10th International Conference on Speech Prosody 2020
25-28 May 2020, Tokyo, Japan

Attention Distribution and Integration of Non-native Segments and Tones by Early Multilingual Speakers

Yi Liu, Jinghong Ning

The Hong Kong Polytechnic University
yi.liu@polyu.edu.hk, chion.ning@polyu.edu.hk

Abstract

It is well acknowledged that tone language speakers are attentive to both segmental and tonal dimensions when processing tonal speech, while less-experienced second language (L2) learners whose first language (L1) has no lexical tones are unlikely to distribute their attention to tonal information. However, it is unclear whether ignorance of the tonal dimension is also prevalent among early and highly experienced multilingual speakers. In the current study, native Cantonese speakers as well as early multilingual speakers whose L1 is Urdu (non-tonal) participated in a series of attention distribution and integration tasks in Cantonese. The multilinguals acquired Urdu first and started learning Cantonese (tonal) and English (non-tonal) simultaneously at an early age. The results showed that although the multilingual speakers could phonologically process Cantonese tones, they were unable to distribute and integrate their attention in a perfectly Cantonese-like way. The multilingual speakers retained their L1 attentional strategy to the processing of L2 tones in highly cognitively-demanding tasks. The results suggested that a language overlap and well-developed L1 and L2 systems both existed for multilingual speakers.

Index Terms: Attention distribution and integration, Cantonese tones, early multilingual speakers, selective perceptual routine.

1. Introduction

In non-tone languages, pitch information is mainly used to express sentential information and lexical accent, while tone language speakers use pitch information as the primary cue to convey lexical and sentential meaning [1]. When perceiving a speech sound, tonal speakers pay simultaneous attention to both segmental and tonal dimensions, in order to acquire valuable linguistic information from the speech. Non-tonal speakers might find it hard to allocate their attention to the tonal dimension because of the absence of sensitivity towards tones when processing a non-native language [2] [1]. In order to become automatic speakers of a tonal language, learners, whose L1 has no lexical tones, not only have to establish phonological categories for non-native tones, but also have to learn how to redistribute their limited cognitive resources to both segmental and tonal dimensions [1].

In the Automatic Selective Perception (ASP) model [3], the feature of speech perception is a purposeful and information-seeking activity. Through this activity, adult listeners can use a highly over-learned and highly automatic program to detect the most reliable acoustic parameters, which specify their first language systems. Such an over-learned and highly automatic program is called a selective perceptual routine (SPR) in the ASP model. For example, tonal speakers can develop an internalized SPR for the distribution of cognitive attention to different linguistic dimensions of speech. With the assistance of such SPR, native speakers are able to automatically extract sufficient information from various linguistic conditions. The ASP model originally predicted perception performance for native speakers and late adult language learners, suggesting that only native speakers are able to utilize SPR in an automatic way, while late adult learners would never establish an perfectly native-like SPR [3]. Recently, more and more neurocognitive and behavior studies on ASP indicate that a more automatic SPR would be available for non-native learners when more L2 proficiency is accumulated [4] [1].

[1] conducted an ABX task where the target syllable in the disyllabic non-words varied along tonal or segmental dimensions. Native Mandarin speakers, non-tonal Dutch speakers (who never learned Mandarin), and Dutch-speaking learners of Mandarin participated in her experiment. These learners had beginner or advanced Mandarin proficiency levels. The main results in [1] demonstrated that Dutch speakers were not able to process Mandarin tones in a phonological way. It was in line with the results of [2], revealing that Mandarin speakers tended to integrate both segment and tone in the processing of stimuli, while native Dutch speakers dominantly relied on the segmental dimension. Moreover, there existed a trend that advanced learners felt it more difficult to ignore tonal information than beginners when processing Mandarin syllables. This showed a developmental trajectory for learners in terms of their automatic SPR.

According to the research on ASP model, L2 SPR is expected to be gradually acquired by non-tonal speakers as they gain L2 experience [4] [1]. However, it remains unclear how sensitive non-tonal speakers are towards tonal information when they have already developed into highly experienced language learners, or even bilingual/multilingual speakers. Individuals can become bilingual/multilingual speakers at any age. Early bilingualism refers to a child who has partially acquired L1 and started L2s in their early childhood. For instance, it is likely for a child to become a bilingual speaker during pre-adolescence, if, at an early age, they move to an environment where the dominant language is not their L1 [5].

Views vary tremendously on whether early multilinguals/bilinguals develop L1 and L2 systems as separate categories or as one merged category. The "two-category" view [6] [7] states that early bilinguals' L1 and L2 are independently developed and stored separately, without necessarily interfering

 10.21437/SpeechProsody.2020-66

with each other. Prior studies in favor of this point of view collected evidence from the comparable performance of native and bilingual speakers. From the "two-category" perspective, the native language experience seems not to play a role in the perception of L2 speech for early bilinguals. In contrast, the "one-category" view [8] [9] advocates the coexistence of two-way stored L1 and L2 linguistic categories, and a common phonetic space (L1/L2 overlap) of the languages for bilingual/multilingual speakers. The robust finding that early bilinguals are not able to discriminate L2 speech in a perfectly native-like manner strongly supports the "one-category" viewpoint. In light of the "one-category" view, the influence of L1 experience will never be ignored for early bilinguals due to the language overlap, and bilinguals should be regarded as unique, configured populations instead of L1 and L2 co-native listeners [8].

In the application of ASP model, previous studies showed how native speakers, late adult learners [3] and experienced learners [4] [1] develop an L2 SPR. With the help of such SPR, native speakers can automatically distribute their limited attention resources to different phonetic dimensions in order to achieve perceptual success. However, it is not clear what early and highly experienced multilingual speakers will be like in developing such a higher-order L2 SPR. The current study aims to fill this gap. Three research questions are being addressed: (1) Can early Urdu-Cantonese speakers successfully discriminate Cantonese lexical tones in a native-like phonological mode? (2) Are they able to distribute their attention to segments and tones and to develop a more integral processing of these two dimensions or can they develop an automatic and Cantonese-like attentional SPR in processing for Cantonese speech? (3) The above two questions can also be interpreted as "Do early and highly experienced bilinguals/multilinguals store two languages in a totally separate way or not? To investigate the above questions, native Cantonese speakers and Urdu-Cantonese speakers were invited to attend a series of cognitively demanding ABX tasks, which were adopted by [1] when studying Mandarin. Urdu speakers were immigrants who arrived in Hong Kong at an early age. In the current case, their L1 is Urdu, and both Cantonese and English were regarded as their L2s.

2. Method

2.1. Participants

Eighteen multilingual speakers whose L1 is Urdu (nine female, nine male) and 20 native Cantonese speakers (10 female, 10 male) were selected as subjects according to their language profiles. Both the native Cantonese speakers (mean age = 11.3 years, SD = 1.8) and multilingual participants (mean age = 11.4 years, SD = 2.2) were year-one students who were studying in local Hong Kong secondary schools. All participants were right-handed and did not suffer from any hearing difficulties or physical disabilities. The multilingual participants were exposed to Urdu during the first one or two years of their lives, and immigrated to Hong Kong at an early age of 2.3 years (SD = 1.1). They started learning Cantonese at the age of 4.5 years (SD = 0.99), and began learning English at the age of 4.6 years (SD = 1.2). The multilingual students only used Urdu at home, and utilized both English and Cantonese as instruction languages at school. At the time of the experiment, all the multilingual participants had been living in Hong Kong continuously for 9.1 years (SD = 1.9). The native Cantonese speakers had learned English and Mandarin before, but they seldom used both languages.

2.2. Stimuli

Two pairs of CVCV nonce words /kasu/-/tafu/ and /biso/-/diso/ (in the current study, /b/ and /d/ referred to voiceless unaspirated stops) were selected with Cantonese Tone 2 (high-rising) or Tone 4 (low-falling) on the initial syllables, and the second syllable for each disyllabic nonce word was neutralized as Cantonese high-level tone. Similar to the stimuli in [1], in each non-word pair, the vowels remained unchanged; only the consonant changed in place of articulation and articulatory manner. Three native Cantonese speakers (two female, one male) were invited to record the disyllables with Boom microphone, in the audio booth at Hong Kong Polytechnic University. The nonce words were provided in Roman script and Cantonese tone marks. These native speakers, who were previously trained in the pronunciation and the Cantonese scripts of the nonce words, were asked to produce the disyllables in a natural speaking speed. The recordings were saved as 16-bit .wav files with a sampling rate of 44,100 Hz.

ABX test was adopted as experimental method, with four conditions, namely, forced-segment condition, forced-tone condition, segment-and-tone condition, and segment-or-tone condition [1]. In the forced-segment and forced-tone conditions, participants were forced to classify the target X along the segmental or tonal dimension, respectively. In the segment and tone condition, participants were required to identify target X that matched with either A or B in both segment and tonal dimensions. In the segment or tone condition, in which target X matched with either the segmental or tonal dimension, participants choose freely between the two. The distribution of attention could be observed from the results.

The nonce word pairs were further arranged for each ABX condition according to the following criteria: (1) The target X contains the same tone or (and) segment as A or B; (2) The stimuli order can be ABX or BAX; (3) The speakers were shuffled in each ABX combination instead of it being produced by the same speaker, in order to increase phonetic variability and listeners' memory load. Thus, for each trial condition, we got 16 ABX stimuli (two non-word pairs × two Cantonese tones × two AB orders × two matches with A or B). The arrangement of stimuli is illustrated in Table 2, which only showed one AB order.

Table 1: *Examples of stimuli in ABX task. The number between syllables represents a Cantonese tone mark.*

Condition	A	B	X
Forced-segment	ka2su1 bi4so1	ta2fu1 di4fo1	ka4su1 di2fo1
Forced-tone	ka2su1 bi4so1	ka4su1 bi2so1	ta2fu1 di4fo1
Segment-and-tone	ka2su1 bi4so1	ta4fu1 di2fo1	ka2su1 bi4so1
Segment-or-tone	ka2su1 bi4so1	ta4fu1 di2fo1	ka4su1 di4fo1

2.3. Procedure

The experiment was carried out through E-prime. Each participant was seated in front of a computer (Lenovo ThinkCentre desktop, i5 core, USB interface: 3.0) in a quiet classroom in the local secondary schools. Instructions were given only in Cantonese and all participants were properly briefed about the task procedure. They were asked to listen to three syllables, A, B, and X and indicate if X sounded more similar to A or B by a mouse click on a "1" or "2" shown on their computer screen. In each trial there was a 600ms interval between standard A and standard B, and X appeared after a 900ms pause [2]. The inter-stimuli interval between two trials was 2,500ms, and if the subject failed to respond within the interval, the stimulus would be shown again later on, so there would be no missing data in the experiment. Four-minute familiarization trials in the segment-and-tone condition were given to the listeners before a formal experiment started. In the formal experiment, ABX stimuli were repeated five times, generating 320 ABX trials (16 ABX stimuli × four conditions × five repetitions) in total for each subject.

3. Results

For the first three trials (forced-segment, forced-tone, segment-and-tone), response rate represents accuracy rate of responses, and for the segment-or-tone trial, response rate stands for segment-based percentage, because there is no "correct" answer in this part. We calculated response rate according to the percentages of "correct" or "segment" responses out of the five responses for each participant and each ABX stimulus. For statistical analysis, both raw data of response rate were natural-logarithmically transformed to achieve better normalcy. On the base of sample size and the distribution of data, the linear mixed-effect model (LMM) was performed in R using the lme4 package [10] in terms of the test field of individual response rate. All p values were corrected with Bonferroni adjustment for multi-comparisons. The LMM model incorporated fixed effects of subject group, experimental trial, as well as their interaction. For random effects, by-subject and by-item intercepts were included

Figure 1 depicts the average response rates for native and non-native children in terms of four ABX conditions, and the statistical results of LMM for response rate is presented in Table 2. According to LMM, there was a significant main effect of subject group (β = 378.67, SE = 54.57, t = 109.81, p < .001), and an experiment trial also exhibited a main effect with β= 123.45, SE = 23.45, t = 45.68, p = .0023. Moreover, LMM revealed an interaction between subject group and experiment trial (β= 245.45, SE = 43.23, t = 95.34, p < .001), suggesting that native and non-native listeners performed differently across the four ABX trials.

The post-hoc Tukey test suggested that both the native (M = 93%, SD = 7) and multilingual (M = 89%, SD = 11) speakers got the highest response rate (all pairwise p < .01) in the segment-and-tone trial, demonstrating the fact that the co-occurrence of tonal and segmental dimensions would certainly facilitate the listeners the most. No statistical difference (p = .76) existed between native and multilingual speakers, which indicated that when both segmental and tonal information was provided, the multilingual speakers could process tones as phonologically as native speakers did.

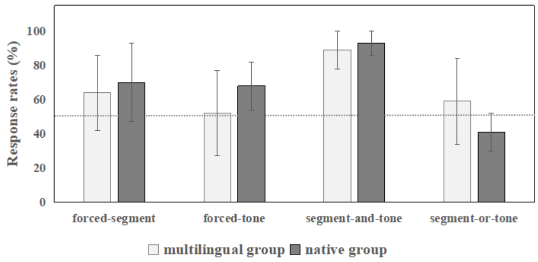

Figure 1: *Mean values of response rates in percentages for the multilingual and native groups. Standard deviations were shown in error bars.*

Table 2: *The results of LMM for response rates.*

Fixed effects	β	SE	t	*p*
Interception	456.455	63.234	123.566	<.001
Subject group	378.678	54.567	109.866	<.001
Experiment trial	123.456	23.456	45.687	.00237
Subject group × Experiment trial	245.453	43.234	95.345	<.001
Random effects	variance	SD		
1|subject	23.4	12.3		
1|item	19.7	10.4		

With regard to the results of the segment-and-tone condition and that of the first two conditions, the native group got far higher accuracy when both tonal and segmental dimensions were matched than in the forced-segment (p = .0014) or forced-tone (p = .0054) trials. Similarly, the multilingual listeners could identify the sounds more accurately (tone-and-segment vs. forced-segment: p < .004; tone-and-segment vs. forced-tone: p < .0012) in the tone-and-segment trial than in the other two trials. In addition, the native group could perceive more accurately than their multilingual counterparts in both the forced-segment condition (p = .033) and the forced-tone condition (p = .003). It thus illustrated that either suprasegmental or segmental mismatch would largely hinder the perception for all listeners, but it was likely that the multilinguals were affected more.

With regard to the results in the tone-or-segment and segment-and-tone conditions, the native group (mean = 41%, SD = 11) showed a lower (p = .013) response rate than the non-native group (M = 59%, SD = 25). Additionally, both the native (p = .021) and multilingual (p = .029) groups obtained much higher response rates in the segment-and-tone condition than they did in the segment-or-tone condition. It thus demonstrated that native listeners redistributed their attention along both segmental and tonal dimensions (both around chance level), while the multilinguals classified the stimuli mainly along the segmental dimensions.

4. Discussion and Conclusion

The current study investigated early multilingual speakers' distribution and integration of attention towards tonal and segmental dimensions when processing Cantonese tones.

The segment-and-tone condition provided both tonal and segmental information to listeners, thus both native and multilingual speakers found it easy to make correct responses to the stimuli. No statistical difference was detected between the results of the native and multilingual speakers in response rates. This is in line with the prior research on bilingualism/multilingualism [6] [8] showing that multilingual speakers are able to process Cantonese tones phonologically as native speakers do when conducting a comparatively less cognitively demanding perception task in the segment-and-tone condition. It also supports the statements in the ASP model, indicating that when an easy perception task is conducted, it is possible for learners to obtain a performance comparable with that of native speakers, because they have enough time and attentional resources to extract sufficient information to make an accurate decision.

Comparing the results of the segment-and-tone condition and that of the segment-or-tone condition, we can see how the multilingual speakers distributed their limited attention resources across different phonetic aspects. In the segment-and-tone condition, native and multilingual listeners could make accurate responses since both tonal and segmental information were matched in the stimuli. However, in the segment-or-tone condition, listeners were only allowed to make responses along one accurate phonetic dimension. The results showed that around 50% of native Cantonese speakers classified the stimuli along the tonal dimension, while the other half depended on the segmental dimension. More time was spent by native speakers in the segment-or-tone condition than in the segment-and-tone condition, which illustrates that more cognitive effort was needed when they were provided with insufficient phonetic information. This is in line with the results of [2] and [1], demonstrating that tonal language speakers (i.e., Mandarin and Cantonese) distribute their attention across both tonal and segmental dimensions when processing native syllables. In comparison, most of (59%) the multilingual speakers classified Cantonese syllables by relying on the segmental dimension, suggesting that multilingual speakers were more attentive (to a statistically significant degree) to segmental information than to tonal information. The different performance of multilinguals from that of the native speakers supports the statements in [8], illustrating that multilinguals should be regarded differently and uniquely from native speakers.

With regard to how the listeners integrated the tonal and segmental dimensions, we compared the results of the first three conditions. The current results showed that native Cantonese speakers obviously had higher accuracy in the segment-and-tone condition than in the other two conditions. This suggests that the mismatch in either phonetic dimension would prevent native Cantonese speakers from making an accurate response, which is in accordance with the results of the tonal speakers' results in [2] and [1]. Previous research [1] illustrated that less-experienced language learners can easily ignore the tonal dimension, which does not convey lexical meanings in their L1, and the mismatch in tonal dimension would not significantly influence their processing of non-native syllables. The performance of multilingual speakers in the current study showed a totally different picture when compared with that of the less-experienced learners in [1], who totally separated the two dimensions, and it is also different from native Cantonese speakers, who tightly integrated the two dimensions. The multilingual speakers slowed down in either forced-tone or forced-segment conditions, and obtained significantly lower accuracy in both, which is different from less-experienced language learners of a tonal language as shown in [1]. In addition, the multilingual speakers showed much lower response rates in most of the conditions compared with native Cantonese speakers. Hence, it is hypothesized that multilingual speakers tended to establish a weaker link between the two dimensions compared with native Cantonese speakers. The multilinguals did not neglect tonal information and did not develop a perfectly Cantonese-like attentional integrality.

In the light of the ASP model, segments are overused in non-native listeners' L1 (Urdu), and their L2 (English). In addition, multilinguals may use a similar attentional strategy when processing a tonal language. As suggested by language learning research, active learners will never stop learning and approaching an L2-like performance [7]. Having started learning Cantonese at an early age, having been immersed in a target language-speaking area, Urdu immigrants had all the positive conditions to acquire a well-developed Cantonese system. This is why it was probable that they exhibited a Cantonese-like performance in the segment-and-tone condition, and exhibited phonological processing in forced-tone and forced-segment conditions with accuracy rates above chance level. Nevertheless, the multilinguals' well-developed L2 system is still deviant from that of native Cantonese speakers, since they are still impeded (but much less influenced than less-experienced learners) by their segment-dependent strategy from their pre-existing languages. Therefore, multilinguals were not more attentive to tonal information than native Cantonese speakers were, and they could not integrate tonal and segmental dimensions as automatically as native speakers did. In this sense, the multilinguals in the current case can be treated as a different and unique group from the native one. The current study is thus similarly to the original ASP [3];) and its extension on experienced language learners [4] [1], suggesting that non-native learners cannot develop an perfectly native-like SPR, but their SPR processing can become increasingly automatic as they accumulate L2 experience.

The current study also supports the "one-category" view in multilingual studies [8] [9], because clear distinctions were found between native and multilingual speakers in tone processing in terms of how phonologically they can perceive tone categories, and how automatically they could use attentional SPR. As claimed by [8], multilinguals may have established well developed L1 and L2 systems, but there is still an overlap between L1 and L2, and we still can see that the segment-dependent strategy they used in Urdu and English systems may hinder the attention distribution and integration of Cantonese segmental and tonal dimensions.

5. Acknowledgements

This work is funded by the Language Fund under Research and Development Projects 2018-19 of the Standing Committee on Language Education and Research (SCOLAR), Hong Kong SAR.

6. References

[1] T. Zou, Y. Y. Chen & J. Caspers. "The developmental trajectories of attention distribution and segment-tone integration in Dutch learners of Mandarin tones". *Bilingualism: Language and Cognition*, vol. 1, pp. 1017-1029, 2017.

[2] B. Braun & E. K. Johnson. "Question or tone 2? How language experience and linguistic function guide pitch processing". *Journal of Phonetics*, vol. 39, pp. 585-594, 2011.

[3] W. Strange, & V. L. Shafer. Speech perception in second language learners: The re-education of selective perception. In J. G. Hansen Edwards, & M.L. Zampini (Eds.), Phonology and second language acquisition. Philadelphia: John Benjamins. 2008, pp.153 – 191.

[4] K. Steinhauer. "Event-related potentials (ERPs) in second language research: a brief introduction to the technique, a selected review, and an invitation to reconsider critical periods in L2". *Applied Linguistics*, vol. 1, pp. 393-417, 2014.

[5] W. Butler. "Bilingualism/Multilingualism and Second-Language Acquisition". In Tejk. Bhatia & William C. Ritchie (Eds.). The Handbook of Bilingualism and Multilingualism: Second Edition. Blackwell Publishing, Hoboken, New Jersey, 2012, pp. 109-136.

[6] K. Amengual. "The perception of language-specific phonetic categories does not guarantee accurate phonological representations in the lexicon of early bilinguals". *Applied Psycholinguistics*, vol. 37, pp. 1221-1251, 2016.

[7] J. E. Flege. "Interactions between the native and second-language phonetic systems". In: P. Burmeister, T. Piske, & A. Rohde (Eds.), An integrated view of language development: Papers in honor of Henning Wode. Trier, Germany: Wissenschaftlicher Verlag, 2002, pp. 217 – 243.

[8] M. Antoniou, M. D. Tyler, & C. T. Best. "Two ways to listen: do l2-dominant bilinguals perceive stop voicing according to language mode?". *Journal of Phonetics*, vol. 40. pp. 582-594, 2012.

[9] C. Pallier, A. Colomé & N. Sebastián-Gallés. "The influence of native language phonology on lexical access: Exemplar-based versus abstract lexical entries". *Psychological Science*, vol. 12, pp. 445 – 449, 2001.

[10] R. H. Baayen, D. J. Davidson & D. M. Bates. "Mixed-effects modeling with crossed random effects for subjects and items". *Journal of Memory and Language.* Vol. 59, pp. 390 – 412, 2008.