Production and Perception of Focus in PFC and non-PFC Languages:

Comparing Beijing Mandarin and Hainan Tsat

Bei Wang*[1], Chenxia Li[1], Qian Wu[1], Xiaxia Zhang[1], Baofeng Wang[1], Yi Xu[2]

[1] Institute of Chinese Minority Languages and Literatures, Minzu University of China, Beijing, China
[2] Department of Speech, Hearing and Phonetic Sciences, University College London, London, UK

bjwangbei@gmail.com, li_200851.yeah.net, mdwuqian@qq.com, Xiaxia523@163.com,
shen8025@126.com, yi.xu@ucl.ac.uk

Abstract

Prosodic marking of focus has been found to be typologically different in terms of existence of post-focus compression in F0 and intensity (PFC). In the current production experiments, we found that PFC showed in Mandarin, but not in Tsat (a language spoken in Hainan, China) or in Tsat-Mandarin (Mandarin spoken by Tsat people). The perception experiments further showed that focus perception in Tsat-Mandarin and Tsat was relatively low (about 30%). An interesting finding was that the perception of focus in Mandarin by Tsat listeners was much lower than that by Mandarin listeners (54.8% vs. 75.6%). In other words, it revealed the difficulty of speakers from a non-PFC language associating PFC with focus perception. It could partly explain why PFC is hard to be passed through language contact.

Index Terms: post-focus compression (PFC), language contact, focus

1. Introduction

Focus is one of the most frequently used communicative functions, which is to highlight certain part of a sentence for some pragmatic reasons, for instance, to make a contrast, to make a correction, or to provide information for a *wh*-question, etc. [1]. In many languages, focus can be marked prosodically with lengthened duration, raised F0, expanded pitch range and a sharp post-focus compression in F0 and intensity (PFC) (e.g., for English: [2], [3]; for Mandarin: [4], [5]). Recently, it has been found that the means of prosodic marking of focus is not universal. In many African languages [6] and languages in South China [7-10], focus is mostly marked with lengthened duration and sometimes raised F0, but not PFC.

It has been found that focus can be well perceived in languages with PFC. For instance, the correct perception of focus was about 90% in Mandarin[10, 11] and Uygur [7] for initial and medial focus. In other languages, such as Finnish[12], Dutch and Italian[13], English[14] etc, focus can also be well perceived. However, in Taiwanese, which lacked PFC, the correct perception of focus was just about 60%[10]. Xu et al.[10] argued that PFC is probably the most effective cue to focus perception. F0 rising and durational lengthening are helpful for focus perception, but with limited power. For instance, in Mandarin the correct perception of final focus was about 60%[10, 11], which also lacks PFC as it is already at the end of a sentence.

From the above literature review, we can see that focus is important for speech communication, and can be marked prosodically. In addition, PFC is important for focus perception. However, why are there so many languages lacking PFC? Can a language without PFC attain it through language contact?

Xu et al.[10] have proposed *the inheritance hypothesis*, according to which all cases of PFC are inherited from a proto-language with PFC. Although such a claim needs support from a large amount of experimental data from various fields, two findings could be taken as supporting evidence. *First*, PFC does not seem to be able to emerge automatically in a language, as there are so many languages in Africa [6] and South China [7] lacking PFC. *Second*, PFC does not transfer easily from one language to another through contact. In Taiwan, PFC has been found to be lost in Taiwan Mandarin through close contact with Taiwanese [10]. It is also found that Cantonese speakers, who were brought up in London and spoke English as their dominant language, did not transfer PFC to their Cantonese [15]. And even PFC in the English of those Cantonese speakers was affected [15]. One of the possible reasons to explain why PFC is hard passed from one language to another may be because, despite its benefit for listeners of a PFC language, PFC is hard to be noticed by listeners of a non-PFC language. This possibility, however, has never been experimentally tested.

The current paper aims to study the production and perception of Tsat language. As introduced in[16], Tsat (also known as Utsat, Utset, Hui, or Hainan Cham) is a language spoken in Sanya, Hainan, China by the Utsuls. It is a member of the Malayo-Polynesian group within the Austronesian language family, and is related to the Cham languages. The language is now spoken by about 4,500 people in Yanglan and Huixin villages in Sanya. Unusually for a Malayo-Polynesian language, Tsat has developed into a full-fledge tone language, as a result of contact with Hlai languages (Tai-Kadai), the monosyllabic and tonal Min dialects of Chinese, and more recently, under the quite intense influence of Mandarin. We aim to answer four questions in this study.

(1) How is focus encoded prosodically in Tsat?
(2) How is focus encoded in Mandarin spoken by Tsat speakers?
(3) How well can focus be perceived in both Tsat and Mandarin by Tsat speakers?
(4) How well can focus in Beijing Mandarin be perceived by Tsat speakers?

The general goal is to find out whether PFC is absent in Tsat as predicted by *the inheritance hypothesis* [10]. And, if it is indeed absent, we expect that PFC also lacks in Mandarin spoken by Tsat speakers (to keep it simple, we call it Tsat-Mandarin in

---

* Corresponding author.

this paper). If so, we further expect that the perception of focus by Tsat speakers is poor for both languages. In addition, we are interested in how well Tsat speakers can perceive focus in Beijing Mandarin. With the new findings, we wish to shed light on theories about the distribution and origin of PFC and to understand why PFC is hard to pass through contact easily.

2. Speech Production Experiment

2.1. Materials

Two target sentences were constructed (see below). The criteria are simple. We chose three target words at sentence initial, medial and final positions with as many sonorant consonants as possible. A paradigm of corrective focus was applied. For instance, to emphasize *Mom* in the target sentence, we used a sentence with a structure as "It is not Daddy. **Mom** asked the younger sister to go out and buy some areca catechu." Below are the sentences used in this study.

Tsat Sentence 1:
ʔa[11]ma[33] i[55] thai[11]mai[33] na:u[32] li[55] nɛ[33].
mother ask younger sister go buy areca-nut
"Mother asked the younger sister to go and buy areca-nut."

Tsat Sentence 2:
ʔa[11]na[11] li[24]ni[33] tu[24] ʔia[33] ma[33] tso[33]mai[33] hui[24].
grandma in the act of take water give granddaughter drink
"Grandma is getting water for the granddaughter to drink."

The Tsat sentences were translated into Chinese as listed below. To compare our results with previous studies in [4, 10], the sentence used in those studies was also added (Sentence 3).

Mandarin Sentence 1:
ma[55]ma[2] tɕiau[51] mei[51]mei[1] tɕhy[51] mai[214] biŋ[55]laŋ[35].
mother ask younger sister go buy areca-nut

Mandarin Sentence 2:
nai[214]nai[4] tʂəŋ[51]tsai[51] na[35] ʂuei[214] kei[214] suən[55]ny[2] xə[55].
grandma in the act of take water give granddaughter drink

Mandarin Sentence 3:
ma[55]ma[55] mo[55] mau[55]mi[55].
mother stoke kitty.

In total, there were 2 (target sentences) × 4 focus conditions × 3 repetitions =24 Tsat sentences and 3 (target sentences) × 4 focus conditions × 3 repetitions =36 Mandarin sentences for each speaker.

2.2. Participants

Eight native speakers of Tsat (3 female and 5 male), aged between 18 and 54 participated in the experiment. All of them were brought up and lived in a Tsat village. They also spoke standard Mandarin as their second native language. They did not report any speech or hearing disorders.

For Beijing Mandarin, four female native Mandarin speakers, aged between 23-38, read three Chinese sentences. They were all from North China, and spoke Beijing Mandarin as their native language.

2.3. Recording procedure

All the speakers were recorded in a quiet room in the Tsat village, using the AudioRec recording software with a Rode NT1-A microphone. The recording was saved in a Lenovo laptop as wav files, with a sampling rate of 44.1 kHz. The focused words were highlighted. A 22-year-old female native Tsat speaker helped the

authors with the recording. Once she noticed any error or inappropriate intonation, she would ask the speaker to read the target sentence again. All the sentences were repeated three times with a random order for each repetition for each speaker.

2.4. Acoustic measurement

The target sentences were extracted and saved as separate wav files. The syllabic boundaries of all the sentences were labeled by hand and the vocal cycles were hand checked. A Praat script, ProsodyPro [17] was used to extract F0 and duration of each syllable. The F0 values were converted from Hz to semitone (st) using formula (1).

$$f_{st} = 12 \times \log_2(f_0 / 50) \qquad (1)$$

2.5. Results

2.5.1. F0

Fig. 1 and Fig. 2 demonstrate the time-normalized F0 contours of the two target sentences of Tsat and Tsat-Mandarin under four focus conditions. The intonational contours of Mandarin sentences read by Mandarin speakers are displayed in Fig. 3. Due to space limitation, we only displayed sentence 1 and 3 for the Mandarin sentences, which were used in the perception experiment. Each value is averaged across 8 speakers and their three repetitions.
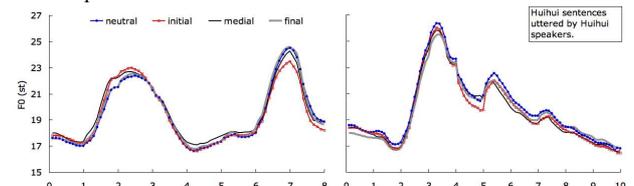

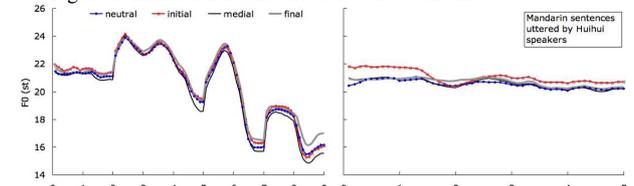Figure 1. Intonational contours of two Tsat sentences

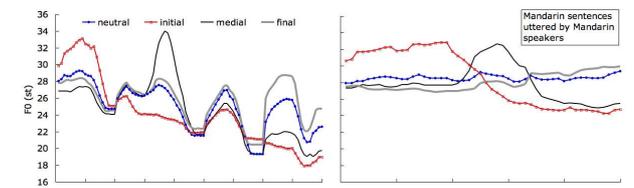Figure 2. Intonational contours of two Tsat-Mandarin sentences

Figure 3. Intonation contours of two Beijing Mandarin sentences

As predicated by *the inheritance hypothesis* [10], Tsat would belong to non-PFC languages. We can see in Fig. 1 and 2 that there is almost no F0 variation due to focus in all the sentences by Tsat speakers. For Mandarin, in contrast, we can see in Fig. 3 that focus is marked with a tri-zone F0 manipulation as reported in[4, 10, 11], that is, F0 is raised in the focused word, and lowered and compressed in the post-focus words, while remaining largely intact in pre-focus words. Sentence 2 shows a very similar pattern to the other two Mandarin sentences for both Tsat-Mandarin and Mandarin. Due to space limitation, we did

not show the statistic results here. But, all the observations above are supported statistically.

2.5.2. Duration

Averaged word duration of the three target words in Tsat and Tsat-Mandarin sentences are displayed in Fig. 4.
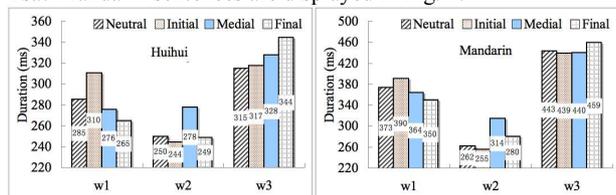

Figure 4. Duration of the three target words under four focus conditions in Tsat and Tsat-Mandarin.

We can see in Fig. 4 that a focused word is lengthened. A two-way measured ANOVA of Tsat showed effect in word position ($F_{(2, 14)}=46.351$, $p<.001$) but not in focus ($F_{(3, 21)}=2.841$, *n.s.*). The interaction between these two variables is significant ($F_{(6, 42)}=11.847$, $p<.001$). Simple effect tests show that the focused word is mostly significantly longer than its counterpart in the other three positions. The same statistical analysis on Tsat-Mandarin showed a very similar pattern. We concluded that Tsat speakers used durational lengthening to mark focus.

The analysis on word duration of Mandarin by native Mandarin speakers also showed durational lengthening on focused words, which is in consistent with previous studies[4, 10, 11].

We can see that focus is marked differently between Tsat, Tsat-Mandarin on one hand and Beijing Mandarin on the other.

3. Speech Perception Experiment

3.1. Materials

Three sets of sentences were used in the perception experiment, which were all from the production experiments.

SET 1:   Two Tsat sentences
SET 2:   Two Tsat-Mandarin sentences (sentence 1 and 3)
SET 3:   Two Mandarin sentences (sentence 1 and 3)
    One reading of the sound files from 4 female Tsat speakers and 4 female Mandarin speakers were used.

3.2. Participants

Two groups of listeners participated in the perception experiment. One group was 15 native Tsat speakers, aged from 18 to 54 and the other was 15 native Mandarin speakers, aged from 20-24, half male and half female. For the Tsat listeners, 6 of them took part in the production experiment as well. The Tsat participants listened to all three groups of sentences, whereas Mandarin participants just listened the Mandarin sentences (set 2 and 3).

3.3. Procedure

The participants were asked to listen to the sentences and judge whether the first word (initial focus), the medial word, the last word (final focus), or none of them (neutral focus) was emphasized. The perception experiment was run with Experiment MFC [11] in Praat. The participants were tested individually, and could only listen to each sentence once each time when the sentence was played. The participants listened to

all the sentences twice, with random order each time. The response from the second trial was analyzed.

Tsat listeners were tested in the Tsat village. And, Mandarin listeners were tested in Beijing.

3.4. Results

The correct perception rate of each condition is displayed in Fig. 5. We can see that the correct perception of Tsat by Tsat listeners is the lowest (30%). For Tsat-Mandarin, the correct perception of focus is a little higher for both Tsat listeners (34.8%) and Mandarin listeners (37.9%). Since there are no reliable acoustic cues for focus perception in these two languages, it is true for listeners in the language background of with or without PFC. And, for Mandarin, the perception of focus is much better than Tsat and Tsat-Mandarin. However, the hitting rate of the Tsat listeners (54.8%) is much lower than that of the Mandarin listeners (75.6%).

Two separate one-way repeated measures ANOVAs showed a main effect of utterance for both Tsat ($F_{(2, 28)}=19.727$, $p<.001$) and Mandarin listeners ($F_{(1, 14)}=243.871$, $p<.001$). It confirms that Mandarin listeners can detect focus with much higher accuracy for Beijing Mandarin than for Tsat-Mandarin. For Tsat listeners, a post-hoc test (S-N-K) showed that the correct perception rate of Mandarin utterances was higher than that of Tsat-Mandarin and Tsat utterances, with no significant difference between the latter two.

For Mandarin utterances produced by Tsat and Mandarin speakers, a mixed one-way repeated measures ANOVA was carried out, with listener as a between-group variable and utterance as a within-group variable. It showed a main effect of utterance ($F_{(1, 28)}=155.015$, $p<.001$) and an interaction between them ($F_{(1, 28)}=14.81$, $p=0.001$). The follow-up paired T tests with Bofenirre adjustment showed that the difference between Tsat and Mandarin listeners on Tsat-Mandarin is not significant ($T_{(14)}=1.125$, *n.s.*), whereas that on Mandarin is significantly different($T_{(14)}=3.996$, $p=.001$).

On average, the correct perception of Tsat and Tsat-Mandarin is 34.2% for all the listeners. It indicates that there is no reliable cue on prosodic marking of focus in Tsat and Tsat-Mandarin. What's interesting to us is that the perception of focus in native Mandarin sentences is only around 55% for Tsat listeners, although there is an obvious F0 variation on marking focus (see Fig. 3).

One possibility can be that the hitting rate of final focus and neutral focus was low for focus perception of Mandarin. To get a clear picture, we calculated the correct perception of initial and medial focus (see Fig.6) separately from that of neutral and final focus (see Fig.7). The former two focus conditions have PFC, where the later two do not.

We can see in Fig. 6 and 7 that the correct recognition of focus between initial + medial and neutral + final is not much different in Tsat and Tsat-Mandarin for both Tsat and Mandarin listeners. For Mandarin, the initial and medial conditions were with an about 10% higher hitting rate than the neutral and final conditions for Tsat listeners, and the difference was about 30% for Mandarin listeners. In turn, the correct perception of initial and medial focus was still relatively low for Tsat listeners (60%), whereas that was 90% for Mandarin listeners. It indicates that Tsat listeners did not seem to have a reliable clue to detect focus.
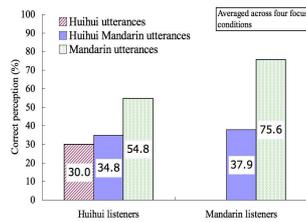
Figure 5. Correct perception of focus of initial and medial focus
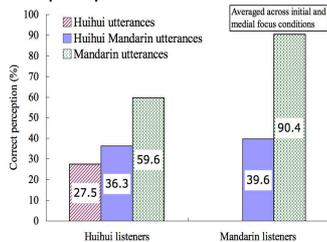

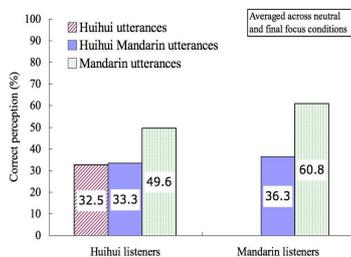Figure 6. Correct perception of focus averaged with initial and medial focus conditions

Figure 7. Correct perception of focus averaged with neutral and final focus conditions

4. Discussion and Conclusions

From the above analysis, we can see that focus in both Tsat and Tsat-Mandarin does not cause any systematic variation in F0 and PFC is clearly absent, as predicted by *the inheritance hypothesis* [10]. There is some duration lengthening in focused words. Although *the inheritance hypothesis* [10] on the origin of PFC still requires experimental studies in various fields and in many languages, at least, the current findings are not against it.

In accordance with the study on Taiwanese and Taiwan Mandarin[10], the correct perception of focus in Tsat and Tsat-Mandarin is very low. The new finding here is that Tsat speakers perceived focus in Mandarin with a much lower correct rate than native Mandarin listeners. It indicates that PFC is probably hard to be noticed by speakers from a non-PFC language. Then, it explains partly why PFC is hard to be learned through language contact easily.

As stated in [18], the genetically closest language to Tsat is the Northern Roglai of Vietnam, which also lacks PFC[19]. Tsat is now radically different from N. Roglai. Tsat has changed in both phonology and syntax. Tsat was changed from sesquisyllabic and atonal to monosyllabic and fully tonal. It is interesting that in the past years, Tsat did not develop PFC automatically, although it had developed many features of Chinese.

5. Acknowledgements

6. References

[1] M. A. K. Halliday, *An Introduction to Functional Grammar*. London: Edward Arnold Ltd, 41 Bedford Square, 1985.

[2] W. E. Cooper*, et al.*, "Acoustical aspects of contrastive stress in question-answer contexts," *Journal of the Acoustical Society of America,*, vol. 77, pp. 2142-2156, 1985.

[3] Y. Xu and C. X. Xu, "Phonetic realization of focus in English declarative intonation," *Journal of Phonetics,* vol. 33, pp. 159-197, 2005.

[4] Y. Xu, "Effects of tone and focus on the formation and alignment of f0 contours," *Journal of Phonetics,* vol. 27, pp. 55-105, 1999.

[5] B. Wang and Y. Xu, "Differential prosodic encoding of topic and focus in sentence-initial position in Mandarin Chinese," *Journal of Phonetics,* vol. 39, pp. 595-611, 2011.

[6] S. Zerbian, "Investigating prosodic focus marking in Northern Sotho.," in *Focus strategies: evidence from African languages*, K. Hartmann and M. Zimmermann, Eds., ed Berlin: Mouton de Gruyter, 2007, pp. 55-79.

[7] B. Wang*, et al.*, "Prosodic encoding of focus in six languages/dialects in China," presented at the 17th ICPhS, Hong Kong, 2011.

[8] W. Gu and T. Lee, "Effects of tonal context and focus on Cantonese F0," presented at the The 16th International Congress of Phonetic Sciences, Saarbrucken, Germany, 2007.

[9] W. L. Wu and Y. Xu, "Prosodic Focus in Hong Kong Cantonese without Post-focus Compression," in *Speech Prosody 2010*, Chicago. USA., 2010.

[10] Y. Xu*, et al.*, "Prosodic focus with and without post-focus compression (PFC): A typological divide within the same language family?," *The Linguistic Review,* vol. 29, pp. 131-147, 2012.

[11] F. Liu and Y. Xu, "Parallel encoding of focus and interrogative meaning in Mandarin intonation," *Phonetica,* vol. 62, pp. 70-87, 2005.

[12] M. Vainioa and J. Jarvikivi, "Tonal features, intensity, and word order in the perception of prominence," *Journal of Phonetics* vol. 34, pp. 319–342, 2006.

[13] E. Krahmer and M. Swerts, "Perceiving focus," *Topic and Focus,* pp. 121-137, 2007.

[14] D. R. Ladd and R. Morton, "The perception of intonational emphasis: continuous or categorical?," *Journal of Phonetics,* vol. 25, pp. 313-342, 1997.

[15] W. L. Wu and L. Chung, "Post-focus compression in English-Cantonese bilingual speakers," presented at the ICPhS, Hong Kong, 2011.

[16] G. Thurgood and F. X. Li, "From Malayic to Sinitic: The restructuring of Tsat under intense contact," in *SEALSXII: papers from the 12th meeting of the Southeast Asian Linguistics Society (2002)*, R. Wayland*, et al.*, Eds., ed Canberra, Pacific Linguistics, 2007, pp. 129-136.

[17] Y. Xu, "ProsodyPro available from," *http://www.phon.ucl.ac.uk/home/yi/tools.html,* 2005-2012.

[18] I. Maddieson and K.-F. Pang, "Tone in Utsat," *Oceanic Linguistics Special Publications: Tonality in Austronesian Languages,* vol. 24, pp. 75-89, 1993.

[19] S. Jannedy, "The effect of focus on lexical tones in Vietnamese," presented at the Proceedings of the 2nd ISCA Workshop on Experimental Linguistics, Athens, Greece, 2008.