INTERSPEECH 2017
August 20–24, 2017, Stockholm, Sweden

The relationship between the perception and production of non-native tones

Kaile Zhang[1], *Gang Peng*[1,2]

[1]Department of Chinese and Bilingual Studies, The Hong Kong Polytechnic University, Hong Kong
[2]Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, China
kaile.k.zhang@connect.polyu.hk, gpeng@polyu.edu.hk

Abstract

To further investigate the relationship between non-native tone perception and production, the present study trained Mandarin speakers to learn Cantonese lexical tones with a speech shadowing paradigm. After two weeks' training, both Mandarin speakers' Cantonese tone perception and their production had improved significantly. The overall performances in Cantonese tone perception and production are moderately correlated, but the degree of performance change after training among the two modalities shows no correlation, suggesting that non-native tone perception and production might be partially correlated, but that the improvement of the two modalities is not synchronous. A comparison between the present study and previous studies on non-native tone learning indicates that experience in lexical tone processing might be important in forming the correlation between tone perception and production. Mandarin speakers showed greater improvement in Cantonese tone perception than in production after training, indicating that second language (L2) perception might precede production. Besides, both the first language (L1) and L2 tonal systems showed an influence on Mandarin speakers' learning of Cantonese tones.

Index Terms: production, perception, lexical tones

1. Introduction

Regarding the relationship between speech perception and production, several theories have proposed different views. The motor theory believes that the elementary events of both speech perception and production are the intended phonetic gestures in a specific module. By sharing the same set of elementary events, speech perception and production are intimately linked. More importantly, this link is not learned, but is innately specified [1]. Opposite to motor theory, psychoacoustic accounts of speech perception suggest that listeners use the basic auditory system to perceive acoustic signals [2, 3]. Although explicit explanations of the relationship between perception and production have not yet proposed, it can be deduced that the two modalities are not connected directly or tightly [4, 5]. The speech learning model (SLM) mainly discusses this issue from the perspective of L2 learning. It specifies that L2 perception precedes production [6]. The claim that production is guided by perceptual representations stored in long-term memory may indicate that perception and production share mental representations and that perceptual learning, at least, is needed to form mental representations. Meanwhile, the SLM also states that the perception and production of L2 may not be in perfect alignment [7].

Except for the above-noted theories, empirical studies have also been carried out to clarify the relationship between perception and production. However, the results of these studies are mixed. It has been found that the perception and production of vowels were correlated [8], which was also the case for consonants [9, 10], while other studies failed to find such correlations [11, 12, 13]. Evidence from the perspective of lexical tones is still lacking. Training regarding Dutch speakers' perception or production of Mandarin tones could be transferred to another modality [14]. English speakers' production of Mandarin tones was improved after perception-only training [15]. The successful transfer of the improvement between perception and production suggests that the two modalities are tightly connected by a module controlling both perception and production or by sharing mental representations [1]. However, one problem with these studies is that the training regarding the two modalities cannot be separated from each other. Usually, an auditory stimulus was played first to the subjects in the production training. For perception training, it was hard to avoid the subjects' implicit imitations. Some studies explored this issue by analyzing the correlation between the learning results of perception and production. German speakers were found to show a strong correlation in their perception and production of Mandarin tones after 40 minutes of intensive training [16]. However, the opposite result was also reported. Some naïve English speakers have a high perceptual sensitivity to the Mandarin tone contrasts, but they are not necessarily highly accurate at tone production, and vice versa [17].

Based on these few studies, it is hard to conclusively say whether non-native tone perception and production are correlated. The present study will further explore this question by investigating Mandarin speakers' perception and production of Cantonese tones. Both Mandarin and Cantonese are tonal languages, while Cantonese has a richer and more complex tone inventory. The detailed description of Mandarin and Cantonese tonal systems can be found in [18]. To avoid the possible demerits of the transferring method, the present study trained Mandarin speakers to perceive and produce Cantonese tones simultaneously and then evaluated whether the learning results of perception and production were correlated.

2. Methods

2.1. Subjects

Thirty-five Mandarin speakers (naïve to Cantonese) from Northern China and 17 native Hong Kong Cantonese speakers were paid to participate in the experiment. All these participants were undergraduates or postgraduates, with no self-reported visual, audio, or cognitive problems. They had never received linguistic, psychological, or musical training. Informed written consents were obtained from all participants before the experiment began.

  http://dx.doi.org/10.21437/Interspeech.2017-714

2.2. Stimuli

Another 12 native Hong Kong Cantonese speakers were recruited to make recordings of 36 Cantonese tonal syllables covering six Cantonese long tones (see Table 1) in a sound-attenuated booth. They were instructed to read each syllable 10 times in a natural way.

Table 1: *The 36 Cantonese syllables.*

Tone	fan/fɐn/	fu/fu/	jan/jɐn/	ji/ji/	se/sɛ/	si/si/
T55	婚	夫	因	醫	些	詩
T25	粉	苦	隱	倚	寫	史
T33	訓	富	印	意	卸	嗜
T21	焚	扶	人	兒	蛇	時
T23	奮	婦	引	耳	社	市
T22	份	負	孕	二	射	事

A small set of five speakers' recordings which were of the best clarity and stability were selected to form the training (two females' and two males' recordings) and the testing (one female's recordings) materials. Finally, 576 utterances (36 tonal syllables × four speakers × four samples) were used as the training materials and 36 utterances (36 tonal syllables × one speaker × one sample) were used as the testing materials.

2.3. Procedures

The whole experiment contained one two-week Cantonese tone training session and two sessions of Cantonese tone tests that were carried out before and after the training. Subjects' proficiency of Cantonese tones was evaluated by the identification and the production tasks. Cantonese subjects only participated in one test session since Cantonese tone training is not necessary for them, while Mandarin subjects finished all the sessions.

2.3.1. Training sessions

In order to train both perception and production at the same time, the present study employed the speech shadowing paradigm. Participants first heard a stimulus played by Praat [19] and then imitated this utterance. Each syllable was played in isolation (i.e., without context). The corresponding traditional Chinese character, the Jyutping, and the tone letter of each syllable were also displayed on the screen for the participants' reference. Each training set contains 576 utterances (36 tonal syllables × four speakers × four samples; about 30 minutes). Subjects finished six sets of training within two weeks in total (one set every two days).

2.3.2. Test sessions

The test sessions were composed of two tasks: the identification task and the production task. In the identification task, the testing stimuli (36 syllables × five repetitions) were randomly played to the subjects. The subjects needed to identify the tonal category of the stimulus played in each trial. The maximum allowable response time was 2,500 milliseconds (ms). In the production task, a traditional Chinese character (36 characters × three repetitions), together with its Jyutping and tone letter, were shown on the screen in each trial. Subjects were asked to read the character in a natural way, and their pronunciations were recorded via a microphone. To enable the subjects to familiarize themselves with the Cantonese tones in the pre-test, a 15-minute familiarization whose procedure was similar to that used in the training sessions was carried out before the experiment.

2.4. Data analysis

The accuracy of the identification task was used to reflect subjects' proficiency in Cantonese tone perception. The acoustic analysis was used to evaluate Mandarin subjects' production of Cantonese tones. First, the pitch height and the pitch slope of each tonal syllable produced by the 35 Mandarin speakers and 29 Cantonese speakers (12 informants and 17 subjects) were calculated based on the method in [18]. Then, a native norm for each tone category was obtained by averaging the 29 Cantonese speakers' productions of the same tone. The distance between a Mandarin subject's utterance and the native norm was calculated based on the equation (1):

$$D = \sqrt{(H_m - H_c)^2 + (S_m \times 10 - S_c \times 10)^2} \quad (1)$$

D, H_m, S_m, H_c, and S_c represent the distance, the pitch height and the pitch slope of the Mandarin subjects' utterances, and the pitch height and the pitch slope of the native norm, respectively. The smaller the distance, the better the pronunciation (i.e., the closer to the native norm). The correlation between the results of the perception and production tasks in the post-test was calculated to see whether learners who were good at tone perception also produced more native-like Cantonese tones.

3. Results

3.1. The identification task

The 35 Mandarin subjects' accuracy in the identification task was submitted to a two-way repeated measures ANOVA with *session* (pre-test and post-test) and *tone* (T55, T25, T33, T21, T23, and T22) as the within-subject factors. The Greenhouse–Geisser method was used to correct the violations of sphericity. The results revealed a significant main effect of *session*, $F(1,34) = 37.091$; $p < 0.001$, and *tone*, $F(5,170) = 120.728$; $p < 0.001$. The accuracy of the post-test ($M = 0.687$, $SE = 0.013$) was significantly higher than that of the pre-test ($M = 0.626$, $SE = 0.016$; $p < 0.001$), suggesting an improvement in Cantonese tone perception after training. The subjects' performance varied a lot across tone categories. T21 ($M = 0.94$, $SE = 0.012$), the only falling tone, and T55 ($M = 0.899$, $SE = 0.015$), the tone with the highest pitch height, were the easiest to acquire. T22 ($M = 0.269$, $SE = 0.024$) was the most difficult. T23 ($M = 0.63$, $SE = 0.03$), T33 ($M = 0.627$, $SE = 0.031$), and T25 ($M = 0.572$, $SE = 0.027$) were somewhat in between.

To see how the Mandarin subjects' tone perception differed from native Cantonese speakers', the accuracy of the Mandarin subjects in the post-test and the accuracy of the Cantonese subjects in the identification task were submitted to a two-way repeated measures ANOVA with *tone* (T55, T25, T33, T21, T23, and T22) as the within-subject factor and *group* (Mandarin and Cantonese) as the between-subject factor. The analysis showed main effects of *tone*, $F(5,250) = 93.641$; $p < 0.001$, and *group*, $F(1,50) = 32.278$; $p < 0.001$, and a significant *tone* × *group* interaction, $F(5,250) = 4.119$; $p < 0.05$. The simple main effect analysis showed that the Mandarin and Cantonese subjects were comparable in regard to identifying T23 and T21. However, for other tones (i.e., T22, T25, T33, and T55), the Mandarin subjects' accuracies were significantly lower than the Cantonese subjects' (*P's* <

0.05). For both the Mandarin and Cantonese subjects, the accuracies of T55 and T21 were highest and the accuracy of T22 was the lowest. The differences between the two groups lay mainly in two tone pairs: T21-T55 and T23-T33. The Mandarin subjects perceived T21 (M = 0.982, SE = 0.005) as being much easier than T55 (M = 0.928, SE = 0.013) and perceived T23 (M = 0.677, SE = 0.031) as being marginally easier than T33 (M = 0.649, SE = 0.029). However, the Cantonese subjects showed the opposite; that is, their accuracy in identifying T55 (M = 0.975, SE = 0.019) was slightly higher than that of T21 (M = 0.973, SE = 0.008) and the accuracy of T33 (M = 0.851, SE = 0.042) was higher than that of T23 (M = 0.786, SE = 0.045).

The confusion matrixes for the Mandarin subjects in the post-test and for the Cantonese subjects are shown in Table 2. As can be seen, tones that were acoustically similar, such as T23-T25 and T22-T33-T55, were easily confused by both groups. However, the Cantonese subjects discriminated three level tones much better than the Mandarin subjects. Specifically, the Cantonese subjects seldom misperceived T55, but the Mandarin subjects sometimes confused it with T33. Besides, the Cantonese subjects occasionally confused T33 as T22, which was acoustically closer to T33. However, Mandarin subjects misidentified it as T55 which exists in Mandarin, showing a strong L1 influence.

Table 2: *The confusion matrixes for (a) the Mandarin subjects in the post-test and (b) the Cantonese subjects.*

(a)

	R_21	R_22	R_23	R_25	R_33	R_55
T_21	98%	1%	1%	0%	0%	0%
T_22	1%	31%	1%	1%	55%	11%
T_23	1%	0%	68%	30%	0%	1%
T_25	0%	0%	41%	57%	0%	1%
T_33	1%	10%	1%	1%	65%	23%
T_55	0%	0%	0%	1%	6%	93%

(b)

	R_21	R_22	R_23	R_25	R_33	R_55
T_21	98%	1%	2%	0%	0%	0%
T_22	1%	49%	1%	1%	47%	1%
T_23	2%	1%	79%	17%	1%	0%
T_25	1%	1%	22%	76%	1%	1%
T_33	1%	14%	2%	1%	82%	1%
T_55	0%	0%	0%	1%	1%	98%

Note: The letter T in the table refers to the target responses and the letter R refers to the responses given by subjects.

3.2. The production task

The distances between the Mandarin subjects' utterances and the native norms (see Table 3) were first calculated based on the method described in 2.4 and were then submitted to a two-way repeated measures ANOVA with *session* (pre-test and post-test) and *tone* (T55, T25, T33, T21, T23, and T22) as the within-subject factors. Significant main effects of *session*, $F(1,34) = 13.326$; $p < 0.05$, and *tone*, $F(5,170) = 13.848$; $p < 0.001$, were obtained. The distance was notably reduced in the post-test (M = 0.664, SE = 0.03), compared with the pre-test (M = 0.742, SE = 0.03; $p < 0.05$). T55, which also exists in Mandarin, was pronounced most closely to the native norm (M = 0.468, SE = 0.032). The distances regarding T23 (M = 0.809, SE = 0.046) and T21 (M = 0.889, SE = 0.074) were comparative larger. T33 (M = 0.634, SE = 0.029), T25 (M = 0.7, SE = 0.036), and T22 (M = 0.718, SE = 0.041) were in between.

Table 3: *The native norm for each Cantonese tone category.*

	T55	T25	T33	T21	T23	T22
Pitch height	4.535	2.726	3.17	1.881	2.444	2.7
Pitch slope	-0.001	0.07	-0.01	-0.08	0.03	-0.013

3.3. The correlation between Cantonese tone perception and production

The correlation analyses were first carried out on the overall perceptual accuracy and production distance in the pre-test and the post-test. The results showed that the perceptual accuracy was negatively and significantly correlated with the production distance in the pre-test, $r(35) = -0.528$, $p < 0.05$, and in the post-test, $r(35) = -0.569$, $p < 0.001$, suggesting that the Mandarin learners who perceived the Cantonese tones better could also achieve a more native-like tone production.

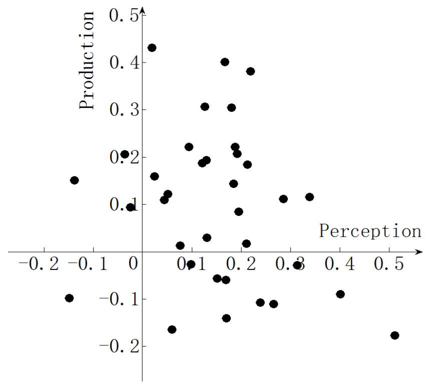

Figure 1: *The performance change of Mandarin subjects' Cantonese tone perception versus production after training.*

However, no significant correlation between the performance change in the perception task and the production task was found, $r(35) = -0.238$, $p = 0.168$. The performance change in the perception task was here defined as the post-test accuracy of the identification task minus the pre-test accuracy divided by one minus the pre-test accuracy [4]. The production change was almost the same: the post-test distance minus the pretest distance divided by zero minus the pretest distance. Each subject's performance change in the two modalities is shown in Figure 1.

4. Discussion

4.1. The correlation between the L2 tone perception and production

The results show that non-native tone perception and production are moderately correlated with each other in both the pre-test and the post-test. Mandarin learners who can perceive Cantonese tones accurately may also produce these tones in more native-like ways. The discrepancy between the results reported by [17] and the present study might be due to the differences in subjects' tonal experiences. The subjects in

[17] were English speakers with little experience in lexical tone processing. Before participating in the experiments, the subjects in [17] only heard each Mandarin tonal stimuli one to five times. However, the subjects in the present study were Mandarin speakers. Although they had never learned Cantonese before, they had gained rich experience in lexical tone processing from their L1. Similarly, the participants recruited in both [15, 16] which found a correlation between tone perception and production were experienced participants who had taken one or two semesters of Mandarin courses. It is possible that the correlation between tone perception and production does not exist at the very beginning of L2 learning but will be gradually formed with the increase in experience of tone processing [17]. This might explain why, even in the pre-test, the subjects' perception and production of Cantonese tones in the present study is already correlated. Further studies which strictly control subjects' language backgrounds need to be carried out to clarify whether the correlation between perception and production is experience-driven.

However, no correlation was observed between the degree of performance change in perception and production, suggesting that an improvement in one modality may not lead to an improvement in another. This is partially in opposition to previous studies (e.g., [14, 15]), which have suggested that learning in one modality could be transferred to another. Significant individual differences were observed in the performance change of two modalities. As can be seen from Figure 1, 21 Mandarin learners showed increased performance in both perception and production, but one Mandarin learner performed worse in regard to two modalities. A total of 10 learners improved only in tone perception and three learners improved only in production. Despite the individual differences, the general trend is still clear: Most Mandarin learners' perception, but fewer learners' production, were improved after training. This tendency may support the SLM which states that perception precedes production [7]. However, perceptual learning cannot always be incorporated into production and not all instances of L2 production have a perceptual origin [7, 15].

4.2. How L1 and L2 tonal systems affect the learning of L2 tones

The results show that Mandarin subjects' learning of Cantonese tones, especially tone perception, is affected by their native tonal system. T21 and T55 are of the highest accuracies in the identification task, since Cantonese T21 might be mapped to Mandarin T21 (the half Tone 3 in Mandarin) and Cantonese T55 might be perceived as a good exemplar of Mandarin T55. The confusion matrix [see Table 2 (a)] shows that T23 was frequently misperceived as T25, suggesting that Mandarin subjects might perceive both Cantonese T25 and T23 as exemplars of Mandarin T25. Therefore, lower accuracies in T25 and T23 were observed in the identification task. This is consistent with the prediction of the Perceptual Assimilation Model: The discrimination of two non-native tones is poor when they are perceived as belonging to a single native category, but if two non-native tones are perceived to belong to two separate native categories, discrimination is expected to be excellent [20, 21]. Mandarin subjects misperceive T33 as T55 which exists in Mandarin, but Cantonese subjects almost never confused T33 with T55; rather, they confused it with T22 which is acoustically closer to T33. Such differences further suggest that native speakers' perception is mainly based on acoustic cues but L2 learners' perception is influenced by their L1 phonological system first [22].

The influence of the L1 tonal system can also be observed in the Cantonese tone production. The production of T55, the tone existing in Mandarin, is comparatively closer to the native norm. T33, T22, and T23 which do not exist in Mandarin were produced much more poorly than others. Although T21 is acoustically similar to Mandarin T21, its pronunciation was the worst. This is probably because Mandarin T21 only exists in continuous speech, but the present study asks Mandarin subjects to produce the tonal syllables in isolation.

The features of the L2 tonal system also play an important role in learning Cantonese tones. Compared with Mandarin, one significant feature of the Cantonese tonal system is that more than one tone category share similar pitch contours. For example, both T25 and T23 are spoken with rising pitch contours. As suggested by the TRACE model, the mental representations of these two similar tones perhaps also share some components at the feature level [23]. When a rising tone is played, it is possible that both the mental representations of T23 and T25 are activated. Mandarin learners who have not formed robust connections between input signals and mental representations may choose either T23 or T25 randomly. Therefore, except for the influence of Mandarin T25, the mutual interference of Cantonese T25 and T23 also led to poorer discrimination of these two tones. However, tones of similar pitch contours do not always interfere with each other. For example, Mandarin subjects seldom confused T22 with T55, since their pitch heights were notably different. It seems that interference from the target tonal system occurs when this tonal system contains several tones of similar pitch contours and similar pitch heights. This might explain why both T25 and T55 exist in Mandarin, but the perception and production of T25 were worse than that of T55.

5. Conclusion

The results of the present study suggest that non-native tone perception and production are moderately correlated with each other. Learners who achieve high accuracy in tone perceptions might also produce the tones more accurately. However, improvement in tone perception and production is perhaps partially separated or constrained by different factors, since the results also show that something acquired via perception does not always lead to a corresponding improvement in production, and vice versa. Besides, Mandarin subjects are affected by both the L1 and L2 tonal systems while learning Cantonese tones.

6. Acknowledgement

This study was supported in part by a grant from the Research Grant Council of Hong Kong (GRF: 14411314), and a grant from National Natural Science Foundation of China (NSFC: 11474300).

7. References

[1] A. M. Liberman and I. G. Mattingly, "The motor theory of speech perception revised," *Cognition*, vol. 21, pp. 1–36, 1985.

[2] K. N. Stevens and S. E. Blumstein, "The search for invariant acoustic correlates of phonetic features," In *Perspectives on the Study of Speech*, P. D. Eimas, J. L. Miller, Ed. Hillsdale, NJ: Erlbaum; 1981. pp. 1–38.

[3] R. L. Diehl and K. R. Kluender, "On the objects of speech perception," *Ecological Psychology*, vol. 1, pp. 1–45, 1989.

[4] A. S. Rauber, P. Escudero, R. A. H. Bion, and B. O. Baptista, "Training Japanese listeners to identify English /r/ and /l/: IV. Some effects of perceptual learning on speech production," *The Journal of the Acoustical Society of America,* vol. 101, no. 4, pp. 2299–2310, 1997.

[5] C. T. Best, "A direct-realist view of cross-language speech perception," In *Speech Perception and Linguistic Experience: Issues in Cross-Language Speech Research*, W. Strange, Ed. Timonium, MD: York; 1995. pp. 171–206.

[6] J. E. Flege, "Second language speech learning: Theory, findings, and problems," In *Speech Perception and Linguistic Experience: Issues in Cross-language Research*, W. Strange, Ed. Baltimore: York Press; 1995. pp. 233–277.

[7] J. E. Flege, "The relation between L2 perception and production,'' in *Proceedings of the XIVth International Congress of Phonetics Sciences*, J. Ohala, Y. Hasegawa, M. Ohala, D. Granveille, and A. Bailey, Ed. Department of Linguistics, University of California at Berkeley, Berkeley, CA; 1999. pp. 1273–1276.

[8] A. S. Rauber, P. Escudero, R. A. H. Bion, and B. O. Baptista, "The interrelation between the perception and production of English vowels by native speakers of Brazilian Portuguese," In *Proceedings of Interspeech 2005*, pp. 2913–2916, 2005.

[9] D. C. Kluge, A. S. Rauber, M. S. Reis, and R. A. H. Bion, "The relationship between the perception and production of English nasal codas by Brazilian learners of English," In *Proceedings of Interspeech 2007*, pp. 2297–2300, 2007.

[10] R. S. Newmana, "Using links between speech perception and speech production to evaluate different acoustic metrics: A preliminary report," *The Journal of the Acoustical Society of America,* vol. 113, no. 5, pp. 2850–2860, 2003.

[11] M. L. Zampini, "The relationship between the perception and production of L2 Spanish stops," *Texas Papers in Foreign Language Education*, vol. 3, no. 3, pp. 85–100, 1998.

[12] S. Peperkamp and C. Bouchon, "The relation between perception and production in L2 phonological processing," In *Proceedings of Interspeech 2011*, pp. 161–164, 2011.

[13] K. Hattori and P. Iverson, "Examination of the relationship between L2 perception and production: An investigation of English /r/-/l/ perception and production by adult Japanese speakers," in *Interspeech Workshop on Second Language Studies: Acquisition, Learning, Education and Technology,* M. Nacano, Ed. Tokyo: Waseda University; 2010. pp. 2–4.

[14] J. Leather, "Perceptual and productive learning of Chinese lexical tone by Dutch and English speakers," in *New sounds 90: Proceedings of the Amsterdam Symposium on the Acquisition of Second-Language Speech,* J. Leather and A. James, Eds., Amsterdam: University of Amsterdam, pp. 72–97, 1990.

[15] Y. Wang, A. Jongman, and J. Sereno, "Acoustic and perceptual evaluation of Mandarin tone productions before and after perceptual training," *Journal of the Acoustical Society of America*, vol. 113, no. 2, pp. 1033–1043, 2003.

[16] H. Ding, O. Jokisch, and R. Hoffmann, "Perception and production of Mandarin tones by German speakers," In *Proceedings of 5th Conference on Speech Prosody*, 100153, pp. 1–4, 2010.

[17] T. Bent, "Perception and production of non-native prosodic categories," Evanston, Illinois: Department of Linguistics, Northwestern University; 2005.

[18] G. Peng, "Temporal and tonal aspects of Chinese syllables: A corpus-based comparative study of Mandarin and Cantonese," *Journal of Chinese Linguistics*, vol. 34, no. 1, pp. 134–154, 2006.

[19] P. Boersma and D. Weenink, "Praat: Doing phonetics by computer (Version 5.3.23) [Computer program]," http://www.praat.org (Last viewed on August 7, 2012), 2012.

[20] C. K. So and C. T. Best, "Cross-language perception of non-native tonal contrasts: Effects of native phonological and phonetic influences," *Language and Speech*, vol. 53, no. 2, pp. 273–293, 2010.

[21] C. K. So and C. T. Best, "Phonetic influences on English and French listeners' assimilation of Mandarin tones to native prosodic categories," *Studies in Second Language Acquisition*, vol. 36, no. 2, pp. 195–221, 2014.

[22] G. Peng, H. Y. Zheng, T. Gong, R. X. Yang, J. P. Kong, and W. S. Y. Wang, "The influence of language experience on categorical perception of pitch contours," *Journal of Phonetics*, vol. 38, no. 4, pp. 616–624, 2010.

[23] J. L. McClelland and J. L. Elman, "The TRACE model of speech perception," *Cognitive Psychology*, vol. 18, pp. 1–86, 1986.