

## **Pitch contours of Northern Vietnamese tones vary with focus marking**

*Annika Tjuka<sup>1</sup>, Hương Thi Thu Nguyễn<sup>2</sup>, Ruben van de Vijver<sup>3</sup>, Katharina Spalek<sup>3</sup>*

<sup>1</sup>Max Planck Institute for the Science of Human History

<sup>2</sup>Humboldt-Universität zu Berlin

<sup>3</sup>Heinrich-Heine-Universität Düsseldorf

Intonation is a means of structuring discourse and one of its functions is to highlight new or contrasting information, i.e., focus. Speakers of different languages use a range of prosodic cues to mark focus. Compared to non-tonal languages such as English, tonal languages use pitch to distinguish lexical tones and focus marking. Determining the interplay between intonation and lexical tone is therefore important. Previous studies found that tonal languages use different strategies to mark focus. For example, some use an increase (e.g., Mandarin Chinese), others a decrease in pitch (e.g., Kammu). The Vietnamese language has six lexical tones and is particularly interesting for examining pitch contours in focus marking. In this article, we present a production study with 70 Northern Vietnamese speakers. Participants read six sentences aloud under two different conditions (narrow/wide focus). In each sentence, focus marked a single noun ('focus item') which occurred in the final position of the sentence and carried one of the six tones. Acoustic analyses of the focus item showed that Vietnamese speakers realized focus with significant differences in pitch at the beginning of the word, but the strategies to increase or decrease pitch varied across tones. Our findings add important insights to the discussion about Information Structure and the role of intonation in tonal languages by analyzing the use of prosodic cues in a complex tone system. The large number of speakers in our study also adds further methodological rigor compared to other studies, which often rely on a few speakers.

**Keywords:** Information structure, focus, intonation, speech production, Vietnamese

# 1 Introduction

Intonation is central to the study of Information Structure in the world's languages. Linguists have been especially interested in the realization of *focus* in sentence processing and production. Focus is used to highlight new or contrastive information in discourse and can be realized with specialized particles, word order, or intonation, thus playing a role at the linguistic levels of syntax, morphology, and phonology (Krifka 2006). Research on prosodic focus traditionally concentrated on languages in which intonation plays a predominant role in structuring information, such as German, English, or Dutch (e.g., Baumann et al. 2007; Gussenhoven 2007; Peters, Hanssen, & Gussenhoven 2014). The marking of focus with phonological features in a tonal language was first systematically studied by Xu (1999). Although it would intuitively make sense that languages with lexical tone make less use of intonation for pragmatic purposes, several languages show a complex interplay between lexical tone and intonation. Speakers of Mandarin Chinese realize focus in a non-sentence-final position with a rise in pitch for the focused word and a fall or compression following the focus (Xu 1999). Studies on Vietnamese showed that in addition to pitch, duration and intensity are used to mark focus (Jannedy 2007; Brunelle, Hà & Grice 2012; Michaud & Brunelle 2016). However, there seems to be variation across low and high tones as well as inter-speaker variation (Brunelle 2017).

From a cross-linguistic viewpoint, Vietnamese is ideal to study prosodic cues for focus marking in tonal languages due to its complex tonal system with six tones. Previous studies did not systematically compare wide versus narrow focus marking across a large number of speakers. Their analyses were based on a limited sample of 2 to 24 speakers. Although differences in focus marking between male and female participants were reported, generalization of these is difficult due to the small number of speakers. Thus, our study adds further insights to focus marking with intonation in Northern Vietnamese with recordings of focus production for each of the six tones from 70 participants (45 female and 25 male) in a controlled setting. The results indicate that the strategies speakers use for marking focus differ for high and low tones.

## 2 Using intonation for focus marking

Prosodic cues for focus marking play a role in the processing of language in that they activate focus alternatives (Krifka 2006; Braun & Tagliapietra 2010; Husband & Ferreira 2016; Yan & Calhoun 2019, Tjuka, Nguyễn & Spalek 2020; Koch & Spalek 2021). It is thus important to investigate how speakers incorporate intonation in utterances to mark focus. The change of pitch patterns for intonation and lexical tone has been described for several tonal languages (for a detailed overview, see Gussenhoven 2004). The following sections provide an overview of the different prosodic cues used for focus marking in tonal languages in general and in Vietnamese in particular.

### 2.1 Focus marking with intonation in tonal languages

Of the world's 7,000 languages, 527 languages have a sufficient description of their phonological system. The majority of these, 307, are non-tonal languages (Maddieson 2013).<sup>1</sup> Tonal languages are subdivided into two classes: Simple versus complex tone systems, 132 and 88 languages, respectively (Maddieson 2013). Many languages of South East Asia, including Vietnamese and Mandarin Chinese, are classified as having a complex tone system.

Mandarin Chinese has a complex tone system with four lexical tones. Research on how Mandarin Chinese speakers encode different types of focus showed a complex interplay between intonation and lexical tone. Xu (1999) found that, in addition to pitch, duration and intensity are used to introduce new or contrastive information. The findings were based on recordings of eight speakers (4 male, 4 female). Although more recent research supports the conclusion that duration and intensity mark focus in Mandarin Chinese, the role of pitch seems to be more intricate (Ouyang & Kaiser 2015). The two functions of the pitch contours for distinguishing tones and focus are

---

<sup>1</sup> Note that the distribution may not represent the overall picture since the sample was based on sufficient description rather than balancing the languages across geography, genealogy, and linguistic diversity.

difficult to separate in an acoustic analysis. For their study, Ouyang and Kaiser (2015) recorded ten speakers (5 male, 5 female). Results of production studies 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 and duration and no change in intensity (Ouyang & Kaiser 2015). The study demonstrated that pitch ranges in the contrastive focus condition were extended for the minimum and maximum bound, but the difference in high versus low tone was not analyzed. According to Chen and Gussenhoven (2008), lowering pitch for low tones may result in a creaky voice which makes the speaker sound raspy and is therefore avoided. However, there is, as of yet, no study that assessed differences in pitch range in low tones further.

Kammu, an Austroasiatic language spoken in Laos and in some parts of Vietnam, Thailand, and China, shows another pattern in focus marking with intonation. There are two major dialect groups: Eastern Kammu versus Northern and Western Kammu. The latter dialects have developed a tone system with two lexical tones in recent years, whereas Eastern Kammu does not use tone to distinguish lexical meaning (Karlsson, House & Svantesson 2012). The study analyzed recordings of ten speakers (7 male, 3 female) for the non-tonal dialect and 14 speakers (6 male, 8 female) for the tonal dialect. The comparison of focus marking in the non-tonal versus tonal dialect revealed that focus is marked with a rising intonation contour, but 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. Karlsson et al. (2012) proposed a hierarchy that speakers of the tonal dialect use to maintain lexical tone before marking phrase-final boundary tone and focus. They reported that the falling lexical tone is in contrast with the rising intonation contour that marks focus and thus, speakers neutralize or use an even lower pitch range to mark focus for the low tone.

The complex interplay of intonation and lexical tone is challenging to tease apart. Another difficulty is that focus types, the position of the focused element, and the complexity of tone systems vary greatly across available studies. Furthermore, most studies could not investigate

individual differences systematically due to the small number of speakers ranging from 2 to 24 participants in the above studies.

## 2.2 Using intonation to mark pragmatic functions and focus in Vietnamese

Vietnamese belongs to the Austroasiatic language family and there are three dialect groups: Northern Vietnamese, Southern Vietnamese, and Central Vietnamese (Hoàng 1989; Vũ 1982). One of the main differences between the dialects is that they vary in their tone inventory. Since the present article investigates focus marking in Northern Vietnamese, which is the standard variety, we will concentrate on the six-tone system. The tones in Northern Vietnamese are expressed by combining pitch and voice quality (see Tab. 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, Nguyễn & Nguyễn 2010; Brunelle & Jannedy 2013). The tones are indicated by a diacritic above the vowel.

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

<b>Tone</b>	<b>Description</b>	<b>Example</b>	<b>Translation</b>
<i>sắc</i>	high-rising	<i>má</i>	mother
<i>ngã</i>	high-falling-glottal	<i>mã</i>	code
<i>ngang</i>	mid-level	<i>ma</i>	ghost
<i>huyền</i>	low-falling	<i>mà</i>	nevertheless, but
<i>hỏi</i>	low-rising	<i>mả</i>	tomb, grave
<i>nặng</i>	low-falling-glottal	<i>mạ</i>	rice seedling

In addition to a complex tone system, Vietnamese speakers make use of intonation to mark different pragmatic contexts (Thompson 1965). Different sentence types – declarative, interrogative, and imperative – are distinguished by changes in global F0 contour, syllable length, and intensity (Đỗ, Trần & Boulakia 1998).<sup>2</sup> The study by Hạ (2012) investigated short utterances in a corpus of telephone calls by 43 Northern Vietnamese participants (20 male, 23 female). When using discourse particles in certain contexts such as backchannels and turn-yielding, the lexical tone is 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 (7 male, 9 female), Brunelle et al. (2012) showed that prosodic cues are used to express the particle *không* ‘empty, no, only’ in different contexts, but they also found 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 and intonation (for an overview, see Michaud & 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; Hole 2013; Erlewine 2017). Intonation, on the other hand, seems to be less systematic but has been described to occur for certain types of focus marking. Jannedy (2007) found in a question-answer paradigm that speakers of Northern Vietnamese 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). The findings were based on two speakers (1 male, 1 female) and thus, exhibit variation in pitch values across speakers. However, inter-speaker variation for duration patterns was not found. Michaud and Vu-Ngoc (2004) showed that words with tone *nặng* receive a rising

---

<sup>2</sup> The authors do not specify the number of speakers on which the results are based.

pitch contour for emphasis, but duration varied across speakers. Their analysis was based on four speakers (3 male, 1 female). In contrast, Miller et al. (2015) found no changes in pitch and phonation in a new information focus versus non-focus condition for tone *săc* and tone *ngã* across nine speakers (2 male, 7 female), but they were expressed with a change in duration and spectral energy to mark focus. These results also reflect some of the findings of focus marking in Southern Vietnamese (Brunelle 2017).

Studies investigating focus marking with intonation are characterized by the fact that they vary greatly in scope. They discuss different types of focus, such as marking new or contrastive information. The position in which the focused element occurs varies from study to study, and not all studies analyze all the tones of the Vietnamese tone system. Furthermore, some studies use individual sentences, others a question-answer paradigm. These disparities make it difficult to define a general strategy for focus marking in Vietnamese. In addition, 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 (for a critical review, see also Xu 2011) based on a sufficient number of speakers to the study of intonation in Vietnamese. Thus, we present a study with 70 Northern Vietnamese speakers producing sentences including a focus item carrying 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-variables instead of standard linear co-variate effects (used by generalized linear models) to capture the nuances of the intonation curve (further details in Section 4).

### 3 Method

#### 3.1 Participants

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 to 39 years ( $M = 25.44$ ,  $SD = 4.64$ ). Forty-five participants were female and 25 were male. 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, Nguyễn & Spalek 2020). Participants were paid 12 euros for their participation.

Table 2. Distribution of variables across the 25 male and 45 female participants.

	Age		Years spend in Germany		Learning German (in years)	Educational level		
	(mean)	(SD)	(mean)	(SD)		(degree)	High school	Bachelor
<b>Male (n = 25)</b>	27.48	6.06	2.58	2.11	3.24	31	12	2
<b>Female (n = 45)</b>	24.31	3.18	2.8	1.5	3.74	10	8	7

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 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 (LEAP-Q, Marian, Blumenfeld, and Kaushanskaya 2007). Out of the original 71 participants, proficiency scores for 59 participants (38 female and 21 male) were collected. The results showed that Vietnamese was the dominant language even for speakers with the highest proficiency in German.



### 3.3 Material

The material for the production study included six short stories with two context sentences (1), followed by a narrow focus question and answer (2) or a wide focus question and answer (3). The Supplementary Material (<https://osf.io/6e8ua/>) includes all sentences with English translations.

- (1) *Lan thấy có tôm, cua và ngao ở chợ. Cô ấy rất thích ăn thủy sản.*

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]<sub>F</sub>.*

What did she buy? She bought [SHRIMPS]<sub>F</sub>. (*narrow focus*)

- (3) *Chuyện gì xảy ra tiếp theo? Cô ấy đã mua tôm.*

What happened? 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 one of the list items (e.g., shrimps). Each tone was realized in the narrow and wide focus condition by each participant.

### 3.2 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 a memory recall experiment described in Tjuka et al. (2020). After the experiment, participants received an oral instruction of the procedure for 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 a written instruction of the task in Vietnamese (see Supplementary Material: <https://osf.io/6e8ua/>). They were instructed to read out loud the target sentences as naturally as possible and imagine the task as a role-playing exercise.

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 instruction 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 minutes for each participant. The recordings were afterwards annotated and analyzed with Praat Version 6.1.27 (Boersma & Weenink 2009).

### 3.3 Pitch Analysis

We conducted the F0 analysis in Praat Version 6.1.27 (Boersma & Weenink 2009). Since pitch is difficult to measure automatically, we determined the estimates for the F0 contour on the basis of 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 F0 object for each of them in time steps of 0.1 seconds with a minimum F0 of 50 Hz and a maximum of 700 Hz.

We normalized the duration of each word, by dividing each word into 100 time steps and extracted for each time slice the mean and its standard deviation of the F0 in Hz as well as the minimum and maximum F0 value in Hz. This procedure resulted in 4,849 data points from 70 speakers.

## 4 Results

The data points were analyzed statistically by applying generalized additive models (GAMs). GAMs are regression models that capture non-linear effects (for a detailed description, see Wieling 2018), and are therefore well suited to analyze the differences in the tone contours contrasting the narrow and wide focus condition in our data set.

In particular, we analyzed the interaction between pitch and lexical tone, which show non-linear effects of focus and tone on F0. 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 *huyệ̀n*, 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 on words in relation to their focus condition.

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 F0 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).

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")). Family: Gaussian, link function: Identity.

<b>Parametric coefficients:</b>					
	<i>Estimate</i>	<i>Std. Error</i>	<i>t value</i>	<i>Pr(&gt; t )</i>	
(Intercept)	197.195	7.759	25.41	<2e-16	***
Sexm	-86.231	4.834	-17.84	<2e-16	***
<b>Approximate significance of smooth terms:</b>					
	<i>edf</i>	<i>Ref.df</i>	<i>F</i>	<i>p-value</i>	
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 = 86078 Scale est. = 373.13 n = 21332

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 4849 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. Figure 1 illustrates the differences between both focus conditions (narrow versus 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. Significant differences in pitch are mainly found at the beginning of the word, except in the word *nhện* for the tone *nặng* (low-falling-glottal). Here, the difference is restricted to the middle of the word.

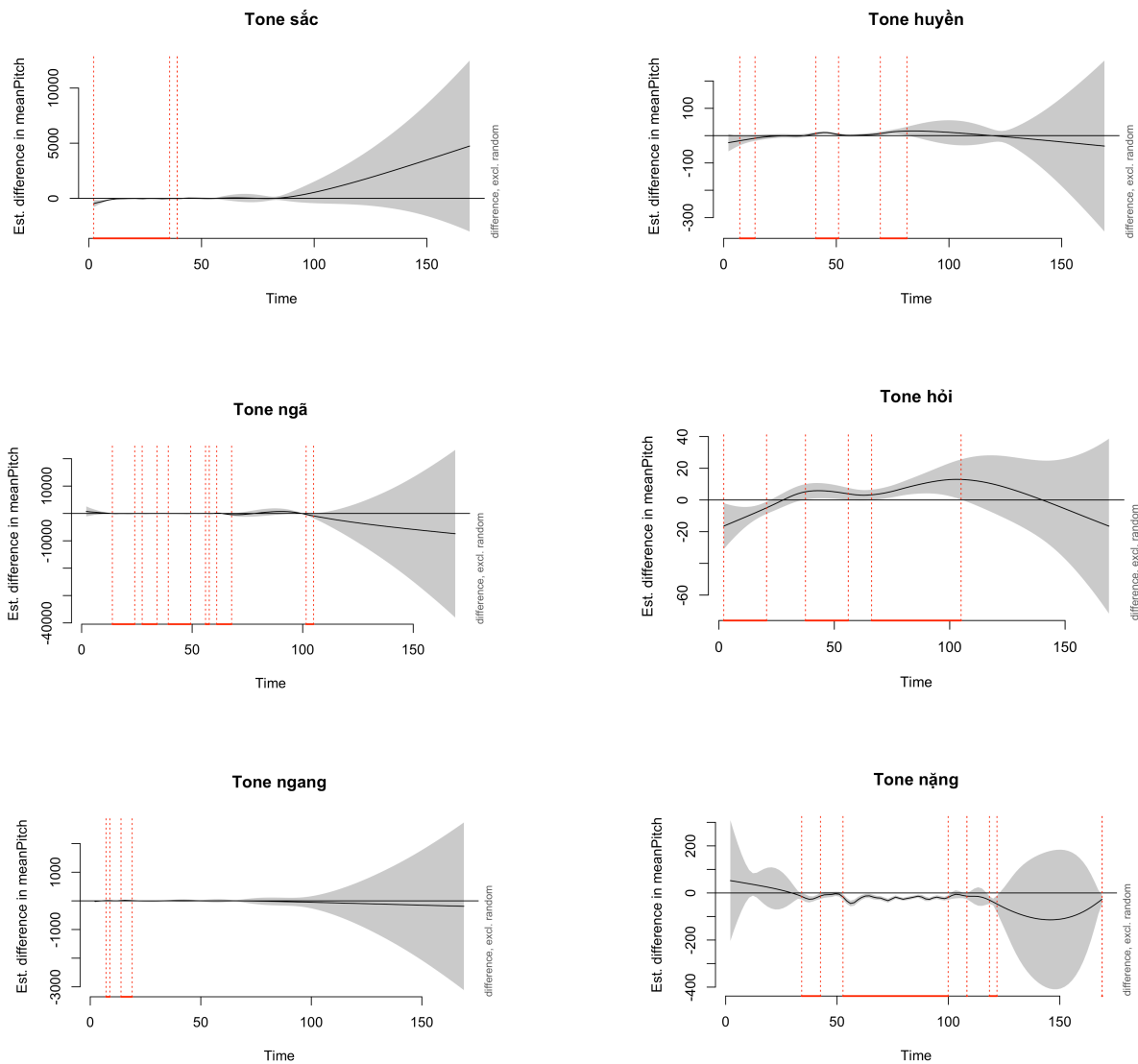


Figure 1. Difference of pitch contours in the narrow versus wide focus condition for each tone based on the best performing model. The high tones *sắc*, *ngã*, *ngang* are given on the left side, and the low tones *huyền*, *hỏi*, *nặng* are on the right side. The gray curve illustrates the variation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

To illustrate the differences in pitch contours for the narrow and the wide focus condition, we created smooth graphs based on a simplified version of our model (see Figure 2).<sup>3</sup> The graphs demonstrate the variation of pitch contours for each tone (left: high tones, right: low tones) in the two focus conditions (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 *vái* for the tone *hỏi* (low-rising). Especially for the tones *sắc* (high-rising) and *ngang* (mid-level), the inter-individual speaker variability in focus marking strategies is large. In comparison, for the tones *ngã* (high-falling-glottal) and *nặng* (low-falling-glottal), a decrease in pitch is used to mark focus. Only the word *cày* for the tone *huyền* (low-falling) is produced with a rise in pitch in the narrow focus condition.

---

<sup>3</sup> Due to the properties of the algorithm built into the *mgcv* package (Version 1.8-40, Wood 2022) that we used to model the results, we applied a different model to represent the pitch contours for both focus conditions (narrow and wide focus). The model also estimates the differences in mean pitch for each tone, so that the scales of the x-axis in the graphs are different.

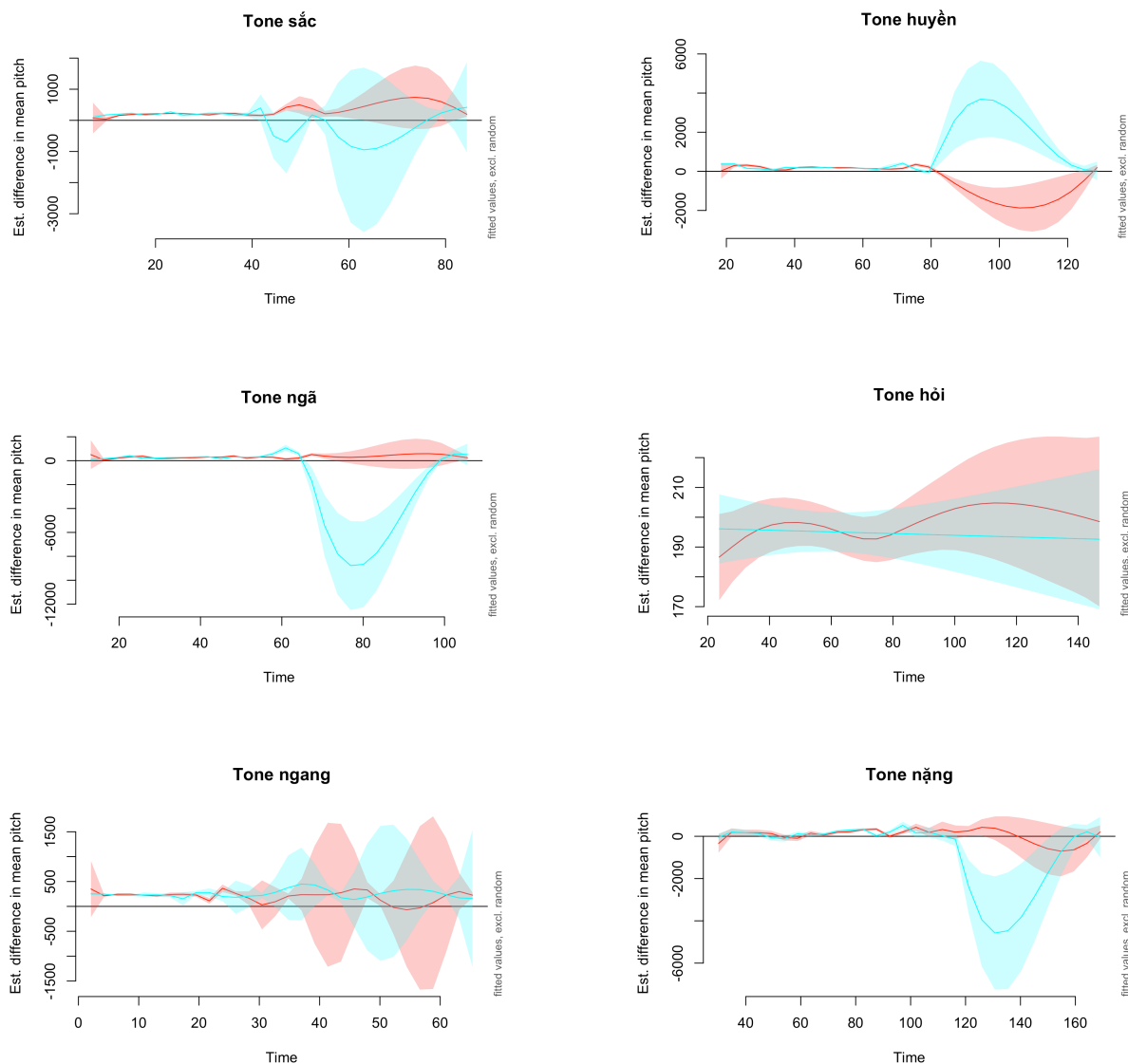


Figure 2. Pitch contours in the narrow (blue) and wide (red) focus condition for each tone. The high tones *sắc*, *ngã*, *ngang* are given on the left side, and the low tones *huyền*, *hỏi*, *nặng* are on the right side. The lines illustrate the production of estimated difference in mean pitch in both focus conditions with confidence bands showing the variation across speakers over time.

## 5 Discussion

In the present study, we examined the pitch contours for marking focus in the six lexicalized tones of Northern Vietnamese. We used a set of sentences introducing a context, question, and target sentences to survey the focus production under two conditions: narrow (new information) versus wide focus. Seventy Northern Vietnamese speakers read out loud the target sentences containing

a focus item with one of the six tones in the last position of the sentence. As of yet, our study is the only study recording focus production in Vietnamese across a large number of speakers and analyzing them with advanced statistic measures. We performed a pitch analysis based on an estimate of the individual pitch range for each speaker (Hirst 2011). To capture the non-linear pitch contours, the differences in F0 in the two focus conditions were then analyzed with generalized additive models (GAMs). The results showed that tones are realized differently in the two conditions. In most parts of the word, but especially at the beginning of the word, pitch contours differed statistically significantly across the focus conditions. To mark narrow focus, Northern Vietnamese speakers used different strategies of increasing and decreasing pitch depending on the lexical tone.

Our results support only partially the findings of earlier studies on focus marking with intonation in Vietnamese. Jannedy (2007) also found that focused items are expressed with a rise in pitch. However, the study investigated focus intonation dependent on the position of the focused element in the sentence, so no direct comparison between words with one of the six tones in the same position was conducted. Although Jannedy's results to some degree do not contradict our systematic study comparing focus production of the same words in the same sentence position, it is difficult to make a direct comparison. Based on our findings, the interplay between intonation and lexical tone is more intricate than previously assumed. The study by Michaud and Vu-Ngoc (2004) focused on one tone (*nặng*) and found that words with this low-falling-glottal contour receive a rising pitch contour for emphasis. These findings are not supported by our GAMs analysis of the word *nhện* for the tone *nặng* in both focus conditions. We found that the word is realized with a lowering of pitch to mark narrow focus. One reason for the different results is that Michaud and Vu-Ngoc (2004) focused, on the one hand, on the effect of glottalization in the production of the tone *nặng*. On the other hand, they examined emphasis, which is not equivalent to focus. They also report high speaker-specific variation so that their small speaker sample may not be able to illustrate a larger pattern.



Miller et al. (2015) claimed that there are no changes in pitch for words with the tones *sắc* and *ngã* to highlight new information. Our findings contradict their results. We found a significant difference in the pitch contours for the narrow versus the wide focus conditions which implies that speakers adapt their pitch to mark focus. In case of the tone *sắc* the differences were limited to the beginning of the word whereas, for the tone *ngã*, pitch ranges varied at the beginning and middle of the word. Both tones were marked with a lowering of pitch in the narrow focus condition although the pattern was stronger for the tone *ngã*. Only the tone *hỏi* did not show strong differences in the realization of pitch for the narrow and wide focus condition, whereas all other tones use a change in pitch to mark focus. It is important to note that these production studies are in general prone to inter-speaker variation due to the low number of data points produced by only a few speakers. Some studies also indicated gender differences, but our results showed no interaction between gender as a main effect and focus. In addition, the literature on the use of intonation for pragmatic function in general, and for focus marking in particular, is still scarce for the Vietnamese language. Thus, further research is needed to establish different strategies of focus marking in different sentence positions and focus conditions.

Our study offers important insights not only for 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 even more complex distinctions. The results of our study are in line with studies on Mandarin Chinese which has a complex tone system of four tones and demonstrates changes in pitch as well as duration and intensity for focus marking (Xu 1999; Ouyang & 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 in the narrow focus condition, illustrating that there are subtle differences in the realization of focus across tonal languages.

Many African languages have simple or complex tone systems. These languages show a large diversity in how intonation is used to mark focus. Some African tonal languages use prosodic cues to indicate new or contrastive information, whereas others show no sign of focus marking with intonation (for an overview, see Zerbian, Genzel & Kügler 2010; Güldemann, Zerbian & Zimmermann 2015). Furthermore, tonal languages that utilize intonation as a focus marker do not all employ the same strategy. For example, speakers of Northern Chichewa use a rise in pitch (Downing 2008), whereas speakers of Akan use deaccentuation to mark focus (Kügler & Genzel 2012). Interestingly, the results of our study show an additional strategy of lowering pitch to mark focus. However, this strategy is restricted mainly to glottalized tones (*ngã* and *nǎng*). Thus, each language may employ particular strategies for focus marking based on its tone system. Of the nearly 2,000 African languages, many remain under-described and it is possible that other strategies will be found. Establishing the similarities and differences across these diverse systems will be a challenge in the future.

The interplay between lexical tone and intonation is a complex phenomenon that requires an intricate analysis and methodological rigor. By restricting our stimuli to a fixed structure and conducting a production study as part of a larger experimental study, we were able to record an unprecedented number of Northern Vietnamese speakers. Our finding that these speakers change their pitch contour to mark narrow focus across all six tones has important implications for our understanding of Information Structure in tonal languages. Northern Vietnamese speakers seem to process focus intonation similarly to non-tonal languages in that prosodic cues can evoke focus alternatives (see also Tjuka, Nguyễn & Spalek 2020). There is still much to be discovered at the intersection of prosody and pragmatics. The present study offered new methods that can be applied by other researchers in their studies on focus intonation in different tone systems.

## **Acknowledgments**

This project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No. GAP-677742, awarded to Katharina Spalek. The authors would like to thank Carsten Schlieve for his technical assistance.

## **Competing Interests**

The authors have no competing interests to declare.

## **Authorship contribution statement**

AT and KS initiated the study, established the procedure for the study, and wrote a first manuscript draft. KS supervised the project and worked closely together with AT and HN while developing the study. HN advised the project as a native Vietnamese speaker and recruited the participants. AT and HN created the Vietnamese stimuli for the recording. AT conducted the recording sessions, annotated the audio files, and conducted an initial data analysis. RV performed the final acoustic analysis for the paper. All authors revised the draft and agree with the final version of the manuscript.

## **Supplementary Material**

The supplementary material including the stimuli, task instruction, and the script for the data analysis can be found under the following link: <https://osf.io/6e8ua/>

## References

- Baumann, S., Becker, J., Grice, M., & Mücke, D. (2007). Tonal and articulatory marking of focus in German. *Proceedings of the 16th International Congress of Phonetic Sciences*, 1029–1032.
- Boersma, P., & Weenink, D. (n.d.). *Praat: Doing phonetics by computer* (5.1.13) [Computer software]. <http://www.praat.org>
- Brunelle, M. (2009). Tone perception in Northern and Southern Vietnamese. *Journal of Phonetics*, 37(1), 79–96. <https://doi.org/10.1016/j.wocn.2008.09.003>
- Brunelle, M. (2017). Stress and phrasal prominence in tone languages: The case of Southern Vietnamese. *Journal of the International Phonetic Association*, 47(3), 283–320. <https://doi.org/10.1017/S0025100316000402>
- Brunelle, M., Hà, K. P., & Grice, M. (2012). Intonation in Northern Vietnamese. *The Linguistic Review*, 29(1), 3–36. <https://doi.org/10.1515/tlr-2012-0002>
- Brunelle, M., & Jannedy, S. (2013). The cross-dialectal perception of Vietnamese tones: Indexicality and convergence. In *Linguistics of Vietnamese* (Vol. 253, pp. 9–34). De Gruyter Mouton. <https://doi.org/10.1515/9783110289411.9>
- Brunelle, M., Nguyễn, D. D., & Nguyễn, K. H. (2010). A laryngographic and laryngoscopic study of Northern Vietnamese tones. *Phonetica*, 67(3), 147–169. <https://doi.org/10.1159/000321053>
- Chen, Y., & Gussenhoven, C. (2008). Emphasis and tonal implementation in Standard Chinese. *Journal of Phonetics*, 36(4), 724–746. <https://doi.org/10.1016/j.wocn.2008.06.003>
- Đỗ, T. D., Trần, T. H., & Boulakia, G. (1998). Intonation in Vietnamese. In D. Hirst & A. Di Cristo (Eds.), *Intonation systems: A survey of twenty languages*. Cambridge University Press.
- Downing, L. J. (2008). Focus and prominence in Chichewa, Chitumbuka and Durban Zulu. *ZAS Papers in Linguistics*, 49, 47–65. <https://doi.org/10.21248/zaspil.49.2008.363>
- Erlewine, M. Y. (2017). Vietnamese focus particles and derivation by phase. *Journal of East Asian Linguistics*, 26(4), 325–349. <https://doi.org/10.1007/s10831-017-9156-y>
- Güldemann, T., Zerbian, S., & Zimmermann, M. (2015). Variation in information structure with special reference to Africa. *Annual Review of Linguistics*, 1(1), 155–178. <https://doi.org/10.1146/annurev-linguist-030514-125134>

- Gussenhoven, C. (1983). Focus, mode and the nucleus. *Journal of Linguistics*, 19(2), 377–417. <https://doi.org/10.1017/S0022226700007799>
- Gussenhoven, C. (2004). *The phonology of tone and intonation*. Cambridge University Press. <https://doi.org/10.1017/CBO9780511616983>
- Gussenhoven, C. (2007). Types of focus in English. In C. Lee, M. Gordon, & D. Büring (Eds.), *Topic and Focus: Cross-Linguistic Perspectives on Meaning and Intonation* (pp. 83–100). Springer Netherlands. [https://doi.org/10.1007/978-1-4020-4796-1\\_5](https://doi.org/10.1007/978-1-4020-4796-1_5)
- Hà, K. P. (2012). *Prosody in Vietnamese – Intonational form and function of short utterances in conversation* [PhD thesis, The Australian National University]. Asia-Pacific Linguistics - SEAsian Mainland Languages E-Series (SEAMLLES).
- Hoàng, T. C. (1989). *Tiếng Việt trên các miền đất nước [Vietnamese in all regions of the country]*. Hanoi National University Publishing House.
- Hole, D. (2008). EVEN, ALSO and ONLY in Vietnamese. In S. Ishihara, S. Petrova, & A. Schwarz (Eds.), *Interdisciplinary Studies on Information Structure* (Vol. 11, pp. 1–54). Universitaetsverlag Potsdam.
- Hole, D. (2013). Focus particles and related entities in Vietnamese. In *Linguistics of Vietnamese* (Vol. 253, pp. 265–304). De Gruyter Mouton. <https://doi.org/10.1515/9783110289411.265>
- Jannedy, S. (2007). Prosodic focus in Vietnamese. In S. Ishihara, S. Jannedy, & A. Schwarz (Eds.), *Interdisciplinary Studies on Information Structure* (Vol. 8, pp. 209–230). Universitaetsverlag Potsdam.
- Karlsson, A., House, D., & Svantesson, J.-O. (2012). Intonation adapts to lexical tone: The case of Kammu. *Phonetica*, 69(1–2), 28–47. <https://doi.org/10.1159/000343169>
- Koch, X., & Spalek, K. (2021). Contrastive intonation effects on word recall for information-structural alternatives across the sexes. *Memory & Cognition*, 49(7), 1312–1333. <https://doi.org/10.3758/s13421-021-01174-1>
- Krifka, M. (2006). Basic notions of information structure. In C. Féry, G. Fanselow, & M. Krifka (Eds.), *The notions of information structure*. Universitaetsverlag Potsdam.
- Maddieson, I. (2013). Tone. In M. S. Dryer & M. Haspelmath (Eds.), *The World Atlas of Language Structures Online*. Max Planck Institute for Evolutionary Anthropology. <https://wals.info/chapter/13>

- Marian, V., Blumenfeld, H. K., & Kaushanskaya, M. (2007). The Language Experience and Proficiency Questionnaire (LEAP-Q): Assessing Language Profiles in Bilinguals and Multilinguals. *Journal of Speech, Language, and Hearing Research*, 50(4), 940–967. [https://doi.org/10.1044/1092-4388\(2007/067\)](https://doi.org/10.1044/1092-4388(2007/067))
- Michaud, A., & Brunelle, M. (2016). *Information Structure in Asia: Yongning Na (Sino-Tibetan) and Vietnamese (Austroasiatic)*. Oxford University Press.
- Michaud, A., & Vu-Ngoc, T. (2004). Glottalized and nonglottalized tones under emphasis: Open quotient curves remain stable, F0 curve is modified. In B. Bel & I. Marlien (Eds.), *Proceedings of Speech Prosody* (pp. 745–748). International Speech Communication Association (ISCA).
- Miller, T., Athanasopoulou, A., Pincus, N., & Vogel, I. (2015). *The effect of focus on phonation in Northern Vietnamese tones*. Poster presented at the 89th Annual Meeting of the Linguistic Society of America, Portland, Oregon.
- Ouyang, I. C., & Kaiser, E. (2015). Prosody and information structure in a tone language: An investigation of Mandarin Chinese. *Language, Cognition and Neuroscience*, 30(1–2), 57–72. <https://doi.org/10.1080/01690965.2013.805795>
- Thompson, L. C. (1965). *A Vietnamese Grammar*. University of Washington Press.
- Tjuka, A., Nguyễn, H. T. T., & Spalek, K. (2020). *Foxes, deer, and hedgehogs: The recall of focus alternatives in Vietnamese*. *Laboratory Phonology: Journal of the Association for Laboratory Phonology*, 11(1), 1–29. <https://doi.org/10.5334/labphon.253>
- Vũ, T. P. (1982). Phonetic properties of Vietnamese tones across dialects. In D. Bradley (Ed.), *Papers in Southeast Asian Linguistics* (Vol. 8, pp. 55–75). The Australian National University.
- 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. *Journal of Phonetics*, 70, 86–116. <https://doi.org/10.1016/j.wocn.2018.03.002>
- Wood, S. (2022). Mixed GAM computation vehicle with automatic smoothness estimation. R package version 1.8-40. <https://cran.r-project.org/web/packages/mgcv/index.html>
- Xu, Y. (1999). Effects of tone and focus on the formation and alignment of F0 contours. *Journal of Phonetics*, 27(1), 55–105. <https://doi.org/10.1006/jpho.1999.0086>

- Xu, Y. (2011). Speech prosody: A methodological review. *Journal of Speech Sciences*, *1*(1), 85–115. <https://doi.org/10.20396/joss.v1i1.15014>
- Yan, M., & Calhoun, S. (2019). Priming effects of focus in Mandarin Chinese. *Frontiers in Psychology*, *10*, 1–16. <https://doi.org/10.3389/fpsyg.2019.01985>
- Zerbian, S., Genzel, S., & Kügler, F. (2010). Experimental work on prosodically-marked information structure in selected African languages (Afroasiatic and Niger-Congo). In M. Hasegawa-Johnson, A. Bradlow, J. Cole, K. Livescu, J. Pierrehumbert, & C. Shih (Eds.), *Proceedings of Speech Prosody* (Vol. 100976, pp. 1–4). [https://www.isca-speech.org/archive\\_v0/sp2010/papers/sp10\\_976.pdf](https://www.isca-speech.org/archive_v0/sp2010/papers/sp10_976.pdf)