

PARALLEL ENCODING OF FOCUS AND INTERROGATIVE MEANING IN THAI LEARNERS' MANDARIN

Zenghui Liu, Lei Zeng, Shufen Liang
Yunnan University, China

zenghui_liu@ynu.edu.cn, zenglei@mail.ynu.edu.cn, liangshufen@mail.ynu.edu.cn

ABSTRACT

This study investigates the parallel encoding of focus and interrogative meaning in Mandarin produced by advanced Thai learners of Mandarin. Mandarin SVO statements and declarative questions with varying information structure were elicited through two picture-mediated tasks. The speakers were native speakers of Thai who were advanced learners of Mandarin. Our data shows that advanced Thai learners of Mandarin vary duration, pitch span, pitch maximum and pitch minimum for encoding focus in statements, similar to native speakers of Mandarin. However, their use of prosodic cues to encode focus in declarative questions is quite limited. The present study thus provides evidence that advanced Thai learners of Mandarin mark focus prosodically in their L2, while they are not native-like in terms of parallel encoding of focus and interrogative meaning.

Keywords: second language acquisition, prosody, prosodic-focus marking, interrogative meaning

1. INTRODUCTION

Prosody plays important roles in speech to convey communicative meanings, such as highlighting new information in a sentence (i.e. focus) and encoding interrogative meanings [6, 7, 9, 14, 16, 18]. Previous studies support that components of intonation, such as focus and interrogative meaning, are defined and organized by individual communicative functions that are independent of each other in languages [10, 15]. For example, in Mandarin, [10, 15] showed that native speakers of Mandarin encode focus by expanding the pitch span and lengthened the duration of the focal constituent, compressing and lowering the pitch of the post-focal constituent, but leaving that of the pre-focal constituent largely unaffected in read speech. Similar results were also reported in semi-spontaneous speech [16]. Further, Liu & Xu [10] revealed an interaction between focus and interrogative meaning in the form of a boost to the pitch raising by the question starting from the focal constituent. Yuan [17] found that focus at the end of a sentence makes statement intonation harder to identify but makes question intonation easier to identify. Together, previous studies revealed a quite

complex mechanism of parallel encoding of focus and interrogative meaning in Mandarin.

Due to the observed use of prosody for encoding focus in languages, such as in Mandarin, the acquisition of prosodic-focus marking in L2 learners has recently received considerable attention [3, 4, 8, 12]. It has been suggested that the acquisition of prosodic focus marking is quite difficult for L2 learners. For example, Chen et al. [3] examined advanced English learners of Mandarin and found that the learners produced focus-related duration changes in a manner similar to native Mandarin speakers. However, those advanced learners did not show native-like patterns of in-focus changes in intensity on Tone 2, mean pitch on Tone 1, and pitch span on Tone 4.

Despite the considerable attention paid on the acquisition of prosodic-focus marking in L2 acquisition, little attention has been paid on the acquisition of parallel encoding of focus and interrogative meaning. The questions that arise for advanced learners of Mandarin are thus: 1) whether and how advanced learners of Mandarin prosodically distinguish statement and declarative sentence-medially; 2) whether and how advanced learners of Mandarin acquire parallel encoding of focus and interrogative meaning. We address these questions by examining the Mandarin production of five advanced Thai learners of Mandarin, who averagely had more than five years of Mandarin-learning experience in China and passed HSK (*Hànyǔ Shuǐpíng kǎoshì* translated as the *Chinese Proficiency Test*) level 5.

2. METHODOLOGY

2.1. Picture-matching & picture-marking game

We adopted and developed a picture-matching and picture-marking game [2, 11, 16] to elicit semi-spontaneously produced statements and declarative questions respectively with varying information structure. In the picture-matching game, three piles of pictures were used: the experimenter and the participant each held a pile of pictures ordered in a certain sequence; the third pile of pictures were scattered around on a table. In the experimenter's pictures (Pile A), there was always something missing, like a subject, an action (verb), an object, or

all three. The participant's pictures (Pile B) all contained a complete event. The participant's task was to help the experimenter with sorting out pictures from her own pile and the third pile (Pile C) that went together.

In the picture-marking game, the participant needed to seek information of various scopes from the experimenter by asking a declarative question in each trial. In the game, two piles of pictures and a marker were used: the first pile of pictures ordered in a certain sequence was put in front of the experimenter and participant (Pile D); the second pile of pictures was stored in a box and put next to the experimenter (Pile E); the marker was held by the participant. In Pile D, there was always information missing, like a subject, an action (verb), an object, or all three. In the top right corner of each picture of the first pile, a small picture was covered by a sticker. This small picture provided the missing part which could be correct or incorrect. The second pile of pictures (stored in the box, Pile E) all contained a complete and correct event. The participant's task was to mark the correctness of the small picture on the sticker.

By using these two experiments, SVO sentences with the same word order and lexical components, which differs in sentence type (i.e. statement vs. question) were elicited.

2.2. Experimental materials

Both statements and declarative questions in four focus conditions were elicited via the picture-marking and picture-matching game: narrow-focus on the subject NP in sentence-initial position (NF-i), narrow-focus on the verb in sentence-medial position (NF-m), narrow-focus on the object NP in sentence-final position (NF-f), and broad focus (BF). The focus condition and sentence type were set up by varied context, as illustrated in examples (1) to (2), where the focal constituents appears in square brackets.

Target sentence: xiǎo māo jiǎn shū
小 猫 剪 书
The cat cuts the book

(1) NF-i in statement:

Ex: Look! The book! There is also a pair of scissors. It looks like someone cuts the book. Who cuts the book?

Pa: [THE CAT] cuts the book.

(2) NF-i in declarative question:

Ex: Look! The book! There is also a pair of scissors. It looks like someone cuts the book. Could you open the sticker and take a look at the small

picture? Then you can ask me a question, I will help you to check in my box.

Pa: [THE CAT] cuts the book?

We included all the Mandarin lexical tones in the experimental design, including Tone 1 (high level tone), Tone 2 (rising tone), Tone 3 (dipping tone) and Tone 4 (falling tone). The SVO sentences were constructed by selecting four subject-noun, verbs and objects in each tone. This resulted in 64 Subject-verb-object combinations. Then, these combinations were combined with 4 focus conditions. In total, we had 512 target sentences, and 256 for each picture-mediated task.

2.3. Participants and procedure

Five female advanced Thai learners of Mandarin participated in our experiments (age range: 21-25, average age = 22.6, SD = 1.62). All participants were advanced Thai learners of Mandarin who had passed the HSK test level 5. The highest level of HSK test is level 6. All participants have more than 5 years of Mandarin-learning experiences. At the time of testing, all the participants were studying in China.

All participants were tested individually by a female experimenter, who was a native speaker of Mandarin, at the Phonetics Laboratory at Yunnan University. The experiments were recorded using a portable ZOOM H1 digital recorder at a 44.1 kHz sampling rate and 16 bit accuracy. Each session was also video-taped.

2.4. Acoustic annotation

The auditory recordings from each participant were first orthographically annotated so that the participant's responses could be selected. A strict selection criterion of the usable data was applied, i.e. a sentence was considered usable only if it contained no self-correction and hesitation and was uttered as a response in the context. In total 96% of the obtained responses (N = 1852) were included in further analysis. The usable sentences were subsequently acoustically annotated in Praat [1]. A textgrid with four interval tiers (word, tone, sentence, comment), and two point tiers (pitch, duration) was created for each target sentence. Every sentence was segmented into words in the "word" tier, then landmarks demarcating verb onset and offset, and the locations of pitch-maximum and pitch-minimum within the verb were added to the "duration" and "pitch" tiers. The landmarks for the onset and offset of verbs were determined according to the information in the waveform and spectrogram.

The pitch values of the pitch landmarks and the time values of the word boundaries were subsequently extracted via Praat scripts. Two measures from these values were calculated: word duration (i.e. offset time minus onset time) and pitch range (i.e. the difference between the maximum pitch and the minimum pitch). We found that the majority of the target verbs with Tone 3 (dipping tone) were produced with either creaky voice or complex tonal patterns. In order to present a comprehensible report, the present study will focus on the results of Tone 1 (high level tone), Tone 2 (rising tone) and Tone 4 (falling tone). In 98 of the usable responses, the pitch values could not be reliably measured. These responses were thus excluded from the analysis.

3. ANALYSIS AND RESULTS

3.1. Statistical Analyses

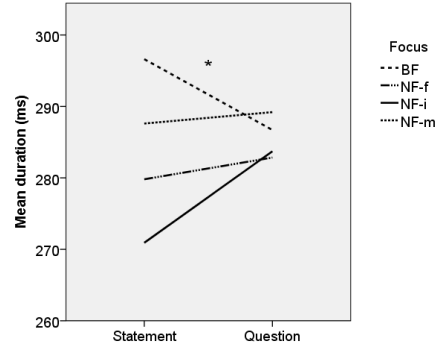
Statistical analyses were conducted using mixed-effects modelling in R [13]. In all models, the SENTENCE TYPE, TONE, and FOCUS were included as fixed factors, while the speaker (i.e. the participants) and sentence (i.e. the order of elicitation) were included as random factors. SENTENCE TYPE had two levels (i.e. “statement” and “question”), FOCUS had four levels (i.e. “BF”, “NF-f”, “NF-i”, and “NF-m”), and TONE referred to the lexical tones of the target verbs, which had three levels (i.e. “Tone 1”, “Tone 2”, and “Tone 4”). Outcome variables were the duration, pitch span, pitch maximum, and pitch minimum of the verbs. Following [5], our models were constructed and evaluated in a stepwise fashion. When building the models, only the factors and interactions that significantly improved the fit of the model were retained until the best fit model was determined.

3.2. Duration

For the analysis of duration, we found that the best-fit model was the model which contained the main effects of TONE, $\chi^2(2) = 34.233, p < .001$, FOCUS, $\chi^2(3) = 20.848, p < .001$, SENTENCE TYPE, $\chi^2(1) = 6.219, p < .05$, and two-way interactions between SENTENCE TYPE and FOCUS, $\chi^2(3) = 11.305, p < .05$, SENTENCE TYPE and TONE, $\chi^2(2) = 21.69, p < .001$. The main effect of SENTENCE TYPE was such that the duration of the target verbs in declarative questions were significantly shorter than their counterparts in statements ($p < .001$). By further exploring the effect of TONE and FOCUS within the sentence type, we found that the duration of the target verbs only varied with focus condition in statement ($p < .05$),

but not in questions, regardless of tones, as shown in Fig. 1.

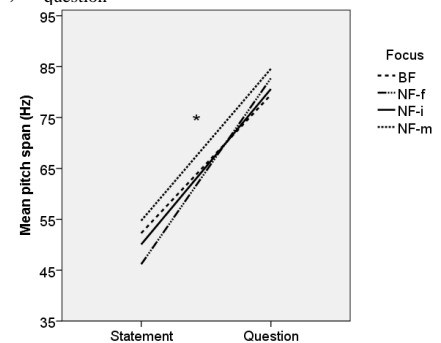
Figure 1: Mean duration of sentence-medial verbs in statement vs. question, $n = 5, N_{\text{total}} = 1852, N_{\text{question}} = 916$.



3.3. Pitch span

For the analysis of pitch span, we found that the best-fit model was the model which contained the main effects of TONE, $\chi^2(2) = 160.87, p < .001$, SENTENCE TYPE, $\chi^2(1) = 283.75, p < .001$, and a two-way interaction between SENTENCE TYPE and TONE, $\chi^2(2) = 8.858, p < .05$. The main effect of SENTENCE TYPE was such that the pitch span of the target verbs in declarative questions were significantly wider than their counterparts in statements ($p < .001$). By further exploring the effect of TONE and FOCUS within the sentence type, we found that the pitch span of the target verbs only varied with focus condition in statement ($p < .05$), but not in questions, regardless of tones, as shown in Fig. 2.

Figure 2: Mean pitch span of sentence-medial verbs in statement vs. question, $n = 5, N_{\text{total}} = 1747, N_{\text{question}} = 889$.

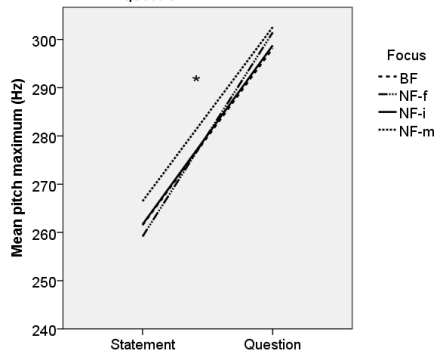


3.4. Pitch maximum

For the analysis of pitch maximum, we found that the best-fit model was the model which contained the main effects of TONE, $\chi^2(2) = 313.66, p < .001$, SENTENCE TYPE, $\chi^2(1) = 537.47, p < .001$, and a two-way interaction between SENTENCE TYPE and

TONE, $\chi^2(2) = 34.541, p < .001$. The main effect of SENTENCE TYPE was such that the pitch maximum of the target verbs in declarative questions were significantly higher than their counterparts in statements ($p < .001$). By further exploring the effect of TONE and FOCUS within the sentence type, we found that the pitch span of the target verbs only varied with focus condition in statement ($p < .01$), but not in questions, regardless of tones, as shown in Fig. 3.

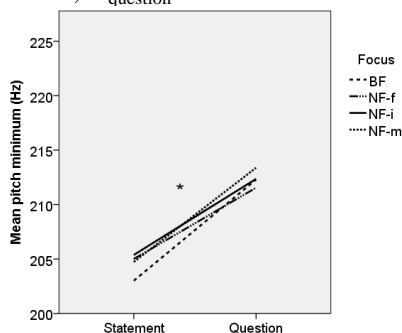
Figure 3: Mean pitch maximum of sentence-medial verbs in statement vs. question, $n = 5$, $N_{total} = 1747, N_{question} = 889$.



3.5. Pitch minimum

For the analysis of pitch minimum, we found that the best-fit model was the model which contained the main effects of TONE, $\chi^2(2) = 249.42, p < .001$, SENTENCE TYPE, $\chi^2(1) = 51.629, p < .001$, and a two-way interaction between SENTENCE TYPE and TONE, $\chi^2(2) = 22.942, p < .001$. The main effect of SENTENCE TYPE was such that the pitch minimum of the target verbs in declarative questions were significantly higher than their counterparts in statements ($p < .001$). By further exploring the effect of TONE and FOCUS within the sentence type, we found that the pitch span of the target verbs only varied with focus condition in statement ($p < .01$), but not in questions, regardless of tones, as shown in Fig. 4.

Figure 4: Mean pitch minimum of sentence-medial verbs in statement vs. question, $n = 5$, $N_{total} = 1747, N_{question} = 889$.



4. DISCUSSION & CONCLUSION

Our results showed that advanced Thai learners of Mandarin shortened the duration of the sentence-medial constituent of Tone 1, 2 and 4 in declarative questions relative to its counterpart in statements, regardless of focus condition. Further, they expanded the pitch span of the sentence-medial constituent of Tone 1, 2 and 4 in declarative questions relative to its counterpart in statements, regardless of focus conditions. Specifically, they raised the pitch maximum and pitch minimum of the sentence-medial constituent of Tone 1, 2 and 4 in declarative questions relative to its counterpart in statements, regardless of focus condition.

Interestingly, our results revealed that advanced Thai learners of Mandarin can vary prosodic cues, including duration, pitch span, pitch maximum and pitch minimum for encoding different focus conditions in statement. However, the use of prosodic cues to parallel encoding focus and interrogative meaning seems absent in advanced Thai learners' Mandarin. Specifically, there is no evidence that advanced Thai learners of Mandarin varied duration, pitch span, pitch maximum, and pitch minimum for encoding focus prosodically in declarative questions. In comparison with previous studies of American English learners of Mandarin [3], our advanced Thai learners of Mandarin successfully master the use of duration and pitch for encoding focus in statement. It might be explained by the fact that our learners were emerged in L2-speaking environment, i.e. China, for more than five years. However, as we only examined the sentence-medial constituent and exclude the Tone 3 from the analysis, and, further analysis on Tone 3 and sentence-final constituent is needed.

To conclude, the present study reveals a confliction in the use of prosody for parallel encoding of different intonation components, i.e. focus and interrogative meaning, in learners' L2. It suggests that the successful use of prosody to encode one component of intonation, e.g., focus, does not guarantee the parallel encoding of different intonation components. Thus, the present study provides evidence that components of intonation, such as focus and interrogative meaning, are defined and organized by individual communicative functions that are independent of each other [10, 15], from an acquisition perspective.

7. REFERENCES

- [1] Boersma, P., Weenink, D. 2006. Praat: Doing phonetics by computer (Version 5.4.01). Retrieved from <http://www.praat.org/>

- [2] Chen, A. 2010. Is there really an asymmetry in the acquisition of the focus-to-accentuation mapping? *Lingua*, 120(8), 1926-1939.
- [3] Chen, Y., Guion-Anderson, S. 2012. Prosodic realization of focus in Mandarin by advanced American learners of Chinese. *The Journal of the Acoustical Society of America*, 131(4), 3234-3234.
- [4] Chen, Y., Xu, Y., Guion-Anderson, S. 2014. Prosodic realization of focus in bilingual production of Southern Min and Mandarin. *Phonetica*, 71(4), 249-270.
- [5] Field, A., Miles, J., Field, Z. 2012. *Discovering statistics using R*. Los Angeles: SAGE Publication Ltd.
- [6] Gussenhoven, C. 2004. *The phonology of tone and intonation*: Cambridge University Press.
- [7] Haan, J. 2002. *Speaking of questions: An exploration of Dutch question intonation*. (Doctoral dissertation), Radboud University, Nijmegen.
- [8] He, X., Hanssen, J., van Heuven, V., Gussenhoven, C. 2011. *Phonetic implementation must be learnt. Native versus Chinese realization of focal accent in Dutch*.
- [9] Ladd, D. R. 1996. *Intonational phonology*. Cambridge: Cambridge University Press.
- [10] Liu, F., Xu, Y. 2005. Parallel encoding of focus and interrogative meaning in Mandarin intonation. *Phonetica*, 62(2-4), 70-87.
- [11] Liu, Z., Chen, A., Van de Velde, H. 2014. *Prosodic focus marking in Bai*. *Proc. Speech Prosody 2014* Dublin, 628-631.
- [12] Nava, E., Zubizarreta, M. L. 2008. *Prosodic transfer in L2 speech: Evidence from phrasal prominence and rhythm*. *Proc. Speech Prosody 2008* Campinas, 335-338.
- [13] R Core Team. 2014. *R: A language and environment for statistical computing*. *R Foundation for Statistical Computing, Vienna, Austria*.
- [14] Xu, Y. 1999. Effects of tone and focus on the formation and alignment of f0 contours. *Journal of phonetics*, 27(1), 55-105.
- [15] Xu, Y. 2005. Speech melody as articulatorily implemented communicative functions. *Speech communication*, 46(3-4), 220-251.
- [16] Yang, A., Chen, A. 2018. The developmental path to adult-like prosodic focus-marking in Mandarin Chinese-speaking children. *First Language*, 38(1), 26-46. doi:10.1177/0142723717733920
- [17] Yuan, J. 2006. Mechanisms of question intonation in Mandarin *Chinese spoken language processing* (pp. 19-30): Springer.
- [18] Yuan, J., Shih, C., Kochanski, G. P. 2002. *Comparison of declarative and interrogative intonation in Chinese*. *Proc. Speech Prosody 2002* Aix-en-Provence