Investigating the variation of intonation contours in Northern Vietnamese tones

Annika Tjuka[1]*, Huong Thi Thu Nguyen[2], Ruben van de Vijver[3] and Katharina Spalek[3]

[1]Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany, [2]Barnim-Gymnasium, Berlin, Germany, [3]Heinrich-Heine-Universität Düsseldorf, Düsseldorf, Germany

Intonation is an instrument for structuring discourse and emphasizing different types of information. In German, for example, pitch is used to highlight focus, while in Vietnamese, different pitch contours distinguish lexical tones. As of yet, the interplay between intonation and lexical tone in relation to information structure has not been sufficiently investigated across languages. Vietnamese has six lexical tones and is particularly interesting for investigating the influence of different intonation strategies on the realization of tones. Here, we present a production study with 70 Northern Vietnamese speakers. The participants read six sentences under two conditions. In each sentence, a word occurring in the final position of the sentence and carrying one of the six tones was pronounced in two different discourse contexts. Acoustic analyses of the intonation contours showed that Vietnamese speakers realized the words with significant differences in pitch at the onset. Yet, the strategies for raising or lowering the pitch varied depending on the tone. Our results show the use of prosodic cues in a complex tone system across a large number of speakers. In addition, the study can serve as a starting point for educational programs that include training on intonation patterns in specific contexts.

KEYWORDS

intonation, pitch, lexical tone, Vietnamese, information structure

Introduction

Intonation is central to the study of information structure in the world's languages. Changes in pitch distinguish interrogative sentences from assertions or new from given information in discourse. Intonation plays a role in emphasizing new or contrastive information, i.e., focus (Baumann et al., 2007; Gussenhoven, 2007; Peters et al., 2014). In non-tonal languages, like German, English, or Dutch, focus is usually marked by an increase in either the maximum F_0 or the F_0 range, a longer duration of the accented syllable in the focalized word, and sometimes also by an increase in intensity. In contrast, tonal languages such as Chinese or Vietnamese use changes in pitch contours to distinguish lexical tones and additionally, to structure information. For example, speakers of Mandarin Chinese realize focus in a non-sentence-final position by a rise in pitch for the focused word and a fall or compression after the focus (Xu, 1999). Studies on Vietnamese have shown that in addition to pitch, duration and intensity are also used to mark focus (Jannedy, 2007; Brunelle et al., 2012; Michaud and Brunelle, 2016), with variation between low and high tones and variation across speakers (Brunelle, 2017). Thus, tonal languages show a complex interplay between lexical tone and intonation (for an overview, see Gussenhoven 2004).

With its complex tone system of six lexical tones, Vietnamese is ideal for studying the realization of lexical tones in different discourse contexts. So far, there is no comprehensive study with a large number of speakers analyzing the variation of intonation contours within a tonal language. This article presents the first production study of F_0 contours for each of the six tones produced by 70 Northern Vietnamese speakers in two different discourse contexts. The large number of speakers contributes further methodological rigor compared to previous studies that relied on a few speakers. Comparing more speakers also makes it possible to analyze individual differences, which were reported in previous studies but not tested systematically due to the small number of speakers. In addition, the study can serve as a starting point for educational programs that include training on intonation patterns in specific contexts.

Pragmatic functions and intonation in tonal languages

Tonal languages use different prosodic cues to structure information. For example, duration and intensity mark focus in Mandarin Chinese, while the role of pitch seems to be more intricate (Ouyang and Kaiser, 2015). The two functions of the pitch contours for distinguishing tones and focus are difficult to separate in an acoustic analysis. Ouyang and Kaiser (2015) recorded ten speakers (five men, five women) and showed that contrastive (i.e., corrective) information was characterized by a change in pitch, duration, and intensity while introducing new information showed less change in pitch or duration and no change in intensity (Ouyang and Kaiser, 2015). Pitch ranges in the contrastive focus condition were extended for the minimum and maximum bound, but the difference in high vs. low tone was not analyzed. Lowering pitch for low tones may result in a creaky voice which makes the speaker sound raspy and is therefore avoided (Chen and Gussenhoven, 2008). Kammu, an Austroasiatic language spoken in Laos and some parts of Vietnam, Thailand, and China, shows another pattern in focus marking with intonation. The Western Kammu dialects have developed a tone system with two lexical tones, whereas Eastern Kammu does not use tone to distinguish lexical meaning (Karlsson et al., 2012). Recordings of ten speakers (seven men, three women) for the non-tonal dialect and 14 speakers (six men, eight women) for the tonal dialect were analyzed. The comparison of focus marking in the non-tonal vs. tonal dialect revealed that focus is marked with a rising intonation contour. However, the lexical tone affects the realization of focus marking in Northern Kammu by neutralizing the pitch rise used for focus intonation in words with a low tone. There is a hierarchy that speakers of the tonal dialect use to maintain lexical tone before marking phrase-final boundary tone and focus (Karlsson et al., 2012). The falling lexical tone was in contrast with the rising intonation contour that marks focus and thus, speakers neutralized or used an even lower pitch range to mark focus for the low tone.

TABLE 1 The Vietnamese tone system with six tones realized in the standard Northern Vietnamese dialect.

Tone	Description	Example	Translation
sắc	High-rising	*má*	Mother
ngã	High-falling-glottal	*mã*	Code
ngang	Mid-level	*ma*	Ghost
huyền	Low-falling	*mà*	Nevertheless, but
hỏi	Low-rising	*mả*	Tomb, grave
nặng	Low-falling-glottal	*mạ*	Rice seedling

Pragmatic functions and intonation in Vietnamese

Vietnamese belongs to the Austroasiatic language family and there are three dialect groups: Northern Vietnamese, Southern Vietnamese, and Central Vietnamese (Vũ, 1982; Hoàng, 1989). One of the main differences between the dialects is that they vary in tone inventory (Brunelle, 2009). Since the present article investigates intonation contours in Northern Vietnamese, the standard variety, we concentrate on the six-tone system. Tones in Northern Vietnamese are expressed by combining pitch and voice quality (see Table 1). There are three high tones—*sắc*, *ngã*, *ngang*—and three low tones—*huyền*, *hỏi*, *nặng* (for a detailed description of tone perception and production, see Brunelle 2009; Brunelle et al. 2010; Brunelle and Jannedy 2013). The tones are indicated by a diacritic above the vowel.

In addition to a complex tone system, Vietnamese speakers use intonation to mark different pragmatic contexts (Thompson, 1965). Different sentence types—declarative, interrogative, and imperative—are distinguished by changes in global F_0 contour, syllable length, and intensity (Đỗ et al., 1998). Hạ (2012) investigated short utterances in a corpus of telephone calls by 43 Northern Vietnamese participants (20 men, 23 women). When using discourse particles in certain contexts such as back-channels and turn-yielding, the lexical tone was overridden by the intonation (Hạ, 2012). However, there is no systematic realization of intonation for specific pragmatic contexts across speakers (Hạ, 2012). Based on an analysis of 16 speakers (seven men, nine women), Brunelle et al. (2012) showed that prosodic cues are used to express the particle *không* 'empty, no, only" in different contexts, but there was an inter-speaker variation due to speaker-specific strategies for using pitch. Thus, intonation contours may not be fully grammaticalized in the Northern Vietnamese dialect.

To mark focus, Vietnamese speakers can use focus particles or intonation (Michaud and Brunelle, 2016). The particles *thậm chí* "even," *chỉ* "only," and *cả* "also" function as syntactic markers and are used systematically to indicate focus (Hole, 2008, 2013; Erlewine, 2017). Intonation, while less systematic, has been described to occur for certain types of focus marking. In a question-answer paradigm, Jannedy (2007) found that two speakers of Northern Vietnamese (one man, one woman) used different intonation contours depending on the position of the focused element. In subject- and verb-focus utterances, a rise in pitch occurred sooner

than in sentential- and object-focus sentences. The focused element was accentuated and lengthened. Furthermore, the participants were able to correctly associate intonation patterns with the respective question, indicating that prosodic cues are used more systematically than expected (Jannedy, 2007). Based on an analysis of four speakers (3 men, 1 woman), another study showed that words with tone *nặng* received a rising pitch contour for emphasis, but duration varied across speakers (Michaud and Vu-Ngoc, 2004). In contrast, Miller et al. (2015) found no changes in pitch and phonation in a new information focus vs. non-focus condition for tone *sắc* and tone *ngã* across nine speakers (two men, seven women), but the tones were expressed with a change in duration and spectral energy to mark focus. In a processing study with Vietnamese speakers, Tjuka et al. (2020) showed that changes in intonation contours for focus marking enhance the recall of focus alternatives (for similar results with German speakers, see Koch and Spalek, 2021).

Studies investigating the interplay between intonation, lexical tone, and information structure in tonal languages vary greatly in scope. In addition, they discuss different types of discourse contexts. For studies on focus, the position in which the focused element occurs varies from study to study, some studies use individual sentences and others question-answer paradigms. These disparities make it difficult to define a general strategy for using intonation contours to mark new or contrastive intonation in Vietnamese or tonal languages in general. Furthermore, the studies discussed here use a small set of speakers sometimes not balanced across gender which contributes to the speaker-dependent variation. It would be desirable to introduce more methodological rigor including statistical analysis (Xu, 2011). Thus, we present a study with 70 Northern Vietnamese speakers, who produce sentences which each include a word that carries one of the six tones in a narrow and wide focus condition. Statistical analysis was performed for each word in each focus condition for male and female participants. Due to the non-linear effects of focus intonation, we used generalized additive models (GAMs) that include smooth functions of co-variates instead of standard linear co-variate effects (used by generalized linear models) to capture the nuances of the intonation curve.

Materials and methods

Participants

Participants were pooled from the large Vietnamese community in (Eastern) Berlin. In total, 71 participants took part in our production study. One participant had to be excluded from further analysis since there were technical issues with the recordings. The remaining 70 participants were native speakers of the Northern Vietnamese dialect aged 19–39 years (*M* = 25.44, *SD* = 4.64). Forty-five participants were women and 25 were men. Table 2 shows a summary of the participants' years spent in Germany, language proficiency, and educational level. The data reported here were part of a larger study, the main results of which have been reported elsewhere (Tjuka et al., 2020). Participants were paid 12 euros for their participation.

All participants were able to converse in at least one other language than Vietnamese, i.e., German or English, or both. However, they grew up in a monolingual household in Northern Vietnam until the age of 15 and acquired English at school. To

control for language attrition, we conducted a *post-hoc* proficiency survey. We used the Vietnamese translation by Phạm and Nguyễn of the Language Experience and Proficiency Questionnaire (Marian et al., 2007). Out of the original 71 participants, proficiency scores for 59 participants (38 women and 21 men) were collected. The questionnaire on language proficiency was administered after the data collection had been completed. We contacted the participants retrospectively and received responses from 59 participants. The 12 missing participants did not respond to our inquiry. The results showed that Vietnamese was the dominant language even for speakers with the highest proficiency in German.

Stimuli

The material for the production study consisted of six short stories with two context sentences (1), followed by one of two types of question-answer pairs. The first type of question was a constituent question (2), thereby putting the corresponding constituent in the answer in narrow focus. The second type of question was a broad inquiry about "What happened next?" (3), focusing on the entire response ("wide focus"). This three-sentence structure is one that we have used often in studies on focus in our lab (Spalek et al., 2014; Gotzner et al., 2016; Koch and Spalek, 2021). In the structure used here, the target item is given in both conditions and the focus is manipulated by the question only. The Supplementary Material includes all Vietnamese sentences with English translations (available here: https://osf.io/6e8ua).

(1) Lan thấy có tôm, cua và ngao ở chợ. Cô ấy rất thích ăn thuỷ sản.
    name see has shrimps crabs and clams at market 3 SG very like eat seafood
    'Lan saw shrimps, crabs, and clams at the market. She loves to eat seafood.' (context)

(2) Cô ấy đã mua gì? Cô ấy đã mua [TÔM]F.
    3 SG PST buy what 3 SG PST buy shrimps
    'What did she buy? She bought [SHRIMPS]F.' (narrow focus)

(3) Chuyện gì xảy ra tiếp theo? Cô ấy đã mua tôm.
    story what happen next 3 SG PST buy shrimps
    'What happened next? She bought shrimps.' (wide focus)

The stories were structured based on Tjuka et al. (2020) in which similar stimuli were used for a memory recall experiment. The first sentence introduced a protagonist and three list items of the same taxonomic category (e.g., shrimps, crabs, clams). The list items were controlled for tone and number of morphemes in that each list item in a particular sentence had the same tone and consisted of the same number of morphemes. The question after the context asked either which item of the list was chosen by the protagonist (narrow focus) or generally what happened next (wide focus). The answer included the target item (e.g., shrimps). Since we modeled the stimuli based on the stories in a previous study on the influence of intonational focus on memory recall, the target item appeared in the sentence-final position. The position may affect the production of the word and further studies need to be carried out to test whether pitch contours of tones vary in different positions. Each tone was realized in the narrow and wide focus condition by each participant.

TABLE 2 Distribution of variables.

	Age		Years in Germany		German	Educational level		
	(mean)	(SD)	(mean)	(SD)	(in years)	(degree)		
						High school	Bachelor	Master
Men (n = 25)	27.48	6.06	2.58	2.11	3.24	31	12	2
Women (n = 45)	24.31	3.18	2.8	1.5	3.74	10	8	7

Procedure

Participants were recruited to take part in an on-site laboratory experiment. They signed an informed consent form and a form about data protection. The production study followed an auditory memory recall experiment described in Tjuka et al. (2020). The stimuli in the experiment were similar, but the sentences used for the production task were new to the participants. Participants were instructed orally on how to do the sentence reading task. The communication between the experimenter and participant was done in German or English, depending on the participant's preference. The participants also received written instructions for the task in Vietnamese (see Supplementary material). They were instructed to silently read the context sentence with the question and to then read out loud the answer, i.e., the sentences with the target item, as naturally as possible. All participants read the six target sentences in both conditions and the sentence order was held consistent across participants. We acknowledge that participants may have attempted to produce different intonation patterns for two conditions when they saw them together on the paper. However, the experimenter encouraged them to think of the task as a role-playing exercise. We did not test the naturalness of the produced sentences in a separate perception study and acknowledge that this procedure can result in unnatural, inflexible sentence productions (Breen et al., 2010; Ouyang and Kaiser, 2015). In a follow-up study, the two conditions of the reading task could be presented separately and the sentences should be repeated multiple times and produced in a random order.

The sentences produced by the participants were recorded with a Sennheiser PC8 headset with an integrated microphone connected to an Olympus digital dictation device WS 853. The microphone was positioned directly in front of the participant's mouth. The task was conducted in a quiet laboratory. The experimenter positioned the piece of paper with the instructions and the sentences on one page directly in front of the participant and started the recording. Participants read out loud each target sentence with a small pause in between. The procedure took no longer than 10 min for each participant. The recordings were afterwards annotated and analyzed with Praat Version 6.1.27 (Boersma and Weenink, 2020).

Pitch analysis

The study aims to analyse differences in F_0 contours as the dependent variable. We conducted the F_0 analysis in Praat Version 6.1.27 (Boersma and Weenink, 2020). Since pitch is difficult to measure automatically, we determined the estimates for the F_0 contour based on the pitch range for each speaker by applying the two-step method proposed by Hirst (2011). We used the raw audio recordings to create an F_0 object for each of them in time steps of 0.01 seconds with a minimum F_0 of 50 Hz and a maximum of 700 Hz. Since there were technical problems with the recording of one participant we excluded their data. Other data were not excluded from the analysis.

Results

The data points were analyzed statistically by applying Generalized Additive Models (GAMs). GAMs are regression models that capture non-linear effects (Wieling, 2018), and are therefore well suited to analyze the differences in the tone contours (instantiated by F_0 contours) co-varying with different information structures, i.e., narrow and wide focus in our data set.

In particular, we analyzed the interaction between pitch and lexical tone, which shows non-linear effects of focus and tone on F_0. To achieve this, we labeled each word in such a way as to provide information about its tone and its focus condition. For example, the word *cày*, which is expressed with the low-falling tone *huyền*, is labeled as "huyen_low_NF" for the narrow condition and "huyen_low_WF" in the wide focus condition. This allowed us to compare the pitch contours of words in different discourse contexts.

We found the best model by gradually increasing its complexity and evaluating whether the increased complexity made the model better in terms of the Akaike Information Criterion (AIC) score. The procedure resulted in the best model consisting of mean F_0 as the dependent variable, and the factors participant sex, a smooth for time by focus_tone, and random effects for speaker and word (see Table 3).

The statistical analysis presented in Table 3 shows that there is an effect of focus on the realization of the tone of a word. The use of GAMs to analyze the 4,849 data points from the pitch contours of 70 speakers allowed us to gain a nuanced and detailed insight into the interplay between intonation and lexical tone in two different contexts. Figure 1 illustrates the differences between both contexts (narrow vs. wide focus) in the pitch contours of each tone (left: high tones, right: low tones). The graphs show that the differences in pitch are restricted to parts of the words. In other words, not the entire word is affected by a change of pitch due to focus marking. Although there is a large variation in pitch ranges at the end of the words, this is likely due to the sentence-final position of the word. Differences in the pitch ranges may also arise from the variation in the segmental makeup of the focus items. The intonation of the initial consonant could affect the F_0 contour at the beginning of the following vowel. Significant differences in pitch are mainly found at the beginning of the word, except in the word *nhện* for the tone *nặng*

TABLE 3 Smooth functions of covariates for GAM of pitch contour (meanPitch ∼ Sex + s(Time, by = tone_Focus, k = 50) + s(Speaker, bs = "re") + s(Word, bs = "re")).

Parametric coefficients:						
		Estimate	Std. Error	*t*-value	*Pr*(>|t|)	
	(Intercept)	197.195	7.759	25.41	<2e-16	***
	Sexm	−86.231	4.834	−17.84	<2e-16	***
Approximate significance of smooth terms:						
		edf	Ref.df	*F*	*p*-value	
	s(Time):tone_Focushoi_low_NF	3.681	4.545	3.515	0.007403	**
	s(Time):tone_Focushoi_low_WF	3.810	4.695	4.846	0.000313	***
	s(Time):tone_Focushuyen_low_NF	7.638	9.275	2.808	0.003105	**
	s(Time):tone_Focushuyen_low_WF	4.307	5.240	3.054	0.010498	*
	s(Time):tone_Focusnang_low_NF	1.002	1.003	28.214	<2e-16	***
	s(Time):tone_Focusnang_low_WF	19.678	22.958	5.374	<2e-16	***
	s(Time):tone_Focusnga_high_NF	25.186	26.433	9.445	<2e-16	***
	s(Time):tone_Focusnga_high_WF	16.321	18.944	6.216	<2e-16	***
	s(Time):tone_Focusngang_high_NF	2.542	3.124	3.161	0.021473	*
	s(Time):tone_Focusngang_high_WF	10.253	11.623	5.113	<2e-16	***
	s(Time):tone_Focussac_high_NF	14.234	16.059	5.534	<2e-16	***
	s(Time):tone_Focussac_high_WF	16.617	18.445	14.719	<2e-16	***
	s(Speaker)	65.381	67.000	50.960	<2e-16	***
	s(Word)	3.736	5.000	15.943	<2e-16	***
	R-sq.(adj) = 0.753 Deviance explained = 75.5%					
	fREML = 86,078 Scale est. = 373.13 *n* = 21,332					

Family: Gaussian, link function: Identity.

(low-falling-glottal). Here, the difference is restricted to the middle of the word.

To illustrate the differences in pitch contours for different discourse contexts, we created smooth graphs based on a simplified version of our model (see Figure 2). The graphs demonstrate the variation of pitch contours for each tone (left: high tones, right: low tones) in the two contexts (blue: narrow focus, red: wide focus). The lines show the different strategies of pitch increase and decrease to mark focus. Almost all words show a striking variation in the pitch ranges, except for the word *vải* for the tone *hỏi* (low-rising). Especially for the tones *sắc* (high-rising) and *ngang* (mid-level), the inter-individual speaker variability is large. In comparison, for the tones *ngã* (high-falling-glottal) and *nặng* (low-falling-glottal), a decrease in pitch is used to mark narrow focus. Only the word *cày* for the tone *huyền* (low-falling) is produced with a rise in pitch in the narrow focus condition.
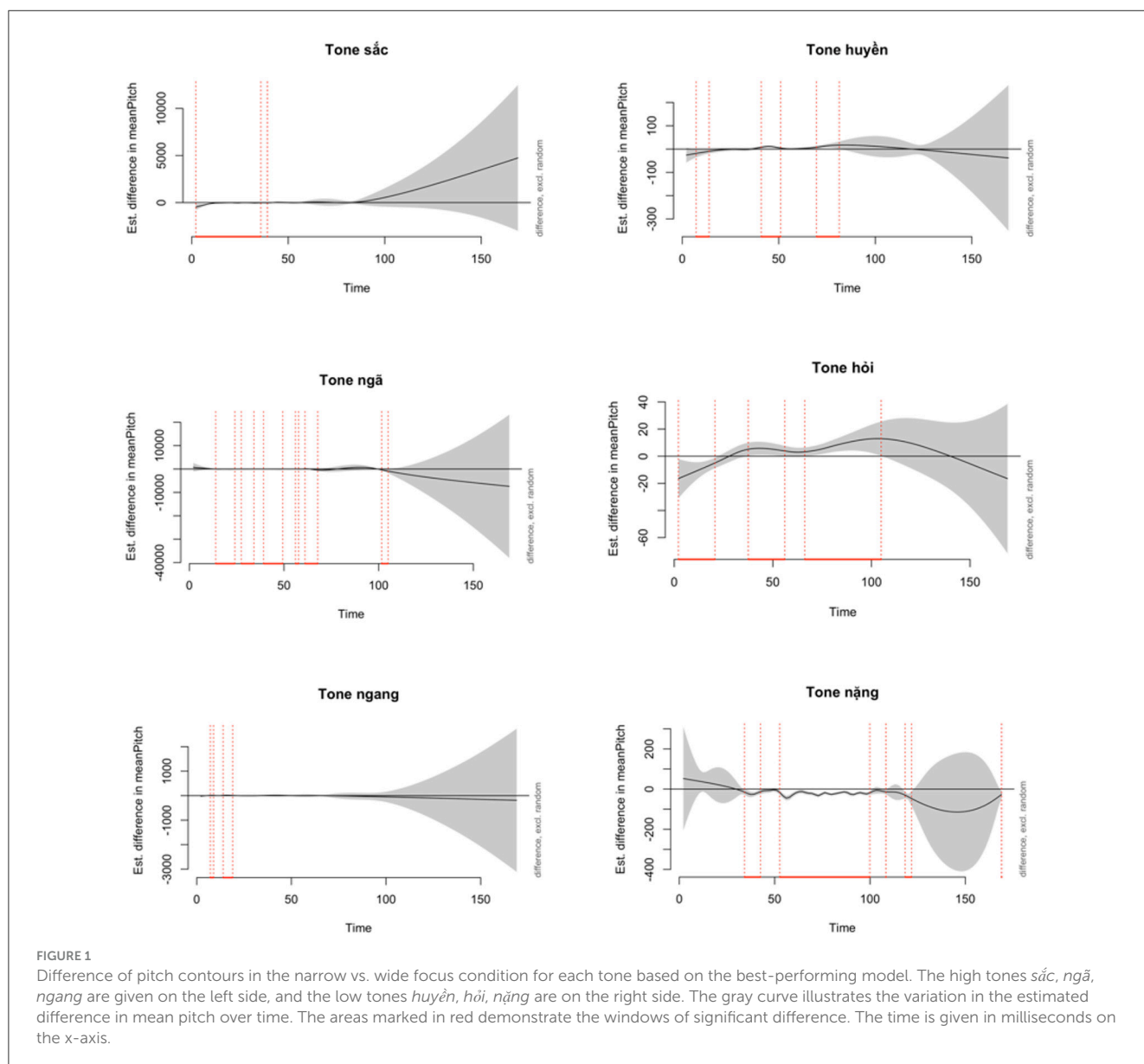
Discussion

In the present study, we examined the realization of lexical tones in different discourse contexts. Seventy Northern Vietnamese

speakers read out loud sentences with a target item for each of the six tones in two contexts and we analyzed the F_0 contours for each item in the two conditions (narrow vs. wide focus). Our study is the only study examining the interplay between intonation and lexical tone in Vietnamese across a large number of speakers and analyzing them with advanced statistical measures. The results showed that the discourse context influences the realization of tones in that different strategies of increasing and decreasing pitch were used depending on the tone and the context. For tone *sắc*, the speakers used a rise in pitch toward the end of the word to mark focus. For tones *huyền*, *ngã*, and *ngang*, the falling pitch contour was enhanced. For tones *hỏi* and *nặng*, the focus marking was produced by a more complex rising-falling pattern. Especially at the beginning of the word, pitch contours differed statistically significantly in the two discourse contexts.

Compared to previous studies examining strategies for structuring information in Vietnamese, our results indicate that the interplay between intonation and lexical tone is more intricate. The study by Michaud and Vu-Ngoc (2004) focused on the tone *nặng* and found that words with this low-falling-glottal tone receive a rising pitch contour for emphasis. These findings are not supported by our GAMs analysis of the word *nhện* for the tone

FIGURE 1
Difference of pitch contours in the narrow vs. wide focus condition for each tone based on the best-performing model. The high tones *sắc, ngã, ngang* are given on the left side, and the low tones *huyền, hỏi, nặng* are on the right side. The gray curve illustrates the variation in the estimated difference in mean pitch over time. The areas marked in red demonstrate the windows of significant difference. The time is given in milliseconds on the x-axis.

nặng. We found that the word is realized with a lowering of pitch to mark narrow focus when it occurs in the sentence's final position. Furthermore, Miller et al. (2015) claimed that there are no changes in pitch to highlight new information for words with the tones *sắc* and *ngã*. Our findings do not support their results. We found a significant difference in the pitch contours in both discourse contexts which indicates that speakers have different strategies to highlight information. In the case of the tone *sắc* the differences were limited to the beginning of the word, whereas pitch ranges varied at the beginning and middle of the word for the tone *ngã*. Both tones were realized with a lowering of pitch to mark narrow focus although the pattern was stronger for the tone *ngã*. Only the tone *hỏi* did not show strong differences in the realization of pitch in the two conditions. Further research is needed to establish a model for the strategies of intonation patterns in different sentence positions and discourse contexts in Vietnamese.

Our study offers important insights not only from a single language but also from a cross-linguistic perspective. As shown by Maddieson (2013), several languages have a tone system either with a distinction between a high and a low lexical tone or more complex distinctions. The results of our study are in line with studies on Mandarin Chinese which has a tone system of four tones and demonstrates changes in pitch as well as duration and intensity for focus marking (Xu, 1999; Ouyang and Kaiser, 2015). However, the study on Kammu by Karlsson et al. (2012) showed a different strategy for the low tone. Here, the word with the high tone was emphasized with a rising pitch contour in the focus condition, but the low tone neutralized the pitch contour. In Vietnamese, the low-falling tone *huyền* is realized with a rise in pitch to mark narrow focus. Some African tonal languages also use prosodic cues to indicate new or contrastive information, whereas others show no sign of using intonation for information structure (for an overview, see Zerbian et al., 2010; Güldemann et al., 2015). Tonal languages that use intonation for structuring discourse do not all employ the same strategy. For example, speakers of Northern Chichewa use a rise in pitch (Downing, 2008), whereas speakers of
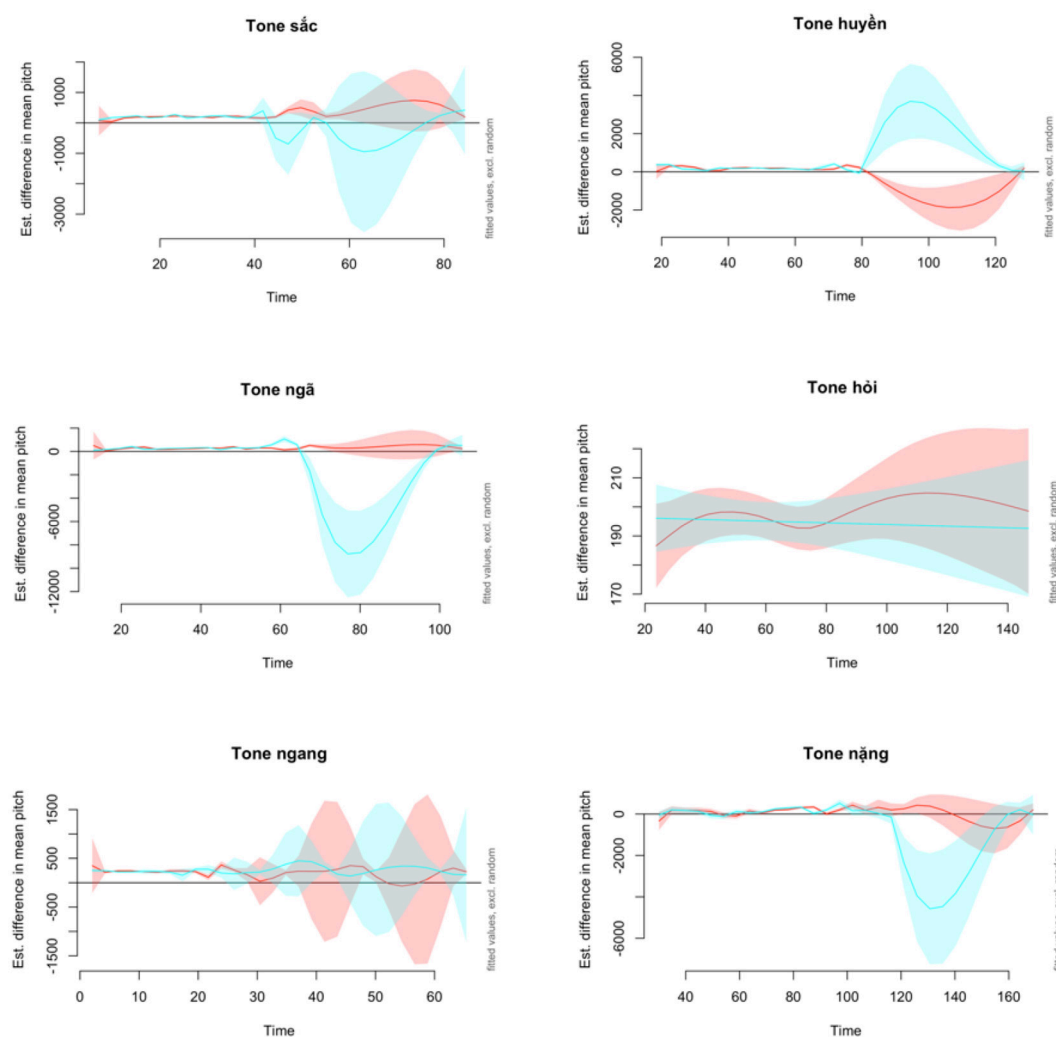
FIGURE 2
Pitch contours in the narrow (blue) and wide (red) focus condition for each tone. The high tones *sắc*, *ngã*, *ngang* are given on the left side, and the low tones *huyền*, *hỏi*, *nặng* are on the right side. The lines illustrate the production of estimated difference in mean pitch in both focus conditions with confidence bands showing the variation across speakers over time. In order to compare the pitch contours in both focus conditions, the time is normalized. This is achieved by treating time as a numeric predictor with 30 equidistant values within the range starting from when the first pitch information became available until the point when the last pitch information was available. The numbers on the *x*-axis represent approximately the time in milliseconds for which pitch information is available.

Akan use deaccentuation to mark focus (Kügler and Genzel, 2012). Our findings show an additional strategy: marking narrow focus with a lowering of pitch restricted mainly to glottalized tones (*ngã* and *nặng*). Thus, each language may employ particular intonation contours for different discourse contexts.

The interplay between lexical tone and intonation is a complex phenomenon that requires a detailed analysis and methodological rigor. We recorded an unprecedented number of Northern Vietnamese speakers in a laboratory condition. Our finding that these speakers change the pitch contours of all six tones in different discourse contexts has important implications for our understanding of information structure in tonal languages and the development of educational programs. The inclusion of different intonation strategies in discourse can help students decipher the meaning of an utterance more effortlessly and integrate a more naturalistic pitch pattern in their speech.

Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: https://osf.io/6e8ua/.

Ethics statement

The studies involving humans were approved by Ethikkommission der Deutschen Gesellschaft für Sprachwissenschaft (DGfS), Prof. Dr. Angela Grimm, Goethe-Universität Frankfurt am Main. The studies were conducted in accordance with the local legislation and institutional requirements. The participants provided their written informed consent to participate in this study.

Author contributions

Funding

Acknowledgments

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

Supplementary Material is available here: https://osf.io/6e8ua. The repository includes all stimuli sentences and instructions in Vietnamese with English translations as well as the scripts used for the pitch analysis.

References

Baumann, S., Becker, J., Grice, M., and Mücke, D. (2007). "Tonal and articulatory marking of focus in German," in *Proceedings of the 16th International Congress of Phonetic Sciences* (Saarbrücken: Pirrot GmbH Dudweiler), 1029–1032.

Boersma, P., and Weenink, D. (2020). *Praat: Doing Phonetics by Computer*. Version 6.1.27. Available online at: http://www.praat.org/

Breen, M., Fedorenko, E., Wagner, M., and Gibson, E. (2010). Acoustic correlates of information structure. *Lang. Cogn. Process*. 25, 1044–1098. doi: 10.1080/01690965.2010.504378

Brunelle, M. (2009). Tone perception in Northern and Southern Vietnamese. *J. Phon*. 37, 79–96. doi: 10.1016/j.wocn.2008.09.003

Brunelle, M. (2017). Stress and phrasal prominence in tone languages: the case of Southern Vietnamese. *J. Int. Phon. Assoc*. 47, 283–320. doi: 10.1017/S0025100316000402

Brunelle, M., Hạ, K. P., and Grice, M. (2012). Intonation in Northern Vietnamese. *Linguist. Rev*. 29, 3–36. doi: 10.1515/tlr-2012-0002

Brunelle, M., and Jannedy, S. (2013). "The cross-dialectal perception of Vietnamese tones: indexicality and convergence," in *Linguistics of Vietnamese, Volume 253 of Trends in Linguistics*, eds. D. Hole, and R. Löbel (Berlin: De Gruyter Mouton), 9–34.

Brunelle, M., Nguyễn, D. D., and Nguyễn, K. H. (2010). A laryngographic and laryngoscopic study of Northern Vietnamese tones. *Phonetica* 67, 147–169. doi: 10.1159/000321053

Chen, Y., and Gussenhoven, C. (2008). Emphasis and tonal implementation in Standard Chinese. *J. Phon*. 36, 724–746. doi: 10.1016/j.wocn.2008.06.003

Đễ, T. D., Trân, T. H., and Boulakia, G. (1998). "Intonation in Vietnamese," in *Intonation Systems: A Survey of Twenty Languages*, eds. D. Hirst, and A. Di Cristo (Cambridge: Cambridge University Press), 398–420.

Downing, L. J. (2008). Focus and prominence in Chichewa, Chitumbuka and Durban Zulu. *ZAS Pap. Linguist*. 49, 47–65. doi: 10.21248/zaspil.49.2008.363

Erlewine, M. Y. (2017). Vietnamese focus particles and derivation by phase. *J. East Asian Ling*. 26, 325–349. doi: 10.1007/s10831-017-9156-y

Gotzner, N., Wartenburger, I., and Spalek, K. (2016). The impact of focus particles on the recognition and rejection of contrastive alternatives. *Lang. Cogn*. 8, 59–95. doi: 10.1017/langcog.2015.25

Güldemann, T., Zerbian, S., and Zimmermann, M. (2015). Variation in information structure with special reference to Africa. *Ann. Rev. Linguist*. 1, 155–178. doi: 10.1146/annurev-linguist-030514-125134

Gussenhoven, C. (2004). *The Phonology of Tone and Intonation. Research Surveys in Linguistics*. Cambridge: Cambridge University Press.

Gussenhoven, C. (2007). "Types of focus in English," in *Topic and Focus: Cross-Linguistic Perspectives on Meaning and Intonation, Studies in Linguistics and Philosophy*, eds. C. Lee, M. Gordon, and D. Büring (Dordrecht: Springer Netherlands), 83–100.

Hạ, K. P. (2012). *Prosody in Vietnamese—Intonational Form and Function of Short Utterances in Conversation*. Canberra, ACT: The Australian National University.

Hirst, D. (2011). The analysis by synthesis of speech melody: from data to models. *J. Speech Sci*. 1, 55–83. doi: 10.20396/joss.v1i1.15011

Hoàng, T. C. (1989). *Vietnamese in All Regions of the Country*. Hanoi: Hanoi National University Publishing House.

Hole, D. (2008). "EVEN, ALSO and ONLY in Vietnamese," in *Interdisciplinary Studies on Information Structure, Volume 11 of Working Papers of the SFB 632*, eds. S. Ishihara, S. Petrova, and A. Schwarz (Potsdam: Universitaetsverlag Potsdam), 1–54.

Hole, D. (2013). "Focus particles and related entities in Vietnamese," in *Linguistics of Vietnamese, Volume 253 of Trends in Linguistics*, eds. D. Hole, and E. Löbel (Berlin: De Gruyter Mouton), 265–304.

Jannedy, S. (2007). "Prosodic focus in Vietnamese," in *Interdisciplinary Studies on Information Structure, Volume 8 of* Working Papers of the SFB 632, eds. S. Ishihara, S. Jannedy, and A. Schwarz (Potsdam: Universitaetsverlag Potsdam), 209–230.

Karlsson, A., House, D., and Svantesson, J.-O. (2012). Intonation adapts to lexical tone: the case of Kammu. *Phonetica* 69, 28–47. doi: 10.1159/000343169

Koch, X., and Spalek, K. (2021). Contrastive intonation effects on word recall for information-structural alternatives across the sexes. *Mem. Cognit*. 49, 1312–1333. doi: 10.3758/s13421-021-01174-1

Kügler, F., and Genzel, S. (2012). On the prosodic expression of pragmatic prominence: the case of pitch register lowering in Akan. *Lang. Speech* 55, 331–359. doi: 10.1177/0023830911422182

Maddieson, I. (2013). "Tone," in *The World Atlas of Language Structures Online*, eds. M. S. Dryer, and M. Haspelmath (Leipzig: Max Planck Institute for Evolutionary Anthropology).

Marian, V., Blumenfeld, H. K., and Kaushanskaya, M. (2007). The Language Experience and Proficiency Questionnaire (LEAP-Q): assessing language profiles in bilinguals and multilinguals. *J. Speech Lang. Hear. Res*. 50, 940–967. doi: 10.1044/1092-4388(2007/067)

Michaud, A., and Brunelle, M. (2016). *Information Structure in Asia: Yongning Na (Sino-Tibetan) and Vietnamese (Austroasiatic)*. Oxford: Oxford University Press.

Michaud, A., and Vu-Ngoc, T. (2004). "Glottalized and nonglottalized tones under emphasis: open quotient curves remain stable, F0 curve is modified," in *Proceedings of Speech Prosody*, eds. B. Bel, and I. Marlien [Nara: International Speech Communication Association (ISCA)], 745–748.

Miller, T., Athanasopoulou, A., Pincus, N., and Vogel, I. (2015). "The effect of focus on phonation in Northern Vietnamese Tones," in *Poster presented at the 89th Annual Meeting of the Linguistic Society of America* (Portland, Oregon).

Ouyang, I. C., and Kaiser, E. (2015). Prosody and information structure in a tone language: an investigation of Mandarin Chinese. *Lang. Cognit. Neurosci.* 30, 57–72. doi: 10.1080/01690965.2013. 805795

Peters, J., Hanssen, J., and Gussenhoven, C. (2014). The phonetic realization of focus in West Frisian, Low Saxon, High German, and three varieties of Dutch. *J. Phon.* 46, 185–209. doi: 10.1016/j.wocn.2014.07.004

Spalek, K., Gotzner, N., and Wartenburger, I. (2014). Not only the apples: focus sensitive particles improve memory for information-structural alternatives. *J. Mem. Lang.* 70, 68–84. doi: 10.1016/j.jml.2013.09.001

Thompson, L. C. (1965). *A Vietnamese Grammar*. Seattle, WA: University of Washington Press.

Tjuka, A., Nguyen, H. T. T., and Spalek, K. (2020). *Foxes*, *deer*, and *hedgehogs*: the recall of focus alternatives in Vietnamese. *Lab. Phonol.* 11, 1–29. doi: 10.5334/labphon.253

Vũ, T. P. (1982). "Phonetic properties of Vietnamese tones across dialects," in *Papers in Southeast Asian Linguistics, Volume 8 of Pacific Linguistics*, ed. D. Bradley (Canberra, ACT: The Australian National University), 55–75.

Wieling, M. (2018). Analyzing dynamic phonetic data using generalized additive mixed modeling: a tutorial focusing on articulatory differences between L1 and L2 speakers of English. *J. Phon.* 70, 86–116. doi: 10.1016/j.wocn.2018.03.002

Xu, Y. (1999). Effects of tone and focus on the formation and alignment of F0 contours. *J. Phon.* 27, 55–105. doi: 10.1006/jpho.1999.0086

Xu, Y. (2011). Speech prosody: a methodological review. *J. Speech Sci.* 1, 85–115. doi: 10.20396/joss.v1i1.15014

Zerbian, S., Genzel, S., and Kügler, F. (2010). "Experimental work on prosodically-marked information structure in selected African languages (Afroasiatic and Niger-Congo)," in *Proceedings of Speech Prosody*, eds. M. Hasegawa-Johnson, A. Bradlow, J. Cole, K. Livescu, J. Pierrehumbert, and C. Shih (Chicago, IL), 1–4.