

Acquisition and Interpretation of Mandarin Speech Prosody by Native Speakers and Cantonese Learners

Xi Chen* and Si Chen†

*The Hong Kong Polytechnic University, Hong Kong, China

E-mail: skyexi.chen@connect.polyu.hk Tel: +852-6047 3819

†The Hong Kong Polytechnic University, Hong Kong, China

E-mail: sarah.chen@polyu.edu.hk Tel: +852-5941 4757

Abstract— This study aims to test the ability to match acoustic cues to different focus types and positions by advanced Cantonese L2 learners of Mandarin under the modalities of auditory-only and visual-auditory. Following the design by [29], participants were instructed to make a 5-Likert response to rate their preferences for the conversations they heard. Results show that visual-aids facilitated the perception of prosody; L2 learners showed fewer difficulties in differentiating narrow and contrastive focus than native Mandarin speakers. These findings provide significances for prosodic perception, second language acquisition and bilingual education.

I. INTRODUCTION

As a crucial clue of representing the information structure, prosody is widely used in real life dialogues. Prosody is often used to suggest what is new, given or emphatic ([19]), so besides locutionary act, prosody exerts more robust influences on the illocutionary act and perlocutionary act. Studies have long tried to unravel how could the acoustic cues affect the prosodic perceptual process ([7] and [27]), while most of the researches concentrated on West Germanic languages ([14]), limited prosodic perception studies carried out on tonal languages.

1.1 Interface Hypothesis

According to the Interface Hypothesis (IH) ([2]), involving prosody, syntax, words, context, and mother tongue (L1) transfer, the processing of information structure becomes a bottleneck for L2 learners. The bulk of studies proved this L2

acquisition hardness in various ways (English: [30]; Mandarin: [8], [28], [40], and [41]). The L1 effect, here means that the different ways of prosody compared with their mother tongue, plays a significant role in L2 prosodic perception (Spanish L2 learners of English: Ref. [22], Chinese L2 learners of English: Ref. [1], English L2 learners of French: Ref. [13], Spanish and Italian: Ref. [25], and French L2 learners of Dutch: Ref. [18]). However, these studies mainly focused on west Germanic languages, but paid little attention to tonal languages.

1.2 Prosodic Features in Mandarin and Cantonese

Mandarin Chinese is the official language of mainland China and is one of the two official languages of Hong Kong ([23]). In mandarin, prosody is usually achieved by post-focus compression and on-focus increase of F_0 ([3], [35], and [41]). Apart from the differences in tone, Cantonese and Mandarin also differ in prosody.

Contrary to the asymmetric natures founded in Mandarin, Cantonese shows identical variations in pre-focus, on-focus and post-focus. Focus always exerts a wide-range increase of F_0 in Cantonese, which means that the influence of focus is neither locally nor wholly ([32]). The expression effect of post-focus in Cantonese is not as distinct as Mandarin. These differences can affect the Cantonese learners' perception of Mandarin prosodic perception.

In addition to these linguistic factors, native speakers also use facial expressions([4], [12], and [20]), head movement ([15]), or

body language ([10]) to emphasis the focus. Native speakers use these clues to decode the prosodic utterances.

1.3 Focus Types

According to the Information Structure theory (IS, [19] and [16]), this research studied broad focus, narrow focus and contrastive focus (see Tab.1) on different positions.

(1) 张生开飞机。(/Zhāng Shēng kāi fēi jī/) (Zhang Sheng drives the airplane).

Sentence (1) is a suitable statement for all of the questions in Tab.1 in text from, they only differ in prosody. (a) elicits “broad focus,” which means that the answer provides a piece of new information ([33] and [35]) that does not use any stress on a particular unit or constituent (see Tab.1 row 1).

Tab. 1 Q-A pairs in Mandarin with different focus types and position

Focus types	Contextual Questions	Corresponding Answers
a. Broad (B)	发生什么事? What happens?	[张生开飞机] _F ¹ . [Zhang Sheng drives the airplane.] _F
b. Narrow-subject N(s)	谁开飞机? Who drives the airplane?	[张生] _F 开飞机。 [Zhang Sheng] _F drives the airplane.
c. Narrow-verb N(v)	张生怎么弄飞机? What does Zhang Sheng do to the airplane?	张生[开] _F 飞机。 Zhang Sheng [drives] _F the airplane.
d. Narrow-object N(o)	张生开什么? What dose Zhang Sheng drive?	张生开[飞机] _F 。 Zhang Sheng drives the [airplane] _F .
e. Contrastive-subject C(s)	刘丽开飞机? Liu Li drives the airplane?	[张生] _F 开飞机。 [Zhang Sheng] _F drives the airplane.
f. Contrastive-verb C(v)	张生坐飞机? Zhang Sheng takes the airplane?	张生[开] _F 飞机。 Zhang Sheng [drives] _F the airplane.
g. Contrastive-object C(o)	张生开巴士? Zhang Sheng drives the bus?	张生开[飞机] _F 。 Zhang Sheng drives the [airplane] _F .

(b)-(c)-(d) exert “narrow focus” respectively on subject (b), verb (c) and object (d). A narrow-focused answer specifies a sentence constituent through prosody. This type of focus could also be used to achieve an “inform” function—which provides the agent, the verb, or the patient. The old information in Tab.1(b)—*drives the airplane*—would be weaken while stressing the subject *Zhang Sheng* ([35] and [40]).

(d)-(e)-(f) are called “contrastive focus” or “corrective focus” ([29] and [35]). It could not be more evident that the function of contrastive focus utterances is for “correcting.” Contrastive-focused utterances are elicited by the wrong expressed sentences. The prominence (noted as [...]_F) is placed on a specific grammatical item while correcting.

1.4 Research Questions

a. Do auditory-only and visual-auditory stimuli yield different outcomes in prosodic perception?

b. Do Cantonese show different patterns in prosody perception compared with native Mandarin? If so, which type of focus is the most difficult one for them to recognize?

c. What kind of suggestion could this research provide to second language acquisition?

II. PRODUCTION EXPERIMENT

This paper concentrates on the perceptual patterns of different language backgrounds, so the purpose of the production part is to record the prosodic stimuli.

1.5 Stimuli Sentences

All of the fifteen basic sentences (Tab.2) in Mandarin were made up of three words structured with the subject (bi-character) + verb (mono-character) + object (bi-character). Each sentence merely contains one tone (Tone 1, 2, 4) ([6], [36], [41], and [43]). We set seven focus conditions for each basic sentence.

¹ The [...]_F refers to the focused constituent in each sentence.

(2) Tab. 2 Fifteen basic Mandarin sentences for production experiment







Basic Sentences	Annotated in Pinyin	Translated in English
1. 张生开飞机。(T1)	Zhāng Shēng kāi fēi jī	Zhang Sheng drives the airplane.
2. 张英摸猫咪。(T1)	Zhāng Yīng mō māo mī.	Zhang Ying strokes the kitten.
3. 张欣织书包。(T1)	Zhāng Xīn zhī shū bāo.	Zhang Xin knits the bag.
4. 张刚吃西瓜。(T1)	Zhāng Gāng chī xī guā.	Zhang Gang eats the watermelon.
5. 张天吹风车。(T1)	Zhāng Tiān chuī fēng chē.	Zhang Tian blows the pinwheel.
6. 刘明拿篮球。(T2)	Líu Míng ná lán qiú.	Liu Ming holds the basketball.
7. 刘宁涂黄油。(T2)	Líu Níng tú huáng yóu.	Liu Ning spreads the butter.
8. 刘同扶盲人。(T2)	Líu Tóng fú máng rén.	Liu Tong helps the blind.
9. 刘平划竹船。(T2)	Líu Píng huá zhú chuán.	Liu Ping rows the (bamboo) boat.
10. 刘文爬长城。(T2)	Líu Wén pá cháng chéng.	Liu Wen climbs the Great Wall.
11. 赵亮看电视。(T4)	Zhào Liàng kàn diàn shì.	Zhao Liang watches the television.
12. 赵月做作业。(T4)	Zhào Yuè zuò zuò yè.	Zhao Yue does the homework.
13. 赵克办护照。(T4)	Zhào Kè bàn hù zhào.	Zhao Ke applies for the passport.
14. 赵丽戴项链。(T4)	Zhào Lì dài xiàng liàn.	Zhao Li wears the necklace.
15. 赵梦画漫画。(T4)	Zhào Mèng huà màn huà.	Zhao Meng draws the comics.

1.6 Methodology

Four native Mandarin speakers from the north of mainland China are recruited (two male-female groups). The stimuli were recorded in a soundproofed booth with a high-quality, head-mounted microphone (*audio-technica AT 2035*), and an external camera (*logitech HD 1080p*). Speakers were instructed to utter complete sentences according to the pictures (examples in Tab. 3 for sentence 1). Trackers or helmets were

ruled out in this process for the most natural facial expressions and the head movements. Despite the questions-eliciting pictures for contrastive focus, no text information were shown on the screen for the speakers. All of the 15 Q-A pairs were featured with the same different focus types with sentence 1. An entire round for the production resulted in 210 trials (15 basic sentences × 7 focus conditions × 2 repetitions). A round was female-ask-male-answer, another was the opposite.

Tab. 3 Eliciting picture samples (sentence 1) for the live dialogues in stimuli production process

Focus types	Contextual Questions	Corresponding Answers
a. Broad (B)	发生什么事? What happens? 	[张生开飞机。] _F [Zhang Sheng drives the airplane.] _F 
b. Narrow-subject N(s)	谁开飞机? Who drives the airplane? 	[张生] _F 开飞机。 [Zhang Sheng] _F drives the airplane. 
e. Contrastive-subject C(s)	刘丽开飞机? Liu Li drives the airplane? 	[张生] _F 开飞机。 [Zhang Sheng] _F drives the airplane. 

1.7 Acoustic Cues Analysis

The following analysis was based on two of the native speakers' recordings (SL, male of 27; WXL, female of 23). Fig.1 displays 20-timepoint-normalized mean F₀ contours under different focus conditions for sentence 1.

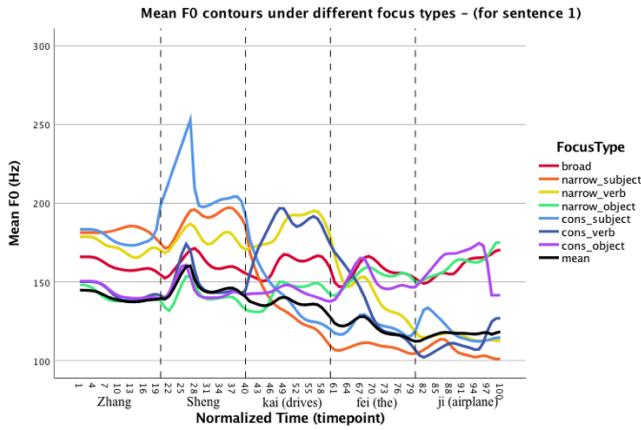


Fig.1 time-normalized F₀ contour

Post-focus features tell us more ([28] and [40]). Although the focused components yield the extension in pitch range, duration, and intensity ([30]), the focus is also realized by the compression of the post-focus components (PFC, [8] and [40]). Besides, the on-focus and post-focus F₀/intensity/duration change (difference between the focused sentence with the broad focused sentence of a sentence component) also changed by focus ([42]). Take on-focus F₀ difference (Fig.2.a) and post-focus intensity difference for instance (Fig.2.b), utterances that not meet the requirements were excluded by two-way ANOVAS

(finally Sentence*post-F₀: (p<0.05); Sentence*post-intensity: p<0.05).

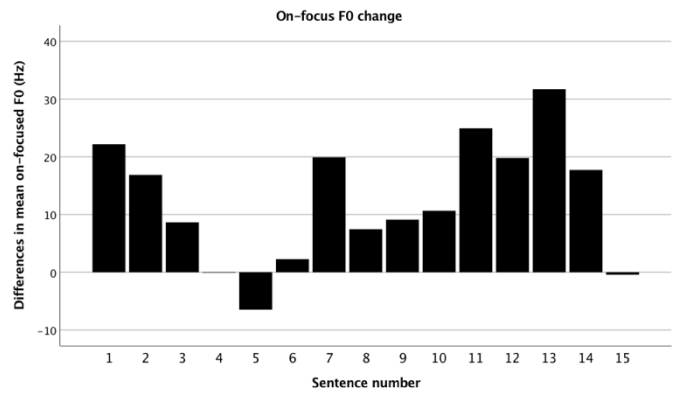


Fig.2.a. On-focus F₀ change

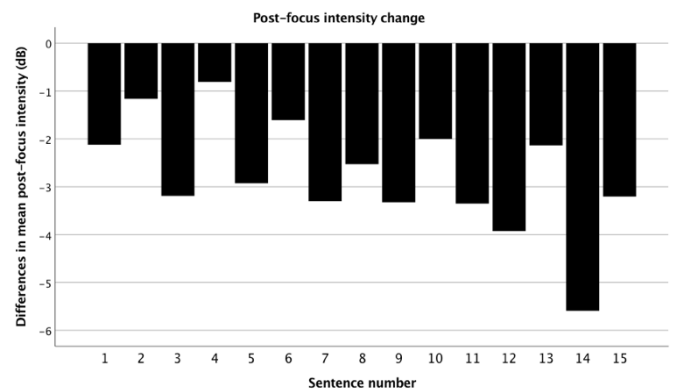


Fig.2.b Post-focus intensity change

Fig.2 Differences between on-focus F₀ / post-focus intensity and the counterpart in broad-focused sentences.

Fig. 2 also verified that the PFC does not always work ([37]). Finally, six sentences were selected for the perception experiments (Tab.4), and sentence 15 was used for practicing.

Tab.4 Basic sentences for perception experiments

Basic Sentences	Annotated in Pinyin	Translated in English
1.张生开飞机。(T1)	Zhāng Shēng kāi fēi jī	Zhang Sheng drives the airplane.
3.张欣织书包。(T1)	Zhāng Xīn zhī shū bāo.	Zhang Xin knits the bag.
8.刘同扶盲人。(T2)	Líu Tóng fú máng rén.	Liu Tong helps the blind.
10.刘文爬长城。(T2)	Líu Wén pá cháng chéng.	Liu Wen climbs the Great Wall.
11.赵亮看电视。(T4)	Zhào Liàng kàn diàn shì.	Zhao Liang watches the television.
13.赵克办护照。(T4)	Zhào Kè bàn hù zhào.	Zhao Ke applies for the passport.
15.赵梦画漫画。(T4)	Zhào Mèng huà màn huà.	Zhao Meng draws the comics.

III. PERCEPTION EXPERIMENT

We used 5-Likert preference multiple choice task following Ref. [29]. The subject groups are native Mandarin speakers (NM) and Cantonese (Hong Kong) learners of Mandarin (L2M).

1.8 Methodology

Participants in each language group were divided into two sub-groups, one received auditory-only (AO) stimuli, while the other received the video-auditory (VA) stimuli. In each experiment, AO and VA stimuli with the form of short Q-A dialogue pair were presented to participants. A pause with a 2.000-second duration was inserted between the question and the answer in each Q-A pair through [5] in *Praat*. Each trial involved two dialogues, the question of these two dialogues were the same, while the two answers differed in the prosodic congruence — one was congruous, and the other was incongruous. The congruous dialogues were combined from the question and answer segments paired initially; the corresponding incongruous dialogues were chosen from the different question-answer pairs, e.g., it could be the broadly focused question with the other six type of focused answers (see (3)). Participants were permitted to listen to the dialogues as many times as they want before making response. All of the experiments were conducted in the E-prime (*2.0 Professional*). The procedure of the task was designed as shown in Fig.3 (AO) and Fig.4 (VA).

(3) Tab.5 Example dialogues in one trial

	Congruous-Dialogue A	Incongruous-Dialogue B
Q	发生什么事? What happens? [Broad focus needed]	发生什么事? What happens? [Broad focus needed]
A	[张生开飞机。] _F [Zhang Sheng drives the airplane.] _F [Broad focus presented]	[张生] _F 开飞机。 [Zhang Sheng] _F drives the airplane. [Narrow focus presented]



Fig.3 Procedure for the auditory-only task



Fig.4 Procedure for the visual-auditory task²

While the dialogue was playing, a dialogue code was shown on the screen (A or B). Participants could control their speed to go for the choosing page to scale the dialogues they just heard/watched; finally, a check page was shown to inspect whether they had payed attention to the dialogues or not. The number button “1 2 3 4 5” used for rating the two dialogues could be translated as the following 5-Likert scale choices:

1. Only A: Only dialogue A sounds natural and satisfactory
2. A preferred: Both of the two are natural, but A is preferred
3. Equal: Two dialogues are equally natural and satisfactory
4. B preferred: Both of the two are natural, but B is preferred
5. Only B: Only dialogue B sounds natural and satisfactory

The perceptual experiment consisted of three blocks — one practice block with the sentence “Zhào Mèng huà màn huà.” (sentence 15 in Tab.4). In block 1 and 2 (192 Q-A pairs in each), two Q-A pairs in each trial were pseudo-randomly chosen from 16 modes of the combining in Fig. 5 (e.g., (3) for testing broad vs. narrow focus).

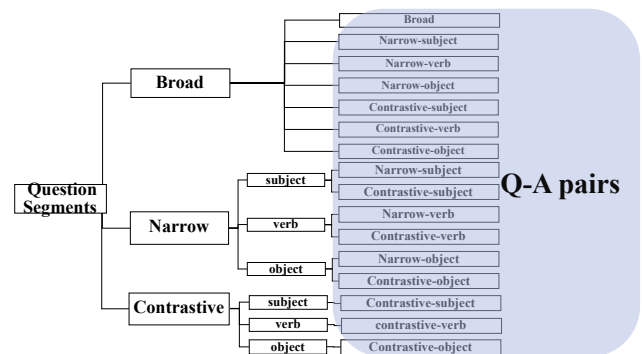


Fig.5 16 modes of Q-A combining pairs

² The screenshot was only approved by the speaker in this paper.

3.2 Participants

17 (1M16F) and 4 (2M2F) native Mandarin speakers (mean age 23.67, who spent most of their time in mainland China and past the PSC (National Proficiency Test of Putonghua) with Level 2A or above were recruited in the perceptual experiments; 6 (2M4F, exclude 1 later) and 4 (2M2F, exclude 1 later) Cantonese (mean age 23.25, who were raised in Hong Kong and started to learn mandarin when they were around ten-year-old) were recruited.

3.3 Statistical Analysis

The correct answer of each trial (including the check task) were randomized, which means that congruous dialogues show either as pair A or B. For the sake of analysis, the scale quantity was normalized to the circumstance of the congruous conversation always presents in pair B. So the congruous pair should always be graded as 5 (means “always”). We used ordinal regression in SPSS to detect the perceptual patterns.

IV. RESULTS

4.1 Different Stimuli Modalities

Ordinal regression showed that in NM group, the estimated parameter for AO was negative (-0.441, p<0.05), suggesting that the visual-auditory stimuli exerted significant higher perceptual scores. In L2M group, the estimated parameter of AO modality was negative (-0.121), either, but with no significance (p>0.05). Visual stimuli did not significantly help L2 learners to understand the prosodic sentences.

4.2 Matched and Mismatched pairs

“Matched” here means that the two dialogues in one trial are identical (e.g., dialogue A: B-B pair; dialogue B: B-B pair). Tab.6 displays both of the groups of natives and L2 learners chose to rate “equal” when they heard the “matched” pairs (P>0.05). Although visual aids helped to narrow the mean

differences compared with 3 (“equal”), the reduction was not significant.

Tab.6 Score rated by NM and L2M groups under different modalities

Ranked Score Matched Focus Type	NM		L2M	
	AO P ³ >0.05	AV P>0.05	AO P>0.05	AV P>0.05
B-B	2.967	2.979	3.000	2.838
N(s)-N(s)	2.994	3.213	3.048	3.000
N(v)-N(v)	2.949	3.021	3.090	3.000
N(o)-N(o)	2.961	3.129	3.051	3.054
C(s)-C(s)	3.018	2.934	3.045	3.057
C(v)-C(v)	3.045	3.000	2.931	3.057
C(o)-C(o)	3.153	2.939	3.024	3.174

4.3 Perceptual Patterns

Both of the NM and L2M tended to choose “equal” (above 45% in all of the two modalities) for most of the time. These two groups’ “prefer” and “always” score percentages jumped over the chance level (20%) under the VA modality.

4.3.1 Distinguish broad focus from other focus types

Fig. 6 represents the perceptual scores of NM and L2M groups under different modalities when distinguishing the broad focus from other focus types.

Perceptual patterns of two groups under different modalities distinguishing broad-focus (B) from other focus types

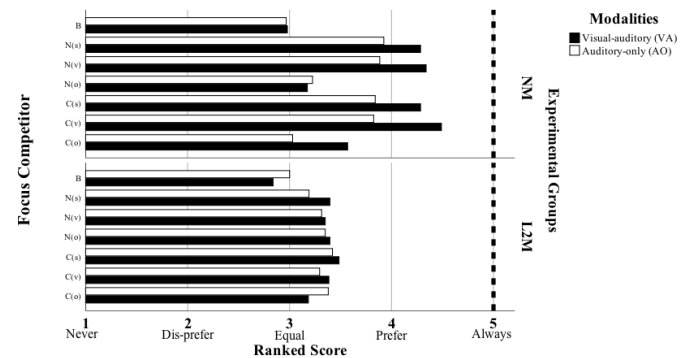


Fig. 6 Perceptual patterns of distinguishing broad focus from other competitors

³ All of the P values for NM and L2M groups under AO and VA modalities.

For NM group under AO modality, listeners would rather to choose the “matched” broad focused dialogues when the competitors were other focus types. Listeners of this experimental group could significantly distinguish the congruous dialogue from the incongruous ones ($p < 0.05$).

As for the VA modality session of NM group, under another significant fitted model ($\chi^2[45]=229.465, p < 0.05$), in terms of the recognition of broad focus, it cannot be more evident that the VA modality exerted higher recognition scores with all of the other focus types despite the narrow focus on the object competitor.

For L2M group under AO modality, listeners showed a slightly preference to the congruous B-B ones compared to the “equal” choice (score=3). But the rated scores were significantly lower than NM group. Expect for the narrow (subject)-focused competitor ($p > 0.05$), L2M group had the ability to differentiate the congruous B-B ones from other incongruous prosodic sentences.

The most exciting thing lies in there is a slight improvement (around 0.01) in each position of the recognition score rates between narrow and contrastive focused condition [N(s)-C(s)=0.596, score=2.98, $p > 0.05$; N(o)-C(o)=0.591, score=2.955, $p > 0.05$; N(v)-C(v)=0.598, score=2.99, $p > 0.05$]. These confirmed that visual clues enhanced the prosodic cues to some extent, which could increase the perception rate for native listeners.

In VA modality session of L2M group, under another fitted model ($\chi^2[45]=155.431, p < 0.05$), most of the incongruous focus types could not be fully recognized by L2M despite the broad focus vs. narrow (subject) focus ($p = 0.036$). An interesting thing seemed to be that the mean score they ranked for the “matched” pairs showed a departure from “equal” (score=3) when they were exposed to the VA stimuli.

4.3.2 Distinguish narrow focus from contrastive focus

As predicted, even native speakers showed difficulties in distinguishing narrow and contrastive focus. Fig. 7 showed the

rated scores by NM and L2M under different modalities when distinguishing narrow focus from contrastive focus.

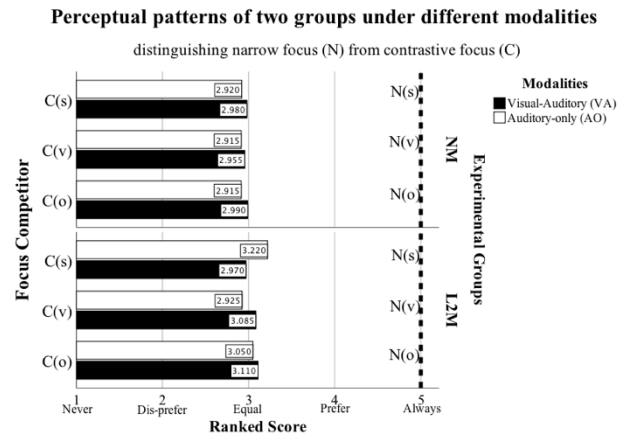


Fig. 7 Perceptual patterns of distinguishing broad focus from other competitors

NM showed bias to “equal” or “dis-preferred” under both of the two modalities. Their results formed a score interval of [2, 3] (lower than 3) —indicated that they may prefer the dialogues featured with more apparent acoustic cues (contrastive focused sentences). NM group failed to distinguish narrow focus from contrastive focus because no significance ($P > 0.05$) was discovered in scoring. What’s more, VA stimuli merely slightly improved the perceptual scores of NM group with no significance.

L2M group could not distinguish the narrow focus from contrastive focus, either ($P > 0.05$). Dramatically, they showed a higher perceptual score compared to NM group, and they formed a score interval of [3,4]. Which means that they scored more closer (mean score ≈ 3.065) to the congruous dialogues than NM (mean score ≈ 2.915). The visual-aid did not provide NM group with a predicted function in differentiating the subject focused sentences.

V. GENERAL DISCUSSION

5.1 Did the stimuli modalities of AO and VA Work Differently?

Generally, the overall results supported that visual-aids facilitate the prosodic perception of native speakers, while confused the second language learners’ acquisition to some degree. Native speakers use facial expression (head movement, lip rounding, etc.) to enhance the speech focus, but even the native

speakers could not always decode this layer of information. Since L2 has the different expression system, the Cantonese learners of Mandarin could not understand the extra information correctly. Visual-aids helped them a little, and sometimes even hinder their understanding in the opposite.

5.2 Could Listeners Distinguish Matched From Mismatched?

They could not tell the difference when the listeners were exposed to the same dialogue pairs. Surprisingly, almost none of the options was chosen above the chance level (20%) except the “equal” option, which means that even native listeners could not always tell the difference between the given two dialogue pairs. This research is quite limited here, we failed to make sure a enough change of stimuli sentences for attracting the listeners. We would involve more prosodic sentences in choosing stimulation in the future.

5.3 How to explain the two groups' different perceptual patterns?

Distinguishing broad focus from other type of focus seems to be the easiest for both of the native Mandarin speakers and the Cantonese learners of Mandarin. Visual-aids help the listeners in most of the time, and this result comes as no surprise.

But things go different in distinguishing narrow focus from contrastive focus as predicted. For native Mandarin speakers, the mis-led tendency of choosing the incongruous contrastive prosody only suggests that native speakers are sensitive to the contrastive focus with the slightly strengthened acoustic cues relative to the narrow focus. However, because of the differences between Mandarin and Cantonese mentioned in the *Introduction* part, L2 learners showed different perceptual patterns ([32]). When Cantonese learners of Mandarin hear the sentences with unfamiliar prosody, they may not show any preference to the options with more obvious acoustic cues. The stronger acoustic cues of contrastive focus failed to deceive L2 learners.

5.4 What contribution could this study make?

The 5-Likert scale paradigm was used to clearly unravel the mapping between the prosodic form and its pragmatic meaning, and this method could keep away of the bias on prosodic speech sentences ([29]).

As described in the Interface Hypothesis, prosodic acquisition exists as a bottleneck of L2 learners. Even their mother tongue comes from the same linguistic family of the second language, and even they start to learn L2 very young, L2 learners show different prosodic perceptual patterns from natives. According to the experiments and the previous theories, several suggestions could be provided in teaching Chinese as a foreign language.

Apart from teaching students with words and grammar, teachers should also pay attention to the prosody of the foreign language. Consider of using more listening activities in teaching—the sound and meaning of a word could be arbitrary ([17]), but the sound of a prosody is not. The experimental paradigm could help the language teachers in designing the class to some extent. Apart from listening, speaking practice also plays an important role in training prosodic perception. Some dialogue-making practice could make the learners get involved in prosody production. Ref. [31] tested that music training improve the learners' sensitivity to speech prosody. Students could find some patterns and rules spontaneously in a joyful musical atmosphere.

REFERENCES

- [1] A. Lee, M. Perdomo, and E. Kaan, “Native and second-language processing of contrastive pitch accent: An ERP study,” *Second Language Research*, pp. 1-25, April 2019.
- [2] A. Sorace and L. Serratrice (2009). “Internal and External Interfaces in Bilingual Language Development: Beyond structural overlap,” *International Journal of Bilingualism*, vol. 13, pp. 195-210, December 2009.

- [3] A.Q. Yang, "The acquisition of prosodic focus-marking in Mandarin Chinese-and Seoul Korean- speaking children," PhD thesis, Utrecht University, Netherlands: Utrecht, 2017.
- [4] B. Granström, D. House, M. Lundeberg, "Prosodic Cues in Multimodal Speech Perception," *Proceedings of the International Congress of Phonetic Sciences (ICPhS99)*, pp. 655-658, 1999.
- [5] C. Darwin, combine-sounds.praat, Online:<http://www.lifesci.sussex.ac.uk/home/Chris_Darwin/Praatscripts/Add2_variable>," 2005-2019.
- [6] C. Shih, "Mandarin third tone sandhi and prosodic structure," in *Studies in Chinese Phonology*, J. L. Wang and N. Smith Eds., Germany: Berlin, 1997, pp. 81-92.
- [7] D. G. Watson, M. K. Tanenhaus, C. A. Gunlogson, "Interpreting pitch accents in online comprehension: H* vs. L+ H," *Cognitive Science*, vol. 32(7), pp. 1232-1244, October 2008.
- [8] F. Liu, Y. Xu, "Parallel Encoding of Focus and Interrogative Meaning in Mandarin Intonation," *Phonetica*, vol. 62(2-4), pp. 70-87, December 2005.
- [9] H. P. Graf, E. Cosatto, V. Strom, and F.J. Huang, 2002. "Visual Prosody: Facial Movements Accompanying Speech," *Proceeding of 5th IEEE International Conference on Automatic Face and Gesture Recognition (FGR'02)*, pp. 381-386, June 2002.
- [10] H. S. H. Fung and P. P. K. Mok, "Temporal coordination between focus prosody and pointing gestures in Cantonese," *Journal of Phonetics*, vol. 71, pp. 113-125, September 2018.
- [11] H. S. Kim, S. A. Jun, H. J. Lee, and L. K. Kim, "Argument Structure and Focus Projection in Korean," *Proceedings of Speech Prosody*, Dresden: Germany, 2006.
- [12] J. Kim, E. Cvejic, and C. Davis, "Tracking eyebrows and head gestures associated with spoken prosody," *Speech Communication*, vol. 57, pp. 317-330, February 2014.
- [13] J. Namjoshi, "The Processing and Production of Prosodic Focus in French by Native and Non-native Speakers," PhD thesis, University of Illinois. Illinois: Urbana, 2015.
- [14] J. Pierrehumbert, and J. Hirschberg, "The meaning of intonational contours in the interpretation of discourse," in *Intentions in Communication*, P. Cohen, J. Morgan, and M. Pollack Eds. United Kingdom: Cambridge, MIT Press, pp. 271-311, 1990.
- [15] K.G. Munhall, J.A. Jones, D.E. Callan, T. Kuratate, and E. Vatioti-Bateson, "Visual Prosody and Speech Intelligibility — Head Movement Improves Auditory Speech Perception," *Psychological Science*, vol. 15(2), pp. 133-137, February 2004.
- [16] K. Lambrecht, "Information structure and sentence form," in *Topic, focus, and the mental representations of discourse referents*, Cambridge University Press Eds. The United Kingdom: Cambridge, 1994, pp. 1-6.
- [17] L. C. Nygaard, A. E. Cook, and L.L. Namy, "Sound to meaning correspondences facilitate word learning," *Cognition*, vol. 112, pp. 181-186, July 2009.
- [18] L. Rasier, J. Caspers and V.J. Heuven, "Accentual marking of information status in Dutch and French as foreign languages. Production and perception," *New Sounds*, in [Proceedings of the 6th International Symposium on the Acquisition of Second Language Speech], pp. 379-385, May 2010.
- [19] M. A. K. Halliday, "Intonation and Grammar in British English" in *Mouton The Hague*, vol. 48, Walter de Gruyter GmH & Co KG Eds, 1967.
- [20] M. Dohen and H. Lœvenbruck, "Interaction of Audition and Vision for the Perception of Prosodic Contrastive Focus," *Language and Speech*. vol. 52, pp.177-206, June 2009.
- [21] M. Dohen and H. Lœvenbruck, "Recognizing Prosody From the Lips: Is It Possible to Extract Prosodic Focus From Lip Features?" in *Visual Speech Recognition: Lip Segmentation and Mapping*, Chapter XIV, Liew, A. W. C., Wang, S. Eds. the United Kingdom: London, 2009, pp. 416-438.
- [22] M. Ortega-Lleiria, L. Coltoni, "Relations between Form-Meaning Associations, Access to Meaning, and L1 Transfer," *Studies in Second Language Acquisition*, vol. 36, pp. 331-353, June 2014.

- [23] P. Chen, "Modern Chinese: History and Sociolinguistics", Cambridge: Cambridge University Press, 1999, pp.1.
- [24] P. K. Mok and D. Volker, "Comparing native and non-native speech rhythm using acoustic rhythmic measures: Cantonese, Beijing Mandarin and English," *Speech and Prosody*, vol. 6-9, pp. 423-426, May 2008.
- [25] P. Mareüil and B. Vieru-Dimulescu, "The Contribution of Prosody to the Perception of Foreign Accent," *Phonetica*, vol. 63, pp. 247-267, February 2006.
- [26] S. A. Jun, "Prosodic Typology Revisited: Adding Macro-Rhythm," *Proceedings of Speech Prosody*, vol. 6, May 2012.
- [27] S. Peppé, J. Maxim, B. Wells, "Prosodic Variation in Southern British English," *Language and Speech*, vol. 43(3), pp. 309-334, September 2000.
- [28] S. W. Chen, B. Wang, and Y. Xu, "Closely related languages, different ways of realizing focus," *Interspeech*, vol. 6-10, pp. 1007-1010, September 2009.
- [29] T.B. Roettger, T. Mahrt, and J. Cole, "Mapping prosody onto meaning — the case of information structure in American English," *Language, Cognition and Neuroscience*, vol. 34 (7), pp. 841-860, February 2019.
- [30] W. E. Cooper, S. J. Eady, and P. R. Mueller, "Acoustical aspects of contrastive stress in question-answer contexts," *The Journal of the Acoustical Society of America*, vol. 77, pp. 2142-2156, June 1985.
- [31] W. F. Thompson, E. G. Schellenberg and G. Husain, "Decoding speech prosody: Do music lessons help?" *Emotion*, vol. 4(1), pp. 46-64, March 2004.
- [32] W. Gu, and T. Lee, "Effects of tonal context and focus on Cantonese F0," *Proceedings of the 16th International Congress of Phonetic Sciences (ICPhS 2007)*, pp. 1033-1036, August 2007.
- [33] W. L. Chafe, "Cognitive Constraints On Information Flow" in *Coherence and grounding in discourse*, R. S. Tomlin Eds, Netherlands: Amsterdam, 1987, pp. 21-22.
- [34] Y. C. Lee, Y. Xu, "Phonetic Realization of Contrastive Focus in Korean," *Speech Prosody*, pp. 10-14, May 2010.
- [35] Y. C. Lee, "Prosodic Focus Within and Across Languages," PhD thesis, University of Pennsylvania. *Publicly Accessible Penn Dissertations*, Paper 1534, January 2015.
- [36] Y. C. Lee, T. Wang, and M. Liberman, "Production and Perception of Tone 3 Focus in Mandarin Chinese," *Frontiers in Psychology*, vol. 7, July 2016.
- [37] Y. K. Yang, S. Chen, and K. Li, "Effects of Focus on Duration and Intensity in Chongming Chinese," in *Proceedings of the 19th International Congress of Phonetic Sciences*, Australasian Speech Science and Technology Association, pp. 3578-3582, 2019.
- [38] Y. R. Chao, "Tone, intonation, singsong, chanting, recitative, tonal composition, and atonal composition in Chinese", *For Roman Jakobson*, pp. 52-59, October 1956.
- [39] Y. Xu, "Effects of tone and focus on the formation and alignment of f₀ contours," *Journal of Phonetics*, vol. 27, pp.55-105, January 1999.
- [40] Y. Xu, "Speech prosody as articulated communicative functions," *Speech Prosody*, pp. 2-5, May 2006.
- [41] Y. Xu, S. W. Chen, and B. Wang, "Prosodic focus with and without post-focus compression: A typological divide within the same language family?" *The Linguistic Review*, vol. 29, pp. 131-147, February 2012.
- [42] Y. Y. Chen and Y. Xu, "Production of Weak Elements in Speech — Evidence from F₀ Patterns of Neutral Tone in Standard Chinese," *Phonetica*, Vol.63, pp. 47-75, August 2006.
- [43] Z. S. Zhang, "Tone and tone sandhi in Chinese," PhD thesis, the Ohio State University, The United States: Columbus, 1998.