

## PRODUCTION AND PERCEPTION OF INTONATION FEATURES BY CANTONESE EFL LEARNERS

Chenyang Zhao<sup>1</sup>; Ziyu Xiong<sup>1,2</sup>; Aijun Li<sup>1,3</sup>

<sup>1</sup>University of Chinese Academy of Social Sciences (UCASS);

<sup>2</sup>Institute of Linguistics, CASS

<sup>3</sup>Corpus and Computational Linguistics Center, Institute of Linguistics, CASS

cyzhao16@mails.jlu.edu.cn, [xiongziyu@163.com](mailto:xiongziyu@163.com), [liaj@cass.org.cn](mailto:liaj@cass.org.cn)

### ABSTRACT

This paper, on the basis of AM intonation theory, aims to investigate Cantonese EFL English learners' production and perception of intonation features, including the phonological features of nuclear accent and boundary tone, as well as the placement of nuclear accent. Altogether 252 sentences, including 136 statements and 116 yes-no questions were collected, from which it was found that learners tended to use H\* and L\* in statements and yes-no questions respectively, and these two tone types were also the ones that are easiest to be perceived. Interactive relationship between learners' production and perception were discussed. The results will give guidance to the EFL intonation teaching for Cantonese learners.

*Index Terms*—Cantonese EFL English Learners, Perception and Production, Phonological Features

### 1. INTRODUCTION

The production and perception of EFL intonation of varied languages are widely studied. A study on HK Cantonese speakers' EFL English question production showed that final rising only covered the very last syllable [1], which is in line with the characteristic of Cantonese question intonation [2]. Another experiment further proved that Cantonese L1 transfer was present in speakers' EFL realization. Their Mandarin question production was featured by the weaker global rising and the stronger final rising compared to native mandarin speakers [3]. A comparison of EFL Mandarin emphatic stress realization between English and Cantonese speakers again proved the importance of L1 transfer effect. Specifically, the prominent F0 raising on the focused syllable of Mandarin was weaker but was still maintained in English speakers' production because of their L1-EFL similarity in stress realization, while it nearly disappeared in Cantonese speakers' due to the absence of F0 expansion in focused syllable in their L1 [4]. L1 transfer effect is also manifested in speakers with other mother tongue background. Italian speakers tended to use lower pitch range and pitch span to realize contrastive focus, which was proved to be transferred in their EFL production of English [5]. However, not every phonetic feature can be successfully transferred into EFL production. According to Gao et al. [6], PFC (post-focus compression) is a feature that is "hard to acquire, yet easy to lose", which was proved by their experimental study on Mandarin speakers' acquisition of EFL English intonation. This phenomenon was also observed in Korean speakers' EFL English intonation [7]. Therefore, transfer effect is not the only cause that leads to EFL learners' deviation.

How does transfer effect and other factors influence the realization of learners' EFL phonological features? In this study, the EFL English production and perception by Cantonese speakers were discussed from phonological aspect.

English is a typical intonation language. According to AM intonation theory, English declarative contour has one or more than one pitch accent, usually realized as H\*, a low phrase accent (L-) and a low boundary tone (L%), while a yes-no question contour is cued by one or more than one low pitch accent (L\*), followed by a high phrase accent (H-) and a high boundary tone (H%) [8]. In English declarative sentences, the pitch contour starts to decline from the first accented syllable, on the contrary, it begins to rise from first syllable of the accented words in yes-no question contour [9]. Cantonese is a Southern dialect of China widely used in part of people from Canton, Guangxi, Hong Kong, and Macao. Unlike Mandarin Chinese, Cantonese has 9 tones, with the first six usually represented by T1-T6, including 6 long tones and 3 checked tones, and among them, T2 and T5 are dynamic tones which had a rising contour, T1, T3, and T6 are static tone, and T4, a low falling, static tone [10]. The tone values of each tone are 55, 25, 33, 21, 23, 22, 5, 3, 2 [11]. Cantonese does not signal the focus by expanding the pitch range, or compressing the pitch range of post-focus syllables, instead, they tend to use duration lengthening instead [2]. Intensity is another cue indicating focal accents in Cantonese, and a high F0 represents a high intensity in a syllable [11]. For syllables with T1 in statements, a high level tone, if a pitch expansion was bestowed on them, listeners would be less possible to perceive them as a stressed syllable compared to those with other tones [11]. For questions, they are encoded by raising the F0 the very end of sentences, instead of doing it globally [12]. Cantonese questions have a rich sentence-final particles (i.e. SPF), the F0 of sentence-final syllable in questions are mainly determined by the pitch of these SPFs [13]. In this study, we will see how these features in Cantonese influence learners' acquisition of English.

### 2. PRODUCTION EXPERIMENT

#### 2.1 Participants selection

We invited altogether 30 participants to take the pre-experiment test. These participants, aged 18-22, were all university students with a Cantonese background, and had passed the CET-4 (College English Test, Band 4). All of them were fluent in their mother tongue, and had no speech or hearing impairment. In order to make sure the intonation proficiency of them were in the same level, a reading test was conducted ahead of the recording. All the participants were required to read the passage *the North Wind and*

*the Sun*. The sounds of production was then rated by experimenters according to Zhao et al. [14] as what is listed below:

In doing the rating, if a phonological feature in a sentence was successfully realized, “1” score would be given, otherwise, they would be marked “0” instead. Scores in all the sentences would be added up as a participant’s performance in intonation production. Participants, whose performance in the middle level, would be eventually recruited and paid to take part in the recording. Altogether 13 participants were finally recruited to take the experiment.

Table 1: Participants’ Production Rating

	Nuclear accent placement	Pitch accent type	Boundary tone	Total
Sentence 1	1	1	1	3
Sentence 2	0	1	0	1
.....				
Total	12	2	1	5

## 2.2 Materials and recording

Altogether 18 sentences, including statements and yes-no questions were prepared which were cited from AESOP-CASS [15]. Eliciting questions were designed so that target sentences with 4 focus conditions, i.e. broad focus, sentence-initial focus, sentence-medial, and sentence-final focus, can be obtained.

Table 2: Examples of recording prompts

Focus condition	elicited question	Target sentence
Initial	Who left the socks on the fence?	<b>Mark</b> left the socks on the fence.
medial	What did mark leave on the fence?	Mark left <b>the socks</b> on the fence.
final	Where did Mark leave-the sock?	Mark left the socks <b>on the fence</b> .
broad	What happened?	Mark left the socks on the fence.
Initial	What departed from France on Sunday?	Did the <b>ship</b> depart from Japan in the afternoon?
Medial	From where did the ship depart from France on Sunday?	The ship departed from <b>Japan</b> in the afternoon.
Final	When did the ship depart from Japan?	The ship departed from Japan in the <b>afternoon</b> .
Broad	What did you say?	The ship departed from Japan in the afternoon.

The recording was conducted in a sound-proof booth. A PC-end software X-perception, was used to record the sentences at a sample rate of 22050 Hz with 16-bit resolution. Altogether 252 target sentences (including 136 statements and 116 yes-no questions) were collected for analysis.

## 3. PERCEPTION EXPERIMENT

After finishing the production, all the participants were asked to take the perception task. The stimuli were the previous 18 sentences read by an American native speaker. This speaker is a female international teacher working for UCASS, and she was born and raised in western-middle American female with no speech and hearing impairment.

In this task, each stimulus utterance was played 3 times to the learners, with 1s interval. The learners were required to pick the stressed words in these sentences. Before this procedure, learners were told that an utterance might have more than one stressed words or no stressed word. This process was conducted through a perception program presented in Figure 1.



Figure 1: Interface of Perception Program

## 4. PROSODIC ANNOTATION AND PERCEPTUAL RATING

Sentences produced by learners and the native English speaker were annotated by using ToBI system. The tonal realization can be labeled by 5 pitch accent types (H\*, L\*, L+H\*, L\*+H, H+!H\*), 2 phrase accent types (H-, L-), and 2 boundary tone types (H%, L%). We pinned down the nuclear accent in each sentence produced by learners and had them labeled with proper accent types. On this basis, we count how many accent types they produced and calculate the percentage of each type.

According to Table 3, we did a statistical analysis on the 13 participants’ nuclear accent perception produced by native speakers. 1 score would be given if a word was perceived as stressed. We added up the scored of all the learners for a certain sentence, and took the word with the highest score as the nuclear accent.

For example, in Table 3, “Can” is the nuclear accent perceived by learners in the sentence “Can Catherine come with Burnell.”. The nuclear accent types perceived by each learners were then counted for the comparison with those in production task.

Table 3: Perceived Accent Rating Chart

	Can	Catherine	Come	with	Burnell
Learner 1	1	1			1
Learner 2	1		1		
Learner 3	1				
.....					
Total	12	4	1	2	1

## 5. RESULTS

### 5.1 Production

Presented below are the results of nuclear accent types counted in production task.

According to AM intonation theory, English declarative contour has one or more than one pitch accent, usually realized as H\*, a low phrase accent (L-) and a low boundary tone (L%), while a yes-no question contour is cued by one or more than one low pitch accent (L\*), followed by a high phrase accent (H-) and a high boundary tone (H%) [8].

In the following part, learners’ phonological representation would be discussed and compared with that of native speakers.

It is noticed that nuclear accents in statements and yes-no question were mostly marked by H\* and L\* respectively. In addition, L tone in statements also accounted for a relatively high portion, so did the H\* in yes-no questions.

5.1.1 Nuclear Accent Type

Table 4: Nuclear Accent Type and Distribution of Learners' Production

statement	Pitch accent type	number	Percentage
	H*	91	66.91%
	L*	40	29.41%
	L*+H	4	2.94%
	L+H*	1	0.74%
Yes-no question	Pitch accent type	number	Percentage
	H*	46	39.66%
	L*	47	40.52%
	L*+H	22	18.97%
	L+H*	1	0.86%

Is the nuclear accent type produced by learners relevant to the focus position? The answer is presented in Figure 2 and Figure 3.

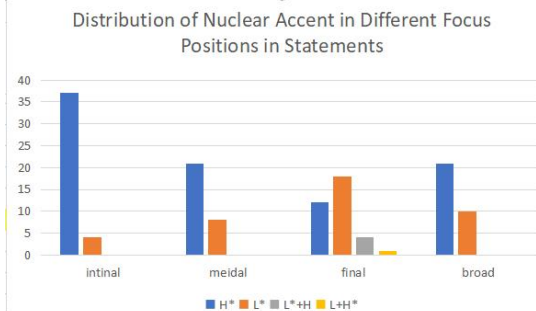


Figure 2: Distribution of Nuclear Accent Types in Different Focus Positions in Yes-No questions

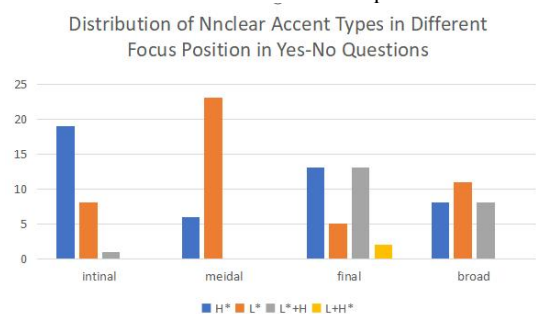


Figure 3: Distribution of Nuclear Accent in Different Focus Position in Yes-No Questions

It is known from Figure 2 that for H\*, it mainly distributed in sentences with focuses at the initial position, while L\* was mostly related to the sentence-final and broad focuses.

While for Yes-no questions, H tone was mostly distributed in sentences with initial and final focus. L\* was highly correlated with sentence-medial focuses.

5.1.2 Placement of Nuclear Accent

As to the nuclear accent placement, we scored learners' realization by the above-mentioned method and counted the corresponding accuracy. Presented in Table 5 are the results.

Table 5: Production Accuracy of Nuclear Accent Placement

position	Production accuracy
initial	67%
medial	80%
final	60%
broad	44%

From table 5, it is clear to see that sentence-initial and sentence-medial focus were relatively easy to pin down in production, while the rest two were harder to produce in a correct position.

5.1.3 Boundary tone Realization

It is known from Table 6 that except sentence-medial statements and sentence-final yes-no questions which were distributed in 3 boundary tone types, all the sentences were realized in 4 types of boundary tones. In addition, most of the statements were realized by low boundary tone (L-L%), except broad-focus ones which were mostly represented by both H-L% and L-L%; for yes-no questions, sentences of all the four focus conditions were encoded by both H-H% and L-H%.

Table 6: Boundary Tone Type Distribution in Production Test

Statement	Boundary tone	percentage		Yes-no question	Boundary tone	percentage	
		count	percentage			count	percentage
Initial	H-H%	3	7.3%	Initial	H-H%	13	46.4%
	H-L%	12	29.2%		H-L%	2	9.2%
	L-H%	2	4.8%		L-H%	11	36.1%
	L-L%	24	58.5%		L-L%	2	9.2%
Medial	H-H%	1	3.4%	Medial	H-H%	10	34.4%
	H-L%	5	17.2%		H-L%	1	3.4%
	L-H%	23	79.3%		L-H%	14	48.2%
	L-L%	4	18.7%		L-L%	4	18.7%
Final	H-H%	5	14.2%	Final	H-H%	16	50%
	H-L%	7	20.0%		H-L%	2	6.2%
	L-H%	3	8.5%		L-H%	14	43.7%
	L-L%	20	57.1%		L-L%	2	7.4%
Broad	H-H%	1	3.2%	Broad	H-H%	9	33.3%
	H-L%	11	35.4%		H-L%	1	7.1%
	L-H%	1	3.2%		L-H%	15	55.5%
	L-L%	18	58.0%		L-L%	2	7.4%

Based on Table 6, we further investigated the interactive relationship between boundary tone and pitch accent. In this part, we ignored the realization of phrase accent of the final prosodic phrase, and only focus on the falling and rising of boundary tones. In this way, boundary tone here was just classified into 2 type, i.e. H% and L%.

According to [16], in yes-no questions, H boundary tone usually goes with a close L pitch accent, and on the contrary, L boundary tone is associated with a close H pitch accent. Therefore, we would survey the interactive relationship between the boundary tone and its close pitch accent, or the final accent in the sentences.

Table 7: Interactive Relationship between Boundary Tone and Pitch Accent

Statement	Boundary tone	Pitch accent	Statement		Yes-no question		
			count	percentage	count	percentage	
	H boundary tone	H*	4/15	26.7%	L*	11/15	73.3%
	L boundary tone	H*	43/121	35.5%	L*	78/121	64.5%
Yes-no question	H boundary tone	H*	40/105	38.1%	L*	64/105	61.0%
	L boundary tone	H*	8/12	66.7%	L*	4/12	33.3%

It is known from Table 7 that both yes-no question and statement were in line with the discovery of [16]. That is to say, the realization of boundary tones of sentences is mostly correlated with the type of the close pitch accent. L tone boundary usually goes with a close H pitch accent, and vice versa.

5.2 perception

In this part, we would discuss the results obtained from perception task.

Like the results in production task, learners were easier to perceive nuclear accent marked by H\* and L\* in statements and yes-no questions respectively. In addition, L+H\* and L\*+H were also prone to be perceived by learners in the sentence types relatively.

It is known from Table 9 that sentences with a sentence-initial and medial focus, were easier to be correctly pin down by learners in perception.

The perception of boundary tone was not conducted since the experiment was done before the explanation of intonation theory to participants. Therefore for learners boundary tone is a strange concept and was hard to test their perception accuracy. We only focus on the distribution of boundary tone types in production.

Table 8: Nuclear Pitch Accent Type Distribution of Learners' Perception

statement	Pitch accent type	number	Percentage
	H*	79	48.47%
	L*	5	3.07%
	L*+H	39	23.92%
	L+H*	40	24.54%
Yes-no question	Pitch accent type	number	Percentage
	H*	29	34.12%
	L*	4	4.71%
	L*+H	33	38.82%
	L+H*	19	22.35%

Table 9: Perception Accuracy of Nuclear Accent Placement

position	Perception accuracy
<b>initial</b>	<b>40%</b>
<b>medial</b>	<b>76%</b>
final	23%
broad	32%

5.3 Relation between production and perception

According to Table 4, it is known that learners could produce 4 types of pitch accent in statements, and among them, H\* accounted for the largest portion, which was in line with their perception. In yes-no questions, the situation was same. Learners were mostly likely to realize nuclear accents by L\*, and were prone to perceive words with L\*+H, which is similar to L\*, as nuclear accents. Henceforth, we can come to a conclusion that EFL learners' production of phonological features is correlated with their perception.

6.DISCUSSIONS AND CONCLUSIONS

From the above discussion, it is known that Cantonese EFL English learners' production and perception were correlated with each other. In statements, learners were prone to produce and perceive H tone on the nuclear accent position, while in yes-no questions, L tone at the same position was easiest to be produced and perceived. Apart from nuclear accents, the production of pitch

accents close to the sentence end was correlated to the type of boundary tones they produced.

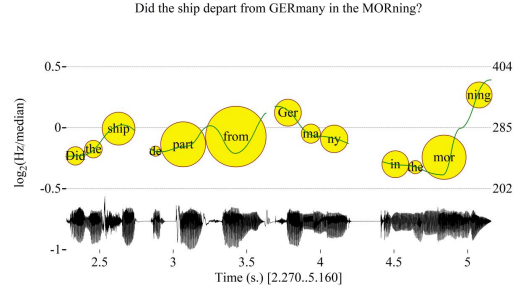


Figure 4: Examples of Accent-Boundary Tone Type Interaction

As is exemplified in Figure 4, in the yes-no question “Did the ship depart from Germany in the morning?”, the syllables were realized as pitch accent, i.e. “Ger” in “Germany” and “mor” in “morning”. “Mor”, as the accent closest to the sentence end, was realized as L tone, and went with a H boundary tone (H-H%), which is in line with the situation described in [16].

In addition, nuclear accent type was also discovered to be relevant to the their position in sentences. In statements, H\* and L\* were found to mainly distributed in sentences with the focus at the initial part, and sentence-final and broad focus respectively. Statements tend to be marked by downward trend, in addition, sentence-final focus is usually realized by duration lengthening, with the F0 seldom changed, therefore, broad and final focus in statement were prone to be marked by L\*. While for yes-no questions, H tone was mostly distributed in sentences with initial and final focus. L\* were highly correlated with sentence-medial focus. This is caused by learners' production deviation. To be specific, learners tend to generate a H\* in yes-no question if a interrogative words are followed by a personal pronoun or a proper name.

As to the realization of boundary tone itself, L boundary tone in statements was discovered to be the frequently used one among the 4 types in any focus conditions. While for yes-no questions, H-H% and L-H% were mostly used in all the 4 focus conditions. L-H% is not a typical boundary tone type of yes-no question. The high frequency of usage is possibly due to the property of Cantonese yes-no question whose final rising starts at the second half of the last syllable [1], leaving the first half of the syllable remaining low. Another phonetic feature of learners' L1 might also be the plausible reasons contributing to the occurrence of H-L% in statements and L-H% in yes-no questions. According to [2], tones Cantonese statements are usually ended with up with a small high tail, which would leads to the a H phrase tone of the final prosodic phrase, and further contribute to the formation of H-L%; as to yes-no questions, it is known that even unlike statements which is characterized by a significant downward trend, yes-no question is still found to be slightly falling till the last second syllable, which may cause an occurrence of L phrase boundary and further a L-H%. These feature is then transferred into learners' English production, contributing to a L phrase accent and a H boundary tone.

Apart from nuclear accent type and that of boundary tone, the accuracy of nuclear accent placement was also conducted in both production and perception tasks. The accuracy of production was still positively correlated that of perception. For learners, it was relatively easy for them to pin down the nuclear accents in sentence-initial and sentence-medial position in both statements and yes-no questions, leaving the rest 2 positions remaining hard for them to find. This is possibly due to the similarity of focus

realization in learners' L1 and L2. Specifically, duration lengthening is used to realized sentence-final and broad focus in English, which is the same as the means usually used by Cantonese speakers in their mother tongue. Learners in the process of L2 production had this feature transferred, but did not notice the difference between the two languages in the degree of lengthening, leaving the production and even the perception of focus in these positions less accurate.

This study provides suprasegmental evidence for EFL language learning models, and also give guidance to the future EFL intonation teaching.

### 7. ACKNOWLEDGEMENT

This work is supported by National Major Social Sciences Foundation of China under Grant 15ZDB103, and Innovation Program of Chinese Academy of Social Sciences.

### 8. REFERENCES

- [1] B. Robert, P. Mok. Final Rising and Global Raising in Cantonese Intonation, *Proceedings of ICPhS 2015*, Hong Kong, 2173-2176, 2015.
- [2] C. Ge. Cross dialectal Research on Chinese Intonation, Doctoral Dissertation, University of Chinese Academy of Social Sciences, 2020.
- [3] W. Gu, L. Liu. Declarative and Interrogative Mandarin Intonation by Native Speakers and Cantonese EFL Learners, *International Speech Communication Association (ISCA Archive)*, 2015.
- [4] W. Gu, Tone, Intonation and Emphatic Stress in EFL Mandarin Speech by English and Cantonese Learners, *Proceedings of ICPhS 2015*, Hong Kong, 2015.
- [5] M. G. Busa, A. Stella. Intonational Variations in Focus Marking in the English Spoken by North-East Italian Speakers, *MEFLP 2012*, Padova, 31-35, 2012.
- [6] W. Gao, Y. Xu, F. Mu. An Experimental Study on Teaching Prosodic Focus to Chinese EFL Learners, *Foreign Language Teaching and Research (bimonthly)*, China Academic Journal Electronic Publishing House, Beijing, 861-873, 2015.
- [7] J. Liu, Y. Xu, Y. Lee. Post-Focus Compression is not Automatically Transferred from Korean to EFL English, *Phonetics and Speech Sciences*, 15-21, 2019.
- [8] Pierrehumbert, J. 1980 The Phonology and Phonetics of English Intonation. Ph.D. Dissertation. MIT.
- [9] S. Chen, P. Mok. Questions Intonation in Hong Kong English: Interaction between Cantonese and English, *ICPhS 2015*, 2015.
- [10] Zee, E. Chinese (Hong Kong Cantonese), in the International Phonetic Association, *Handbook of the International Phonetic Association*, Cambridge, U.K.: Cambridge University Press, 56-60, 1999.
- [11] W. L. Wu, Y. Xu, Prosodic Focus in Hong Kong Cantonese without Post-focus Compression, *Speech Prosody 2010*, 2010.
- [12] D. R. Ladd, Intonation Phonology, Cambridge: *Cambridge University*, 1996.
- [13] S. P. Law, The syntax and Phonology of Cantonese Sentence-final Particle, Boston: *Boston University*, 1990.
- [14] C. Zhao, A. Li, Z. Xiong. Using Multimodal Methods in Teaching Intonation of Chinese EFL Learners. *Proceedings of OCOCOSDA 2020*, Yangon, 2020.
- [15] Y. Jia, A. Li, C. Tseng. Construction of English learners from Chinese dialectal regions, *Chinese Journal of Phonetics*, 38-45, 2013.
- [16] C. Cao, A. Li, A Study of Production and Perception of Focus by Cantonese EFL English Learners, *Phonetic Journal of China*, Beijing, 2022 (manuscript).

[This paper was published at OCOCOSDA 2022]