Directional asymmetry in lexical tone perception

Ratree Wayland, Si Chen, Fang Zhou, et al.

Citation: Proc. Mtgs. Acoust. **39**, 060005 (2019); doi: 10.1121/2.0001300
View online: https://doi.org/10.1121/2.0001300
View Table of Contents: https://asa.scitation.org/toc/pma/39/1
Published by the Acoustical Society of America

Volume 39 http://acousticalsociety.org/

178th Meeting of the Acoustical Society of America

San Diego, CA

2-6 December 2019

Speech Communication: Paper 1aSC5

Directional asymmetry in lexical tone perception

Ratree Wayland
Department of Linguistics, University of Florida, Gainesville, Florida, 32611-5454; ratree@ufl.edu

Si Chen, Fang Zhou, and Yitian Hong
Department of Chinese and Bilingual Studies, The Hong Kong Polytechnic University, Hong Kong, HONG KONG; sarah.chen@polyu.edu.hk; fang-zf.zhou@polyu.edu.hk; hongyt@link.cuhk.edu.hk

Directional asymmetries in cross-linguistic vowel discrimination, in which a change from a less peripheral vowel to a more peripheral vowel was found easier to detect than the reverse direction, has been well documented. However, the perceptual processes underlying the phenomenon remain to be fully understood. This study explored asymmetries in Mandarin lexical tone discrimination by native Mandarin and native Cantonese listeners. Both groups of listeners found a change from Mandarin tone 1 to all other Mandarin tones easier to detect than the reverse direction. However, the opposite was true for Mandarin tone 3. Neither processing load nor training had an impact on the results. Stimulus dynamicity, phonological underspecification and language-independent acoustic salience were discussed as possibly to account for the observed asymmetric pattern.

Published by the Acoustical Society of America

 https://doi.org/10.1121/2.0001300

1. INTRODUCTION

A change from a less peripheral vowel to a more peripheral vowel is easier to detect than the reverse (e.g., Polka and Bohn 2003, 2011, Masapollo, Polka and Ménard 2017, Masapollo et al., 2017, 2019). According to the Natural Referent Vowel (NRV) framework (Polka and Bohn 2003, 2011), asymmetric vowel perception arises from a universal, speech-specific bias, favoring vowels produced with extreme articulatory postures and, thus, a greater degree of convergence of acoustic energy in a narrow spectral region (Masapollo et al., 2019). For instance, extreme tongue height and fronting gestures during the production of a high front vowel /i/ result in close proximity between the second, the third and the fourth formants (F2, F3 and F4), whereas F1 and F2 converge for the lowest back vowel /ɑ/ and the highest back vowel /u/, giving rise to well-defined spectral peaks in these frequency regions. According to the NRV, listeners are attuned to these auditorily salient 'focal' vowels such that a change from a less focal vowel is more discriminable than the reverse direction. More recently, the stimulus dynamic hypothesis was proposed to account for acoustic salience of focal vowels (Masapollo et al., 2019). According to this hypothesis, dynamic spectral change (i.e., dynamic formant trajectories) rather than spectral proximity drives asymmetries in vowel discrimination. Specifically, extreme articulatory gestures involved in extreme vowel production result, not only in greater focalization of acoustic energy, but also in a greater rate of acoustic energy movement (i.e., greater spectral dynamics). By this account, a change from a less focal vowel to a more focal vowel is easier to detect than the opposite direction due to the focal vowel's steeper formant trajectories. Support for the stimulus dynamics hypothesis was reported in Masapollo et al. (2019) using nonspeech tonal stimuli.

Importantly, the NRV framework views asymmetries in vowel discrimination as a reflection of the listeners' speech-specific bias that emerge when they perceive speech, rather than their sensitivity to raw acoustic energy (i.e., psychoacoustic bias) at a lower-level, auditory processing (Polka & Bohn, 2011; Polka, Ruan and Masapollo, 2018). Finding that asymmetries in vowel discrimination emerged at a longer interstimulus interval (ISI = 1,500ms), but not at shorter ISIs (500-ms and 1,000-ms) when demand on processing load is reduced was interpreted to support this view (Masapollo, Franklin, et al., 2018).

An alternative, but not mutually exclusive, prototypicality account suggests that vowel phonemic status, rather than its acoustic salience, plays a role in determining vowel perceptual asymmetric pattern. For example, 6–8 and 10–12 months old English-learning infants were better at detecting the change from /y/ to /u/ (rather than from /u/ to /y/), and from /ʏ/ to /ʊ / (rather than from / ʊ/ to /ʏ/) (Polka and Werker, 1994). According to Polka and Werker (1994), these results are consistent with the perceptual magnet effect previously observed in within-category vowel discrimination among infants and adults reported by Kuhl (1991) and Kuhl et al. (1992) and suggest that a more prototypical or native-like vowel (i.e., /u/ and /ʊ/) rather than a less native-like vowel (i.e., /y/ and /ʏ/) acts as a perceptual magnet and thus reduces psychoacoustic distance between it and other members of the same category. However, the finding that both English and German learning 6-8- month and 12-month-old infants prefer a change from /y/ to /u/ (a non-English contrast) and from /ɛ/ to /æ/ (a non-German contrast) rather than the opposite direction reported by Polka and Bohn (1996) suggests a language-universal, rather than a language-specific bias behind directional asymmetries in vowel discrimination.

Directional asymmetries have also been documented in both infant and adult perception of lexical tones. For example, Tsao (2008) found that a stimulus change from the background Mandarin T1(55) to the target Mandarin T3 (214) was easier than the reverse among one-year-old

Mandarin learning infants. According to the NRV's stimulus dynamics account, this finding suggests that Mandarin infants are more attuned to a more dynamic Mandarin T3 than a static Mandarin T1. On the contrary, in an ERP study, Politzer-Ahles et al. (2016) found that Mismatch Negativity (MMN) was attenuated among both native and non-native Mandarin listeners when Mandarin T3 was the standard and other Mandarin tones the deviant in comparison to the reverse. By the stimulus dynamics account, this would imply that Mandarin T3 is less acoustically salient than other Mandarin tones. In addition, Yeung et al. (2013) reported that 4- and 9-month-old Mandaring learning infants are better at Cantonese tone discrimination after being familiarized with tone 25 than with tone 33. By the NRV account, this finding suggests that static tone is more salient than contour tone for Cantonese-learning infants.

It is difficult to directly compare these studies since different groups of listeners and stimuli were used. Thus, asymmetric patterns of lexical tones remain unclear, the underlying processes and stimulus characteristics driving these processes also remain to be explicated.

The goal of this study is to fill this research gap by examining asymmetries in lexical tone perception in Mandarin among native Mandarin and native Cantonese listeners. Due to differences in their tonal inventories, Mandarin speakers are more sensitive to pitch contour while native Cantonese speakers are more sensitive to pitch height (Peng et al., 2012). If the NRV's stimulus dynamics hypothesis is correct, a change from a less dynamic F0 pattern to a more dynamic F0 contour should be more easily detectable than the reverse, and this should be true for both groups of speakers. However, if asymmetries reflect an experience-dependent processing bias favoring natively salient F0 dimensions, different asymmetric patterns are expected between the two groups.

We also investigated the effects of processing load or processing level on tonal asymmetric patterns by varying the interstimulus-intervals (ISI). A shorter ISI (250ms) places a weaker demand on the working memory and activates an auditory mode of processing where stimuli are processed according to their auditory attributes, whereas a longer ISI (1,000ms) imposes a greater demand on the working memory and activate a phonological mode of processing, inducing the influence of the native tonal categories. According to the NRV framework, asymmetries in non-native tone discrimination should diminish at a shorter ISI.

Finally, we examined the effects of training on perceptual asymmetries. Here we are interested in learning if short-term exposure to Mandarin tones would affect asymmetric responses from the Cantonese listeners. Greater familiarity with Mandarin tones after training may decrease or increase asymmetries depending on whether the training leads to increased low-level sensitivity to acoustic properties of Mandarin tones or to a greater mapping between Mandarin and Cantonese tones at the phonological level.

2. THE CURRENT STUDY

The overall design of this study involved administering an AX categorial discrimination task to adults native speakers of Mandarin and native speakers of Cantonese before and after a brief training using naturally produced Mandarin tone stimuli. The study was guided by three research questions.

1. Are asymmetric patterns the same or different between Mandarin and Cantonese listeners?
2. Do asymmetric patterns vary as a function of ISI?
3. Does increased exposure to Mandarin tones affect tonal asymmetric patterns among Cantonese speakers?

A. METHODS

I. SUBJECTS

Twenty native Cantonese speakers (ten females; ten males) aged between 20 and 25 (Mean ± SD: 22.2 ± 2.09) and twenty native Mandarin speakers (ten females; ten males) aged between 21 and 30 (Mean ± SD: 24.65 ± 2.18) participated in the study. All Cantonese subjects were born and raised in Hong Kong, having stayed in no other places except Hong Kong. Their formal language learning of Mandarin started after 6 years old. Mandarin subjects were born in Northern parts of China and stayed in their hometown before the age of 18. They have not received any formal training of Cantonese and have been in Hong Kong for less than 4 years. No participants reported any problem of speaking, hearing, or reading and none have received formal musical training for more than 10 years or within the past 5 years.

II. STIMULI

Stimuli were /ni/ syllables produced with all 4 Mandarin tones by a female and a male speaker. Three tokens of each tone were produced by each speaker. All possible pairings of the 4 tones were presented for discrimination in both directions separated by two different ISIs: 250ms and 1,000ms. The 250ms ISI activates a nonlinguistic acoustic processing whereas the 1,000ms ISI reflects a linguistic, phonetic processing (Werker & Tees, 1984). Same pairings were also included. The categorical AX, same different discrimination was used. Both stimuli were always physically different but were produced by the same speaker. Two blocks of stimuli, one for each ISI, were presented. Each block contained 144 trials (72 same trials and 72 different trials).

III. PROCEDURE

Pre-post tests

After filling out a language background questionnaire, each participant was given a set of 10 practice trials with correct feedback provided. The goal of the practice was to familiarize the participant with the task. The stimuli used in the practice session were produced by a different female and a different male speaker who did not produce the test stimuli. In addition, an ISI of 500ms was used. After completing the practice session, each participant took the pre-test which was comprised of two blocks of trials. Trials in each block were randomly presented and the presentation of the two blocks were counterbalanced across subjects. After hearing each trial, the participant was asked to decide if the two stimuli comprising each trial carry the same or different tone by pressing the button “1” or “2”, respectively, on the keyboard. The post-test, which is identical to the pre-test, was administered after a brief training session. All three sessions were administered on the same day and lasted approximately 1.5 hour. Breaks were given as needed.

Training

After the pre-test, a brief training session was administered to all participants. During training, each participant listened to 60 trials each of Mandarin tones 1 and 4 produced by the same two speakers who produced the test trials. The participants were told that they would hear Mandarin tone 1 (or tone 4) and that their task was to try their best to learn to recognize the tone and to click ‘next’ when they were ready to hear the next trial. Trials were randomly presented. Half of the participants were trained on tone 1 first and the other half on tone 4 first.

B. RESULTS

D-prime scores [d' = z(H) -z(F)] were computed for each contrast and for each subject based on the proportion of 'hits' (i.e., correct response on 'different' trials) and false alarms (i.e., incorrect response on 'same' trials).

To examine asymmetric patterns, d-prime data for each tone for each listener group were separately submitted to linear mixed-effects models with ISI (250ms and 1,000ms), Training (pre- and post-) and Position [left (A) and right (X)] as fixed factors and intercept for subjects as random factor using R (R Core Team, 2012) and lme4 (Bates, Mächler, Bolker and Walker, 2015). The R syntax for these models is as follows:

```
lmer(dPrime ~ Training+ISI+Position +Training:ISI+Training:Position+ISI:Position
        +(1|Subject), REML = FALSE, data=data.M)
```

I. MANDARIN LISTENERS

For Mandarin listeners, the analyses yielded a significant main effect of Position (i.e., as the first (left) or second (right) member of the AX pair) only for tone 1 [χ^2 (1) = 17.15, p < .001] and tone 3 [χ^2 (1) = 23.41, p < .001]. As shown in Figure 3, these results suggested that a change from tone 1 to all other tones was significantly easier to detect than the opposite direction. On the other hand, a change from tone 3 to all other tones was significantly more challenging than the reverse direction. No directional asymmetries were found for tone 2 and tone 4 and no other significant main effects or interactions were found.

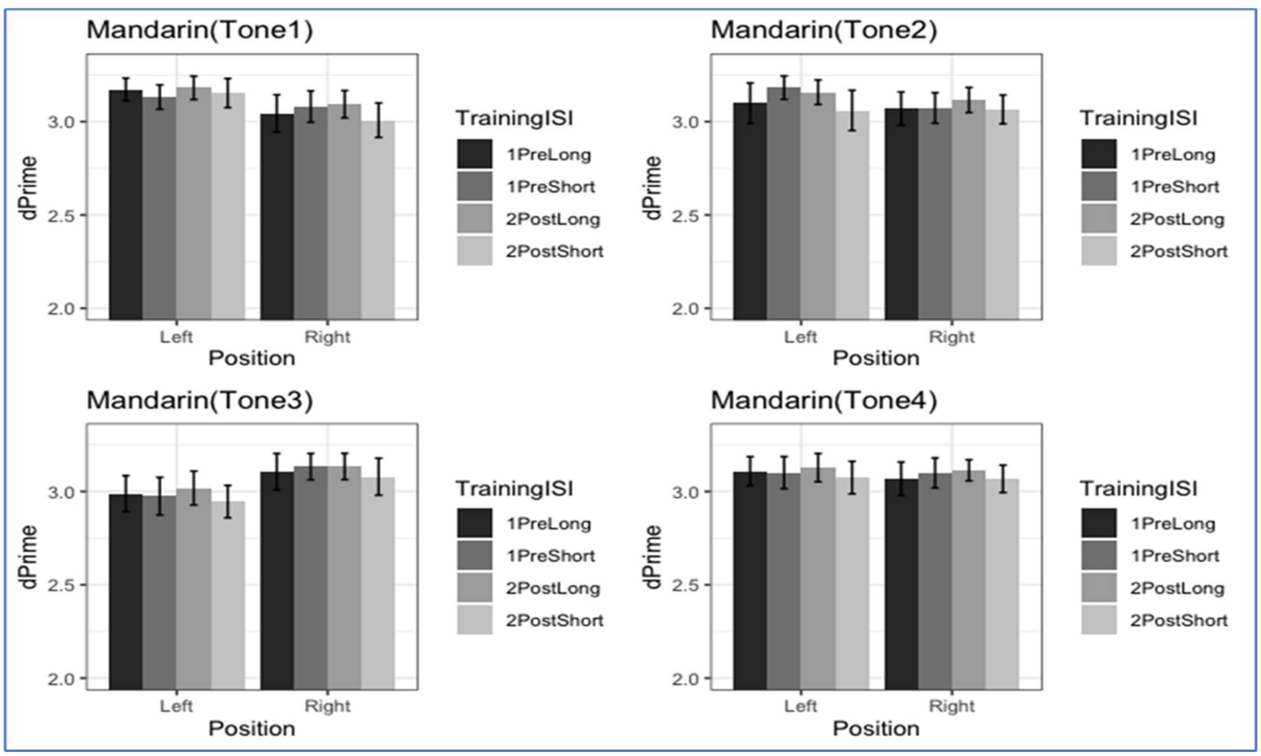

Figure 3. D-prime scores for 4 Mandarin tones when they are heard as the first (left) or the second (right) of the AX pair before and after training under both short (250ms) and long (1,000ms) for native Mandarin listeners.

I. CANTONESE LISTENERS

For the Cantonese listeners, a significant main effect of Position was found for Mandarin tone 1 only [χ^2 (1) =11.65, $p < .001$.]. Like Mandarin listeners, Cantonese listeners found it significantly easier to detect a change from this tone to all other Mandarin than the reverse. However, unlike Mandarin listeners, significant main effects for both Position [χ^2 (1) =16.63, $p < .001$] and Training [χ^2 (1) =5.44, $p < .05$] were significant for tone 3. Like Mandarin listeners, Cantonese listeners found a change from Mandarin tone 3 to all other Mandarin tones significantly more difficult to detect than the reverse, and this became worse after training.

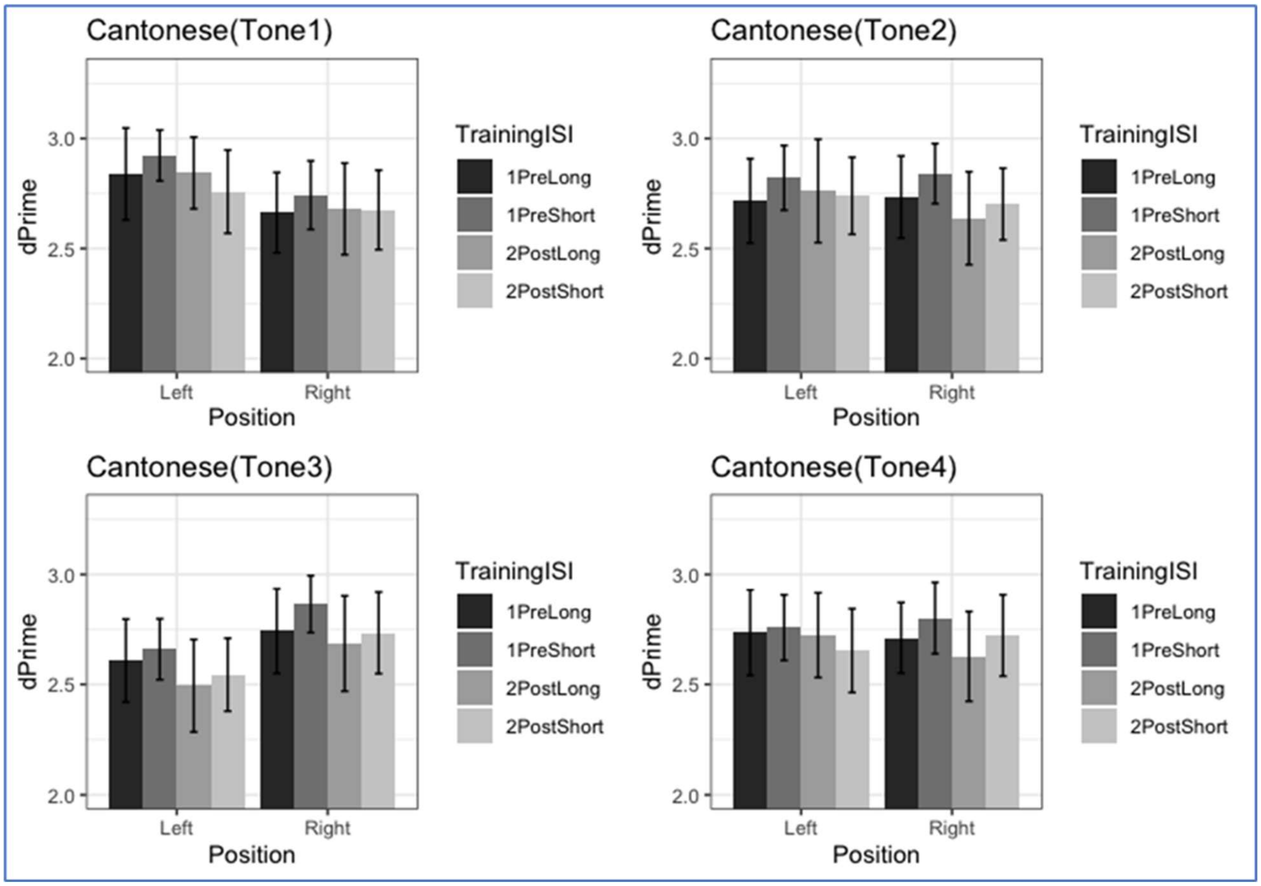

Figure 4. D-prime scores for 4 Mandarin tones when they are heard as the first (left) or the second (right) of the AX pair before and after training under both short (250ms) and long (1,000ms) for native Cantonese listeners.

3. DISCUSSION

Asymmetric response pattern is common in vowel perception among both infants and adults, but asymmetries in lexical tone perception has been relatively understudied. However, the underlying processes accounting for this ubiquitous phenomenon remains to be fully understood. Based on cross-language vowel discrimination results, a Natural Vowel Reference (NRV) framework proposed that proximity of vowel formants and/or steeper formant movement (i.e., greater spectral dynamics reflecting greater acoustic energy change) is critical cue to observed asymmetries in vowel perception. Specifically, by this account, a change from less focal vowel to a more focal vowel produced with a greater concentration or a greater movement of acoustic energy (e.g., the stimulus dynamic account) is easier to detect than the reverse direction. More importantly, this effect does not reflect the listeners' sensitivity to raw acoustic properties of the stimuli but emerged only when the stimuli are perceived as speech.

In this present research, we explored the role of language background, processing level or load and training on lexical tone perceptual asymmetries among both native (Mandarin) and non-native (Cantonese) listeners. Our results demonstrated that, like vowels, lexical tone discrimination was more successful when presented in one direction compared to the reverse. Specifically, we found that both Mandarin and Cantonese listeners found a change from Mandarin tone 3 to all other tones significantly less discriminable than the reverse. This finding may be interpreted as being consistent with the stimulus dynamic account since in citation form (as is produced in this study) or in phrase-final position, this Mandarin tone is described as a dipping tone with a low-falling then rising F0 contour and is, therefore, more dynamic than all other Mandarin tones (Politzer-Ahles et al. 2016). It should be pointed out, however, that thus far, support for the stimulus dynamic hypothesis has been reported only for non-speech tonal stimuli with a linear falling F0 trajectory (Masapollo et al, 2019). In that study, steepness of falling tone stimuli was manipulated and a change from a falling tone with a shallower slope to a falling tone with a steeper slope was easier to discriminate than the reverse. On the other hand, Mandarin tone 3 exhibits both a falling and a rising contour. Its initial fall is shallower than Mandarin tone 4 whereas its final rise is less steep than Mandarin tone 2. If degrees of frequency fall (or rise) is a measure of dynamicity, as was the case in Masapollo et al, (2019), then it can be argued that Mandarin tone 3 is less dynamic than either Mandarin tone 2 and 4, and this result would be considered incompatible with the stimulus dynamic hypothesis.

However, this finding is consistent with Politzer-Ahles et al. (2016)'s result that Mismatch Negativity (MMN) was attenuated when Mandarin T3 was the standard compared to when other Mandarin tones were the standard. These authors proposed that lexical underspecification of Mandarin Tone 3 among Mandarin listeners may account for this finding. That is, with some of its features left unspecified, Mandarin Tone 3 may be less discriminable from other tones.

However, the underspecification account cannot explain why native Cantonese speakers also found a change from Mandarin tone 3 to other tones harder to detect than the reverse since they have no access to underlying phonological representation of this tone (unless a cross-linguistics underspecification for Mandarin T3 is assumed). It is more likely that the Cantonese listeners' lack of sensitivity to pitch direction (as opposed to pitch height) accounts for this finding. Unable to detect the falling-rising pitch contour of Mandarin tone 3, Cantonese listeners may have paid more attention to other pitch dimensions, particularly height and/or average pitch, rendering it less distinct compared to other tones. Surprisingly, this became worse after a brief exposure (i.e., training) to Mandarin tones 2 and 3. The reason as to why this was the case is not readily clear. It is possible that confusion between these two tones increased rather than decreased after the brief training, and this, in turn, exacerbated an existing lack of distinction between tone 3 and other tones. Further research is needed to test this hypothesis.

On the other hand, the Cantonese listeners' sensitivity to pitch height may explain why they found a change from Mandarin tone 1 (high-level) to all other tones easier to detect than the reverse. Like Cantonese listeners, Mandarin listeners also found a change from Mandarin tone 1 to other tones easier to detect than the reverse. This result is also consistent with Tsao (2008)'s report that a change from the background Mandarin T1 (high-level) to the target Mandarin T3 (213) was easier than the reverse. Mandarin tone 1, or more generally a high-level tone, is a good perceptual anchor point cross-linguistically in lexical tone discrimination due likely to its distinct pitch height. Further research involving native speakers of other tonal languages is required to substantiate this claim.

This result is also consistent with the stimulus dynamics hypothesis since, as a level tone, Mandarin tone 1 is acoustically less dynamic than other Mandarin tones and a change from a less dynamic tone to a more dynamic tone is predicted by the hypothesis. However, future research with level tones of different height is needed to determine whether pitch shape or pitch height was responsible for the observed discrimination ease between Mandarin tone 1 and all other Mandarin tones.

Asymmetric patterns did not vary as a function of the ISI, suggesting a lack of memory load or processing level on perceptual processes underlying asymmetries in lexical tone perception, at least among native speakers of lexical tone languages. This result was unexpected according to the NRV account of asymmetries in vowel discrimination. By this account, asymmetries in vowel discrimination are derived from speech-specific phonological rather than basic psychoacoustic processes. Therefore, asymmetries were expected to emerge under a longer ISI when a greater demand was placed on phonological working memory and to diminish under a shorter ISI when the stimuli can be processed at the lower, auditory level of processing. It is possible that the longer ISI of 1000ms that we used was not long enough for the differential effect to be observed. An ISI of 1,500ms was used in Masapollo, Franklin, et al. (2018). However, the fact that asymmetric responses emerged even under the two shorter ISIs of 250ms and 1,000ms was inconsistent with the finding of Masapollo, Franklin, et al. (2018) and the NRV account. It is possible that, unlike vowel processing, these two ISIs activate phonemic processing, particularly among native speakers of a tone language. Further research focusing on different processes involved in vowel and lexical tone discrimination under different processing load among both native and non-native speakers will shed more light on this issue.

4. CONCLUSION

This study established that, like vowels, asymmetries are also found in lexical tone perception among adult native (Mandarin) and non-native (Cantonese) listeners. Both groups found a change from a high-level tone (Mandarin tone 1) to all other Mandarin tones to be significantly better to detect than the reverse direction, whereas a change from Mandarin tone 3 to other tones was significantly more challenging compared to the opposite direction. Further research is needed to ascertain whether the stimulus dynamics hypothesis previously proposed to account for asymmetries in vowel perception or language-dependent phonological representation and/or language-independent, acoustic salience account for the observed asymmetric patterns.

ACKNOWLEDGMENTS

This study was supported by two grants awarded to the first two authors: 1) The effects of training on directional perceptual asymmetry in lexical tone perception (88DW), Department of Chinese and Bilingual Studies, The Hong Kong Polytechnic University; 2) Perceptual Asymmetry in Mandarin Tone Processing and the Training Effects (ZVNV), Faculty of Humanities, The Hong Kong Polytechnic University

REFERENCES

Bates, D., Mächler, M., Bolker, B., & Walker S (2015). Fitting Linear Mixed-Effects Models Using lme4. Journal of Statistical Software, **67**(1), 1–48. doi: 10.18637/jss.v067.i01.

Kuhl, P. K. (1991). Human adults and human infants show a "perceptual magnet effect" for the prototypes of speech categories, monkeys do not. Perception & psychophysics, **50(2)**, 93-107.

Kuhl, P. K., Williams, K. A., Lacerda, F., Stevens, K. N., & Lindblom, B. (1992). Linguistic experience alters phonetic perception in infants by 6 months of age. Science, **255(5044)**, 606-608.

Masapollo, M., Franklin, L., Morgan, J., & Polka, L. (2018). Asymmetries in vowel perception arise from phonetic encoding strategies. The Journal of the Acoustical Society of America, ***143*(3),** 1919-1919.

Masapollo, M., Polka, L., & Ménard, L. (2017). A universal bias in adult vowel perception–By ear or by eye. Cognition, **166**, 358-370.

Masapollo, M., Polka, L., Molnar, M., & Ménard, L. (2017). Directional asymmetries reveal a universal bias in adult vowel perception. The Journal of the Acoustical Society of America, **141(4)**, 2857-2869.

Masapollo, M., Zhao, T. C., Franklin, L., & Morgan, J. L. (2019). Asymmetric discrimination of nonspeech tonal analogues of vowels. Journal of Experimental Psychology: Human Perception and Performance, **45(2)**, 285.

Peng, G., Zhang, C., Zheng, H. Y., Minett, J. W., & Wang, W. S. Y. (2012). The effect of intertalker variations on acoustic–perceptual mapping in Cantonese and Mandarin tone systems. Journal of Speech, Language, and Hearing Research, **55**, 579–595

Politzer-Ahles, S., Schluter, K., Wu, K., & Almeida, D. (2016). Asymmetries in the perception of Mandarin tones: Evidence from mismatch negativity. Journal of Experimental Psychology: Human Perception and Performance, **42(10)**, 1547.

Polka, L., & Bohn, O. S. (1996). A cross-language comparison of vowel perception in English-learning and German-learning infants. The Journal of the Acoustical Society of America, **100(1)**, 577-592.

Polka, L., & Bohn, O. S. (2003). Asymmetries in vowel perception. Speech Communication, **41(1)**, 221-231.

Polka, L., & Bohn, O. S. (2011). Natural Referent Vowel (NRV) framework: An emerging view of early phonetic development. Journal of Phonetics, **39(4)**, 467-478.

Polka, L., & Werker, J. F. (1994). Developmental changes in perception of nonnative vowel contrasts. Journal of Experimental Psychology: Human perception and performance, **20(2)**, 421.

Polka, L., Ruan, Y. F., & Masapollo, M. (2018). Understanding vowel perception biases—It's time to take a meta-analytic approach. Manuscript submitted for publication.

R Core Team (2012). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Tsao, Feng-Ming. The effect of acoustical similarity on lexical-tone perception of one-year-old Mandarin-learning infants. *中華心理學刊* 50, no. 2 (2008): 111-124.

Werker, J. F., & Tees, R. C. (1984). Phonemic and phonetic factors in adult cross-language speech perception. The Journal of the Acoustical Society of America, **75(6)**, 1866-1878.

Zhao, T. C., Masapollo, M., Polka, L., Ménard, L., & Kuhl, P. K. (2019). Effects of formant proximity and stimulus prototypicality on the neural discrimination of vowels: Evidence from the auditory frequency-following response. Brain and language, **194**, 77-83.