



Intonational phrase marking in Southern Vietnamese

Marc Brunelle¹

¹ Department of Linguistics, University of Ottawa, Canada

marc.brunelle@uottawa.ca

Abstract

Earlier studies of Vietnamese prosody, which, to a single exception, looked at the Northern dialect, portray a language in which intonation is weak, variable and idiosyncratic. The intonational targets typically proposed in autosegmental-metrical models (boundary tones, pitch-accents) have remained elusive. This may very well be due to the fact that almost all previous experimental work was conducted on relatively unnatural read speech; in fact, research conducted on natural conversations has revealed that intonation may surface in peripheral areas of grammar where lexical tone plays a low functional role, like backchannels and repair utterances. In this short paper, an eight-hour corpus of Southern Vietnamese spontaneous speech is used as a basis for investigation. I first show that, just as in Northern Vietnamese, there is little evidence for strongly phonologized boundary tones in Southern Vietnamese. I then put forward evidence that although an effect of intonation is seen in discursive monosyllabic utterances, this effect seems more optional and less categorical than in Northern Vietnamese.

Index Terms: tone, intonation, discourse markers, acoustics, final particles, Vietnamese

1. Introduction

Vietnamese is a language in which each syllable has its own underlying lexical tone. It has no tone reduction rules, which means that even on the surface, a Vietnamese sentence is heavily specified for tones. This high tone density is further complicated by the large tone inventory of the language. For instance, the standard dialect, which is based on the Hanoi variety, features six contrastive tones that combine pitch and voice quality contrasts [1-7]. This raises the old question of the cohabitation of intonation and tone in tonal languages [8].

Vietnamese dialects can be fairly divergent from one another, but to the exception of some central dialects, they nonetheless maintain a high level of mutual intelligibility. In this paper, I will compare new evidence on Southern Vietnamese intonation with the results of previous research on Northern Vietnamese intonation. Although the two dialects exhibit significant segmental differences [9] and have different tone inventories (cf. §2), they share similar syllable structures and basic phonotactics. Intonational differences between the two varieties are possible, but if they exist, they are inconspicuous enough that they have never been reported in the literature and are not apparent to naïve speakers.

So far, the investigation of Vietnamese intonation has mainly focused on the phonetic properties that distinguish communicative and expressive functions. Studies seem to converge on two descriptive generalizations: 1) intonation is realized through a combination of the overall pitch register of sentences and of phrase-final targets (which previous authors rarely interpret as boundary tones) [10-14], and 2) the phonetic

cues used to mark intonation are *f*₀, intensity, voice quality and duration [13-18]. However, an experiment conducted on controlled Northern Vietnamese sentences has recently shown that there is substantial inter-speaker variation in intonation: speakers exhibit variable (if weak) intonational strategies, but none of these strategies is shared by all speakers [12, 19]. This suggests that intonational targets may not be phonologized, perhaps because lexical tones monopolize pitch and leave little room for intonation. In fact, the high prevalence of sentence-final particles in Vietnamese is often considered connected to the absence of strongly conventionalized intonational cues, even if this is rarely explicitly articulated [4].

On the other end, it has been shown that in Northern Vietnamese telephone conversations, intonation on discourse markers (such as the backchannels *ờ* and *ừ* and the repair utterances *đạ* and *vâng*, which are all forms of acknowledgement) can obscure lexical tone targets, and that some tones are more affected than others [20-22]. In slow speech, the lexical tone of discourse markers is concatenated with an intonational boundary tone, and the effect of intonation seems to be even more pronounced in fast speech, where lexical tones seem largely lost. One could interpret this as evidence for covert intonational targets: when the lexical identity of a word matters, as in regular sentences, the lexical tone wins out. However, when a word is just a semantically bleached discourse marker, its lexical tone loses its importance and the effect of intonation is allowed to surface.

In this paper, a corpus of spontaneous Southern Vietnamese speech is used to address three research questions:

- 1) Is there evidence for phonologized phrase-final intonational targets in spontaneous Southern Vietnamese?
- 2) Are intonational targets clearly marked in the absence of a phrase-final particle?
- 3) Is there any evidence that a “covert” intonation surfaces in discourse marking monosyllabic utterances (as in Northern Vietnamese)?

2. Methods

The acoustic results reported in this study were extracted from a corpus of Southern Vietnamese speech. Southern Vietnamese was chosen because it is an easily accessible and fairly homogeneous dialect whose tones, contrary to those of the standard Northern dialect, do not have voice quality contrasts, which greatly facilitates automatic pitch tracking. It has seven tones, as shown in Figure 1, but two are restricted to checked syllables (syllables closed by an obstruent) and could be considered allophones of other tones. Impressionistically, final particles also seem less common in Southern Vietnamese, which could force the realization of communicative functions by intonational means.

For lack of space, statistical models will not be presented in this 4-page paper. Mixed models and Kolmogorov-Smirnov tests fitted on the data will be reported in further publications.

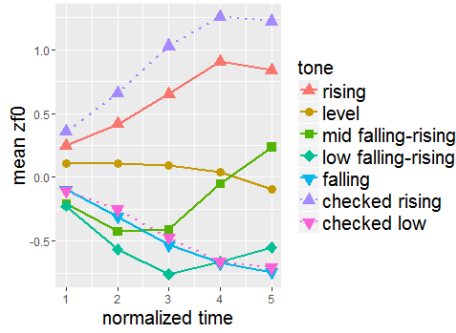


Figure 1: The tones of Southern Vietnamese (speaker- and time-normalized)

2.1. The corpus

A corpus of Southern Vietnamese speech was collected in Hồ Chí Minh City in 2013-2014 with speakers originally from Hồ Chí Minh City or the Mekong Delta. It consists of:

- Two televised interviews between a show host and a singer (3 speakers, all in their 20s). The style of these interviews is spontaneous, but semi-formal.
- Two televised comedy skits (8 speakers, apparently ranging from their 20s to their 50s). The intonation used in these recordings is natural, but not spontaneous and somewhat exaggerated.
- Four spontaneous conversations between same-age and same-sex friends (two men in their 20s, two women in their 20s, two men in their 60s and two women in their 50s). These recordings sound very natural.

Each of these eight recordings lasts about one hour. The recordings add up to 472 minutes of speech. There is a total of 86,232 syllables grouped into 26,160 intonational phrases in the corpus. The corpus was entirely transcribed and annotated by a team of native Vietnamese speakers (all students or researchers in linguistics) and by the author. Four types of annotated information are relevant for this paper:

- The lexical tone and identity of each syllable.
- Intonational phrase boundaries. Intonational phrases are defined as syntactic clauses, following most models of the prosodic hierarchy [23, 24]. However, several syntactic clauses were considered to be merged into a single intonational phrase when there was no pause, final lengthening or intonational event marking the presence of a boundary between them (rare). On the opposite, clauses interrupted by pauses or hesitations were treated as several intonational phrases (common).
- The presence or absence of a final particle at the end of intonational phrases
- The modality or communicative function of intonational phrases (IPtypes). The types that are sufficiently frequent to be reported here are continuative, declarative, imperative, exclamative and interrogative. Three types of interrogatives were originally distinguished (alternative, yes-no and open questions), but they have been merged because of their lack of significant differences. Phrases containing words under focus are not included here as our current typology of focus is not reliable enough (but they seem to have a boosted f0). Monosyllabic utterances used as discourse markers were classified following their own set of communicative functions: acknowledgment, polite acknowledgement, yes answer, backgrounding of previous information, emphatic agreement, incredulity, backchannels and topic closure.

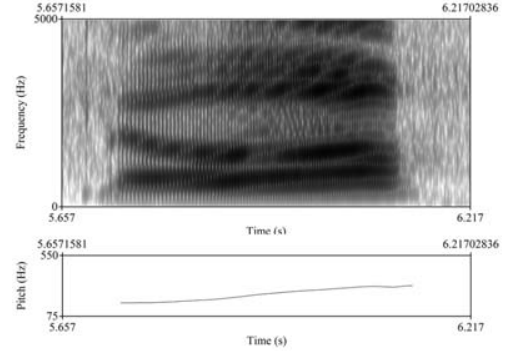


Figure 2: Exclamative intonation on the word *được* [d̪ək̚] 'to be able to', that should bear a checked low tone (female speaker in her 50s)

2.2. Acoustic analysis

The recordings were measured automatically with Praat scripts. The main indicator reported here is f0 as measured over five equidistant points in the voiced portion of syllables. In order to facilitate comparisons across speakers, f0 measurements were z-normalized over the entire corpus (results are therefore reported in zF0). Durational results are not systematically reported because duration does not appear to distinguish phrase modalities (although phrase-final lengthening and lengthening of syllables under focus do seem fairly systematic). Intensity is left out because loudness changes make it difficult to control in our spontaneous recordings.

3. Results

It is easy to find dramatic examples of intonation overriding lexical tones in the corpus. A non-ambiguous one is given in Figure 2: the word *được* [d̪ək̚] 'to be able to' should bear a low checked tone, but this tone is replaced with a rising intonation in an exclamative one-word utterance. Even more common is the presence of weak phrase-final rises or falls that do not obscure the patterns of contrasts of lexical tones, but slightly modify their contours. There is thus no doubt that intonational effects are frequent in the language. However, is this intonational variation conventionalized or is variation so idiosyncratic and contextual that no clear patterns emerge?

3.1. Intonation in phrase-final syllables

A look at tone contours on syllables inside (right panel of Fig. 3) and at the end of intonational phrases (left panel of Fig. 3) reveals that there are no important differences in their heights, shapes or ranges of variation. The only notable difference is that

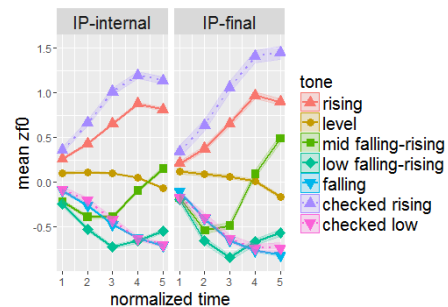


Figure 3: Realization of the tones inside and at the end of intonational phrases (speaker- and time-normalized). Error bars report the 95% confidence interval.

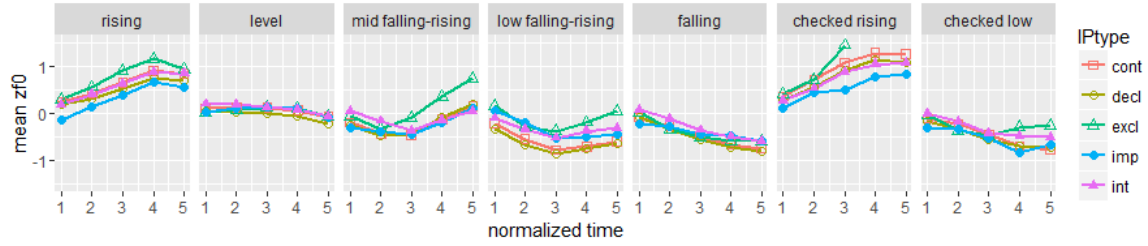


Figure 4: Realization of the tones in the final syllable of different types of intonational phrases (final particles excluded). Mean values are only reported when $n > 20$.

the overall f_0 range at tone offsets is a little wider at the end of intonational phrases. This is not caused by an intonational effect, but is rather due to the greater duration of syllables in phrase-final than phrase-internal positions (z-normalized duration of 0.66 vs. -0.28), which allows a fuller realization of f_0 contours. Note, however, that Figure 3 does not break down phrases by intonational pattern.

Figure 4 therefore reports the effect of modality on each of the seven tones on the last syllable of intonational phrases, after excluding final particles. Categories (or sampling points) with less than 20 tokens are excluded. We see that overall, intonational types have a very limited effect on tone contour or height, suggesting that there is no single conventionalized strategy to realize intonation on the last syllable of intonational phrases. The only systematic pattern that seems to emerge is that the final syllables of exclamative phrases (transparent triangles) tend to have a higher f_0 for most tones (focal phrases, which are not reported here, seem to obey the same tendency).

Although this short paper only focuses on phrase-final effects, there is no clear evidence that the f_0 range of entire phrases is shifted upwards or downwards to realize communicative functions. An exception, is that the f_0 raising found in the last syllable of exclamative and focal phrases seem to be attested phrase-medially as well, but this will be reported elsewhere.

3.2. Intonation in discourse markers

Is there as little intonational variation in discourse markers as there is on the final syllable of longer utterances? In order to test this out, the three Southern Vietnamese monosyllabic discourse markers ∂ [ɤ], \dot{u} [i] and $d\dot{a}$ [ja], which are the only ones that appear often enough to allow an analysis (246, 182, 152 occurrences, respectively), were extracted from the corpus and investigated in more detail. The former two, which are produced with falling tones in citation form, are informal and could roughly be translated as “hmm”, while the latter, which has a low falling-rising tone in citation form, is a more formal and polite form of acknowledgement. Results are reported in

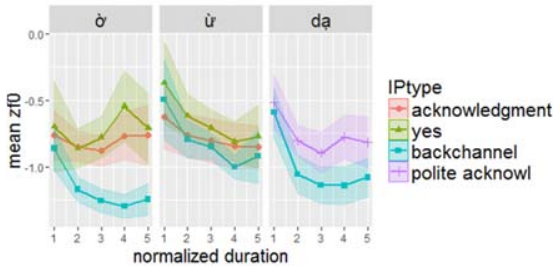


Figure 5: The tone contours of ∂ , \dot{u} and $d\dot{a}$ for IP types with more than 20 tokens. Error bars represent the 95% confidence interval for each category.

Figures 5 and 6, classified according to the functions listed in §2.1.

Figure 5 gives the f_0 contours of the three particles under study when produced with the four most common communicative functions. While the f_0 contour of \dot{u} is not very variable and matches the contour of the falling tone in citation form, the two other particles are more variable. The particle ∂ is not as falling as expected when used for acknowledgement and as a “yes” answer, but it still falls close to the normal range of realization of the falling tone. The contour of the particle $d\dot{a}$ in polite acknowledgement is also similar to that of a typical low falling-rising tone. However, the two particles seem to share a fairly low realization in backchannel utterances (blue squares).

Another way of considering possible intonational variation in backchannel particles is to look at the distribution of their onsets and endpoints when used with various communicative functions (Figure 6). Overall, we see that no symbol cloud is circumscribed to a specific area of the chart, which means that communicative functions are not categorically distinguished by their height or slope. Two tendencies nonetheless emerge. First, a high proportion of monosyllabic utterances that convey emphatic agreement (full green diamonds) are realized with a high endpoint. This is reminiscent of the raised f_0 found on the final syllables in exclamative phrases. Second, most monosyllabic utterances used as backchannels (empty blue circles) are produced with a low endpoint, as in Figure 5. More generally, and despite the lack of categorical groupings, a close inspection of individual monosyllabic utterances in Figure 6 reveals that, even if this seems to be less systematic and structured than in Northern Vietnamese [16-18], the expected

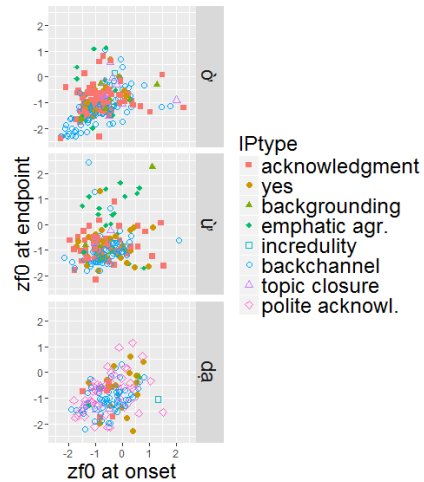


Figure 6: Distribution of ∂ , \dot{u} and $d\dot{a}$, by communicative function

falling slope of the tones of the three particles is often overridden by intonation: the tokens located above a diagonal line running from the bottom left to the upper right corner could all be described as having a rising f0 slope.

4. Discussion

The results obtained in this study hint that there are no clear conventionalized strategies for the intonational marking of phrase modalities and communicative functions in Southern Vietnamese. This is evidenced by the absence of marked and/or categorical differences between the contours of tones realized on the final syllable of intonational phrases with different modalities (Fig. 4). One difference that seems to emerge is that final syllables have a slightly higher f0 in exclamative mode, but this does not seem to affect the shape or direction of tone contours or to compromise tonal identity. It would therefore be difficult to formalize this as a boundary tone. Furthermore, no evidence was found that intonation is more pronounced in phrases without final particles, suggesting that there is no functional trade-off between intonation and final particles.

Relatively clear intonational patterns were expected to surface in Southern Vietnamese monosyllabic utterances in the same fashion as in Northern Vietnamese [16-18]. Results partly confirm these expectations. Although the intonation of monosyllabic utterances cannot be categorically classified by communicative function, they tend to have a lower f0 when used as backchannels and a high f0 when used for emphatic agreement. The expected falling tonal contour of a significant minority of monosyllabic utterances used for emphatic agreement even seems to be overridden by a rising intonation. One could speculate that the main difference between Northern and Southern Vietnamese monosyllabic utterances is not the nature of the intonational marking of their communicative functions, but the fact that this marking is more optional in Southern Vietnamese.

The picture that emerges from these results is that Southern Vietnamese does have intonation, but that its intonational patterns are not very phonologized, i.e. cannot be classified into discrete grammatical units with categorical realizations and clear semantics. This seems in line with previous results on Northern Vietnamese [12]. However, even if Vietnamese intonation is not very phonologized, there is still good evidence that it is structured in two ways: individual speakers seem to have their own ways of realizing communicative functions or modality [19] and/or the optional intonational patterns that are often found in individual utterances seem to reflect the emergence of universal biological codes [25-27]. More specifically, the relatively higher f0 associated with exclamatives, emphatic agreement and focus, and the relatively lower f0 associated with backchannels could be a reflection of the effort code, which links emphasis with greater vocal effort.

That said, the apparent lack of phonologized intonation in (Southern) Vietnamese does not mean that there are no grammaticalized ways of conveying communicative functions and phrasal modality in this language: if we were to adopt the view that intonation is not strictly tonal, final particles could even count as a highly conventionalized way to convey intonation [28].

5. Conclusions

Although there are numerous instances of intonation being superimposed over lexical tones in Southern Vietnamese, this study of a corpus of spontaneous speech confirms previous reports obtained from read Northern Vietnamese speech that there is little evidence that the language uses discrete and

categorical intonational targets at the end of intonational phrases. The two systematic intonational patterns uncovered in this study, a higher f0 in the final syllable of exclamatives and emphatics, and a lower f0 in backchannel utterances, could both be attributed to a weak effect of a universal effort code [23-24] rather than to phonological targets. Furthermore, these intonational effects are of an order of magnitude that seems too weak to be attributed to conventionalized intonational targets (such as boundary tones). This strengthens the hypothesis that the intonation of Vietnamese has at best reached a low level of phonologization [12].

6. Acknowledgements

Special thanks to Phạm Thị Thu Hà for her help in coordinating and overseeing the annotation of the corpus, to three anonymous reviewers and to TAL participants for their feedback and comments. This project was made possible by a research grant from the Social Science and Humanities Research Council of Canada (435-2012-0468).

7. References

1. Vũ, T.P., *The Acoustic and Perceptual Nature of Tone in Vietnamese*. Ph.D. Australian National University, 1981.
2. Brunelle, M., D.D. Nguyễn, and K.H. Nguyễn, *A Laryngographic and Laryngoscopic Study of Northern Vietnamese Tones*. *Phonetica*, 2010. **67**(3): p. 147-169.
3. Brunelle, M., *Tone perception in Northern and Southern Vietnamese*. *Journal of Phonetics*, 2009. **37**: p. 79-96.
4. Earle, M.A., *An Acoustic Phonetic Study of Northern Vietnamese Tones*. Speech Communications Research Laboratory Monograph. Vol. 11. 1975, Santa Barbara: Speech Communications Research Laboratory Inc. 214.
5. Michaud, A., *Final Consonants and Glottalization: New Perspectives from Hanoi Vietnamese*. *Phonetica*, 2004. **61**: p. 119-146.
6. Phạm, A.H., *Vietnamese Tone: Tone is not pitch*. University of Toronto, 2001.
7. Kirby, J., *Dialect experience in Vietnamese tone perception*. *Journal of the Acoustical Society of America*, 2010. **127**(6): p. 3749-3757.
8. Chao, Y.R., *Tone and intonation in Chinese*. *Bulletin of the Institute of History and Philology, Academia Sinica*, 1933. **4**: p. 121-134.
9. Brunelle, M., *Vietnamese (Tiếng Việt)*, in *The Handbook of Austroasiatic Languages*, M. Jenny and P. Sidwell, Editors. 2015, Brill. p. 909-954.
10. Hoàng, C.C., *Bước đầu nhận xét về đặc điểm ngữ điệu tiếng Việt (trên cơ liệu thực nghiệm)*. *Ngôn Ngữ*, 1985(3): p. 40-49.
11. Trần, H.M., *Tones and Intonation in South Vietnamese*, in *Series A - Occasional Papers #9, Papers in Southeast Asian Linguistics No.1*, Đ.L. Nguyễn, H.M. Trần, and D. Dellinger, Editors. 1967, Linguistics Circle of Canberra: Canberra.
12. Brunelle, M., K.P. Hạ, and M. Grice, *Intonation in Northern Vietnamese*. *The Linguistic Review*, 2012. **29**(1): p. 3-36.
13. Nguyễn, T.T.H. and G. Boulakia, *Another look at Vietnamese intonation*. *Proceedings of the XIVth International Congress of Phonetic Sciences*, 1999: p. 2399-2402.
14. Đỗ, T.D., T.H. Trần, and G. Boulakia, *Intonation in Vietnamese*, in *Intonation systems: A Survey of Twenty Languages*, D. Hirst and A.D. Cristo, Editors. 1998, Cambridge University Press: Cambridge. p. 395-416.

15. Vũ, M.Q., Đ.Đ. Trần, and É. Castelli, *Intonation des phrases interrogatives et affirmatives en langue vietnamienne*. Journées d'Étude de la Parole, 2006: p. 4.
16. Michaud, A., *Prosodie de langues à tons (naxi et vietnamien), prosodie de l'anglais : éclairages croisés*. Ph.D. Paris 3 - Sorbonne Nouvelle, 2005.
17. Trần, Đ.Đ., *Synthèse de la parole à partir du texte en langue vietnamienne*. Ph.D. Institut national polytechnique de Grenoble, 2007.
18. Mạc, Đ.K., *Génération de parole expressive dans le cas des langues à tons*. Ph.D. Grenoble INP, 2012.
19. Cangemi, F., et al., *Individual specificity, redundancy and the evolution of phonological systems: Intonation in a tone language*, in *LabPhon 15*. 2016: Ithaca.
20. Hà, K.P. and M. Grice. *Modelling the Interaction of Intonation and Lexical Tone in Vietnamese*. in *Speech Prosody 2010*. 2010. Chicago.
21. Hà, K.P., *Prosody in Vietnamese: Intonational Form and Function of Short Utterances in Conversation*. Ph.D. University of Cologne, 2012.
22. Hà, K.P., *Prosody of Vietnamese from an interactional perspective: ở, ừ and vâng in backchannels and requests for information*. *Journal of the Southeast Asian Linguistics Society*, 2010. **3**(1): p. 56-76.
23. Nespor, M. and I. Vogel, *Prosodic Phonology*. 1986, Dordrecht/Riverton: Foris Publications.
24. Selkirk, E., *The Syntax-Phonology Interface*, in *The Handbook of Phonological Theory*, J. Goldsmith, J. Riggle, and A.C.L. Yu, Editors. 2011, Blackwell. p. 435-484.
25. Ohala, J., *The frequency code underlies the sound-symbolic use of voice pitch*, in *Sound Symbolism*, L. Hinton, J. Nichols, and J. Ohala, Editors. 1994, Cambridge University Press: New York. p. 325-347.
26. Chen, A., C. Gussenhoven, and T. Rietveld, *Language-specificity in the perception of paralinguistic intonational meaning*. *Language and Speech*, 2004. **47**(4): p. 311-349.
27. Gussenhoven, C., *The Phonology of Tone and Intonation*. 2004, Cambridge: Cambridge University Press. 355.
28. Hyman, L.M. and K.C. Monaka, *Tonal and non-tonal intonation in Shekgalagari*, in *Prosodic categories: production, perception and comprehension*, S. Frota, G. Elordieta, and P. Prieto, Editors. 2011, Springer. p. 267-289.