-lexical utterances such as *uh-huh, um* and *hmm*, pitch height appears to correlate with degree of interest rather than with different functions in discourse: non-lexical backchannels in American English are produced with a low or flat pitch, and they often “start high and go down as the topic winds down”. Compared to words, non-lexical items have a tendency to be produced lower in pitch. The same tendency was also found by Mori (2005) for Japanese backchannels.

Results from a recent study on the prosody of backchannels in German everyday telephone conversations (Müller 2011) – the same type of data as those in the Northern Vietnamese corpus – show that 86% (121 out of 140 tokens) of monosyllabic “continuers” are produced with a low level or falling pitch. Falling-rising pitch is found more in disyllabic rather than in monosyllabic continuers (85.4%, 41 out of 48 tokens). In some studies of German based on auditory analysis, *hm* as a discourse particle is realised with a number of different intonational patterns. For example, when it is produced with a falling pitch, it signals acknowledgment, whereas when realised with a falling-rising pitch, it functions as a confirmation or indicates unrestricted agreement. Reproductions of short forms have also been observed. Crucially, there seems to be no systematic occurrence of falling contours for these (Schmidt 2001, Ehlich 1986). Interestingly, as will be shown in section 4 below, the Northern Vietnamese corpus exhibits a number of reduplications of backchannels occurring with the falling pitch contour (e.g. *vâng vâng* ‘yes, yes’). The significance of these differences across languages will be discussed in terms of their implications for intercultural communication.

To highlight the intonational patterns use in backchannels, studies based on other types of data are also briefly considered here, namely those using collaborative task-oriented dialogues. Backchannels in these studies tend to be produced with final high pitch. For instance, Benus et al. (2007) found that *mmhm* in American English is prosodically marked by high pitch, analysed in the Autosegmental-Metrical framework (Ladd 1996) as a high boundary tone H%. Caspers (2000) found similar intonational patterns in backchannels in Dutch map task dialogues. Results from the study by Savino (2011) also using map task dialogues showed both types of pitch contours – falling and high/rising – on backchannels in Italian map task dialogues depending on whether the listener intends to take the floor or not. However, this type of backchannels is, by definition, not the same as the one investigated in the current thesis. In task-oriented dialogues, they primarily have a dual function, namely acknowledgment and at the same time, elicitation of information. Therefore, we shall focus on studies using spontaneous naturally occurring data (Gardner 2003, Ward 2004, Mori 2005 and Müller 2011) which are most similar to the data used in the current study.

The following analysis shows that just as in the majority of the cases found in English, Japanese and German spontaneous naturally occurring data, backchannels in Northern Vietnamese everyday telephone conversations are commonly produced with a low level or falling pitch. This intonational pattern appears to significantly influence the pitch contour, seemingly masking the lexical tone of the words. However, the studies mentioned above show that English and German can also use a high pitch on backchannels, especially to
Convey interest. Therefore, the current study also looks at what happens when Northern Vietnamese speakers produce a rise instead of a fall in backchannels.

4.2 Case-by-case analysis

4.2.1 Falling/low level pitch contour on backchannels carrying a lexical low-falling tone

Example (4.1) is a conversation between speaker F2 and speaker F7 talking about F2’s daughter. F7 signals that she is paying attention to F2’s talk without taking a turn (line 2 and 4). The pitch contour produced in these two utterances is a falling contour (Figure 4-1 below), which is the same as the citation form of the lexical low-falling tone. The transcription includes word-for-word glosses and sentence translation (see also Appendix B).

Example (4.1): S8-Family ((F2 and F7 are friends, both aged 54))

1 F2 ò nhung mà nó có việc lu ò BÒ nhó có ó nhà đâu
   Yes but she often work busy she be at home NEG
   Yes. But she is quite busy and doesn’t stay at home.
   suốt ngày nó đi o cái việc của Nó ý mà
   everyday she go eh CLASS work of she y mà
   She goes out every day to do something for her work.
   -> 2 F7 ó.
   Yes.
   3 F2 ó nó vê a được bón tuần
     Yes she come back ah Được four weeks
     Yes. She has come back for four weeks.
   -> 4 F7 <<p>> ó.
     Yes.
   5 F2 thế nhu[ng mà:],
     so but
     However, ...

Figure 4-1: Pitch contour on two cases of ò functioning as a backchannel (marked by arrows) produced by speaker F7.
In almost all other conversations, ō and ừ functioning as backchannels also involve a falling pitch contour which is similar to the citation form realisation of the lexical low-falling tone (huyền A2).

Below is an example (example (4.2)) including backchannels with a low level pitch rather than a fall (see also Figure 4-2). This is also an alternative citation form realisation for the tone huyền, as it is often produced with a low-level pitch, particularly by male speakers, if it is near the bottom of their pitch range.

**Example (4.2): S10-Exam notes ((M9 is a workmate of M3’s wife. The children of both speakers have just taken the entrance examination for the same secondary school. M9 informs M3 about his son’s marks))**

(,...)

1 M9 ṣvang(-) thì là của HDA là là lớp A16 Yes so be of name of M3’s son be be class A16
(-) bôn năm học sinh giỏi thì four year pupil excellent TOP-MARKER ...
as an excellent pupil during the last four years,
họ thường là tâm điểm they reward be eight points
they rewarded him with eight points.

→ 2 M3 <<t> ờ. >
Yes.

3 M9 với cả điểm môn văn là 10,8 and mark subject literature be 10,8
and the mark for literature is 10,8.

→ 4 M3 <<p> 10,8 một nửa <<t, creaky> ở. > 10,8 one half =yes
10,8. A half. Yes.

5 M9 vang điểm môn văn là 10,8 (. ) điểm yes mark subject literature be 10,8 mark
môn ngoại ngữ là 11,5
subject foreign language be 11,5

The mark for foreign language is 11,5.

---

Figure 4-2: Low level pitch realised on two cases of ō functioning as a backchannel (see arrows) produced by speaker M3.
The backchannels in lines 2 and 4 are produced with a mean pitch of 121Hz and 96Hz, respectively. The second backchannel is produced with such a low pitch that the creaky voice of speaker M3 can clearly be perceived.

The function of the falling and low level pitch contours as prosodic features of backchannels in Northern Vietnamese can be further shown by cases in which the token vâng does not retain the citation form of its lexical high-level tone (ngang A1) when produced in backchannel utterances. This suggests that the intonationally motivated pitch contours identified can override the lexical tone.

4.2.2 Falling/low level pitch contour on the backchannel vâng (lexically high-level)

In Northern Vietnamese, vâng is used as an acknowledgment when addressing elders. In the corpus, this word occurs frequently in backchannel utterances and is commonly produced with a falling/low level pitch contour. In other words, producing this form either with a falling pitch contour or in the lower region of the voice range may be used to signal a speaker’s attention to the current talk. An example is given in (4.3).

Example (4.3): S5-Sister talk ((F2 and F6 are sisters. They talk about F2’s daughter))

(....)

1 F6 =mừng cho nó
glad for she
I’m happy for her.

-> 2 F2 <<t> vâng.>
Yes.

3 F6 bao giờ nó học xong hangs hay
when she learn finish then be in question
when she finishes studying, then it’ll be possible...

-> 4 F2 <<t> vâng.>
Yes.

5 F6 rồi nó lên thăm các bác sau
then she come visit PLURAL aunt afterwards
...for her to come to visit us.

-> 6 F2 <<t> vâng>
Yes.

Figure 4-3 exhibits the pitch contours on vâng when produced as a backchannel by speaker F2. The first two backchannels have a falling pitch and the last one has a low level pitch.
4.3 The non-lexical item *mm*

In the corpus, the semantically empty item *mm* can be found in backchannels, although less frequently than *ờ*, *ừ*, and *vâng*. It is also found in repair initiations (see Chapter 6). Since *mm* has no lexical tone, we should expect to find evidence of the presence of intonation alone. This means that intonational contours on *mm* as backchannels should also look like the ones in utterances *ờ*, *ừ*, and *vâng*. In fact, we found the low falling or low level pitch contour in 90 cases: 47 in isolation, 14 repetitions, and 29 in overlaps plus repetitions (see also below). Example (4.4) illustrates this intonational contour used in a telephone call made by two male speakers, M3 and M8. They are friends, aged 56 and 60. During M8’s explanation, M3 signals that he is listening and acknowledges the information provided by M8 with a low fall on *mm* (see Figure 4-4).

**Example (4.4):** S9-Children ((M8 and M3 are friends, aged 56 and 60. M8 explains that his son cannot go to school at the age of 6 years and 8 months and must wait until he is 7))

```
(....)
1 M8   nếu mà::i: đên a::: sinh tự tháng chính [dố vê (.) trước
   if        until ah born from September turn to front
   If he were born before September,
2 M3   [dù ( ) tròn tuổi            [ô.
      enough round age        yes
      Reach the age of 7,      yes
3 M8   là được sau [thì là тогда
      be ok after then be not ok
      he could go to school, born after that he cannot.
-> 4 M3   [ô.               mm. mm.
      Yes.  Yes.  Yes.
5 M8   sau là phải sang năm
      after be must turn to year
      Born after that he has to wait till next year.
-> 6 M3   mm.
      Yes.
```

Figure 4-3: Falling/low level pitch on three cases of *vâng* functioning as a backchannel (see arrows) produced by speaker F2.
Prosody in Vietnamese

4.4 Quantitative analysis

The study looked at 516 backchannel tokens in the 95-minute-corpus. The average frequency of occurrence is approximately 5 backchannels every minute. The most frequently occurring backchannels are ở (n = 216), followed by vâng (n = 123); ở was used 71 times. Interestingly, speakers used more than 100 mm utterances. Table 4-1 provides an overview of the pitch contours found across the backchannel tokens investigated.

Table 4-1: Frequency of backchannel tokens occurring in the telephone corpus and their intonation contours.

<table>
<thead>
<tr>
<th>Contours</th>
<th>Tokens</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>ở</td>
<td>ừ</td>
</tr>
<tr>
<td>falling/low-level</td>
<td>166</td>
<td>53</td>
</tr>
<tr>
<td>non-falling/non-low</td>
<td>50</td>
<td>18</td>
</tr>
<tr>
<td>Total</td>
<td>216</td>
<td>71</td>
</tr>
<tr>
<td>Percentage</td>
<td>41.9%</td>
<td>13.8%</td>
</tr>
</tbody>
</table>

We also carried out an analysis of these tokens with respect to their distribution in overlapping talk [+overlapped] or at the ‘transition-relevance place’ [-overlapped], and their occurrence in reduplications [+reduplicated] or in monosyllabic form [-reduplicated]. The results raise interesting issues with regard to cross-language comparisons.
Table 4-2: Distribution of backchannel tokens (n = 516) pooled across intonation contours, positions [± overlapped] and forms [± reduplicated].

<table>
<thead>
<tr>
<th></th>
<th>Falling/low-level (n = 422)</th>
<th>Non-falling/non-low (n = 94)</th>
<th>Total (percentage)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>+reduplicated</td>
<td>- reduplicated</td>
<td>+reduplicated</td>
</tr>
<tr>
<td>- overlapped</td>
<td>91</td>
<td>196</td>
<td>14</td>
</tr>
<tr>
<td>+overlapped</td>
<td>34</td>
<td>101</td>
<td>4</td>
</tr>
<tr>
<td>Total</td>
<td>125 (24.2%)</td>
<td>297 (57.6%)</td>
<td>18 (3.5%)</td>
</tr>
</tbody>
</table>

Table 4-2 pools all tokens giving an overview of the distribution of the tokens across the intonation contours as well as positions/forms. With respect to the overlapped backchannel, 31.4% of the investigated backchannels (n = 162 out of 516) are produced during the turn of the current speaker. The amount of overlapping is considerably higher than in a comparable German telephone conversations (Müller 2011) in which only about 8% (n = 16 out of 188) of the investigated backchannels signalling attention involved overlapping. This suggests potential communicative inefficiency in intercultural communication between speakers of Northern Vietnamese and speakers of German (see Heinz 2003 for comparative analysis of German and American English backchannels).

While there seems to be no systematic occurrence of reduplication of short forms with falling pitch contours in German (Schmidt 2001, see Ehlich 1986), 125 falling backchannels of this type are found in the Northern Vietnamese corpus. That is, approximately 29.6% of all falling backchannels (125 out of 422) or, in other words, 24.2% of all investigated backchannel tokens (125 out of 516), involve reduplicated backchannels. Furthermore, falling contours in reduplicated backchannels in German, where they occur, are often related to disaffiliation or even disagreement (Schmidt 2001), whereas those in Northern Vietnamese appear, instead, to convey increased agreement or affiliation.

Overall, the majority of backchannels in Northern Vietnamese were produced with a falling or low pitch contour (n = 422), i.e. approximately 82% of the investigated backchannels. A small number (n = 94) were produced with a non-falling or non-low pitch. Concrete sequential analyses of cases with the latter intonational pattern show that pitch variation can also be used in the language to convey other pragmatic functions in discourse. These are turn-yielding, requesting information during telephone openings and providing clear confirmation or agreement. These three functions will be dealt with in the following chapter.
5 Other interactional contexts

This chapter looks at what happens when speakers produce a rise instead of the expected fall on the acknowledgment tokens investigated in Chapter 4. The results show that this pitch contour, which can be seen as marked, conveys a number of other interactional contexts, namely turn-yielding, request for information during telephone openings, and confirmation or agreement.  

5.1 Turn-yielding

5.1.1 Background

To signal that one is at the end of his or her turn and that the other speaker has the right/obligation to take the next turn, the speaker can use a number of turn-yielding techniques. Sacks et al. (1974: 716f.) report on some strategies: a) using the first part of an adjacency pair such as a question in a question-answer sequence; b) repeating parts of the prior utterance with question intonation or question words like What?, Who? etc., a technique analysed as repair initiation; c) using tag questions which commonly come after the transition-relevance place, especially when the other speaker does not voluntarily take over the turn.

A number of studies deal with the relation between prosodic features and turn-yielding techniques. For example, Local et al. (1986) found radically different pitch patterns when comparing turn yielding across two English varieties: i) in London Jamaican a “narrow fall on last syllable, start no higher than previous syllable; no greater pitch movement earlier”, ii) in Tyneside a “step up at the end of turn; [which] ends higher, and is no narrower, than any earlier step up”. In a study on Ulster English, Wells and Peppé (1996) found that, among other parameters, a bowl-shaped pattern (falling-level-rising) is used for turn exits (the authors use instead the term ‘turn delimitation’). Results from this study (Wells and Peppé 1996) show a significant difference in the use of prosody in turn exits between Jamaican English and the other dialects. The authors therefore emphasise factors which are language- and dialect-specific, and are involved in the ways interlocutors convey their turn ending as well as yield the new turn. In comparison to English (dialects), Standard German makes use of both a rising and a falling pitch contour to convey a possible turn ending (Selting 1995).

In Northern Vietnamese, we have found that speakers use a high/rising pitch to signal that they are exiting their turn (sequence) as well as yielding the turn to other interlocutors. In openings during telephone calls, the high/rising intonation pattern can also be used to request the reason for the call or to request the identification of the caller. This function is referred to in the thesis as ‘requests for information during telephone openings’. It affects the pitch contour and deviates from the citation form of the lexical tone. Interestingly, precisely the same words which are used in backchannels (ơ, ĩ and ĕ) are used in turn

---

4 Chapters 4 and 5 are an extended version of my article “Prosody of Vietnamese from an interactional perspective: ơ, ĩ and ĕ in backchannels and requests for information”, published in the Journal of the Southeast Asian Linguistics Society 3:1 (2010), 56-76.
40 Other interactional contexts

exits. The role of intonation can be clearly illustrated by comparing the same lexical item used as a backchannel and then as a turn exit.

5.1.2 Case-by-case analysis

The rising pitch contour which Northern Vietnamese speakers use to signal the end of their turns is presented in example (5.1). This example is from the same conversation between the two male speakers from example (4.4) above.

Example (5.1): S9-Children ((M8 and M3 are friends, aged 56 and 60. M8 explains that his son cannot go to school at the age of 6 years and 8 months and must wait until he is 7))

(\ldots )

1 M8 nếu mà ::::: đến a ::::: sinh từ tháng chín [dó vē (. ) trưc
   if until ah born from September turn to front
   If till uh if he were born before September,
2 M3 [dù ( ) trờn tuổi [ô.  
   enough round age yes
   Reach the age of 7, yes.
3 M8 là được sau [thì là thôi  
   be ok after then be not ok
   he could go to school, born after that he cannot.
4 M3 [ô. mm. mm.  
   Yes. Yes. Yes.
5 M8 sau là phải sang năm 
   after be must turn to year
   Born after that he has to wait till next year.
6 M3 mm.  
   Yes.
- > 7 M8  
   \ldots 
   Yes.
8 M3 thế là cùng ( ) chăm mặt một năm nhờ  
   so be also late one year SENT-PART
   That means he'll go to school almost a year late, right?
9 M8 ((laughs) ô.)  
   Yes.

During M8’s explanation, M3 signals that he is listening and acknowledges the information provided by M8 with a low fall on ô and mm (lines 2, 4 and 6). M3 takes the floor after the acknowledgment produced by M8 with a rise in pitch (line 7, see Figure 5-1). Note that the rise is on ô which has a lexical low-level/falling tone.
Figure 5-1: Rising pitch contour on ơ (lexical low-level/falling) produced by speaker M8 yielding his turn to speaker M3.

The following example (5.2) illustrates this intonational contour used in a more complex way, namely by both interlocutors in adjacent positions in talk as well as in a longer utterance.

**Example (5.2):** S4-Friend talk ((F2 and F4 are friends. They are talking about F2’s daughter))

```
1 F4  về  lâu chưa
     come back long Q-PART
     How long has she been there?
2 F2  mới về  được hai ngày,
     she new come back ĐƯỢC two day
     She has just been there for two days.
3 F4  à thể à  ủ thể tốt rồi
     oh I see OK so good already
     Ok. That’s good!
-> 4 F2  ơ,
     Yes.
-> 5 F4  ủ,
     Yes.
6 F2  thế bày là không biết là bà nội ơ đi đâu
     so say be NEG know be grandma uh go where
     I don’t know where you are,
     chưa thấy bà nội đâu cả=
     not see grandma somewhere
     I haven’t seen you for a long time.
```
Beginning with the longer utterance nó mới về được hai ngày ‘She has just been there for two days’ produced by speaker F2 (line 2), the rise occurs at the end of the utterance (marked by the dotted circle in Figure 5-2) indicating that F2 already wishes to exit her turn at this point and signals that F4 should go on. However, F4 just acknowledges the information (line 3) and does not go on saying more about the topic they are talking about. F2 then produces the acknowledgment token ơ with a rising pitch (line 4) indicating that F4 should continue. The latter uses the same strategy to signal that F2 should say more by producing the acknowledgment token ừ with the same rising intonation contour again (line 5). Both speakers F2 and F4 produce a rise on their acknowledgment tokens to exit their turns (see arrows in Figure 5-2). After ừ produced by F4, F2 starts a new turn which includes a new topic.

This case provides evidence that the high/rising intonation contour as a turn-exit/yielding technique also occurs in longer utterances. The last syllable ngày ‘day’ in the utterance nó mới về được hai ngày in line 2 has a lexical low-level tone. Yet, it is produced with a final rising intonation contour when the speaker wishes to yield the turn to the other interlocutor. This result supports the analysis of intonation in Northern Vietnamese discourse in terms of pitch targets (low or high) at the edges of the utterances leading to final falls or rises. Further information concerning this issue is provided in Chapter 7.

### 5.1.3 Quantitative analysis

A rise was also found in 27 short utterances in total. Table 5-1 provides the distribution of the utterances investigated, which occur in turn-yielding contexts.

<table>
<thead>
<tr>
<th>Token</th>
<th>ơ</th>
<th>ừ</th>
<th>vâng</th>
<th>mm</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>- overlapped</td>
<td>10</td>
<td>6</td>
<td>2</td>
<td>3</td>
<td>21</td>
</tr>
<tr>
<td>+overlapped</td>
<td>4</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>6</td>
</tr>
<tr>
<td>Total</td>
<td>14</td>
<td>7</td>
<td>3</td>
<td>3</td>
<td>27</td>
</tr>
</tbody>
</table>
The corpus also yields examples of a rise in the response tokens hà (‘pardon?’), lexically falling-(rising)) and dạ (‘yes’, lexically falling-glottalised) in responses to vocatives such as ơi ‘hey’. One example is given below for dạ.

Example (5.3): S35-Meeting 3 ((M3 and M19 are friends. M19 is some years older than M3. M19 calls to ask M3 to join a meeting))

((phone rings))

1 M3 ạlô- Hello?
2 M19 â V dây â ah first name of M3 that Q-PART Hi V?
(-)
3 M3 â anh  M â ah form addressing males first name of M19 Q-PART Ah, hi M!
(-)
4 M19 â anh V ơi ah form addressing males name of M3 VOC-PART Hey V!
-> 5 M3 dạ?
6 M19 bây giờ anh lên ngay chào a::: now you go immediately place ah (1.8s) ((addressing another speaker in the background)) bốn ba Hoàng Văn Thái nhớ forty three Hoàng Van Thai street ok Please come to Hoàng Van Thai Street Nr. 43, ok? ((....))

Speaker M3 produces a rise on the utterance dạ? to convey the meaning of ‘yes, go on!’ yielding the speaking turn to speaker M19. This example provides further support for the hypothesised function of the high/rising pitch as an intonational strategy for turn-yielding.

The high/rising pitch has also been found in telephone openings for requesting for the reason of the call or for the identification of the caller. We now turn to this context.

5.2 Request for information during telephone openings

5.2.1 Background

Openings in telephone calls are discussed in detail in Schegloff’s work on American English (1968, 1986). Although cultural variation can play a role in setting up conversations on the telephone, four standard opening sequences are identified (Schegloff 1986): a summons/answer sequence, an identification sequence, a greeting sequence, and an exchange of how-are-you sequences (see also Couper-Kuhlen 2001). In everyday communication between friends, family members, or acquaintances, these sequence types can overlap (e.g. identification and greeting sequences) owing to a high(er) degree of intimacy between participants.

In contrast to face-to-face conversation, telephone calls are commonly made at a distance. Somewhere at the beginning of the call, the caller must, therefore, give the reason why he or she is calling. Schegloff (1986) calls the position in which this occurs the anchor position. It is also the first topic of the conversation (see Sacks 1992). With respect to everyday and natural home settings, it is sometimes the case that the caller simply calls up and asks Hi, how are you? If the person receiving the call requests the reason for the call, he or she often gets the answer No reason, just felt like calling (Sacks 1992: 74). In
other words, intimate friends and relatives can find it unnatural to be asked explicitly for the reason of the call.

In the next section we present how speakers of Northern Vietnamese request the reason for the call during telephone openings. Examples from the corpus show that in addition to explicit formulation such as *What's up?*, speakers use a final high/rising pitch contour on acknowledgment words. This intonation contour can also be used to request the identification of conversational partners.

### 5.2.2 Case-by-case analysis

The rise is used in a total of five telephone openings: three utterances containing ờ, one ừ utterance and one vâng utterance. Among these utterances, two are responses to yes/no-questions in a new turn. Even though these utterances are not backchannels, they are still considered here. The results will show that on exactly the same tokens which occur as backchannels with a falling contour, a rising pitch contour is used to convey requests for information during telephone openings. Two examples of this contour are presented below and two cases in which this intonation pattern is not used are analysed. We argue that conversational partners orient themselves to the high/rising pitch contour as a request for identification or the reason for the call.

#### 5.2.2.1 Final high/rising pitch on acknowledgment words carrying a lexical low-level/falling tone

Example (5.4) is an excerpt from a telephone call in which the person who answers the phone (speaker M3) is not the person to whom the caller (speaker M12) wants to talk. The latter thus asks to speak to a third person, speaker F2.

**Example (5.4):** S19-Address ((M12 is a friend of M3, F2’s husband. He calls and talks first to M3 before asking to talk with F2. M3 tells F2 who is calling. F2 picks up the phone))

```
((...))
1 F2 anh T ã form addressing males first name of M12 Q-PART
   Hi T?
2 M12 VĂNG mè mú T yes man form addressing female friends F2’s first name
   a Q-PART
   Yes. Hi T?
3 F2 ↑Ứ có gì đây yes have what SENT-PART
   Yes. What’s up?
-> 4 M12 (-) ((laughs))
   (.)
-> 5 F2 <<p> ờ ?> Yes?
6 M12 ă cái bà cái bà DTNM,
   so CLASS woman CLASS woman name abbreviation of a third person
   nhà chồng nào nhỏ house place which SENT-PART
   Mrs. M, where does she live?
```

In this example, speaker F2 asks explicitly *ừ có gì đây* ‘What’s up?’ (line 3). This sentence frames the context as a request for the reason of the call. However, immediately after the request, the caller (M12) just laughs. He gives the reason why he is calling after
F2 produces the acknowledgment word ờ (line 5) with a final rising pitch contour (see Figure 5-3).

![Figure 5-3: Pitch contour on ờ (lexically low-level/falling) produced by speaker F2 requesting the reason for the call.]

### 5.2.2.2 Final rising pitch contour on the acknowledgment token vâng carrying a lexical high-level tone

Extract (5.5) shows that vâng carrying a lexical high-level tone can also be produced with a high/rising pitch contour when functioning as a request for identifying information (Figure 5-4). Note that vâng is used here as an answer to a yes/no question asking for confirmation of identification and can therefore be considered as a turn on its own, in contrast to other cases presented. Despite this sequential difference, the example still provides evidence of the rise functioning as a request for the identification of the caller.

**Example (5.5): S5-Sister talk** ((F2 and F6 are sisters. F6 is the mother-in-law of F5. F2 calls to talk with F6))

```
1 F5  alô
    Hello?
2 F2  à H à
    ah first name of F5 Q-PART
    Are you H?
-> 3 F5  vâng?
    Yes?
4 F2  ü::::ô đi T  ñây thế nào khỏe không con
    yes aunt first name of F2 here how well KHÔNG you
    Yes, aunt T here.  So, are you well?
5 F5  à đi à [con khỏe à
    ah aunt Q-PART I well POL-PART
    Ah, hello aunt. I’m well.
6 F2  [Ü:
    Yes.
```
5.2.3 Deviant cases

The corpus provides examples showing that speakers orient themselves to the high/rising pitch realised on backchannels or responses to yes/no questions, and interpret them as requests for the reason of the call. Below are two examples which illustrate how recipients respond differently when the high/rising pitch is not used. Example (5.6) is an excerpt from a call involving two male speakers M3 and M9:

**Example (5.6):** S17-Funeral ((M9 is a workmate of M3’s wife, speaker F2. He calls to talk with F2))

```
1  M3   alô
    Hello?

2  M9  <<all> <t> (vâng)> yes
   anh V  â>
   address form for male first name of M3 FOL-PART
   Yes. Are you V?

3  M3  â  H  â
   ah first name of M9  Q-PART
   Ah, are you H?

4  M9  âm vâng em H  â día  â
   ahm yes I first name of M9 here FOL-PART
   Yes. H here.

-> 5  M3  [Ô::
    Yes.

6  M9  [<<t> vâng >
    Yes.
    (---)

7  alô?
    Hello?

8  M3  ô anh đang nghe đẩy
    yes I CONT hear here
    Yes. I'm listening.
```
In line 5 M3’s token is produced with a considerable final falling pitch (see arrow in Figure 5-5). The response of M9 (line 7) as the recipient of this turn of M3 is different from those in unmarked cases illustrated in examples (5.4) and (5.5) above. He says alô ‘Hello?’ again signalling that something has gone wrong in the conversation. This indicates that M9 orientates himself to the typically occurring final rising pitch contour on the acknowledgment token used to request the reason for the call. After M3 confirms that he is still on the phone and listening, M9 explains why he is calling. Although there is an overlap at the beginning of the two turns (lines 5 and 6, see Figure 5-5), the final falling pitch contour is still perceivable and it triggers the response of M9.

![Figure 5-5: Final falling pitch contour as deviant case for requesting the reason for the call produced by speaker M3.](image)

The second deviant case (example (5.7)) does not exhibit any final high/rising pitch contour either, namely on the token ở used when acknowledging information during a telephone opening. This is the position in which the utterance could otherwise be used as a means of requesting the reason for the call. The excerpt is from a call between two colleagues, speakers M3 and F16. The former called to ask the latter about the work they were planning.
On the acknowledgment word of speaker F16 in line 3, a final falling pitch contour is produced. After that a relatively long series of sequences are exchanged (lines 4 to 11). Of particular interest is the turn in which M13 gives the reason for his call (line 12): he starts with the phrase 'cháu muốn hỏi ... 'I’d like to ask ...' In comparison to other cases where callers directly give reasons for the calls after a prosodic request on the part of the receiver, the account for the call by M13 is launched first with this phrase functioning as a pre-sequence. Example (5.7) illustrated as a deviant case suggests that as long as no final rising pitch contour is realised on🇨🇳/🇷🇺 and no explicit questions are asked, there is a relatively long series of inserted sequences until the reason for the call is given. Furthermore, the way M13 explains why he is calling indicates that his explanation has not been directly made relevant by F16, i.e. in terms of so-called ‘conditional relevance’ of adjacency pairs such as a question from one speaker expects an answer from the other (Hutchby and Wooffitt 2008). Instead it appears to be more of an unconstrained, self-initiated action of accounting for the call.
5.3 Confirmation/agreement

Among the remaining non-falling/non-low utterances (62 in total), 36 were found to involve confirmation or agreement. This function is analysed as emerging from the context, that is, the backchannels are used to signal not only attention but also the listener’s opinions on what has been said.

5.3.1 Case-by-case analysis

The following extract (example (5.8)) involves a conversation between speakers F2 and F7. F2 tries to explain that, although their daughters were both born in 1981, her daughter’s birthdate counts as 1980 according to moon calendar and that of F7’s daughter to 1981 according to sun calendar. F7 shows agreement with her friend by using a high/rising pitch on mm (line 4) (see Figure 5-6 below).

Example (5.8): S8-Family ((F2 and F7 are friends, both aged 54))

\[ (...) \]

1 F2 â the thi nó sang năm tâm một rồi
Ah so TOP-MARKER she turn to year 81 already
That means she was born in 1981.
(.)

2 F7 mm::
Yes.

3 F2 [con này tháng một năm tâm: một cho nên là
My daughter January year 81 so be
nó vẫn còn ở cái: năm cũ mà:
she still in CLASS year old SENT-PART
My daughter was born in January 1981, so her birthdate still counts as 1980, the previous year, you know!
(.)

4 F7 mm, Yes.
(.)

5 F2 [ô. ( )
Yes?

6 F7 [dây [ô.
You see.

7 F2 [lâu lâu có gặp
long long I see
con

\[ D \]
address form of female friends name of third person
không,
KHÔNG
Have you seen D recently?

8 F7 [ô. <<all> chua>
uh not yet
Uh, no I haven’t.

9 F2 [thè à
I see.

10 F7 [mai...
Tomorrow...

---

That is, 94 (total amount of non-falling/non-low utterances) - 5 (requests) - 27 (turn-exits) = 62.
5.3.2 Quantitative analysis

As for the remaining (62 – 36 =) 26 utterances with non-falling/non-low pitch, acoustic analysis was not possible because of the bad quality of the speech signal due to, for example, background noise and laughter. They are grouped into the category “Others” as shown in Table 5-2. The table also provides the frequency of different functions across backchannel utterances characterised by a non-falling/non-low contour.

Table 5-2: Distribution of pragmatic functions across non-falling/non-low backchannels.

<table>
<thead>
<tr>
<th>Function</th>
<th>Turn-yielding function</th>
<th>Confirm./ Agr.</th>
<th>Others</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Turn exit</td>
<td>Request</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ờ</td>
<td>14</td>
<td>3</td>
<td>18</td>
<td>14</td>
</tr>
<tr>
<td>ừ</td>
<td>7</td>
<td>1</td>
<td>8</td>
<td>2</td>
</tr>
<tr>
<td>vâng</td>
<td>3</td>
<td>1</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>mm</td>
<td>3</td>
<td>0</td>
<td>7</td>
<td>6</td>
</tr>
<tr>
<td>Total</td>
<td>27</td>
<td>5</td>
<td>36</td>
<td>26</td>
</tr>
</tbody>
</table>

5.4 Interim conclusions

This chapter has shown that Northern Vietnamese uses intonation on discourse particles to convey communicative meanings. Table 5-3 below gives an overview of the results, along with those found in Chapter 4.

(1) In backchannel utterances, the falling or low level pitch contour is frequently used on acknowledgment tokens ờ and ừ. This realisation appears to be identical to the lexical tone of these two tokens. What is striking is that the token vâng is also produced with a falling or low level pitch indicating that its lexical tone is overridden by the intonation. This issue is addressed in more detail in Chapter 7. The pitch patterns found on mm tokens provides evidence that we are dealing with specific intonational contours used in this interactional context, namely the low or falling pitch contour in backchannels.

(2) On the same tokens ờ, ừ, vâng and mm, a final high/rising pitch contour can be used to request the identification of a caller and/or the reason for the call during telephone
openings. There is a tendency for this intonational pattern to also be used for conveying agreement or confirmation. No utterance involving \textit{mm} was found in requests for information. In Chapter 7, it is proposed that in these contexts the lexical tone of the words is overridden or masked by the intonation (as for \textit{vâng} in backchannel context).

Table 5-3: Pitch contours on \textit{ờ}, \textit{ừ}, \textit{vâng} and \textit{mm} in different interactional contexts.

<table>
<thead>
<tr>
<th>Word</th>
<th>Stylised contour in citation form</th>
<th>Backchannel</th>
<th>Request Turn-yielding</th>
<th>Confirmation /agreement</th>
</tr>
</thead>
<tbody>
<tr>
<td>ờ, ừ</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>vâng</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>mm</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Despite the fact that in informal telephone conversations asking explicitly for the reason for the call may give an unnatural or sometimes even impolite impression, people who receive the call do not answer passively and wait until the reason for the call is given (Sacks 1992). In requesting the reason for the call as well as the caller’s identification, Northern Vietnamese speakers tend to produce a final rising contour on \textit{ờ/ừ} and \textit{vâng}. The organisation of requests using this pitch contour is further shown by the analysis of deviant cases, which indicate that recipients do indeed orient themselves to the identified pitch contours. These identified contours trigger specific reactions on the part of the speakers in talk.

The results in (2) support those found in other conversation-analytical studies. Firstly, short acknowledgment utterances and positive responses to yes/no questions are ideal sequential environments for prosody to be used in. In this environment prosody often conveys more actions than just acknowledging information, as discussed in Raymond (2010). The author argues that this is because the means available to speakers are minimal (short utterances) and highly structured (in terms of the organisation of adjacency pairs) while the contingencies (turn-taking rules and indication of an ongoing action) are frequently complex. Secondly, studies of the structure of telephone openings as a particular type of conversation (Schegloff 1986, 1998) show that exchanges in this phase of conversation are both highly structured and very complex. The use of prosody in greetings or the use of names in summons or identification sequences potentially convey more than simply greeting or addressing a person (see Raymond 2010). The rising pitch found in the acknowledgment tokens \textit{ờ}, \textit{ừ} and \textit{vâng} both as backchannels (no turn status) and as responses to yes/no questions (as turns on their own) reflects that prosody is flexible as a contextualisation cue. The rising pitch is used to contextualise the same context (e.g. request for information during telephone openings) in different utterances (backchannels and responses to yes/no questions).
This chapter addresses the function of prosody in conveying the activity of repair initiation in Northern Vietnamese. Results show that Northern Vietnamese speakers initiate repairs by using a high or rising pitch contour on question words and particles. Further evidence from longer utterances and from the non-lexical item *mm* is provided. We also found that if the high/rising contour is not used, speakers interpret the token in question as an acknowledgment. Controlled dialogues recorded using role playing are also analysed.\(^6\)

### 6.1 Background

As first mentioned in the introduction in Chapter 1, repair is a universal mechanism dealing with problems of hearing, producing or understanding utterances in conversation. A repair sequence consists of three phases: a trouble source turn, initiation of repair and performance of repair. The current analysis focuses on the second phase, i.e. initiation of repair, which is referred to in Conversation Analysis as a next turn repair initiator (hereafter NTRI), i.e. a repair initiated in the next turn by the listener due to problems understanding an utterance. The initiators/utterances investigated in the study are so-called ‘open class initiators’ (see Drew 1997): *hả*, *gi*, *sao*, *oi* and *mm*, which can all be translated after a mishearing as ‘pardon’.\(^7\)

Open class initiators can be associated with different types of problems occurring in the trouble source turn. They are typically associated with hearing problems, problems which occur due to overlapping talk and which are solved by the repetition of the prior turn. An example is:

```
1 M1  öz  độ dài nó đang chạy ở ngoại dã
   yes  wait wait it CONT run in out IMP-PART
   Yes, wait wait, it’s running outside,
2   phòng bao [niêu]
   room how much/many
   in which room are you?
3 M2  [(  )choi cua ông là gì?
   play of you be what
   What’s your nickname?
4 M1  hà?
   Pardon?
5 M2  tên người chơi của ông là gì
   name person play of you be what
   What is your nickname?
```

---

\(^6\) This chapter is a more elaborate version of my article Hạ (to appear) “Prosodic means in repair initiation as an activity in Northern Vietnamese conversation” in the volume *Linguistics of Vietnamese – an international survey* (Mouton De Gruyter) edited by D. Hole and E. Löbel.

\(^7\) The yes/no question word *hả* has a lexical falling (-rising) tone (hỏi C1), *gi* ‘what’ is lexically low-level/falling (huyền A2); *ai* ‘who’, *sao* ‘why/how’, and the vocative particle *oi* are lexically high-level (ngang A1); the particle for politeness *dạ* ‘yes’ is lexically falling-glottalised (nặng B2), and the non-lexical item *mm* does not bear a lexical tone.
The repair sequence in this example shows the sequential organisation of hearing problems, which includes:

(i) an overlap in speech as the trouble source (lines 2 and 3)
(ii) an NTRI hà (line 4)
(iii) a repetition of the previous turn (line 5).

Complementary to the open class initiators, there are specific repair initiators which are used when a speaker fails to identify a referent, e.g. ai ‘who?’ There can also be partial repetitions including a question word such as ‘move what?’ or ‘the what?’ This problem might be interpreted as involving “finding a referent for a word” rather than involving “the meaning of the word” itself (Sidnell 2010: 128). Since the current study investigates predominantly one-word utterances, those which are longer than one word are not dealt with thoroughly in this chapter. They will however be mentioned briefly below as providing evidence for the intonational pattern investigated.

The prosody of repair has been investigated in a number of languages. For instance, standard German uses pitch height and loudness for repair initiation as an activity type. A high overall pitch movement is used on the wh-question word *was* ‘what’, when the speaker has hearing problems. An even higher overall pitch movement and a higher overall loudness can be used on *was*, when what was previously uttered is contrary to the listener’s expectation (Selting 1996). Rhythm and tempo in other-repair initiations have also been found to contribute to the interpretation of the trouble source in American English repair sequences, either as a mishearing problem of the hearer initiating a repair from the speaker or a potential face-threat faulting the speaker of the trouble source (Couper-Kuhlen 1992).

In a study by Curl (2005), different phonetic realisations of repetitions in repair are found to convey different structural/sequential functions of trouble source turns in relation to the prior talk. Specifically, “upgraded” repetitions that are produced louder, longer, and with expanded pitch ranges, and different articulatory mechanisms indicate that the trouble source turn fits structurally/sequentially with the prior turn. In other words, it forms a sequential pair with the prior talk (e.g. greeting-greeting or thanking turn – rejecting the thanking as unnecessary). “Nonupgraded” repetitions, i.e. repetitions that are quieter, shorter, and with compressed pitch ranges and very similar articulatory gestures, signal that the trouble source turn does not form a sequential pair with the prior talk. The author argues – in line with Couper-Kuhlen (1992) – that the upgraded/more precise production of repetitions is a rejection of the role of troublemaker proposed by conversational partners initiating repairs, rather than a response to a mishearing/routine problem.

The analysis below focuses on intonation contours without going further into the types of trouble sources (Sidnell 2010, Couper-Kuhlen 1992) or the interrelationship between phonetic realisation and sequential organisation (Curl 2005). The aim is to address the question of whether Northern Vietnamese makes use of intonation in constructing actions in talk and if so, how intonation contours influence the lexical tones present in the language.

### 6.2 Case-by-case analysis

Repair initiations in Northern Vietnamese dialogues were found to be produced with a final high pitch leading to a rise in the signal. In order to illustrate this intonational pattern,
six extracts will be presented – each featuring one of the investigated lexical words mentioned above.

The first extract is a conversation between two male speakers talking about a computer game, the transcription of which is an extended version of the one discussed above.

**Example (6.1):** S2-Computer game 2 ((M1 and M2 are friends, both aged 11))

1. M1 \(\uparrow\) (0.5s) đợi đợi nó đang chạy ở ngoài đâu
   Yes wait wait it CONT run in out IMP-PART
   yes, wait wait, it’s running outside

2. 

3. phòng bao [nhieu]
   room how much/many
   in which room are you?

4. M2 [(approx. 0.2s) chơi của ông là gì?]
   playing of you be what
   What’s your nickname?

\(\rightarrow\) 5. M1 (--) hà?
   pardon?

6. M2 (--) tên người chơi của ông là gì:
   name person play of you be what
   What is your nickname?

7. M1 (--) ăm ở cụ tí
   Uhmm ah Cutí
   Eh uh Cutí.

A few milliseconds after the trouble source turn in line 4, which overlaps with the prior turn of speaker M1, speaker M1 initiates a repair using a final high pitch on the question particle hà (line 5). The mishearing of M1 is evident from the next turn (line 6) where M2 repeats his prior utterance. Figure 6-1 provides the F0 contour produced on hà.

![Figure 6-1: High/rising pitch contour on yes/no question particle hà (lexically falling (-rising)) produced by speaker M1 initiating a repair after mishearing.](image)

It is important here to recall that the lexical falling (-rising) tone (hỏi C1) is often produced in Hanoi Vietnamese with merely the falling component, both in isolation as well as in spontaneous speech. The final high pitch used on hà in the context of a next turn repair initiator brings about a new pragmatic function for this word, namely to ask for information as in ‘pardon?’ or ‘what did you say?’
The high/rising pitch contour can also be identified on the wh-question word *gi* ‘what’, which has a low-level/falling lexical tone (huyền A2), when speakers use it to initiate repairs, here again, after mishearing caused by overlapping speech. Example (6.2) illustrates a repair initiation produced on this word, which is additionally characterised by fast speech tempo (indicated in the transcription as <<all>>) and has a duration of only 10 ms (see Figure 6-2). This is an excerpt from a conversation of two male speakers talking about a school friend who has just passed an entrance exam for a university.

**Example (6.2): S6-Entrance exam ((M4 and M5 are friends, both aged 19))**

1  M5  tôi chưa biết [nố thí (approx. 0.4s)  
   I NEG know she take exam  
   I don’t know about the exam ( ).

2  M4  [nố độ trường gi]  
   she pass university what 
   In which university did she pass the exam?

3  (0.9s)  

4  M5  <<all> gi>?

What?

5  M4  (. . nố độ trường gi)  
   she pass university what 
   In which university did she pass the exam?

6  M5  ( . . . tôi cũng chưa biết nó độ trường gi ý)  
   I also NEG know she pass univ. what ý  
   I don’t know in which university she passed the exam either.  
   ( . . . )

Figure 6-2: High/rising pitch contour on wh-question word *gi* (lexically low-level/falling)  
produced by speaker M5 initiating a repair after a hearing problem.

In this example the lexical low-level/falling pitch of the word *gi* is absent in the signal indicating that the final high pitch functioning as a repair initiation can mask the lexical tone of the word.

The next example is an excerpt from a call between the family of speaker M1 and the person who collects monthly payments for electricity, speaker F1, whose age is believed to be between 25 and 35 years:

In this example the lexical low-level/falling pitch of the word *gi* is absent in the signal indicating that the final high pitch functioning as a repair initiation can mask the lexical tone of the word.

The next example is an excerpt from a call between the family of speaker M1 and the person who collects monthly payments for electricity, speaker F1, whose age is believed to be between 25 and 35 years:
Example (6.3): S3-Electricity ((M3 is the father of M1 and F23))

1 F1 <<t> vang> hai trieu bon tram ba mui mot nghin
   POL-PART
   Yes. In total 2,431,000 dong.
   (1.2s)
2 M1 cho ca ba nhai y a
   POL-PART
   for all three family y
   Uh no.
   (1.3s)
3 M3 <<t> o khoang > ((addressing M1))
   No.
   (0.8s)
4 F1 <<all> da?>
   Pardon?
   (1.3s)
5 M3 <<t> con > ((addressing M1))
   You!
   (0.8s)
6 F23 vang chi di ti nhai
   POL-PART
   Yes. Please wait a moment.
   (1.3s)
7 <<t> vang vang.>
   Yes. Yes.
   (0.8s)

Figure 6-3: High/rising pitch contour on politeness particle da (lexically falling-glottalised)
produced by speaker F1 initiating a repair after a problem understanding the
prior utterance of speaker M3 (glottalisation marked on F0 contour as ?).

In line 5, speaker F1 does not appear to understand why M1 asks cho ca ba nhai y a ‘For
three families?’ (line 3). She indicates this problem and produces the word da to initiate a
repair. In Northern Vietnamese, da is used to acknowledge information (‘yes’) and is the
polite form used when addressing elders or people one meets for the first time. It has a
lexical falling-glottalised tone but is produced in the context of a next turn repair initiator
with high pitch after the glottalisation (see arrow in Figure 6-3).

Similarly to the example including the word ha, this high pitch contour gives a new
pragmatic function to da, namely that of asking for information as in ‘pardon’ or ‘what did
you say?’ This is an important point concerning the function of prosody in Vietnamese
repair: when produced with a final high/rising pitch, *đạ* indicates problems in hearing or understanding utterances that require some sort of resolution, for example, a repetition or clarification from other interlocutors or alternatively just ignorance of the trouble source turn (line 3) as is speaker M3’s and speaker F23’s response in lines 4, 6, 8 and 9. In this context the word *đạ* is not used to mean ‘yes’ anymore.

Like *gi* in example (6.2), the word *đạ* was produced at a fairly fast speech tempo with a duration of 19 ms. The falling portion of the lexical tone might be analysed as absent in the contour indicating that the lexical pitch is masked when produced at a fast speech rate. These examples suggest a role of speech rate in how much of lexical tone is masked or overridden by intonation. This issue will be addressed again in Chapter 7.

Similarly, the wh-question words *ai* ‘who’, *sao* ‘why/how’ and *ơi* ‘vocative particle’, all bearing a lexical high-level tone (ngang A1), are produced with a final high/rising pitch when initiating a repair after problems understanding utterances. Three extracts for these words are given in examples (6.4), (6.5) and (6.6).

**Example (6.4):** S14-Metro ((F10 and F12 are friends, aged 28))

((...))

1 F10 tao vi [goi] cho con N nói báo là (...) thụ sau
   I just phone to 3rd-PERS-SG N she say that/be Friday
tối thụ sau nó bay ra đây
   evening Friday she fly to here
   I’ve just called N. She told me that she is coming Friday
evening.

2 F12 [thẹ̀ à]
   so Q-PART
   I see.

3 (0.5s)

4 F10 là ngày hai chín
   be day twenty nine
   It’s the 29th.

-> 5 F12 *ai?* à con N *N* tàm
   who ah 3rd-PERS-SG N *N* what
   Who? Oh, N, isn’t it?

6 F10 c. Yes.

In line 1, speaker F10 tells speaker F12 about a third person named N who is coming back to Hanoi from Saigon. F12 is not sure whether she has understood about whom F10 is talking. Thus, she initiates a repair by asking *ai* ‘Who?’ (line 5) with a rising pitch contour (see arrow in Figure 6-4). The second question *à con *N* tàm? ‘Oh, N, isn’t it?’ of F12 in the same line indicates that she has misheard the name of the person talked about in the conversation.
Prosody of one-word repair initiations

Figure 6-4: Rising pitch contour on wh-question word *ai* (lexically high-level) produced by speaker F12 initiating a repair after mishearing.

Example (6.5) below is an excerpt from a conversation between two female friends regarding dinner plans. Speaker F10 calls to make sure that speaker F11 goes to the dinner they have both been invited to. The latter appears to have changed her plans (line 2) leading to a request for clarification from the former (line 3). In this request, speaker F10 produces the word *sao* ‘Why?’ (lexically high-level) with a final high/rising pitch to elicit/initiate a repair from F10 (see Figure 6-5 below).

Example (6.5): S13-Dinner ((F10 and F11 are friends, aged 28))

1. F10  nay the thu ba toi den nha
   hey so Tuesday coming go house
   con
   form addressing female friends k eat rice SENT-PART
   Hey, so next Tuesday we go to K’s place to have dinner, ok?

2. F11  mm quen mat khong bao chuyen vao thu hai
   uh forget NEG say change to Monday
   Uh, I forgot to say that we should change to Monday

   -> 3. F10  sao,  Why?
   (0.5s)

4. F11  mm (0.5s) quen ( )
   uh forget
   Yes, I forgot.

5. F10  sao sao nhu the nafo ro ra xem nao
   why why how clarify see SENT-PART
   So what happened?
   [ban a thu ba ban a
   busy Q-PART Tuesday busy Q-PART
   Are you busy on Tuesday?

6. F11  [o tho di tao ( )
   ok let me
   Ok. Let me (think about it).
Prosody in Vietnamese

Figure 6-5: Final rising pitch contour on wh-question word *sao* (lexically high-level) produced by speaker F10 initiating a repair after a problem understanding the utterance produced by speaker F11.

In the same conversation (see example (6.6)), speaker F11 produces the final high pitch not only on the word *gi*, lexically low-falling, but also on *oi*, lexically high-level, to convey that she has problems understanding (see Figure 6-6 below), probably due to differing expectations which are indicated by the long pauses (1.4s and 0.9s) after the trouble source turns in lines 4 and 6. These problems appear to be solved by the repetition and clarification provided by speaker F10 in line 8.

**Example (6.6): S13-Dinner ((F10 and F11 are friends, aged 28))**

1 F11 thời được rồi ok
ok ok
(0.5s)
2 F10 Thế có được không ok không?
*So CÓ ok KHÔNG ok KHÔNG*
*Is that ok for you?*
3 F11 rồi rồi rồi rồi rồi
ok ok ok ok ok
4 F10 Thế mm::: bảo cả anh
*so uh say also form addressing elder males*
H nhỏ
*name of a person SENT-PART*
*So, tell H, too, ok?*
(1.4s)
-> 5 F11 gì?
*What?*
6 F10 bảo cả anh H đến mẹ K
tell also H come mother name of another friend
nâu cơm cho cả anh H mà,
cook rice for also form addressing males H SENT-PART
Tell H to come, too, K’s mother cooks for him, too.
(0.9s)
-> 7 F11 <<f> cí?>
Pardon?
(---)
8 F10 mẹ K nấu <<f> cơm cho cả anh
mother K cook rice for also form addressing males
Prosody of one-word repair initiations

Figure 6-6: Final rising pitch contour on the vocative particle ơi (lexically high-level) and the question word ɡì (lexically low-level/falling) produced by speaker F11 initiating a repair after a problem (supposedly due to a problem with expectations) understanding the utterance produced by speaker F10.

In sum, the six investigated words have underlying lexical tones and distinct meanings (hả, falling (-rising), ‘yes/no question particle’; ɡì, low-level/falling, meaning ‘what’; ɡa, falling-glottalised, meaning ‘yes’; ɡi, high-level, meaning ‘who’; sao, high-level, meaning ‘how/why’ and ɡi, high-level, ‘vocative particle’), but the final high/rising pitch used in repair initiations gives each a new pragmatic function, namely that of asking for information. The independence of the lexical and pragmatic meanings is best shown in the case of đạ which is used to mean ‘yes’ when produced with a falling pitch (ending with glottalisation) and means ‘pardon’ when produced with a final high/rising one. While in fast speech the rise in repair initiations can mask the lexical tones of the words, in normal or slow speech there may be some trace of the lexical tone in the initial part of the word (see the realisations of hả in Figures 6-1 and ɡì in Figure 6-6). We will address this issue again in Chapter 7.

6.3 Quantitative analysis

In total, we found 39 one-word other-repair initiations. The majority of cases of repair initiations deal with overlapping talk and problems understanding. Table 6-1 below provides the frequency of the words occurring in the investigated context. More than 50% of the cases investigated involve the yes/no question word hả, followed by the wh-question word ɡi ‘what’ (12.8%), then the politeness particle đạ and the wh-question word ɡi ‘who’ (each 7.7%) and finally the vocative particle ɡi (5.1%) and the wh-question word sao ‘why/how’ (2.6%). In comparison to backchannel tokens, which were much more frequent in the corpus, this frequency shows that conversations in the corpus generally went smoothly and that speakers successfully managed the flow of their talks.
Table 6-1: Frequency of investigated tokens functioning as repair initiations in the corpus.

<table>
<thead>
<tr>
<th>-Token-</th>
<th>Count</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>hà</td>
<td>22</td>
<td>56.4%</td>
</tr>
<tr>
<td>gì</td>
<td>5</td>
<td>12.8%</td>
</tr>
<tr>
<td>dạ</td>
<td>3</td>
<td>7.7%</td>
</tr>
<tr>
<td>ai</td>
<td>3</td>
<td>7.7%</td>
</tr>
<tr>
<td>mm</td>
<td>3</td>
<td>7.7%</td>
</tr>
<tr>
<td>oi</td>
<td>2</td>
<td>5.1%</td>
</tr>
<tr>
<td>sao</td>
<td>1</td>
<td>2.6%</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>39</strong></td>
<td><strong>100%</strong></td>
</tr>
</tbody>
</table>

### 6.4 Longer repair initiation utterances

In the corpus, repairs are also found to be initiated in utterances longer than one word, for example in open class initiators 1) cái gì cơ? ‘what?’ or 2) sao sao như thế nào? ‘What? What do you mean?’ They also involve specific repair initiators including question words, such as 3) ý nghĩa của các gì? ‘Meaning of the what?’ or 4) mua cái gì? ‘Buy the what?’ or 5) ghi hoàn cảnh của ai? ‘Write the situation of whom?’ Interestingly, the final contour of these utterances, namely the pitch on the last syllable, is also high, regardless of lexical tone: the particle cơ in 1) which is used to indicate intimacy or a close relationship between participants has a high-level lexical tone as does ai ‘who’ in 5); the word nào ‘which/what’ in 2) has a low-falling lexical tone as does the question word gì in 3) and 4). This high pitch produced at the end of longer utterances provides evidence for the function of the rise as a conveyor of repair initiations, not only on the investigated tokens in isolation, but also on any word as long as it is final in a repair initiation.

Example (6.7) illustrates a longer utterance including the question word gì ‘what’ produced in the corpus as a specific repair initiation (see also Figure 6-7). Speaker M1 is reading out homework to speaker M11 who has presumably missed a class. The latter writes it down while talking.

**Example (6.7): S20-Homework 1 ((M1 and M11 are school friends, aged 11))**

```
1 M1  ý nghĩa của các chi tiết ( ) tương (-)
     meaning of CLASS detail (ima)gination
     miraculous
     Meanings of the details of miraculous (ima)gination.
     (3.1s) ((M11 writes when talking))

-> 2 M11  ý nghĩa của các gì?
    Meaning of CLASS what
    Meaning of what?

3 [của các
    of CLASS
    (0.5s)
    à cửa ý nghĩa của các chi tiết tương tương kí ảo
    Ah of meaning of CLASS detail imagination miraculous
    Ah, no, meanings of the details of miraculous imagination.
    (1.8s)
    đọc thiếu
    read lack
    I read incompletely.
```
6.5 The non-lexical item \textit{mm}

In the corpus, three utterances including \textit{mm} occur in the context of repair initiation. Here we found a high/rising pitch used when the speaker has problems understanding or mishearing a previous utterance of the other interlocutor. Since the item \textit{mm} does not bear a lexical tone, results in this section provide evidence that the high/rising intonation contour is used to convey repair initiations. One of the utterances is given in example (6.8). It is an excerpt from the same conversation as before, between speakers M1 and M11. Here overlapping talk affects the understanding of utterances. In this example, speaker M11 produces the word \textit{mm} in his initiation of repair, again, with a high/rising pitch (see Figure 6-8 below).

**Example (6.8): S20-Homework 1 ((M1 is reading the homework to M11))**

1 M1 cách hai dòng xong rồi gạch đầu dòng [cười dừa tài]
   Leave two line then bullet point the competition
   Leave two lines, then put the next bullet point “the competition”!

2 M11 [rõí]  
   \( \text{(1.0s)} \)
   \( \text{Ok.} \)

   \( \rightarrow 3 \)
   \( \text{mm?} \)
   \( \text{Pardon?} \)
   \( \text{(1.6s)} \)

4 M1 <\h> câu a câu gạch đầu dòng  ý
   You ah you put bullet point  ý
   You just put a bullet point!

5 M11 [rõí]  
   \( \text{Ok.} \)

6 M1 [cách môt a cách hai dòng ra rồi gạch đầu dòng
   Leave one ah leave two line out then bullet point
   Leave one, no, two lines and then next bullet point.
6.6 Repair initiations vs. backchannels

As mentioned in Chapter 2 referring to the methodologies in CA, spontaneous speech does not always provide a sufficient number of deviant cases through which the phenomenon in question can be proven to be relevant. To validate the pitch contour found in repair initiations, controlled dialogues were recorded using role playing. Four female speakers, aged between 24 and 28 years, participated in the recordings. They had not participated in the recordings of telephone conversations. These dialogues show that speakers actually produce the word dự with a different pitch depending on whether it is used in the context of acknowledging information ‘yes’ or the context of repair initiation ‘pardon?’. Thus, this result reveals the relevance of the high/rising pitch in conveying repair initiations in Northern Vietnamese talk.

The instructions of the reading task were given orally by the author. The dialogue used in the recording involves roles for two speakers, one playing a teacher, and the other a student. The latter has not done well on a test and receives suggestions from the former:

```
T  lần này làm bài chưa tốt lần sau em có gang time this do homework not good time next you try
  This time you haven’t done very well on the test, try again next time.
-> S  dự.
  Yes.
  ((The student feels sad and becomes thoughtful ... Suddenly the teacher continues ...))
T  em về nhà xem và học lại những bài sau You go home read and learn again these lesson following
  Now go back home and revise these exercises.
-> S  <<f> dự? >
  Pardon?
T  em về nhà xem và học lại những bài sau You go home read and learn again these lessons following
  Now go back home and revise these exercises.
```

The dialogue is constructed so that the student first acknowledges the teacher’s comment about her test (the first part of the dialogue) and then initiates a repair due to a
problem hearing or understanding a subsequent utterance of the teacher (the second part of the dialogue). In the transcription above, turns exhibiting these two different actions are indicated by arrows.

Results show that all speakers produced the word *đã* with a falling pitch ending with a glottal stop to convey acknowledgments and with a (final) high/rising pitch to initiate a repair. With respect to acknowledgments, the falling pitch with glottalisation appears to be consistent with the citation form of the lexical tone of the word (nặng B2). The dichotomy in pitch used on the word *đã* suggests that the use of different intonational patterns on the same word conveys different actions in conversation, namely the falling pitch contour for acknowledging information and the high/rising pitch (after the glottalisation) for initiating a repair due to, for example, problems hearing. The pitch contours on *đã* produced in the two different interactional contexts by one of the speakers are shown in Figure 6-9.

![Figure 6-9: Different pitch contours produced on *đã* in different interactional contexts – a falling contour when acknowledging information and a high/rising contour when initiating a repair due to a problem hearing (glottalisation marked on F0 contour as ?).](image)

In Figure 6-9 the falling component of the second *đã* can be explained by the type of conversation. Here it appears in a dialogue elicited by the use of role playing, where participants tend to speak more carefully than in spontaneous speech. This could lead to the preservation of lexical tone. Nevertheless, the high/rising pitch contour is clearly the determining factor in eliciting the repetition of the teacher. This reveals that in this dialogue only the prosody of *đã* determines the subsequent reaction of the recipient: either the teacher believes that the student has understood her advice, or she has to repeat her advice so that the message is conveyed. The identified pitch patterns used on *đã* can thus be described as contextualisation cues for different activities of interlocutors, namely the falling pitch for acknowledging information and the high/rising one for repair initiations due to problems in hearing or understanding utterances.

### 6.7 Interim conclusions

This chapter has shown that intonation plays an important role in Northern Vietnamese: pitch can be used as a contextualisation cue that is relevant in conversational interaction. Differences in pitch can lead to different interpretations of utterances, ultimately resulting in different recipient responses.
Table 6-2: Lexical tone contours of investigated words and their pitch contours in repair initiations (glottalisation marked by ʔ).

<table>
<thead>
<tr>
<th>Word</th>
<th>Lexical tone/Citation form</th>
<th>Stylised tone contour</th>
<th>Repair initiation in slow/careful speech</th>
<th>Repair initiation in fast speech</th>
</tr>
</thead>
<tbody>
<tr>
<td>hà [ha]</td>
<td>falling(-rising)</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>gi [zi]</td>
<td>low-level/falling</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>đa [za]</td>
<td>falling-glottalised</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>ai [ai]</td>
<td>high-level</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>sao [sao]</td>
<td>high-level</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>oi [yɪ]</td>
<td>high-level</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>mm</td>
<td></td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

Table 6-2 summarises the pitch contour used in repair initiations for each of the words examined in this chapter. All repair initiations have a high or rising pitch. In slow and more careful speech the falling components of the lexical tones (in the cases of hà, gi and đa) appear to remain in the signal; in fast speech tempo they appear to be masked by the intonation. The property of lexical tones that does remain is the glottalisation, as shown in the case of đa. Results from spontaneous speech support previous claims made on the basis of word lists read aloud (Phạm 2003b) that voice quality plays an important role and is, in fact, more stable than pitch. A detailed analysis of the interaction between lexical tone and intonation is provided in the next chapter.

The relevance of the high pitch is particularly clear in the case of the word đa ‘yes’ which is produced with different pitches depending on its function in the discourse. Speakers interpret the utterance đa as an acknowledgment of information when it has a falling pitch contour (default function) and as a repair initiation when it has a final high pitch. The latter can be seen as deviating from the former. By producing this word in a repair initiation, the speaker indicates that what s/he produces deviates from the default function of the word.

With respect to the relationship between the two main phenomena investigated, backchannels and repair initiations appear to be closely related in the way that if one is employed, the other is not available as an option. If a speaker produces ừ, ơ or vâng ‘yes’ as a backchannel acknowledging information, s/he does not/cannot at that moment say, for instance, gi ‘what’ to convey a mishearing and initiate a repair. In this respect, the relationship of backchannels and repair initiations in Northern Vietnamese is analysed as mutually exclusive.
Results from Chapters 4 and 6 exhibit the relationship of these two phenomena in terms of their prosody. Chapter 4 shows that backchannels in Northern Vietnamese are commonly produced with a low or falling pitch contour (on acknowledgment words), and Chapter 6 shows that other-repair initiations in the language occur with a high or rising pitch contour (on question words or particles). These opposing intonational patterns (low/falling versus high/rising) can be best observed in cases featuring the non-lexical items *mm*, which was found in both interactional contexts.
Interaction between lexical tone and intonation

In this chapter we provide an analysis of the interaction between the intonational contours used in the contexts investigated in Chapters 4, 5 and 6, and the contours of the lexical tones. A tentative proposal will be presented as to how intonational tones interact with lexical tones. While in backchannels and other interactional contexts (request for information, turn-yielding and agreement/confirmation) lexical tones appear to be overridden by intonational tones (in that little trace of the lexical tone is found in the signal), both the intonational and the lexical tones are found in certain cases of repair initiation, in particular those with slow or careful speech. In an autosegmental model, lexical and intonational tones are associated with the word and phrase, respectively. Cases in which the lexical tone appears to be absent are analysed as a delinking of this tone.

7.1 Data and methods

In this chapter, words in interactional contexts (backchannel, request for information, turn-yielding, agreement/confirmation and repair initiation) are descriptively compared with those in citation form produced by the same speakers. Even though these two data conditions (spontaneous vs. read speech) are quite different from each other, the contours in citation form are meant to be a benchmark with which the contours of utterances in telephone conversations may be compared. Since it was not possible to recruit all of the speakers who participated in the recordings of telephone calls, data for only six of them (speakers M3, F2, F7, M1, F11, and F12) are presented here. To elicit words in citation form, a frame sentence was used:

\[
Tớ
dạy
tớ X
do cho
dế.
\begin{align*}
\text{[short pause] } & X. \\
\text{(I teach the word X to the child. X)}
\end{align*}
\]

Speakers were instructed to read the frame sentence in a “normal/not fast” speech rate and replace X with the target words investigated in Chapters 4, 5 and 6: ơi, ừ, vâng, hà, già, đa, ai, sao and ơi. After producing a word in position 1 (within the sentence) and position 2 (after a short pause at the end of the sentence), they took another short pause (roughly 20ms long) and began the next sentence including the next target word. Each sentence was repeated three times. In total, each speaker produced 27 tokens (9 words × 3 repetitions). Target words were embedded in the frame sentence (position 1) instead of being read aloud separately in a list in order to avoid listing intonation. In this position they were produced in a neutral manner and we expect an effect of progressive tonal coarticulation.

---

8 This chapter extends work presented in Hạ and Grice (2010).
on the target word, namely from the lexical low-level/falling tone on the word từ (see Brunelle 2009b). By contrast, their production in position 2, that is at the end of the sentence, is expected to involve the intonation of a declarative which lowers the pitch of the lexical tones (see figures illustrated below), on the one hand, and is, due to the pauses before and after, not influenced by tonal coarticulation from neighbouring syllables, on the other. Words in both positions were extracted for the analysis.

In order to illustrate the tone-intonation interaction, the mean F0 contours of three citation-form repetitions for each word in position 1 as well as position 2 are descriptively compared with those of words spoken in telephone conversations. All contours of the words were time-normalised and measured at 20 equidistant time points. Where target words were produced only once by a given speaker in the telephone corpus (see Table 7-1 below) their contours are directly compared with the mean contour of the citation forms. Since no comparison across speakers is intended, the F0 values are presented in Hertz, not in semitones. Tokens produced in overlapping talk or in reduplications as well as those accompanied by laughing or noise are excluded from the analysis.

Table 7-1: Occurrence of utterances in telephone conversations selected for the analysis of tone-intonation interaction.

<table>
<thead>
<tr>
<th>Token ▶</th>
<th>กระทรวง</th>
<th>or/ừ/vâng</th>
<th>hà</th>
<th>gi</th>
<th>ai</th>
<th>Bộ</th>
<th>Repair initiation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Speaker ▼</td>
<td>Backchannel</td>
<td>Turn-yielding</td>
<td>Request</td>
<td>Conf./Agr.</td>
<td>Repair initiation</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M3</td>
<td>16/0/11</td>
<td>1/1/1</td>
<td>1/0/0</td>
<td>1/0/1</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>F2</td>
<td>23/2/4</td>
<td>1/1/1</td>
<td>1/0/0</td>
<td>2/2/0</td>
<td>3</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>F7</td>
<td>4/0/0</td>
<td>2/0/0</td>
<td>1/0/0</td>
<td>0/1/0</td>
<td>2</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>M1</td>
<td>0/0/1</td>
<td>0/0/0</td>
<td>0/0/0</td>
<td>0/0/0</td>
<td>4</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>F11</td>
<td>0/0/0</td>
<td>0/0/0</td>
<td>0/0/0</td>
<td>0/0/0</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>F12</td>
<td>0/0/0</td>
<td>0/0/0</td>
<td>0/0/0</td>
<td>0/0/0</td>
<td>2</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Total</td>
<td>43/2/16</td>
<td>4/2/2</td>
<td>3/0/0</td>
<td>3/3/1</td>
<td>12</td>
<td>1</td>
<td>2</td>
</tr>
</tbody>
</table>

Table 7-1 provides information about the occurrence of the tokens included in the analysis. We see that the speakers did not produce all target words in the telephone corpus. Specifically, the requests for information produced by these speakers do not include ừ or vâng, and their repair initiations do not include dạ or sao. In the cases of speakers F11 and F12, the mean F0 contours of citation forms are compared only with those in repair initiations (gi and oĩ, or hà and ai), since data for these speakers in the telephone corpus provides only repair initiations.

7.2 Backchannels and other interactional contexts

7.2.1 Backchannels

Recall that in backchannels, the corpus analysis (Chapter 4) revealed that ṇ, ṩ and vâng are produced with a fall or a low fall, regardless of the lexical tone. Figures 7-1 (speaker M3), 7-2 (speaker F2) and 7-3 (speaker F7) show the contours of ơ and vâng in citation form in position 1 (neutral) and position 2 (declarative), alongside the contours in backchannels (in orange) for comparison. Since only two backchannels on ṩ (produced by speaker F2) are analysed in this chapter, their contours are illustrated below along with other interactional contexts.
Figure 7-1: Speaker M3, mean F0 values on \( \acute{o} \), lexically low-level/falling (diamonds), and \( \text{vâng} \), lexically high-level (triangles). Backchannels are in orange.

Figure 7-2: Speaker F2, mean F0 values on \( \acute{o} \), lexically low-level/falling (diamonds), and \( \text{vâng} \), lexically high-level (triangles). Backchannels are in orange.
Interaction between lexical tone and intonation

Figure 7-3: Speaker F7, mean F0 values on ờ, lexically low-level/falling. Backchannel is in orange.

Table 7-2 below provides information about the F0 values taken as a reference for the intonational patterns illustrated above. In this table, these values are given at three out of the 20 time points measured in backchannel contours produced by speakers M3, F2 and F7: the first time point (start), the tenth (mid) and the twentieth (end).

<table>
<thead>
<tr>
<th>Speaker</th>
<th>Token</th>
<th>Context</th>
<th>start (Hz at 1st time point)</th>
<th>mid (Hz at 10th time point)</th>
<th>end (Hz at 20th time point)</th>
</tr>
</thead>
<tbody>
<tr>
<td>M3</td>
<td>ờ</td>
<td>Position 1</td>
<td>103.2</td>
<td>106.1</td>
<td>98.9</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Position 2</td>
<td>100.7</td>
<td>99.3</td>
<td>88.1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>BC</td>
<td>111.6</td>
<td>114.8</td>
<td>104.3</td>
</tr>
<tr>
<td></td>
<td>ừ</td>
<td>Position 1</td>
<td>100</td>
<td>133.5</td>
<td>126.7</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Position 2</td>
<td>102.9</td>
<td>126.2</td>
<td>121.7</td>
</tr>
<tr>
<td></td>
<td></td>
<td>BC</td>
<td>110.8</td>
<td>113.8</td>
<td>108.5</td>
</tr>
<tr>
<td>F2</td>
<td>ờ</td>
<td>Position 1</td>
<td>138.6</td>
<td>138.8</td>
<td>120.9</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Position 2</td>
<td>132.6</td>
<td>127.4</td>
<td>110.5</td>
</tr>
<tr>
<td></td>
<td></td>
<td>BC</td>
<td>154.8</td>
<td>152.5</td>
<td>138.9</td>
</tr>
<tr>
<td></td>
<td>ừ</td>
<td>Position 1</td>
<td>150.6</td>
<td>193.6</td>
<td>183.2</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Position 2</td>
<td>130.4</td>
<td>173.2</td>
<td>155.9</td>
</tr>
<tr>
<td></td>
<td></td>
<td>BC</td>
<td>159.3</td>
<td>157.7</td>
<td>150.5</td>
</tr>
<tr>
<td>F7</td>
<td>ờ</td>
<td>Position 1</td>
<td>148.3</td>
<td>148.9</td>
<td>130.9</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Position 2</td>
<td>138.8</td>
<td>128.9</td>
<td>114.1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>BC</td>
<td>154.8</td>
<td>152.5</td>
<td>138.9</td>
</tr>
</tbody>
</table>

The pitch contour on this word will be illustrated below along with other interactional contexts (see Figure 7-5).
First, as can be seen in the figures above and Table 7-2, tokens produced in position 1 appear to have a higher overall F0 than in position 2. An exception is the case of the start point of vâng produced by speaker M3, which is slightly lower due to the tonal coarticulation caused by the word từ as well as the micro-prosody on the voiced fricative [v] in vâng. The lower pitch in position 2 may be due to finality, reflecting a low boundary tone characteristic of declaratives in this position (see Nguyễn and Boulakia 1999, Vũ et al. 2006, Brunelle et al. 2012) or may alternatively be influenced by a declination or some other downtrend.

Second, the F0 values in backchannels decrease over the second half of the syllable (from point 10 to 20). Thus, the contour in this context is a fall. As can be seen in Figures 7-1 and 7-2 above, the backchannel contours for both words fall; ờ and vâng have different lexical tones (ờ being low-level/falling and vâng being high-level), yet as backchannels they appear to have almost the same pitch movement/shape, indicating a possible delinking of the lexical tone in this interactional context.

Let us now consider the token ờ produced by speaker F7 (see Figure 7-3). The contour overlaps with and is nearly identical to the citation form contour in final position (position 2). That is, both contours (citation form for a lexically low-level toned word and backchannel) have the same shape, which is unsurprising, since the lexical tone has a similar citation form contour to what appears from the other examples to be the intonation contour.

7.2.2 Other interactional contexts

Chapter 5 shows that speakers might also use the words ờ, ừ and vâng (all meaning ‘yes’) to ask who is calling (request for identification), why the person is calling (request for the reason for the call), or to give the floor to the other speaker (turn-yielding context). In contrast to backchannels, which are produced with a (low) fall, these contexts are found to be conveyed by a rise. In other words, the exact same word (ờ, ừ or vâng) has quite a different pitch contour, depending on the context.

The following figures present the contours on ờ and ừ (both with a lexical low-level/falling tone) in the various interactional contexts produced by speakers M3, F2 and F7. In all interactional contexts (except backchannel in Figure 7-5) the contours rise towards the end of the utterance indicating a high pitch target at the right edge of the intonational phrase. This rise can be seen in Figure 7-4 in requests, turn-yielding context and agreements or confirmations produced on ờ by speaker M3; in Figure 7-5 in agreements or confirmations, turn-yielding context produced on ừ by speaker F2; in Figure 7-6 in turn-yielding context and requests produced on ờ by speaker F7; and finally, in Figure 7-7 in turn-yielding context and agreements or confirmations produced on ờ by speaker F2.
Interaction between lexical tone and intonation

Figure 7-4: Speaker M3, mean F0 values on ż, lexically low-level/falling, in citation form, in request for information, turn-yielding and agreement/confirmation contexts.

Figure 7-5: Speaker F2, mean F0 values on yr, lexically low-level/falling, in citation form, in backchannel, turn-yielding and agreement/confirmation contexts.
Interestingly, the rising contour in these contexts was also found on vâng, which bears a lexical high-level tone. This is illustrated in the figures below. Here we can see that vâng in different functional contexts also has different contours. In agreements or confirmations (speaker M3, Figure 7-8) or in turn-yielding context (speaker F2, Figure 7-9) this word (lexically high-level) is produced with a high rise, whereas in backchannel context it is produced with a low fall. This indicates that backchannels do not appear to provide information as to the lexical tone on the word.
Interaction between lexical tone and intonation

![Graph](image1)

Figure 7-8: Speaker M3, mean F0 values on vâng, lexically high-level, in citation form, in agreement/confirmation and backchannel contexts.

![Graph](image2)

Figure 7-9: Speaker F2, mean F0 values on vâng, lexically high-level, in citation form, in turn-yielding and backchannel contexts.

7.3 Repair initiations

After dealing with backchannels and other interactional contexts produced on ờ, ừ and vâng we now turn to repair initiation, the context in which the function words hà, gi, ai and sao (‘yes/no question particle’, ‘what’, ‘who’ and ‘why/how’, respectively) as well as the particles đa and oí (‘politeness particle’ and ‘vocative particle’, respectively) occurred (see Chapter 6). In this context, in particular when the speaker has misheard what the other participant has said, a rising contour was found. The analysis of controlled dialogues showed that in careful speech the lexical tone remains in the contour, that is, the rise initiating a repair is preceded by the lexical tone (see Chapter 6, section 6.6). The analysis
in this section includes the words hả, gì, ai and ơi, since the telephone data from the six speakers involved in the recordings of citation forms only provides these utterances (see Table 7-1).

Figure 7-10: Speaker M1, mean F0 values on hả, falling (-rising), in citation form and repair initiations.

Figure 7-11: Speaker F2, mean F0 values on hả, falling (-rising), in citation form and repair initiations.
Let us first consider words starting with a lexical low pitch: hà and gì. Figures 7-10, 7-11 and 7-12 above present the mean F0 contours of hà (falling (-rising)) produced by speakers M1, F2 and F7 in repair initiations and in citation forms. In the case of speaker M1 (Figure 7-10), the repair initiation contour falls in the first 30% of the signal and rises during the last 70%. The rise reflects the function of repair initiation, whereas the fall can be analysed as a trace of the lexical tone of the word. The same tendency can be observed in the data of speakers F2 (Figure 7-11). However, in the case of speaker F7 (Figure 7-12) it is more the lack of a final fall, rather than preservation of an initial fall, that appears to be used to convey the repair initiation context.

A trace of the lexical tone in the first portion of the signal can be observed in the data produced by speaker F11 (Figure 7-13), whose repair initiation involves the question word gì ‘what’. The word has a lexical low-level/falling tone. The first 50% of the signal presents the same falling pitch as in citation form. The rise does not occur until the second half of the contour.
The rise functioning as a cue for repair initiations can also be found on words with a
lexical high pitch such as the question word *ai* ‘who’ and the vocative particle *ői* as
illustrated in Figures 7-14 and 7-15. While the contours on words bearing a low lexical
tone rise towards the end of the word, the rise on words bearing a high lexical tone appears
to start from a higher point. This can be seen by comparing the utterances *gi* (lexically low-
level/falling) and *ői* (lexically high-level) produced by speaker F11 and plotted together in
Figure 7-16.
Interaction between lexical tone and intonation

Figure 7-16: Speaker F11, mean F0 values on ơi, lexically high-level, (glottalised region due to the production of the vowel-vowel sequence “từ ơi” in position 1 is marked by the dotted line) and گی, lexically low-level/falling, in citation form and repair initiations.

An excursus on speech rate

As reported in Chapter 6, speech rate can play a role in how much lexical tone is overridden or masked by intonation. Repair initiations can often be produced in fast speech. In this particular context, the intonation is more likely to completely mask the lexical tones of the function words and particles investigated. For convenience, two examples illustrated in Chapter 6 are given again here for the question word گی ‘what’ which generally bears a lexical low-level/falling tone and the politeness particle دا which bears a falling-glottalised tone.

When گی is used to initiate a repair at a fast speech rate, as illustrated in Figure 7-17, the contour on this word rises and the lexical low or falling pitch seems to be completely lost. The same tendency is found for دا: in fast speech, the low or falling component of the lexical falling-glottalised tone disappears, as shown in Figure 7-18. However, the glottalisation remains. In slow or careful speech, as reported in Chapter 6, section 6.6, dealing with controlled dialogues, the lexical tone of the word appears to be discernible in the signal, as illustrated in Figure 7-19.

Compared to the contours presented above in which a trace of the lexical tone is still present in the signal, these examples indicate that speech rate can play a role in the interaction between lexical tone and intonation; namely, in repair initiations produced at a fast speech rate, the lexical tone of the word can be completely masked by the intonation.
Figure 7-17: Pitch contour on *gi*, lexically low-falling, as a repair initiation in a telephone conversation produced by speaker M5.

Figure 7-18: High/rising pitch contour on politeness particle *dạ*, lexically falling-glottalised, produced by speaker F1 initiating a repair after a problem understanding the prior utterance of speaker M3 (glottalisation marked on F0 contour as ?).
Figure 7-19: Different pitch contours produced on *da* in different interactional contexts – a falling contour when acknowledging information and a high/rising contour when initiating a repair due to a problem hearing (glottalisation marked on F0 as ?).

7.4 **Autosegmental representation of tone-intonation interaction**

This section provides an autosegmental representation of the interaction between tone and intonation in the interactional contexts investigated in the current study. A schematic representation of the model is presented in Figure 7-20. Since the current study focuses only on monosyllabic one-word utterances, intermediate nodes in the prosodic structure between the intonation phrase (IP) and the syllable are left out of the representation.

**Backchannels.** The interaction between tone and intonation can be modelled as a delinking of the lexical tone in the case of backchannels so that it is only the intonational tones which are realised. Backchannels are analysed as having two boundary tones, a low boundary tone at the beginning (%L) and at the end of the intonation phrase (L%). For ô and ɾ, the backchannel pitch contour is identical to the lexical low-level/falling tone. However, this is not the case for vâng. The model accounts for the fact that regardless of the lexical tone, backchannels are typically produced as low level or falling.
Prosody in Vietnamese

Repair initiation
slow & careful speech (left), fast speech (right)

Figure 7-20: Proposal for modelling the interaction between lexical tone and intonation used in investigated interactional contexts (heavy crossed line marks tone delinking, $t_{\text{lex}} =$ lexical tone, TRN = tonal root node).

In the following, Figure 7-21 and Table 7-3 provide again the tonal representation proposed in the current thesis and the phonetic properties of the six lexical tones in Northern Vietnamese (see also Appendix A for an overview).

Figure 7-21: Representation of lexical tone in the simplified model used in this thesis and an illustration of the tone nặng
Interaction between lexical tone and intonation

Table 7-3: Tonal properties of six-tone system in Northern Vietnamese based on auditory impression (breathiness not considered).

<table>
<thead>
<tr>
<th>Tone name</th>
<th>ngang (A1)</th>
<th>huyền (A2)</th>
<th>sắc (B1)</th>
<th>nặng (B2)</th>
<th>hỏi (C1)</th>
<th>ngã (C2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pitch</td>
<td>h</td>
<td>l</td>
<td>lh</td>
<td>(h)lʔ</td>
<td>hl(h)</td>
<td>1ʔh</td>
</tr>
<tr>
<td>Citation form</td>
<td>high-level (or mid-level)</td>
<td>low-level (or falling)</td>
<td>rising</td>
<td>falling-glottalised</td>
<td>falling-(rising)</td>
<td>rising-glottalised</td>
</tr>
<tr>
<td>Stylised contour</td>
<td>···················</td>
<td>···················</td>
<td>···················</td>
<td>···················</td>
<td>···················</td>
<td>···················</td>
</tr>
</tbody>
</table>

Figures 7-22 and 7-23 present the modelling of the interaction between lexical tone and intonation used in the investigated contexts. This model includes two types: intonation overrides lexical tone and intonation follows lexical tone.

**Intonation overrides lexical tone**

**Backchannel**

**Request, turn-yielding, agreement/confirmation**

**Repair initiation in fast speech**

Figure 7-22: Modelling the interaction between the lexical tones of investigated tokens and the intonation used in investigated contexts, where the intonation overrides the lexical tone.
7.5 Interim conclusions and discussion

We have compared the pitch contours of discourse particles and function words used with different functions in different pragmatic contexts in telephone conversations with those in citation form. The pragmatic contexts investigated include backchannel, turn-yielding, request for information, agreement or confirmation, and repair initiation. This analysis has been able to provide evidence for two types of interaction between intonation and lexical tone: i) intonation can override the lexical tone; and ii) the lexical tone remains and occurs in sequence with an intonational tone. The major findings are as follows.

First, the low/falling intonation contour conveying backchannels is modelled with an initial low boundary tone (%L) and a final low boundary tone (L%). The contour consistently overrides the lexical tones of the words ờ and vâng leading to an analysis whereby the lexical tone is delinked. In the case of ờ, the intonation is identical to the lexical low-level/falling tone of the word and is therefore not particularly striking. But in the case of vâng, the lexical high-level tone can clearly be seen to be overridden as it is produced in backchannels with a low/falling intonation.

Second, in other interactional contexts (turn-yielding, request for information or agreement/confirmation) there is a tendency for a rise to be used. This contour is analysed to have an initial low boundary tone (%L) and a final high boundary tone (H%), such that the interpolation between the two overrides the lexical tone of the word.

Third, the data for repair initiations in slow or careful speech indicate that lexical and intonational tones are realised in sequence with the intonational boundary tone in final position. However, unlike for these other contexts, the lexical tone in repair initiations produced at a fast speech rate can be delinked before the final high boundary tone.

The analysis is based on limited data so far, precluding a complete analysis at this point. Nonetheless, the findings in this chapter contribute to the growing field of research on intonation, in particular with regard to typological differences across languages in terms of how the inherent conflict of different kinds of tones (accent tones, lexical tones and intonational edge tones) is resolved (particularly in monosyllabic utterances).

As reported in Ladd (1996, 2008), a number of intonation languages can be classified into groups, depending on the way they resolve a particular conflict: Southern Standard British English, for example, has been established as a ‘compression’ language, that is, a string of accent and boundary/edge tones stretched across a longer utterance is fully realised, even in monosyllabic utterances. For instance, the combination of L* H L H% (L is a phrase accent, Grice et al. 2000) in questions can stretch over the utterance driving instructor?! to express a strongly challenging or contradicting echo question. The whole combination can also be associated to a single syllable such as Sue?! By contrast,
Hungarian makes use of ‘truncation’, that is, either a phonetic reduction or a phonological deletion of tones. The typical tonal sequence for question intonation in the language is L* H L% (H is a phrase accent, Grice et al. 2000), but it is reduced to L* H when the question is monosyllabic. The boundary tone L% is lost. The same is true for L* H L% contours in Palermo Italian (Grice 1995). German – in comparison with English – tends to avoid compression. Like Hungarian, it also makes use of truncation. For example, in the surname Schmidt (L*) H used to answer the phone, it is the accent tone that is lost, not the boundary tone (Ladd 1996: 133-135).

For Southeast Asian tone languages, there is a tendency towards truncation. In Thai the lexical tone of final particles is found to override the boundary tones. For instance, in statements ending with the high tone particle *kháp ‘male, polite’, it is the low boundary tone (L%) characteristic of declaratives that is lost and the high tone of the particle that remains (Pittayaporn 2007). In Northern Vietnamese, the current study provides evidence that in backchannels and repair initiations (the latter in fast speech) it is the lexical tone that is lost, giving priority to the realisation of intonational boundary tones.

The findings in this chapter further show that differences might also exist within a single language, here (Northern) Vietnamese. These results are supported by our recent study on the realisation of utterance type and affect in Northern Vietnamese (Brunelle et al. 2012). There we showed that, in addition to gradual effects on F0 and intensity, speakers also use discrete intonational boundary tones to distinguish statements from questions or commands, and neutral from marked attitudes. However, a great degree of inter-speaker variation was apparent, indicating that the intonation patterns investigated are perhaps not (yet) fully grammaticalised; instead they merely enhance the other strategies used for marking communicative functions.

Comparing initial results from Brunelle et al. (2012), in which we found inter-speaker variation in intonational strategies, with those from the current study, which has shown that discrete intonational tones are used to convey functions in discourse such as backchannel and repair initiation, we assume that the use of intonational tones in Northern Vietnamese (likely) depends on the communicative functions being expressed. The use of such tones appears to be more stable and consistent in discourse-management related functions than in distinguishing utterance types (e.g. statements or questions) and affect (e.g. neutral or marked). In other words, while intonational tones appear to enhance other strategies marking utterance type and affect (given the large number of sentence particles in Vietnamese), they are linguistically relevant in their own right for discourse management functions. Different intonational tones can lead to different interpretations of utterances (backchannel vs. request for information, acknowledgment vs. repair initiation, etc.).
This chapter presents two perception experiments testing how native listeners of Northern Vietnamese interpret the naturalness and the affective meanings of politeness and dominance in backchannels when the final pitch levels are modified from low to high. The results show no clear tendency in the perception of naturalness leading to the interpretation that listeners perceived both backchannels ending with a high and a low pitch as equally natural. This appears to be compatible with the results found in the corpus analysis (Chapters 4 and 5) that backchannels are predominantly produced with a low pitch; they might also be produced with a final high pitch to acknowledge information and at the same time serve as requests for information, turn-yielding and agreements/confirmations. In the case of dominance, listeners tend to judge backchannels ending with a low pitch as less dominant. However, these judgments appear to depend on the context: this is the case for backchannels as signals of attention but not completely for backchannel as signals of agreement. In the case of politeness, different pitch levels are found to have a significant influence in both contexts: the low endings are interpreted as more polite, whereas the high endings are perceived as less polite. The findings indicate that Northern Vietnamese might be an exception to the intonational universals proposed across a large number of languages as derived from the Frequency Code (Ohala 1983, Gussenhoven 2004) that low pitch is associated with a lower degree of politeness and a higher degree of dominance. The study has implications for cross-cultural misunderstandings and foreign language acquisition.

8.1 Introduction

A number of studies on intonational meaning have found that speakers with different language/cultural backgrounds associate an intonational form with certain meaning both in similar and different ways. With respect to similarities across a large number of languages, high and low pitch, for example, have been established as conveying questions or statements, respectively. The similarities have so far been accounted for by intonational meanings derived from three biological codes which are considered to have an influence on the vibration of the vocal folds, and therefore on pitch. The Frequency Code (Ohala 1983, Gussenhoven 2004), which makes reference to the relationship between the size of the larynx and the vibration of the vocal folds across speakers. This code is claimed to explain the association of high pitch with questions and affective meanings such as friendliness, politeness and submissiveness as well as the association of low pitch with confidence and dominance (Uldall 1964, Haan et al. 2002, Gussenhoven 2004, Chen 2005). The Effort Code refers to the relationship between the energy expended on speech production and articulatory precision, explaining the expression of focus, in terms of linguistic meanings, or surprise, in terms of affective meanings (Gussenhoven 2004). The Production Code makes reference to the relationship between subglottal pressure and the exhalation phase of the breathing process (Gussenhoven 2004). This code is claimed to
explain the association of high pitch with the beginning of the utterances and low pitch with the end.

However, perception studies have shown that listeners with different language backgrounds, even if the languages are closely related and have similar intonation systems, can differ in their perception strategy (e.g. Gussenhoven 2002, Chen 2005). An experiment from Chen’s (2005) study on universal and language-specific perception of paralinguistic intonational meanings shows that Chinese, Dutch and Hungarian listeners perceive stimuli as questions when they hear a higher peak, a later peak and a higher end pitch. However, Dutch listeners appear to be more sensitive to peak height conditions than Chinese listeners, presumably due to the fact that in contrast to Dutch, Chinese makes use of neither peak raising nor delay for expressing interrogatives. Another experiment from the same study shows that British English and Dutch listeners appear to make different interpretations when conflicting meanings are derived from biological codes (Chen 2005). While the former listener group perceives a high pitch register with a higher degree of friendliness as derived from the Frequency Code, the latter perceives it with a higher degree of emphasis as derived from the Effort Code.

Differences in the interpretation of intonational meanings may therefore lead to misunderstandings between speakers of different languages or from different cultures (see Gussenhoven 2002, 2004: 81, Ding et al. 2010). Potential misunderstandings also exist between speakers of different dialects of the same language. Wells and Peppé (1996) found that in Ulster English, finality is conveyed by a rising tone and not a falling one as in Southern Standard British English.

A number of studies have shown that backchannel behaviour is also language and culture specific. Heinz (2003) found in her work on German and American English telephone data that German speakers tend to produce fewer backchannels in overlap than American speakers. Wichmann (2000) found that in British English, backchannels which are produced with a falling tone plus a rapid speech tempo can convey non-supportiveness, such as in the case where the hearer attempts to end the current speaker’s turn/topic. The same tendency can be seen in the study by Ward (2004) on American English, which shows that by using a high pitch on backchannels, the hearers can show interest in the talk and indicate that the speaker should say more about the current topic. These results illustrate the different layers of meaning that backchannels can convey. The listener is hearing and has understood what has been said, and the way s/he responds reflects his/her emotional state and attitude (Allwood et al. 1992, Grahe and Bernieri 1999, Schröder et al. 2006).

In this chapter we therefore examine how Northern Vietnamese listeners perceive backchannels in their language in terms of naturalness, politeness and dominance. We conducted two experiments, a forced choice task and a semantic scaling task, using stimuli whose final pitch was manipulated from low to high. We aim to address the question of whether Northern Vietnamese follows the intonational universals proposed for a number of languages as derived from the Frequency Code.

**Hypotheses**

Our hypotheses are based on the results found in the corpus analysis (Chapter 4) and on the proposed universal characteristics of paralinguistic meanings of intonation as derived from the Frequency Code. Backchannels in Northern Vietnamese are commonly produced with a low/falling pitch contour but they might also be produced with a final rise when speakers acknowledge information and at the same time request information during telephone openings, yield the turn to conversational partners or provide a confirmation/agreement.
I. Both a final low and a final high pitch on backchannels in Northern Vietnamese sound natural.

II. The higher the final pitch on backchannels the more polite the listener sounds;

III. The lower the final pitch on backchannels the more dominant the listener sounds.

Hypotheses II and III are illustrated in Figure 8-1.

![Figure 8-1: Hypotheses of the perceived degrees of politeness and dominance relative to the pitch level.](image)

To test these hypotheses, two experiments were carried out. The first experiment is a forced-choice task to determine whether a final high or a final low pitch is judged as more natural, more polite or more dominant. The second experiment is a semantic scaling task (see Chapter 2, section 2.1.2) testing perceived degrees of these meanings.

### 8.2 Stimuli preparation

For pitch manipulation, three source backchannels ờ (×2) and ừ (×1) with good sound quality (i.e. clear speech without noise or laughter) were chosen. They stem from the three following excerpts of spontaneous telephone conversations. In order to reduce the length of the experiments, we excluded utterances in brackets from the stimuli data set.

**Excerpt 1**

F2 ừ nhưng mà nó cứ lười bu: nó có ở nhà suốt ngày nó đi: ơ việc của nó ý mà
Yes. But she is quite busy and doesn’t stay at home. She goes out every day to do something for her work.

F7 ờ
Yes.

(F2 ờ nó về: được bốn tuần)
Yes. She has come back for four weeks.

**Excerpt 2**

F2 đi vài ngày cho nó sống di vào các siêu thị của nó thich LẮM
So go there, and have fun. The stores are really great,

F7 ờ
Yes.

(F2 bên ý mày xem mày xem nó cũng ngăn nắp lắm khôn g như mình đâu)
then you will see it is very tidy there, not like here.

**Excerpt 3**

F2 ừ nó bảo nó điều xe nếu bà đi tôi sẽ điều xe
Yes, he told me he would bring his car to help (give me a lift),

F4 ừ
Yes.

(F2 thế xong báo thôi thông cảm thời chú T. chưa biết thẳng ấy nó ở cái chỗ nào
The pitch contours of the source backchannels were first stylised in Praat with three target points. Since contour endings are more likely associated with affective interpretations derived from the Frequency Code (Haan et al. 2002, Gussenhoven 2004), the last 50% of the word (in terms of duration) was manipulated creating four contours with endings at four pitch levels, level 0 being the lowest, level 3 being the highest, as illustrated in Figure 8-2. The difference of each level is 30Hz. Stylised source backchannels are illustrated with solid line and manipulated backchannels with dashed lines.

The manipulated backchannels were embedded in the excerpts above and tested in different ways: in Experiment I only two extremes of the manipulated backchannels, the highest and the lowest (levels 3 and 0), were used to test which pitch contour (falling or rising) is more natural, more polite and more dominant. At the same time participants became accustomed to the manipulated pitch range of the signals. In Experiment II, the perception of gradient pitch variation at all four levels was tested. Since the contexts in Excerpts 1 and 3 appear to be the same (backchannels signalling attention), only one of them was chosen for this experiment. As stimuli, Excerpts 1 and 2 were used, the former including a backchannel signalling attention and the latter including a backchannel signalling agreement.

8.3 Experiment I: A forced-choice task
8.3.1 Design and methods

Subjects

49 second-year students from the faculty of Vietnamese studies at the Hanoi University of Education (aged from 20 to 24 years, 5 males and 44 females) took part in the experiment. They come from different provinces in Northern Vietnam. All speak English as a foreign language; their proficiency varies from beginner to intermediate. Some of them are additionally studying Korean as a foreign language. None had participated in a perception experiment before.

Stimuli and procedure

The two extremes (the lowest and highest pitches) of the manipulated backchannels were embedded in the three excerpts above, and each excerpt was judged for 3 meanings – naturalness, politeness and dominance. In total, participants listened to 9 pairs of stimuli (3 excerpts \(\times\) 3 meanings). These were presented to subjects on 9 slides appearing on a screen, as illustrated in Figure 8-3. The dialogue with the low backchannel is presented in

---

10 The reason why semitones were not chosen for the pitch manipulation is that the experiment does not aim at cross-speaker comparison, as will be shown in detail in the description of the experiments below. In experiment II only the stimuli produced by speaker F7 were used.
A and the one with the high ending is presented in B. Participants chose A or B as the preferred stimulus and wrote their answers on a piece of paper. The answers were collected before participants proceeded with Experiment II. The experiment took about 5-10 minutes and was run in a quiet room at the faculty of Vietnamese studies at the Hanoi University of Education using two computers and two sets of headphones.

8.3.2 Results

Figure 8-4 provides the number of judgments for the low (A) and high endings (B) in backchannels for all subjects across the three investigated meanings. There are a total of 147 judgments (3 excerpts × 49 subjects) per meaning. Beginning with naturalness, judgments of the low pitch were assigned just above chance level (55.2%, n = 81 out of 147); 44.8% (n = 66 out of 147) of judgments for naturalness were associated with the high pitch. For politeness, the judgments of the low pitch add up to 66.7% (n = 98 out of 147), considerably above chance, leaving the judgments of the high pitch amounting to 33.3% (n = 49 out of 147). Finally, for dominance only 38% of the judgments (n = 56 out of 147) were assigned to the low pitch, while judgments for the high pitch were considerably above chance level (62%, n = 91 out of 147).

Results from an one-sample chi-square test show that there is a significant difference between the judgments of low and high pitch in the perception of politeness ($\chi^2 = 16.333;\)
df = 1; p < 0.001) and dominance ($\chi^2 = 8.333; df = 1; p < 0.01$), but not in the perception of naturalness ($\chi^2 = 1.531; df = 1; p > 0.05$). This means, the choice of the low ending (i.e. falling) pitch for politeness and the high ending (i.e. rising) pitch for dominance is significantly above chance. This result does not support Hypotheses II and III. Yet, for naturalness – as expected – the difference in values did not reach significance, suggesting that both contours sound equally natural. This confirms Hypothesis I.

8.4 Experiment II: A semantic scaling task

8.4.1 Design and methods

Subjects, stimuli and procedure

In total, we have 8 stimuli (two excerpts $\times$ 4 manipulated pitch levels). These stimuli were presented three times each in an automatically randomised order. They were each preceded with a sound to draw the attention of the participants. When judging one meaning, each subject listened to 24 stimuli in total (8 stimuli $\times$ 3 repetitions). All participants (the same as in Experiment I) judged naturalness first. Twenty four of them then judged politeness and 25 dominance. Each participant therefore judged two meanings, either naturalness and politeness, or naturalness and dominance, with 48 stimuli in total (2 meanings $\times$ 24 stimuli). The experiment took about 15-20 minutes. Using Praat scripts, three 5-point-Equal Appearing Interval Scales (EAI) were arranged, in which 1 stands for “not natural”, “not polite”, “not dominant” and 5 for “natural”, “polite” and “dominant”. The scales are visualised on the screen (see Figure 8-5). Participants judged the stimuli by clicking on a number (1-5) on the scale. They could listen to the stimuli once again if required.

![Figure 8-5: Judgment scale for naturalness used in Experiment II.](image)

8.4.2 Results

Since participants differ in their scoring styles, results were normalised (in terms of z-scores) (Grabe et al. 1997, Rietveld and Chen 2006). Based on the mean z-scores, one repeated measures ANOVA was performed on each data set (naturalness, politeness and dominance) with the two factors – LEVEL and CONTEXT. The two-way interaction between LEVEL and CONTEXT was calculated. For significant values at the 0.05 level, effect sizes were considered; and post-hoc tests were carried out. We found no significant inter-factorial effect of Instruction, i.e. no effect of the testing order of naturalness-politeness or naturalness-dominance ($F_{1, 47} = 1.72, p > 0.05$). Therefore, results from the repeated measures ANOVAs are reported separately for each set of data.
Naturalness

In the data set for naturalness, no main effect was found for LEVEL ($F_{1,81,86.76} = 0.864$, $p > 0.05$) or CONTEXT ($F_{1,48} = 2.671$, $p > 0.05$) and no significant interaction was found between LEVEL and CONTEXT ($F_{3,144} = 1.809$, $p > 0.05$). Nonetheless, there appears to be a tendency for judgments in each context as illustrated in Figure 8-6.

![Naturalness](image)

Figure 8-6: Mean z-scores across all subjects judging for naturalness in two backchannel contexts (context 1: paying attention and context 2: agreement) in the semantic scaling task.

In context 2 (signal of agreement) the higher pitch tended to be perceived as most natural, while this picture is different in context 1 (signal of attention). Both the low end pitch (level 0) and the high end pitch (levels 2 and 3) tended to be perceived as natural. This seems to be compatible with the results from the corpus analysis (see Chapter 4) which show that backchannels signalling attention in Northern Vietnamese are commonly produced low or low falling. However, speakers can also produce a rise on the backchannel tokens when at the same time they request information during telephone openings, yield the floor to conversational partners, or confirm or agree with the information being discussed (Chapter 5).

However, we should keep in mind that while doing the task, subjects might have had different interpretations of the term ‘naturalness’ which may have an effect on their perception. The intended interpretation in this experiment is the appropriateness of the backchannel in context (in terms of whether the hearer in the dialogues has an appropriate reaction to the speaker or whether the former sounds strange in the way s/he reacts to the latter, see Schröder et al. 2006 for perception of affect bursts). Nonetheless, ‘naturalness’ might have been interpreted in terms of the sound quality of the manipulated backchannels.

Politeness

In the data set for politeness, a main effect was found for LEVEL ($F_{2,18,50.2} = 6.36$, $p < 0.05$, $\eta^2 = 0.217$), but not for CONTEXT ($F_{1,23} = 0.25$, $p > 0.05$). The interaction between LEVEL and CONTEXT does not reach significance ($F_{3,69} =1.19$, $p > 0.05$). A post-hoc test shows significant differences at the 0.05 level between level 0 and level 3, and level 1 and level 3, as illustrated in Figure 8-7. This indicates that across both contexts the low end pitches were perceived as significantly more polite than the higher ones. The result does
not confirm Hypothesis II that higher pitch leads to a higher perceived degree of politeness. On the contrary, it shows the opposite tendency: the low ending on Northern Vietnamese backchannels using ờ is perceived as significantly more polite than the high ending.

![Politeness](image)

**Figure 8-7**: Mean z-scores across 24 subjects judging for politeness in two backchannel contexts (context 1: paying attention and context 2: agreement) in the semantic scaling task.

Let us consider the observed tendency in the judgment of naturalness above, where the high ending (level 3) in context 1 was perceived among other levels as most natural. Interestingly, this level is perceived as significantly less polite than lower endings (levels 0 and 1). A high end pitch on backchannels in Northern Vietnamese signalling attention might not sound unnatural. However, it appears to be perceived as less polite. This is an important point for learners of Northern Vietnamese as a foreign language regarding the acquisition of communicative skills and specifically how to sound natural whilst still being polite in speech.

**Dominance**

In the data set for dominance we found a main effect for CONTEXT (F1, 24 = 10.46, p < 0.05, η² = 0.304), but not for LEVEL (F1.72, 41.1 = 1.76, p > 0.05). The interaction between LEVEL and CONTEXT does not reach significance (F3, 72 = 0.789, p > 0.05). The effect of CONTEXT shows that the judgments for dominance were different depending on whether the backchannel was a signal of attention, context 1, or an agreement, context 2, as can be seen in Figure 8-8.
Dominance

Figure 8-8: Mean z-scores across 25 subjects judging for dominance in two backchannel contexts (context 1: paying attention and context 2: agreement) in the semantic scaling task.

All the mean z-scores for context 1 are positive; those for context 2 are negative. That means listeners tended to perceive a higher pitch ending on backchannels signalling attention as more dominant than a higher pitch ending on backchannels signalling agreement. Though the tendency of the judgments in context 2 is not exactly the same as in context 1, the higher the final pitch the more dominant the backchannel sounds across both contexts.

8.5 Interim conclusions and discussion

In this chapter we present how native listeners of Northern Vietnamese interpret the naturalness, politeness and dominance in backchannels when the end pitch varies. As we expected, there is no clear result in the perception of naturalness indicating that listeners judged both a low and a high pitch as equally natural. This is compatible with the results from the corpus analysis (Chapter 4). The perception of politeness and dominance does not confirm the hypotheses: native listeners of Northern Vietnamese did not perceive a low end pitch as less polite. On the contrary, they interpreted the low end pitch in backchannels signalling attention as well as agreement as significantly more polite than the high end pitch.\textsuperscript{11} Although listeners did not reliably interpret the end pitch levels for dominance, there is a tendency for the high ending to be interpreted as more dominant. These results suggest that the pitch reflecting politeness and dominance in Northern Vietnamese may not be derived from the Frequency Code as proposed for a large number of languages (Ohala 1983, Gussenhoven 2004). High pitch seems to be perceived in Northern Vietnamese with emphasis and interpreted as having too much energy or being too emphatic, resulting in a higher degree of perceived dominance and a lower degree of perceived politeness. An

\textsuperscript{11} A production study on the phonetic profile of Korean polite and informal speech registers (Winter & Grawunder, 2012) shows the same tendency as in Northern Vietnamese that Korean speakers lower their voices, among others, to convey politeness.
explanation for this tendency might be culturally motivated. A final high pitch can be linguistically interpreted as a question. When addressing people in a higher position, questioning is generally not very acceptable in Vietnam. The results from the current study are in line with those found in Chen’s (2005) study showing that different speech communities may associate pitch variation with certain meanings in opposite ways supporting a theory of intonational meaning that also considers language-specific perception.

From the interactional point of view, the acknowledgment tokens ờ, ừ or vâng produced with a high/rising pitch can convey (strong) confirmations or agreements as found in the telephone corpus analysis. When producing a high pitch on these tokens, one implies that s/he already knows about the content of the talk (and confirms the information). If this is signalled in the role of a hearer during the talk of the current speaker, i.e. the role of paying attention, it tends to give the impression of being overbearing. As a result, the hearer might often be interpreted as not being polite.

The results from the current study imply possible misunderstandings in intercultural communication, here in speech between Northern Vietnamese speakers and speakers from other languages. The association of a final rise/high pitch with politeness has been found in a number of languages. In British and American English, the high pitch in backchannels is interpreted as more supportive and more interested (Wichmann 2000, Ward 2004). The final rise in Australian English is reported to be a conveyor of politeness, friendliness and attentiveness (Ogden and Routarinne 2005). The results from the current study appear to be the opposite: the low end pitch in Northern Vietnamese is associated with a higher degree of politeness.

Another contribution of the current study is that it provides information for learners of Vietnamese, in particular for speakers of Western Germanic languages. Backchannels in Northern Vietnamese are commonly produced with a low ending. A high ending may appear to be natural but it can be perceived as impolite. Inefficiency and misinterpretation in speech with native speakers might happen, if non-natives transfer the prosodic patterns from their first language (L1) to Northern Vietnamese speech (L2). The way of giving listener feedback politely in a large number of languages is to raise the pitch, whereas in Northern Vietnamese it is the low pitch that is perceived as more polite.

Limitations need to be mentioned. The present study does not consider duration of backchannels, a parameter which presumably plays a role in the perception of the investigated meanings. The target word vâng as a backchannel used to address elders, as well as the position of the backchannels in different discoursal contexts (before/during/after a turn or an activity) should also be considered. It will be necessary to test subjects with various sociocultural backgrounds to ascertain whether the results found in this study can be generalised to other groups of subjects.
In this section the contents of the thesis are summarised. The introduction in Chapter 1 describes the motivation and research questions of the study. In addition to a large number of particles, Northern Vietnamese has a complex lexical tone system which is characterised by pitch and voice quality. The thesis addresses the following questions: i) which functions at the utterance level can prosody (in terms of pitch) convey in the language? ii) to what extent does it manifest itself, given that every single syllable/word in the language already has a lexical pitch?

Chapter 2 provides an overview of analytical frameworks including a) Conversation Analysis, b) Contextualisation Theory and c) the interactional and experimental approaches of investigating intonation. To investigate the form and function of intonational contours in discourse, a corpus of naturally occurring telephone conversations in Northern Vietnamese is analysed. The verification of results is achieved through the analysis of sequential organisation of talk, deviant cases and additional speech materials (controlled dialogues). The interactional approach serves as the basis for the telephone corpus analysis, whereas the experimental approach provides theoretical background for the perception experiment. Before presenting the results in the next chapters, an overview of the lexical tones as well as of the intonation research to date in Vietnamese is given. Here a simplified representation of the six lexical tones in Northern Vietnamese is proposed.

Chapter 3 describes the telephone corpus and the methods used to record, transcribe, annotate and analyse the data. The characteristics of the target utterances under investigation are given here including segmental composition, lexical tones and glosses. In Chapters 4 and 5 we provided evidence that pragmatic functions at the utterance level and lexical meanings of words can be conveyed simultaneously by the same means (pitch). Low/falling pitch contours are found on the utterances ờ, ừ and vâng (‘yes’) to convey the context of backchannel. High/rising pitch contours are found on the same words to convey the contexts of turn-yielding, request for information, and agreement/confirmation.

Chapter 6 deals with repair initiation, the second main phenomenon investigated in the thesis. In this context we found a high/rising pitch contour on the utterances hà, đi, đa, ai, sáo and ỏi (‘yes/no question particle’, ‘what’, ‘politeness particle’, ‘who’, ‘why/how’, and ‘vocative particle’, respectively) to convey mishearing or problems understanding. Evidence for these intonational contours was provided by the analysis of the non-lexical item mm occurring in the above contexts.

Chapter 7 provides an analysis of the interaction between lexical tone and intonation. The intonational contours, where they occur, can completely override the lexical tone of the words investigated as analysed in backchannels, other interactional contexts (turn-yielding, request for information, and agreement/confirmation) and repair initiations in fast speech. In these contexts, only intonational tones are realised. Intonational tones can also co-occur with lexical tones, in which case the lexical tone is followed by the intonational tone. This is the case in repair initiations in slow or careful speech.
Chapter 8 looked at intonation from the point of view of perception. We found that native listeners of Northern Vietnamese do have affective/paralinguistic interpretations when the final pitch levels on backchannels are modified. They interpret a low final pitch in this context as more polite and less dominant than a high final pitch. This result indicates that Northern Vietnamese is an exception to the intonational universals proposed for a large number of languages as derived from the Frequency Code (Ohala 1983, Gussenhoven 2004).

The present study contributes to research on intonation, especially the intonation of Vietnamese, in several ways. First, we explored interactive linguistic and affective functions of intonation in Northern Vietnamese within two theoretical and methodological approaches: a) Conversation Analysis – (sequential) analysis of naturally occurring data, and b) Laboratory Phonology – acoustic analysis and perception experiments. This combination not only ensures the validity of the phenomena investigated, but also provides a qualitative and quantitative analysis of these phenomena.

Second, we provide a model analysing the interaction between intonation and lexical tone within Autosegmental Phonology, a theory which has successfully accounted for tonal phenomena in a variety of languages, ranging from intonation languages (e.g. English, German) and pitch-accent languages (e.g. Japanese, Serbo-Croatian) to tone languages with simple tone systems (e.g. Chichewa, Hausa) and more complex ones (e.g. Mandarin Chinese, Cantonese).

Finally, results shed light on within-language functions of intonation in Vietnamese (e.g. backchannels versus requests for information), on the one hand, and provide possible accounts for cross-language perception of paralinguistic intonational meanings, on the other. While in a number of languages high pitch is associated with politeness and low pitch with dominance, in (Northern) Vietnamese the perception of these intonational patterns appears to be reversed indicating a language-specific aspect of the relation between intonational form and paralinguistic meaning.
References


Caspers, Johanneke, 2000, Melodic characteristics of backchannels in Dutch map task dialogues. InterspeechISCA 2:611-614.


———1996, Towards an interactional perspective on prosody and a prosodic perspective on interaction. In Couper-Kuhlen, Elisabeth and Selting, Margret, eds. 11-56.


Ding, Hongwei and Jokisch, Oliver and Hoffmann, Rüdiger, 2010, Perception and Production of Mandarin Tones by German Speakers. Speech Prosody, Chicago.


References


_____2010, Prosody in interaction: State of the art. In Bart-Weingarten, Dagmar et al., eds. 3-40.


Selting, Margret et al., 2009, Gesprächsanalytisches Transkriptionssystem 2 (GAT2).


Appendix A

An overview of Northern Vietnamese lexical tones with traditional names and alphanumeric codes

<table>
<thead>
<tr>
<th>Stylised contour</th>
<th>Citation form (description of tone)</th>
<th>Traditional name</th>
<th>Alphanumeric code</th>
<th>Diacritics in orthography</th>
<th>Example</th>
<th>Gloss</th>
</tr>
</thead>
<tbody>
<tr>
<td>••••••••</td>
<td>high-level/mid-level</td>
<td>ngang</td>
<td>A1</td>
<td>none</td>
<td>la</td>
<td>bawl</td>
</tr>
<tr>
<td>••••••••</td>
<td>low-level/low-falling</td>
<td>huyền</td>
<td>A2</td>
<td>`</td>
<td>là</td>
<td>to be</td>
</tr>
<tr>
<td>••••••••</td>
<td>rising</td>
<td>sắc</td>
<td>B1</td>
<td>´</td>
<td>là</td>
<td>leaf</td>
</tr>
<tr>
<td>••••••••</td>
<td>falling-glottalised</td>
<td>nảng</td>
<td>B2</td>
<td>.</td>
<td>là</td>
<td>strange</td>
</tr>
<tr>
<td>••••••••</td>
<td>falling (-rising)</td>
<td>hỏi</td>
<td>C1</td>
<td>´</td>
<td>là</td>
<td>exhausted</td>
</tr>
<tr>
<td>••••••••</td>
<td>rising-glottalised</td>
<td>ngã</td>
<td>C2</td>
<td>~</td>
<td>là</td>
<td>unclean</td>
</tr>
</tbody>
</table>
Appendix B

Transcription conventions

[ speech overlaps
[ quick and direct connection of new turns or units
( ) micro pause
( ), (--) , (---) short pause, long pause, longer pause, from approx. 0.25-0.75, to approx. 1 second
; , ::, :::: lengthening, according to duration
((laugh)) description of laughing, explanation of the context or events
((…)) omission in the transcript
( ) incomprehensible passages according to duration
⇒ investigated line in the transcript
ACCENT accented token

Final pitch movement
? high rising
, mid rising
- level
. (deep) falling

Noticeable pitch skips
↑ up
down

Shifted pitch register
<<t> > low pitch register

Loudness and speech tempo
<<f> > forte, loud
<<p> > piano, quiet
<<all> > allegro, fast