Cognitive basis of individual differences in speech perception, production and representations: The role of domain general attentional switching

Jinghua Ou¹ · Sam-Po Law¹

Published online: 31 January 2017
© The Psychonomic Society, Inc. 2017

Abstract This study investigated whether individual differences in cognitive functions, attentional abilities in particular, were associated with individual differences in the quality of phonological representations, resulting in variability in speech perception and production. To do so, we took advantage of a tone merging phenomenon in Cantonese, and identified three groups of typically developed speakers who could differentiate the two rising tones (high and low rising) in both perception and production [+Per+Pro], only in perception [+Per–Pro], or in neither modalities [–Per–Pro]. Perception and production were reflected, respectively, by discrimination sensitivity d' and acoustic measures of pitch offset and rise time differences. Components of event-related potential (ERP)—the mismatch negativity (MMN) and the ERPs to amplitude rise time—were taken to reflect the representations of the acoustic cues of tones. Components of attention and working memory in the auditory and visual modalities were assessed with published test batteries. The results show that individual differences in both perception and production are linked to how listeners encode and represent the acoustic cues (pitch contour and rise time) as reflected by ERPs. The present study has advanced our knowledge from previous work by integrating measures of perception, production, attention, and those reflecting quality of representation, to offer a comprehensive account for the underlying cognitive factors of individual differences in speech processing. Particularly, it is proposed that domain-general attentional switching affects the quality of

perceptual representations of the acoustic cues, giving rise to individual differences in perception and production.

Keywords Individual differences · Speech perception · Speech production · Perceptual representations · Attentional switching/shifting

Introduction

One ubiquitous theme in speech sciences concerns the relation between characteristics of the acoustic signal and our phonetic percepts. A growing body of work has focused on the individual differences in the perception of the acoustic cues that define sound contrasts, and shown that this relation may vary as a function of factors such as language background (Escudero & Boersma, 2004; Schertz, Cho, Lotto, & Warner, 2015), cue weighting strategy (Beddor, McGowan, Boland, Coetzee, & Brasher, 2013; Kong & Edwards, 2011), as well as cognitive abilities (e.g. attention: Janse & Adank, 2012; working memory: Francis & Nusbaum, 2009; cognitive processing style: Stewart & Ota, 2008; Yu, 2010). Focusing on native sound contrasts, the purpose of the present investigation was to examine the cognitive factors that give rise to individual differences in perception and production of speech sounds.

When perceiving speech, listeners need to first decode the auditory signal and transform this time-varying input into accurate phonemic representation (Cutler & Clifton, 1999). The way in which listeners use speech cues to activate phonemes may to some degree be modulated by higher level cognitive processes. For example, the operation of attention in speech perception is proposed to optimize the signal-to-noise ratio in the encoding of acoustic cues (Gordon, Eberhardt, & Rueckl, 1993). Mattys and Wiget (2011) found decreased accuracy in discriminating voice onset time (VOT) differences in the

✉ Sam-Po Law
splaw@hku.hk

¹ Division of Speech and Hearing Sciences, the University of Hong Kong, Hong Kong, SAR, People's Republic of China

presence of a cognitive load due to dual task compared to a single-task condition, suggesting that detailed phonetic analysis would be compromised if attentional resources are reduced. It appears that attention improves contrast sensitivity and signal segmentation by engaging the attentional focus onto relevant information to intensify the signal (e.g. Francis & Nusbaum, 2002; Goldstone, 1998; Nosofsky, 1986).

Attentional ability has also been argued to be necessary for deriving stable representations. Developmental studies have provided support for the idea that attention is important for fine-tuning perceptual representations (Conboy, Sommerville, & Kuhl, 2008; Jusczyk, 2002). Theories concerning the mechanism of attention in pre-lexical processing during speech perception have been developed from research on developmental disorders, i.e., specific language impairment, and developmental dyslexia. For instance, the “Sluggish Attentional Shifting” (SAS) hypothesis proposes that the flexibility of the attentional system influences the quality of phonological representations (Hari & Renvall, 2001; Lallier et al., 2010; Ruffino et al., 2010). Specifically, when individuals with sluggish attentional shifting process sequences of rapidly presented stimuli, their automatic attentional system cannot disengage fast enough from one item to encode the next, leading to confusion of consecutive sounds, thereby hindering adequate processing of salient acoustic cues, and resulting in unstable and variant phonemic representations in the mental lexicon. This hypothesis predicts individuals with efficient attentional shifting have sharper categorical perceptual representations; in contrast, those with less efficient attentional shifting have more graded perceptual representations.

Although the foregoing work has provided evidence for the role of attention in speech processing, many of these studies focused on perception, and relatively little is known about the production aspect. Lev-Ari and Peperkamp (2014) have recently examined both perception and production, and tested whether individual differences in attentional inhibitory skill can lead to individual differences in perception and production of voicing in words that have a voicing neighbor among native French speakers. Results revealed that individuals with lower inhibition skill responded faster to synthesized stimuli of a shorter, intermediate pre-voicing in a lexical decision task, rather than prototypical stimuli of longer voicing values. The authors argued that these individuals habitually experience greater activation of the competing neighbor, and consequently the competing voicing feature would have an intermediate voicing value in their representation of the word, which would therefore be activated more efficiently by tokens with a similar voicing value. However, the inhibitory skill did not predict the individual differences in pre-voicing duration in production. This study has offered evidence for the role of cognitive abilities in influencing the efficiency of speech perception. The findings also imply that cognitive skills may affect the properties of representation. As perceptual representation of a

speech sound is considered by some to constitute a crucial link between perception and production (e.g., Guenther, 1995; Lotto, Hickok, & Holt, 2009), it is reasonable to expect that the quality of representation mediates the degree of distinctiveness in perception and production. The current investigation assessed the possibility of attentional ability as a potential source of individual differences in perceptual representations [reflected by event-related potentials (ERPs)], further leading to variability in the perception and production of speech sound contrasts. Specifically, we focused on the lexical tone contrasts in Cantonese spoken in Hong Kong (HKC).

Lexical tones refer to syllabic-level pitch patterns that are used to contrast meaning in words of identical segmental compositions (Bauer & Benedict, 1997). There are six contrastive tones for non-stopped syllables in HKC, namely T1 (high level tone), T2 (high rising tone), T3 (mid-level tone), T4 (low falling/extra low level tone), T5 (low rising tone) and T6 (low level tone). The pitch contours of these six tones are shown in Fig. 1. The complex tonal system is reported to be undergoing a sound change—tone merging (Bauer, Cheung, & Cheung, 2003; Fung & Wong, 2010, 2011; Law, Fung, & Kung, 2013; Lee, Chan, Lam, van Hasselt, & Tong, 2015; Mok, Zuo, & Wong, 2013). Previous studies have shown that, among the six tones, T1 is the most resistant to confusion, while three tone contrasts are merging: T2 vs. T5, T3 vs. T6 and T4 vs. T6. Specifically, T2/T5 have undergone an extensive merging in the community, and a significant number of native speakers can no longer distinguish the contrast between the two tones in perception and/or production. Most speakers can discriminate T3/T6 in perception, while this contrast is not maintained in production. The opposite pattern has been observed for T4/T6 where distinction in production is preserved, whereas perception is not. This protracted phenomenon is realized in the form of variable patterns of perception and production among native speakers. To investigate the variability in native tone perception and production with respect to individual differences in cognitive functions, Ou, Law, and Fung (2015) recruited three groups of participants differing in perception and/or production, and assessed their ability of

Fig. 1 Time-normalized fundamental frequency (F0) patterns of the six Cantonese tones on the syllable/fu/produced by a male native speaker

attention and working memory in both auditory and visual modalities using a battery of objective tools. The three participant groups represented, respectively, the pattern of good perception and good production of all Cantonese tones [+Per+Pro], that of good perception of all tones but poor production of specifically the distinction of the two rising tones (T2 and T5) [+Per–Pro], and that of good production of all tones but poor perception of specifically the distinction of the two low level tones (T4 and T6) [–Per+Pro]. The findings revealed that the three participant groups differed in their ability of attention switching. The group difference was largely driven by the significantly lower score in [–Per+Pro]. Furthermore, discrimination latencies were shown to be longer among individuals with non-distinctive production [+Per–Pro] or perception [–Per+Pro] compared to the control group [+Per+Pro], and were negatively associated with individuals' cognitive abilities of working memory and attentional switching that were independent of modality. The authors proposed that individuals with lower attentional switching ability and working memory capacity may have less distinctive representations, which may take longer to be accessed or recognized by incoming acoustic information during speech perception. The overall findings have demonstrated a link between individual differences in native tone perception and production, and domain-general cognitive functions, in particular attentional switching. However, it is arguable that response latency may reflect a multitude of cognitive, motor, and perceptual processes involved in performing the discrimination task, and thus may not necessarily correspond to one's discrimination ability. Moreover, tone production in the study was based on auditory transcription only.

To address the above issues, Ou and Law (2016) subsequently conducted a more detailed examination of the differences in perception and production of T2/T5 between the [+Per+Pro] and [+Per–Pro] groups from Ou et al. (2015), using ERPs obtained in a passive oddball paradigm for perception and acoustic measurements for production. The ERP measures included (1) an early component reflecting sensitivity to the rise time of sound amplitude envelope (Carpenter & Shahin, 2013; Thomson, Goswami, & Baldeweg, 2009), which has been reported to contribute to tone recognition (Fu & Zeng, 2000; Whalen & Xu, 1992; Zhou, 2012); (2) the mismatch negativity (MMN, see Näätänen, Paavilainen, Rinne, & Alho, 2007 for a review) indicating auditory discrimination sensitivity to the primary acoustic dimension of tone contrasts, i.e., pitch contour/height (Gandour, 1983; Khouw & Ciocca, 2007); and (3) P3a indexing involuntary attentional switching (Polich, 2003). For production, the differentiations between T2 and T5 in terms of pitch offset and amplitude rise time were measured acoustically. The results demonstrated that the amplitudes and peak latencies of MMN and P3a to T2/T5 did not differ between the two participant groups, as expected given their comparable accuracies of tone

perception. On the other hand, the two groups differed significantly in the brain responses to the subtle acoustic cue of rise time. These perceptual differences were further shown to be associated with acoustic differences in producing the two rising tones with respect to pitch offset and amplitude rise time differences, indicating an association of individual differences in speech production and perceptual sensitivity. Nonetheless, Ou and Law did not include cognitive measures, and thus the role of cognitive abilities in speech processing and phonological representation (reflected by ERP measures) remained to be explored.

While Ou et al. (2015) and Ou and Law (2016) have demonstrated the role of cognitive functions in individual differences in native speech perception and production with behavioral measures, and the close and subtle relationship between tone production and neural correlates of perception (i.e., pitch contour and rise time), more convincing evidence would come from associations among cognitive functions, neural correlates of perceptual representations of the acoustic cues, and behavioral measures reflecting the distinctiveness of tone perception and production. In addition, it is important to point out that in Ou et al. (2015), the three patterns of tone perception and production were based on performance on two tonal contrasts, i.e., T2/T5 and T4/T6. A more rigorous approach would be to compare variations in perception and production of the same tone contrast. In this light, the present study extends the works of Ou et al. (2015) and Ou and Law (2016) by incorporating a sensitivity measure of perception, acoustic measures of production, ERP measures reflecting tone representation as well as cognitive measures, and significantly adding a participant group exhibiting poor perception and production of T2/T5. The inclusion of this group has provided the critical end of the perception-production continuum of the T2/T5 contrast, and thus a more complete picture of individual differences in native tone perception and production among typically developed individuals. Note that the pattern of good production of all tones but poor perception of specifically the T2/T5 distinction was not represented, as this pattern is rarely observed, based on a large scale survey of the HKC tone merging (Fung & Wong, 2010, 2011). Combining the behavioral and neural measures of the two previous studies, the present study employed discrimination accuracies, responses time (RT) and a discrimination sensitivity index for assessing tone perception performance, the MMN, P3a and an early ERP component to stimulus rise time as neural correlates of T2/T5 perceptual processing, acoustic measurements of pitch offset and amplitude rise time for evaluating tone production, and performance on tests of attention and working memory in the visual and auditory modalities for reflecting cognitive abilities. Data from the new participant group were compared with those representing the patterns of good perception and production as well as good perception and poor production in Ou and Law (2016).

The rich set of observations has enabled us to examine the hypothesis that individual patterns of tone perception and production are linked to how listeners encode and represent speech cues (i.e. rise time and pitch contour), and, more importantly, that differences in speech representation are related to individual differences in cognitive abilities, attentional ability in particular. Given the findings in Ou et al. (2015), it is expected that, compared with [+Per+Pro] and [+Per–Pro], [–Per–Pro] would show longer RT in tone perception as well as lower scores in cognitive measures tapping into attention switching. Additionally, based on the results in Ou and Law (2016), we predict that [–Per–Pro] would differ from the [+Per+Pro] and [+Per–Pro] groups in terms of a smaller distinction of T2/T5 pitch offset and rise time in tone production, and MMN and/or P3a amplitude and latency to the pitch contrast of T2/T5. For ERP responses to rise time, which were found to correlate with difference in production, [–Per–Pro] is predicted to differ from [+Per+Pro], but not necessarily [+Per–Pro]. The perceptual differences in MMN and/or P3a, and ERP to rise time are expected to associate with differences in production across groups. Lastly, if cognitive abilities affect the quality of perceptual representations, we expect to see a relationship between performance on specific cognitive tasks and behavioral and neural measures of tone perception and production.

Methods

Ethics statement

All participants gave informed consent in compliance with an experimental protocol approved by the University of Hong Kong Research Ethics Committee for Non-Clinical Faculties (Ref. # EA261113), and were paid for their participation in the study.

Participants

Three groups of native Cantonese speakers were included in the present study. Participants were recruited on a rolling basis upon initial recruitment, until our recruitment targets for each group were reached (approximately $n = 20$ for each group). At the time of writing, a total of 168 participants were recruited. These speakers were selected into one of the three groups according to their ability to perceive and produce the tone contrast T2/T5 in HKC. The tasks were a tone discrimination task and a tone production task (details were described in Ou et al., 2015; Ou & Law, 2016). Participants who could distinctively perceive and produce all six Cantonese tones were classified into the [+Per+Pro] group, and those who could perceive all tones but failed to produce T2 and T5 distinctively were selected into the [+Per–Pro] group (selection criteria and

participant characteristics for the [+Per+Pro] and [+Per–Pro] groups were reported in Ou et al., 2015; Ou & Law, 2016). After the initial screening, the selected participants were invited back for two phases: a behavioral session for a series of cognitive tasks, and an EEG session. The two phases progressed in sequence, and, ideally, all selected participants would complete both testing sessions. However, some of the participants in Ou et al. (2015) did not take part in the EEG session (they either refused to come back due to logistic or motivational issues, or could not be contacted), and the sample were reduced to 13 [+Per+Pro] and 14 [+Per–Pro] participants. To ensure adequate statistical power, seven additional [+Per+Pro] and five additional [+Per–Pro] participants from the same pool described above were selected. Thus, the final sample in Ou and Law (2016) comprised 20 [+Per+Pro] (female = 8) and 19 [+Per–Pro] (female = 11) participants. These participants did not differ from those studied by Ou et al. (2015) in any relevant aspect (specifically, the selection criteria, i.e., perception and production of T2/T5). For the [+Per+Pro] group, the accuracy rate for discriminating between T2 and T5 in Ou et al. (2015) was .99, while that for the additional tasks was .98, and the production accuracies for T2 and T5 were .99 for both the original and additional cohorts. Regarding the [+Per–Pro] group, the perception accuracies for the original and additional participants were both .98, and the production accuracies were .63 for the original and .58 for the additional ones. None of the differences approached significance (all P s > .24).

A third participant group [–Per–Pro] who could not distinguish T2 and T5 in both perception and production was included in the present study to compare with the [+Per+Pro] and [+Per–Pro], all of whom had complete cognitive and EEG data. The [–Per–Pro] group comprised 19 native speakers of Cantonese (female = 10), all born and raised in Hong Kong. No speaker reported a history of hearing abnormalities. They were all right-handers according to the Edinburgh Handedness Inventory (Oldfield, 1971). As this study aimed to compare data of this participant group with reported data of the [+Per+Pro] and [+Per–Pro] groups (Ou & Law, 2016), observations of all three groups are presented here for ease of comparison. Table 1 shows the characteristics of the three participant groups in terms of their performance on tone perception, tone production, and musical background. The three groups were matched in age [$F(2,57) = .528, p = .592$], years of formal education [$F(2,57) = .312, P = .733$] and musical background in onset and duration of training [onset: $F(2,57) = 2.24, P = .116$; duration: $F(2,57) = 1.27, P = .289$].

Compared with [+Per+Pro] and [+Per–Pro], the [–Per–Pro] group did not perceive and produce T2 and T5 distinctively. As shown in Table 1, the accuracy score and sensitivity index d' of T2–T5 discrimination were significantly lower in the [–Per–Pro] group than the other two groups, based on results of ANOVAs [$F(2, 57) = 315.813, p < .001, \eta^2_{\text{partial}} = .92$] for

Table 1 Background information on participants in [+Per+Pro], [+Per-Pro] and [-Per-Pro] groups

	[+Per+Pro]		[+Per-Pro]		[-Per-Pro]	
	(N = 20)		(N = 19)		(N = 19)	
	M	SD	M	SD	M	SD
Age	22.00	0.59	21.24	0.84	21.20	2.84
Years of education	17.10	0.67	17.00	0.12	17.50	1.10
Tone perception						
Distinguishing T2/T5						
Accuracy ^a	0.99	0.02	0.98	0.01	0.59	0.09
Sensitivity (d') ^a	6.59	0.48	6.37	0.81	2.74	0.64
All other tones						
Accuracy	0.99	0.01	0.97	0.01	0.95	0.05
Sensitivity (d')	6.53	0.23	6.24	0.35	6.40	0.26
Tone production ^b						
T2–T5 offset difference ^a	2.95	0.38	1.11	0.59	0.49	0.95
T1 mean pitch height	4.62	0.52	4.71	0.43	4.65	0.50
T3 mean pitch height	3.06	0.60	3.20	0.51	3.15	0.46
T4 mean pitch height	1.09	0.32	1.10	0.47	1.13	0.24
T6 mean pitch height	2.11	0.75	2.25	0.67	2.10	0.50
Musical background						
Onset	6.75	6.37	5.62	4.70	7.55	4.25
Duration	5.35	5.78	5.00	4.82	4.57	4.04

M mean, SD standard deviation

^a Group means differed statistically at $P < .05$

^b Pitch in normalized T-value

accuracy, and [$F(2, 57) = 206.243, P < .001, \eta^2_{\text{partial}} = .88$] for d'. Bonferroni post-hoc comparisons revealed these effects were respectively attributed to the significantly lower accuracy and discrimination sensitivity in [-Per-Pro] than the other two groups (all $P < .001$). No significant difference was observed between [+Per+Pro] and [+Per-Pro] (both $P = 1.00$).

Following Ou and Law (2016), the production data were first transcribed by a native Cantonese speaker; this was followed by an acoustic analysis. The acoustic properties of T2 and T5 produced by the [+Per+Pro] were taken as a reference to verify the status of the [-Per-Pro] participants, and the pitch offset was taken as an acoustic index for distinctive production of the two rising tones. For a participant to be regarded as poor in distinguishing T2 and T5 in production, his or her pitch offset difference (T2 pitch offset minus T5 pitch offset) had to be at least 2.5 SDs below the mean of the pitch offset difference of the [+Per+Pro] group. An omnibus ANOVA revealed a significant main effect of group, [$F(2, 57) = 69.297, P < .001, \eta^2_{\text{partial}} = .71$]. Pairwise contrasts showed a gradient of the pitch offset difference across groups ([+Per+Pro] > [+Per-Pro] > [-Per-Pro]). [+Per+Pro] demonstrated a larger difference than the other two groups (both $P <$

.001), and [+Per-Pro]'s difference was greater than that of [-Per-Pro] ($P = .019$). The acoustic index of the level tones (T1, T3, T4 and T6)—mean pitch heights—were comparable among the three groups [all $F(2, 57) < 1.4, P > .255$]. Figure 2 demonstrates the distribution of perceptual sensitivity d' and acoustic distinctions of T2–T5 pitch offset for each participant group.

Procedures

Cognitive measures

Following Ou et al. (2015), the current study obtained a comprehensive set of cognitive measures tapping into attention, working memory and inhibitory control, in order to evaluate these higher cognitive processes as potential sources of individual differences in tone perception, production and representation. As reviewed above, attention has been shown to play a role in signal optimization during early processing stage of speech perception (e.g. Gordon et al., 1993; Mattys & Wiget, 2011). However, these studies did not measure individuals' attentional ability, and it remains unclear which components of attention may influence the acoustic analysis during perception. In this study, we employed a published test battery—Test of Everyday Attention (Robertson, Ward, Ridgeway, & Nimmo-Smith, 1994)—to cover four basic components of attention: selective attention, divided attention, sustained attention and attentional switching (see Cohen, 1993, p. 311). Individuals' working memory capacity was also assessed in auditory and visual modalities respectively by the auditory digit span backward test and the subscales of visual processing speed—symbol search, coding and cancellation in the Wechsler Adult Intelligence Scale-IV (WAIS-IV; Wechsler, 2010). In addition, an auditory stream segregation

Fig. 2 Scatterplot of production distinction of T2–T5 Pitch offset vs. perceptual sensitivity d' to T2 and T5 for the three participant groups

task, and the Flanker task were used to measure individuals' attentional shifting and inhibitory control respectively, given that individual differences in these cognitive skills have been found to be related to the quality of phonological representation (e.g., Lallier et al., 2010; Lev-Ari & Peperkamp, 2014). Further details regarding the stimuli and procedures for the cognitive measures can be found in Ou et al. (2015).

EEG experiment

A passive oddball task was used to examine the neural processing and representation of different acoustic cues associated with tone contrasts. Again, details of the stimuli, procedures, EEG data recording and processing have been described in Ou and Law (2016). Briefly, the experiment consisted of four oddball conditions of different Standard/Deviant pairs, including T2/T5 and T5/T2 as two experimental conditions, and two control conditions, i.e., T1/T2 and T1/T5. The four oddball conditions were presented in separate blocks, each of which consisted of 535 trials. The standard stimuli were presented in 85% of the trials, and each deviant occurring on 15% (or 80) of the trials. During the task, participants were told to watch a silent movie while completely ignoring the auditory stimuli. The entire experiment lasted about 100 min.

Data analyses

Perception and production of tones Apart from the measures used for participant selection including accuracy score, discrimination sensitivity d' and T2–T5 pitch offset differences, RTs were obtained from the tone discrimination task, and T2–T5 rise time differences (T2 rise time minus T5 rise time) were extracted from tone production. To compare performance of [–Per–Pro] with those of [+Per+Pro] and [+Per–Pro], one-way ANOVAs were performed on the overall RTs and the T2–T5 rise time difference. The RTs of tone perception with incorrect responses and outlier responses greater than ± 3 SD of the group mean were removed. As the [–Per–Pro] participants failed to reliably perceive the distinction between T2 and T5, RTs to these two tones were not valid observations of discrimination. Thus, trials involving T2 and T5 were excluded in the calculation of the individual overall RT for all participants. Post hoc comparisons were conducted if a significant group effect was found.

Cognitive measures A one-way multivariate analysis of variance (MANOVA) was used to test for significant differences among the three groups of participants in terms of their performance on the cognitive tasks. Given the significant correlations between attentional switching ability and visual working memory with RTs in tone discrimination reported in Ou et al. (2015), we focused on cognitive measures related to

these two skills in the present study. Three composite scores were computed by summing the z-scores of the test of everyday attention (TEA) visual attention switching subtests (TEA visual), the auditory attention switching subtests (TEA auditory), and the three working memory subtests in WAIS (visual WM), respectively. Four dependent variables were included in the multivariate model: TEA visual, TEA auditory, visual WM and stream segregation threshold. Univariate ANOVAs were conducted following findings of significant main effects or interactions. To control for the occurrence of Type I errors, the significance level of each ANOVA was adjusted to 0.0125, which was obtained based on 0.05 divided by 4 (the number of ANOVAs run).

ERP analyses As in Ou and Law (2016), non-parametric perceptual analyses were first carried out for each block to identify significant ERP components reflecting responses to contrasts in pitch height/contour to different tone pairs and rise time between T2 and T5. Subsequently, conventional analyses were performed to examine whether the three groups differed in the magnitude and latency of the ERP components. Based on previous studies, data from the Fz and FCz electrodes were selected for statistical analyses, where the strongest mismatch effects were usually observed (e.g. Chandrasekaran, Krishnan, & Gandour, 2007; Tong et al., 2014; Tsang, Jia, Huang, & Chen, 2011).

MMN and P3a to pitch height/contour The latency of MMN was defined as the most negative peak during the time window of 100–250 ms post-divergence point of the respective condition, and the latency of P3a as the most positive peak following the individual MMN peak. The average amplitude was computed of the 100–250 ms time window for MMN, and of the 300–500 ms time window for P3a respectively, then averaged across the two selected electrodes. To verify the presence of components at the two selected electrodes, paired-samples t -tests were performed between the difference wave and the dummy wave for each component of interest. Furthermore, to compare the neural responses of [–Per–Pro] with those of [+Per+Pro] and [+Per–Pro], separate one-way ANOVAs were conducted for mean amplitude and peak latency for each component of interest in each of the four conditions. The Greenhouse-Geisser adjustment to the degrees of freedom was used when necessary to correct for the violations of sphericity associated with repeated measures.

ERPs to rise time The grand averaged ERPs for all occurrences (both standard and deviant) of T2 and T5 were computed to elucidate any differences in brain response to rise time between T2 and T5. The mean amplitudes at the Fz and FCz electrodes were measured in the time windows of 50–150 ms post vowel onset where rise time differed between the two stimuli, and were subject to a two-way mixed-design

ANOVA with condition (T2, T5) as a within-subject factor and group ([+Per+Pro], [+Per-Pro] and [−Per-Pro]) as a between-subject factor. Similarly, Greenhouse-Geisser adjustment was applied to correct for any violations of sphericity associated with repeated measures.

Analyses among measures of tone perception, production and cognitive functions To examine the relationship between perception (reflected in neural and behavioral measures) and production, Pearson product-moment correlation coefficients were first computed among all participants. Thus perception measures, including the overall RTs in the tone discrimination task, and neural responses to rise time and pitch height/contour of T2 and T5, i.e., responses to rise time of T2 and T5, peak latencies and mean amplitudes of MMNs and/or P3a to T2/T5 and T5/T2, and production of T2–T5 rise time difference were entered into the correlation analysis. To reduce the number of correlations, only those measures with significant differences between groups were analyzed. The key correlations emerged were further subject to a partial correlation to examine the relationship between perception and production while controlling for effects of group status. As covariates in partial correlation must be continuous measures, a composite of accuracy scores from the tone discrimination and production tasks (which were used to define group status) was computed and entered as the covariate. The value of the composite for [+Per+Pro] was thus high, while that for [−Per-Pro] was low.

The cognitive measures were used to predict (1) behavioral performance in tone perception—overall response latency in tone discrimination and the sensitivity index d' to T2–T5 discrimination; (2) acoustic parameters in tone production—the T2–T5 pitch offset difference and the T2–T5 rise time difference in tone production; and (3) brain responses to tone contrasts—responses to rise time of T2 and T5, peak latencies and mean amplitudes of MMNs and/or P3a to T1/T2, T1/T5, T2/T5 and T5/T2, using a canonical correlation analysis. With multiple dependent variables, the canonical correlation, as a multivariate technique, can limit the probability of committing Type I error by allowing for simultaneous comparisons among the variables rather than requiring separate statistical tests be conducted (Tabachnick & Fidell, 1996; Thompson, 2000). If the multivariate test was significant, follow-up univariate tests would be conducted to examine the contribution of each predictor variable to each of the predicted variables. To reduce the number of predictor and predicted variables, only those cognitive measures found to differ significantly across participant groups in the aforementioned MANOVA test would serve as predictors, and only those behavioral and neural measures with significant differences across groups would be taken as predicted variables.

Results

Results of the [+Per+Pro] and [+Per-Pro] groups, as mentioned before, have been reported in Ou and Law (2016). They are included here to provide comparisons with the [−Per-Pro] group, and to allow us to examine the relationships among individual differences in cognitive abilities, tone perception and production.

Perception and production of tones

For the overall RT, ANOVA revealed a significant main effect of group [$F(2, 57) = 8.435, P = .001, \eta^2_{\text{partial}} = .23$] with a significantly shorter response latency in [+Per+Pro] ($M = 1046.18$ ms, $SD = 80.19$) than [+Per-Pro] ($M = 1191.96$ ms, $SD = 155.03, P = .018$) and [−Per-Pro] ($M = 1216.83$ ms, $SD = 204.27, P = .001$). [+Per-Pro] and [−Per-Pro] did not differ from each other ($P = .852$).

For tone production, in addition to the significant group differences in T2 and T5 F0 offset difference ($P < .001$), a main effect of group on T2–T5 rise time difference was also observed, [$F(2, 57) = 44.260, P < .001, \eta^2_{\text{partial}} = .63$]. Results showed that T2–T5 rise time difference was significantly larger in [+Per+Pro] ($M = 29.6$ ms, $SD = 18.7$) than the other two groups (both P s $< .001$). No significant difference was observed between [+Per-Pro] ($M = 1.6$ ms, $SD = 4.71$) and [−Per-Pro] ($M = 2.02$ ms, $SD = 3.5, P = 1.00$).

MANOVA of cognitive measures

Results of the one-way MANOVA revealed a significant multivariate main effect for group, Wilks' Lamda equals 0.595 [$F(8, 104) = 3.86, P = .001, \eta^2_{\text{partial}} = .23$]. Given the significance of the overall test, the univariate main effects were examined. The means and standard deviations of the four cognitive measures and results of the univariate ANOVAs are given in Table 2. At the adjusted level of 0.0125 for P -value, a significant main effect of group was found for the TEA visual composite [$F(2, 57) = 10.47, P < .0005, \eta^2_{\text{partial}} = .28$] and the auditory stream segregation threshold [$F(2, 57) = 5.117, P = .009, \eta^2_{\text{partial}} = .16$]. TEA visual composite reflects individuals' efficiency of attention switching in the visual modality, and Bonferroni post-hoc comparisons revealed better performance of [+Per+Pro] ($P = .004$) and [+Per-Pro] ($P < .0005$) than [−Per-Pro] in switching their attention between task demands. The stream segregation task evaluates individuals' efficiency of attention shifting in the auditory modality, and post-hoc analysis showed that [+Per+Pro] exhibited smaller auditory thresholds than [−Per-Pro] ($P = .007$), whereas no significant differences between [+Per+Pro] and [+Per-Pro] ($P = .235$), or [+Per-Pro] and [−Per-Pro] ($P = .524$) were observed. None of the other measures reliably differed among groups ($P > .186$).

ERP results

Cluster-based permutation tests

The results of the cluster-level permutation test revealed several significant clusters in different conditions in the three participant groups (see Fig. 3). Clusters with appropriate scalp distributions in the interval of 100 to 250 ms post divergence point were interpreted as MMN, and those in the interval of 300–500 ms post divergence point as P3a components. For both the T2/T5 and T5/T2 conditions, significant clusters were also observed in the interval of 50–150 ms post vowel onset.

MMN In the [+Per+Pro] group, the nonparametric statistics revealed significantly greater negativities for the difference waves relative to dummy waves in the conditions of T1/T5, T1/T2, and T2/T5. These effects were distributed mainly in the fronto-central area, with significant time windows typical of MMNs between 100 and 166 ms (post-divergence point unless specified otherwise) for T1/T2 ($P < .001$), between 100 and 166 ms for T1/T5 ($P = .006$), and between 150 and 200 ms for T2/T5 ($P = .015$). A significant negative cluster in the time window between 150 and 238 ms was observed in the T5/T2 condition ($P = .044$) but with a centro-parietal distribution, which was hence not considered as an MMN. In the [+Per–Pro] group, MMNs were also elicited in the T1/T2 (110–166 ms, $P = .006$), T1/T5 (104–154 ms, $P = .025$) and T2/T5 (150–200 ms, $P = .015$) conditions, but no significant negative cluster was observed in the T5/T2 condition. In the [–Per–Pro] group, MMNs were elicited in the T1/T2 (100–160 ms, $P = .009$) and T1/T5 (94–156 ms, $P = .012$) conditions, but no significant negative cluster in the time window of

Fig. 3 Grand-averaged difference waves and dummy waves in each condition at the Fz and FCz electrodes for the three participant groups. Significant clusters are marked in grey. The time window for each component was defined at the pitch divergence point in each condition

MMN was observed in the two experimental conditions (Fig. 3).

P3a The contrast between T1 and T2 elicited a significant positive cluster immediately following the MMN in the time window of 300–400 ms for [+Per+Pro] ($P = .025$) and 342–404 ms for [+Per–Pro] ($P = .039$), which can be considered P3a, but not for [–Per–Pro]. No significant positive clusters were found in the other conditions.

Early components In the two experimental conditions, the contrast between T2 and T5 elicited significant early clusters during the time period from vowel onset to the pitch divergence point (100–250 ms) where the amplitude rise time differed between the two stimuli. All three groups exhibited an early positive-going cluster in the T2/T5 condition in the time window between 62 and 154 ms for [+Per+Pro] ($P = .015$), between 64 and 144 ms for [+Per–Pro] ($P = .025$) and between 56 and 162 ms for [–Per–Pro] ($P = .001$). For the T5/T2 condition, an early negative-going component was observed in [+Per+Pro] (36–176 ms, $P = .039$) and [–Per–Pro] (38–114 ms, $P = .005$), but not in [+Per–Pro].

To summarize, compared with [+Per+Pro] and [+Per–Pro], [–Per–Pro] did not show a P3a in the control condition T1/T2; moreover, no MMN was elicited to the T2/T5 contrast in [–Per–Pro].

Table 2 Performance on subtests of test of everyday attention (TEA) visual, TEA auditory, visual working memory (WM), and auditory stream segregation of the three participant groups

Measures	[+Per+Pro]		[+Per–Pro]		[–Per–Pro]		P^a
	M	SD	M	SD	M	SD	
TEA visual							.0005
Visual elevator accuracy (VE1)	10.75	2.77	12.16	2.22	9.95	2.74	
Visual elevator timing (VE2)	14.20	2.80	13.95	2.07	11.47	1.95	
TEA auditory							.875
Elevator counting with distractor (ECD)	12.20	1.67	11.94	1.90	12.31	1.70	
Elevator counting with reversal (ECR)	15.00	4.05	14.95	4.36	13.74	3.46	
Visual WM							.059
Symbol search (SS)	15.15	2.72	15.84	2.43	14.47	2.58	
Coding (CD)	17.15	2.51	16.68	2.05	15.68	2.38	
Cancellation (CA)	12.15	3.15	12.47	2.50	11.95	3.04	
Auditory stream segregation (AS)	58.28	33.87	78.46	32.76	94.11	38.45	.009

^a P values for univariate ANOVAs

Fig. 4 Averaged event-related potentials (ERPs) to all occurrences (both standards and deviants) of T2 and T5 at Fz and FCz electrodes for the three participant groups

observed for T2 rise time ($M_{[+Per+Pro]} = 1.43$, $SD = 2.13$, $M_{[+Per-Pro]} = .60$, $SD = 1.62$, $M_{[-Per-Pro]} = 1.07$, $SD = 1.31$; $P = .335$).

Correlation analysis

Relationships between perception and production

Five measures of tone perception and production that were significantly different among the three participant groups, including the overall RT, mean amplitude of the ERP to T5 rise time, mean amplitude and peak latency of the MMN to T2/T5, and production of T2–T5 rise time difference, were entered into the correlation analysis. For simplicity, we focused on the results of partial correlation here (for readers who are interested in the results of the Pearson product-moment correlation, please refer to Appendix 1). Upon controlling for group status, partial correlations showed that discrimination RT was not related to production of T2–T5 rise time difference ($r = -.20$, $P = .132$) or the neural response to T2/T5 (MMN amplitude: $r = .15$, $P = .278$; MMN latency: $r = .13$, $P = .323$); however, correlations between production and the brain responses remained significant (ERPs to T5 rise time: $r = .28$, $P = .038$; T2/T5 MMN latency: $r = -.27$, $P = .040$, see Fig. 5).

The role of cognitive functions in tone perception and production

Cognitive measures of TEA visual composite and auditory stream segregation threshold reached significance in MANOVA, thus the two sub-scores in the TEA visual composite—the accuracy score in the Visual Elevator task (VE1), the timing score in the Visual Elevator task (VE2)—as well as the auditory stream threshold (AS), were taken as predictors in the canonical correlation analysis. Pearson product-moment correlations between the three cognitive measures and the different measures of perception and production were considered prior to the canonical analysis.

Given the significant Pearson product-moment correlations (refer to Appendix 2 for detailed results), a canonical correlation analysis was then conducted using the three cognitive measures—VE1, VE2 and AS predictors of the four perception and production variables—discrimination sensitivity d' , T2–T5 pitch offset difference, T2–T5 rise time difference and T2/T5 MMN latency, to evaluate the multivariate relationship between the two variable sets. Results of the multivariate test revealed that the overall proportion of the variance accounted for in the four dependent variables by the predictor variables was significant, Wilks' Lambda = .517, $F(12, 135) = 3.188$, P

Fig. 5 Scatter plots with regression lines of ERPs to T5 rise time and T2T5 MMN latency as a function of production distinction of rise time between T2 and T5 for the three participant groups

$< .0005$, $\eta^2 = .483$, indicating that the full model explained about 48% of the variance shared between the two variable sets. Given that the multivariate test reached significance level, follow-up univariate regressions were conducted to determine the unique relationships between each of the predictor and predicted variables. As shown in Table 4, the three cognitive variables accounted for significant variance in all four dependent variables. Specifically, VE2 and VE1 significantly predicted the discrimination sensitivity d' , accounting for 33.7% [$F(3, 54) = 10.693$, $P < .0005$] of the variance. The VE1 and VE2 can be taken to index attentional switching in the visual modality. The higher the scores on these two measures, the more sensitive a participant is to detect the difference between T2 and T5. With respect to the production measures, AS was confirmed as the significant predictor, accounting for 15.9% and 13.7% of the variance, respectively, for the pitch offset difference [$F(3, 54) = 4.589$, $P = .006$], and the rise time difference [$F(3, 54) = 4.026$, $P = .012$]. The AS task assesses attentional shifting in the auditory modality, with smaller threshold in auditory segregation associated with higher production distinction between of T2 and T5. Lastly, all three cognitive measures, VE2, VE1 and AS were significant predictors of the latency of T2/T5 MMN accounting for 32.1% of the variance [$F(3, 54) = 8.504$, $P < .0005$]. As mentioned before, these cognitive measures can be taken to indicate the ability of attentional switching/shifting, and higher cognitive performance was associated with faster neural responses to the pitch contrast between T2 and T5. Figure 6 provides scatterplots that represent the relationship between each of the predicted variables and their significant predictor(s).

Table 4 Canonical correlation analysis

Variable	B	SE B	β	t	P
d'					
VE1	.114	.043	.286	2.653	.010
VE2	.192	.045	.460	4.223	.000
AS	-.005	.003	-.195	-1.794	.078
T2–T5 production pitch offset difference					
VE1	.073	.056	.158	1.299	.199
VE2	.116	.059	.238	1.943	.057
AS	-.010	.004	-.305	-2.489	.016
T2–T5 production rise time difference					
VE1	-.0006	.000	-.101	-.882	.415
VE2	.002	.000	.226	1.883	.074
AS	-.0001	.000	-.324	-2.605	.012
T2T5 MMN latency					
VE1	-6.392	3.087	-.232	-3.671	.043
VE2	-11.946	3.254	-.416	-2.071	.001
AS	.460	.224	.232	2.048	.045

Discussion

By examining behavioral and neural measures of tone perception and acoustic measures of tone production, as well as cognitive performances across attention and working memory tasks, the present investigation compared results from participants who cannot distinguish the two Cantonese rising tones in perception and production, [–Per–Pro], with those of [+Per+Pro] and [+Per–Pro] reported previously (Ou & Law, 2016). The [–Per–Pro] group has provided the critical end of the perception-production continuum, and thus formed a more complete picture of the relationship between native lexical tone processing and cognitive functions among typically developed individuals. Relative to [+Per+Pro] and [+Per–Pro], the non-distinctive perception and production of T2–T5 of the [–Per–Pro] participants were confirmed by their lower discrimination sensitivity in tone perception, and a smaller distinction of pitch offset and rise time in tone production; additionally, the lack of sensitivity to the pitch contour contrast between T2 and T5 on the part of [–Per–Pro] was verified by the ERP measures in amplitude and peak latency in the MMN time window, as well as the neural responses to T5 rise time. Moreover, [–Per–Pro] differed from the other groups in the ability of attentional switching/shifting in both auditory and visual modalities. This difference was further shown to be associated with the differences in behavioral discrimination sensitivity, pitch offset and rise time in tone production, as well as the latency of neural responses to T2/T5. Overall, based on the current findings, we propose that speech perception and production are contingent upon the degree of distinctiveness of perceptual representations (as reflected by ERPs), the quality of which is affected by the individual differences in modality-independent attentional switching/shifting.

The overall findings of behavioral tone perception and production are consistent with our expectations that the [–Per–Pro] group was poorer in each of the measures than one or both of the other two groups. Particularly, in tone perception, [–Per–Pro] showed longer discrimination latency of tone contrasts (excluding trials of T2 and T5) than that of [+Per+Pro], reflecting lower efficiency in discriminating between tones in general. The non-distinctive production in [–Per–Pro] has been demonstrated by a smaller distinction of pitch offset than those of [+Per+Pro] and [+Per–Pro]. Interestingly, [+Per–Pro]’s difference was greater still than that of [–Per–Pro] although both groups were classified as having non-distinctive production, and thus a gradient of difference in pitch offset has been shown across groups ([+Per+Pro] > [+Per–Pro] > [–Per–Pro]). On the other hand, the differentiations of rise time in tone production were comparable for [–Per–Pro] and [+Per–Pro], and both showed a smaller distinction than that of [+Per+Pro].

The differences in perception among the three participant groups are revealed in more detail by their neural responses to

Fig. 6 Scatter plots with regression lines of the four predicted variables—discrimination d' prime, production of pitch offset,

production of rise time, and T2/T5 MMN latency—as a function of their respective significant cognitive predictor(s) of all participants

the two rising tones—MMN to pitch height/contour and ERPs to rise time. As expected, the [-Per-Pro] group exhibited reliably smaller and slower neural responses in the experimental condition of T2/T5, compared with [+Per+Pro] and [+Per-Pro], compatible with the pattern of behavioral measures. Specifically, the results of the cluster-based permutation test revealed statistically reliable clusters corresponding to the MMN to the T2/T5 contrast in both [+Per+Pro] and [+Per-Pro], whereas the T2/T5 contrast did not elicit a reliable MMN in [-Per-Pro]. These cluster-level permutation results were

consistent with the significant main effect of group in the ANOVA test, where lower MMN amplitudes to T2/T5 were observed in [-Per-Pro] than [+Per+Pro]. Apart from the mean MMN amplitudes, a main effect of group was also observed in the peak latency of MMN to T2/T5, which was largely driven by the slower responses in [-Per-Pro] than the other two groups. However, contrary to expectations, differential brain responses between [-Per-Pro] and the two groups with distinctive perception, [+Per+Pro] and [+Per-Pro], were also observed in the control condition of T1/T2, while comparable

behavioral performances in discriminating the two tones were obtained among the three groups. Although an MMN to T1/T2 was confirmed in [–Per–Pro] in the permutation test, these participants exhibited a smaller amplitude in Fz and FCz to T1/T2 than the other groups. Given that the size of MMN has been found to correlate with individual difference in auditory discrimination ability (e.g. Díaz, Baus, Escera, Costa, & Sebastián-Gallés, 2008; Jakoby, Goldstein, & Faust, 2011), the present findings suggest that auditory discrimination sensitivity, indexed by the MMN, was in general weaker in [–Per–Pro] compared with the other two groups, even to sound contrasts that they could distinguish with high accuracies behaviorally. Moreover, both the permutation and ANOVA tests confirmed the absence of P3a to T1/T2 in [–Per–Pro], whereas robust P3a components were observed in both [+Per+Pro] and [+Per–Pro]. This finding seems convergent with Law et al. (2013), in which individuals with non-distinctive perception of T4/T6 showed a tendency of weaker P3a in all conditions, as compared with controls. The elicitation of P3a depends on sufficient difference between the deviant and the standard stimuli; thus, the P3a is usually described as an index of involuntary switch of attention upon change detection (Berti, Roeber, & Schröger, 2004; Escera, Alho, Winkler, & Näätänen, 1998). In the present study, the lack of a P3a to T1/T2 in the [–Per–Pro] participants may be due to lower than optimal encoding and detection of tone contrasts in general on the one hand, and a less flexible attentional mechanism (as will be elaborated below) related to the generation of P3a on the other hand.

Regarding the early neural responses to the rise time of sound amplitude, the results conform to our expectation based on Ou and Law (2016), where stronger brain responses to T5 rise time were seen in individuals with distinctive T2 and T5 production than those without. In the present study, the ANOVA test of the averaged ERPs to T2 and T5 revealed a significant group difference in ERPs to T5, but not T2. Specifically, larger ERP responses to T5 rise time were observed among individuals with distinctive production [+Per+Pro] compared to those without, i.e., [–Per–Pro] and [+Per–Pro], whereas the latter two groups did not differ. On the whole, the results of behavioral and neural measures of tone perception and acoustic measures of tone production have revealed that not only pitch contour/height, but also rise time of sound amplitude envelope contribute to distinctive perception and production of rising tones in HKC (see Kong and Zeng (2006) for the role of both cues in Mandarin tone recognition).

Significant correlations were found among different measures of T2–T5 perception and production, even when group status was controlled for, reinforcing the close relationship between the two modalities of speech processing (see Fig. 5). Specifically, the correlation between production of T2–T5 rise time and ERPs to T5 rise time reached significance

($r = .280$), so did that between production of T2–T5 rise time and T2/T5 MMN latency at $r = -.270$. Thus, a link between production distinctiveness and perceptual sensitivity persists independent of group status. This observation contributes to the previous literature that has mainly focused on segmental contrasts and found that the distinctiveness of speakers' production was related to their discrimination of the contrasts (e.g., Newman, 2003; Perkell et al., 2004, 2008). Furthermore, the findings of Ou and Law (2016) that the production distinctiveness index correlated positively with the perceptual encoding of T5 rise time, as reflected by the ERP amplitude, were replicated in the present study with additional participants of a different perception and production pattern, in line with the proposal that sensitivity to rise time contributes to the formation of perceptual representations (Goswami et al., 2002; Thomson et al., 2009). Taken together, the results support the hypothesis that individual differences in speech perception and production are mediated by the quality of perceptual representations, as reflected by the strength of neural responses to different acoustic cues.

Our observations regarding cognitive performances of the participants with different patterns of perception and production of T2 and T5 further demonstrate that individual differences in speech processing may be determined by some higher level cognitive function, such as attention switching/shifting (Hari & Renvall, 2001; Lallier et al., 2010). Specifically, individuals' abilities to switch their attentional focus in the visual modality (measured with VE1 and VE2) can account for differences in behavioral discrimination sensitivity (d'). VE2 and VE1, together with the ability of attention shifting in the auditory domain (measured with AS), are associated with how fast an individual's auditory system detects the pitch differences between T2 and T5 (reflected by MMN latency). It is noted that among the three significant neural measures reflecting tone representations, i.e., ERPs to T5 rise time, MMN amplitude and MMN latency to T2/T5, only MMN latency was significantly related to the cognitive measures. This may have to do with the fact that the magnitude of brain responses, i.e., ERPs to T5 rise time and MMN amplitude lack a speed component, and thus have less in common with one's efficiency of allocating attentional resources. The current findings of contribution of attention to speech processing are compatible with our previous study where visual attention switching could predict tone discrimination efficiency (Ou et al., 2015), as well as research that also found a role of attention in modulating perceptual sensitivity to speech sounds (e.g., Astheimer & Sanders, 2009; Díaz et al., 2008; Jesse & Janse, 2012). The significance of attention switching/shifting in influencing the distinctiveness of speech representations has been hypothesized in the SAS hypothesis (Hari & Renvall, 2001; Lallier et al., 2010; Ruffino et al., 2010). While the SAS hypothesis originally set out to account for the poor phonological decoding skills among children with

developmental dyslexia, the present findings point to the possibility that this hypothesis could be extended to typically developed adults. Our results are compatible with the claim that individuals with efficient attentional shifting have sharper categorical perceptual representations; in contrast, those with less efficient attention shifting have more graded perceptual representations. This is reflected in the distinctive perception and production in [+Per+Pro] compared with non-distinctive perception and production in [−Per−Pro]. According to this account, attentional shifting exerts its effect on the temporal computation of the sensory input, and efficient acoustic processing and segmentation of the speech signal require rapid engagement of attention (Facoetti et al., 2010; Renvall & Hari, 2002). Attentional shifting can be considered a mechanism in which automatic attentional focus disengages from a stimulus to process a rapidly successive one, in order to capture every single stimulus of the acoustic signals. Sluggish attention shifting may thus result in impoverished sensory analysis, further leading to degraded auditory representation (Hari & Renvall, 2001).

Another possible role of attentional shifting in speech processing is its involvement in cue weighting to improve contrast sensitivity with discriminant (strong) cues being given priority (Gordon et al., 1993). A related account comes from the attention-to-dimension model (e.g., Francis & Nusbaum, 2002; Goldstone, 1994; Nosofsky, 1986), in which the operations of attending and ignoring represent shifts of attention to or away from acoustic dimensions. As the distribution of cues in the acoustic signal varies across different phonetic contexts—a cue that is highly predictive of a particular contrast in one context may be less useful in another—it is possible that listeners naturally shift their attention between cues while they shift between talkers, phonetic contexts, and speaking rate as part of the process of context normalization. Individuals with higher ability of attentional shifting may be better able to adaptively change listening strategy and switch attention to relevant acoustic cues, thus maintaining the distinctiveness between categories.

Although the present findings reveal a link between attentional switching and individual differences in speech perception and production, we recognize that different sets of cognitive tasks were involved in relation to speech perception in Ou et al. (2015) and the current study. In the former, group differences were observed for VE2 only, in contrast with VE1, VE2 and AS in the present investigation. The three cognitive measures were further demonstrated to predict differences in perception and production in the present study, whereas in Ou et al. (2015), working memory in the visual domain (as measured with the Coding and Cancellation subtests in WAIS-IV), and attentional switching in the auditory domain (indexed by the Elevator Counting with Reversal subtest in TEA), predicted differences in tone discrimination RT. We suggest that the different findings regarding group

differences in cognitive measures could be related to the two studies comparing individuals with different perception and production patterns, i.e. [+Per+Pro], [+Per−Pro] and [−Per+Pro] in Ou et al. (2015) vs. [+Per+Pro], [+Per−Pro] and [−Per−Pro] in this study. Moreover, Ou and colleagues examined two tone contrasts, the rising tones of T2/T5 and the level tones of T4/T6, while the present study, with a more rigorous design, focused on the T2/T5 contrast only.

In addition, the present study did not find a relationship between working memory and speech perception and representation. The null finding could possibly be due to the different approaches employed in data analyses in the two studies. In Ou et al. (2015), only one predicted variable of tone discrimination RT was used, and its relationship with cognitive measures was explored by correlations and multiple regressions in two steps, regardless of the significance of group difference in specific cognitive abilities. First, in order to reduce the number of predictor variables, three composite scores for TEA visual subsets, TEA auditory subsets and visual WM subsets were computed. Composite(s) that were significantly correlated with RT were entered into multiple regressions. Subtest scores comprising the composite(s) were put into a second multiple regression to delineate which cognitive component(s) contributed to RT. The present investigation, on the other hand, employed measures of T2-T5 perception (both behavioral and neural) and production that were significantly different among the three participant groups as predicted variables. Predictors were restricted to cognitive measures that differed among groups, i.e., TEA visual composite (including VE2 and VE1) and AS. The relationship between cognitive functions and measures of tone perception and production was examined with correlations followed by the canonical correlation analysis. It is worth mentioning that the current approach of analysis shows a lack of association between the TEA auditory composite and tone processing performance. It is plausible that tasks with a speeded component, such as the TEA visual switching task and the stream segregation task, are better able to capture individual differences in attentional switching, compared with the TEA auditory switching task, which is not timed.

We are also aware that the lack of contribution of inhibitory skills as measured by the Flanker task to speech perception apparently diverged from Lev-Ari and Peperkamp (2014). In that study, inhibitory control was shown to predict the ease with which participants recognize words (reflected by recognition time) in a lexical decision task. Hybrid representations as a result of accumulated co-activation were hypothesized. Spoken word recognition is proposed to involve a process whereby the perceptual input activates lexical representations, and the competition among these activated representations then determines the response (e.g., Allopenna, Magnuson, & Tanenhaus, 1998; McQueen, Norris, & Cutler, 1994). Most speech perception models (e.g., McClelland & Elman, 1986;

Norris, 1994) posit that inhibitory mechanisms are involved in boosting the activation of the word that closely matches the input, while at the same time suppressing the less probable candidates. Eventually selection occurs when the level of activation is sufficiently strong to support one of the alternatives. In the present study, the oddball paradigm in the EEG experiment, which can be considered a discrimination task where the deviant is compared against the standard, was used to assess the distinctiveness between speech representations. As selection among activated representations is not required in such a task, inhibitory control may have a less important role to play. We further suggest that different processes underlie the experimental tasks in the two studies, and cognitive skills can influence representations through different mechanisms. The inhibitory skill account in Lev-Ari and Peperkamp explained its relation with perceptual performance but not production, whereas attentional shifting in the auditory modality (as measured with AS) in this study predicted performance in both perception and production.

The present overall findings that attentional switching/shifting mediates speech processing in a domain-general nature are congruent with some previous studies where the relationship is modality-independent (e.g., George & Coch, 2011; Janse, 2012; Ou et al., 2015), yet inconsistent with others that found a modality-specific relationship (e.g., Moore, Ferguson, Halliday, & Riley, 2008; Kraus, Strait, & Parbery-Clark, 2012). Nonetheless, a growing number of neuroimaging studies has evidenced that the cortical circuitry engaged in the auditory switching of attention is similar to that engaged during visual attention orientation (Diaconescu, Alain, & McIntosh, 2011; Larson & Lee, 2013; Smith et al., 2010). This suggests that the attentional control network could be supramodal, engaging both visual and auditory modalities (Knudsen, 2007; Lee, Larson, Maddox, & Shinn-Cunningham, 2014). However, the underlying mechanism of this executive network, and the extent to which it operates unimodally depending on particular task of speech processing, remain unclear and are beyond the scope of the present investigation.

Given the specific linguistic context in which the present investigation was conducted, our findings can be said to contribute to the line of research focusing on sound change at the individual level. Previous studies have identified several aspects of systematic heterogeneity among individuals, e.g., speech perception (Beddor, 2009), speech production (Baker, Archangeli, & Mielke, 2011), the mapping between perception and production (Stevens & Reubold, 2014) and cognitive processing style (Yu, 2010), which have elucidated how sound change occurs and progresses within a particular community. For instance, Beddor (2009) has proposed that sound change can arise from idiosyncratic perception

grammar; specifically, individual variation in perceptual analysis of coarticulatory variation would result in variation in phonological representations. The present findings have advanced our knowledge from previous works by demonstrating that individual differences in the cognitive ability of attentional switching/shifting influence the variability in perception and production during the process of tone merging in HKC. The quality of tonal representation may vary from person to person as a function of individual differences in attentional switching/shifting. Individuals with better attention switching/shifting skills are proposed to have more distinctive tonal representations, whereas those with less than optimal attention skills tend to have less distinctive ones. Further, the differences in the distinctiveness of the representations may lead to variation in tone perception and production, with more precise representations rendering more distinctive perception and production; on the other hand, representations with larger variance may give rise to speech variants in a community. One may also ask if the relationship between attention switching/shifting and tone merging could generalize to other sound change phenomena, such as the neutralization of initials *n/l* and codas *n/ng* mergers in HKC. Future work is necessary to investigate whether individual differences in attention switching/shifting associate with perceptual sensitivity and production distinction of other phonetic features.

Besides the main findings, there is an interesting observation in the results that deserves further discussion. Results of the EEG experiment revealed an asymmetrical MMN pattern to the contrast between T2 and T5, that is, the MMN response was present in T2/T5 but not T5/T2 (as was also observed in Ou & Law, 2016). The asymmetric MMN pattern was reported in studies using vowels or consonants (e.g., Eulitz & Lahiri, 2004) as well as Mandarin tones (Politzer-Ahles, Schluter, Wu, & Almeida, 2016). Asymmetry has been hypothesized to occur as a result of phonological underspecification (e.g., Cornell, Lahiri, & Eulitz, 2011, 2013). According to this hypothesis, MMNs are larger when the standard corresponds to a sound that is believed to be fully specified in the phonology of a language than when it is underspecified. Hearing a sound as a deviant in an oddball paradigm activates all its features, but the underlying representation is accessed when a sound is the standard stimulus. According to the underspecification account, underspecified features (e.g., coronal place of articulation) are not stored in the abstract memory representations in the lexicon. For instance, when the standard is [t], the coronal articulation feature is underspecified in long-term memory representations, an incoming deviant sound like [k], which has a different articulation feature, will thus not strongly clash with the memory representation, resulting in reduced MMN. In the case of Cantonese rising tones, it is plausible that T5 (low rising) is

underspecified compared to T2 (high rising), as T2 is suggested to be more phonologically marked with a higher tonal register (Yip, 2002). However, future study is needed to investigate such a proposal.

In conclusion, the current findings have reinforced the role of attentional flexibility in speech perception observed in Ou et al. (2015), and the close relationship between perception and production (Ou & Law, 2016). More importantly, we have extended previous research by considering measures of both speech perception and production, and examining their relationships with cognitive functions, to offer a more comprehensive account for the role of cognitive abilities in speech processing. Although causality cannot be determined from corre-

lations, there is support for interpreting the direction in this relationship. Specifically, our results suggest that the ability of switching/shifting attention in a domain general fashion may affect the quality of perceptual representation, which is assumed to be critical for the distinctiveness in speech perception and production.

Acknowledgements This research was supported by a Small Project Fund at the University of Hong Kong (Project titled “Neural correlates and cognitive capability associated with individual variations in tone perception and production in Cantonese—an event-related potential (ERP) study”). We are grateful to the editor and two anonymous reviewers whose insightful suggestions have appreciably improved the manuscript.

Appendix 1

Table 5 Pearson product-moment correlations between tone perception and production across all participants

	Tone perception overall RT	T2–T5 production rise time difference	T5 rise time ERP amplitude	T2/T5 MMN amplitude
T2–T5 rise time production	–.425**			
T5 rise time ERP amplitude	–.165	.420**		
T2/T5 MMN amplitude	–.278*	–.221	–.043	
T2/T5 MMN latency	.304*	–.312*	–.285*	.211

* $P < .05$, ** $P < .01$

Appendix 2

Table 6 Pearson product-moment correlations among perception and production measures and three cognitive measures (VE1, VE2 and AS) across all participants

	Tone perception overall RT	T2–T5 discrimination sensitivity d'	T2–T5 production pitch offset difference	T2–T5 production rise time difference	T5 rise time ERP amplitude	T2/T5 MMN amplitude	T2/T5 MMN latency
VE1 ^a	–.047	.313*	.181	.078	.033	.111	–.259*
VE2 ^b	–.186	.499**	.288*	.269*	.055	.083	–.458**
AS ^c	.163	–.273*	–.346**	–.352*	–.176	–.206	.301*

* $P < .05$, ** $P < .01$

^a Correct responses in the Visual Elevator task

^b Time-per-switching in the Visual Elevator task

^c Auditory segregation thresholds in the Attentional Shifting task

References

- Alloppenna, P. D., Magnuson, J. S., & Tanenhaus, M. K. (1998). Tracking the time course of spoken word recognition using eye movements: Evidence for continuous mapping models. *Journal of Memory and Language*, 38(4), 419–439.
- Asstheimer, L. B., & Sanders, L. D. (2009). Listeners modulate temporally selective attention during natural speech processing. *Biological Psychology*, 80(1), 23–34.
- Baker, A., Archangeli, D., & Mielke, J. (2011). Variability in American English s-retraction suggests a solution to the actuation problem. *Language Variation and Change*, 23, 347–374.
- Bauer, R. S., & Benedict, P. K. (1997). *Modern Cantonese phonology*. Berlin: Mouton De Gruyter.
- Bauer, R., Cheung, K. H., & Cheung, P. M. (2003). Variation and merger of the rising tones in Hong Kong Cantonese. *Language Variation and Change*, 15(2), 211–225.
- Beddor, P. S. (2009). A coarticulatory path to sound change. *Language*, 85(4), 785–821.
- Beddor, P. S., McGowan, K. B., Boland, J. E., Coetzee, A. W., & Brasher, A. (2013). The time course of perception of coarticulation. *The Journal of the Acoustical Society of America*, 133(4), 2350–2366.
- Berti, S., Roeber, U., & Schröger, E. (2004). Bottom-up influences on working memory: Behavioral and electrophysiological distraction varies with distractor strength. *Experimental Psychology*, 51(4), 249–257.
- Carpenter, A. L., & Shahin, A. J. (2013). Development of the N1–P2 auditory evoked response to amplitude rise time and rate of formant transition of speech sounds. *Neuroscience Letters*, 544, 56–61.
- Chandrasekaran, B., Krishnan, A., & Gandour, J. T. (2007). Mismatch negativity to pitch contours is influenced by language experience. *Brain Research*, 1128, 148–156.
- Cohen, R. (1993). *The neuropsychology of attention*. New York: Plenum.
- Conboy, B. T., Sommerville, J. A., & Kuhl, P. K. (2008). Cognitive control factors in speech perception at 11 months. *Developmental Psychology*, 44(5), 1505.
- Cornell, S. A., Lahiri, A., & Eulitz, C. (2011). “What you encode is not necessarily what you store”: Evidence for sparse feature representations from mismatch negativity. *Brain Research*, 1394, 79–89.
- Cornell, S. A., Lahiri, A., & Eulitz, C. (2013). Inequality across consonantal contrasts in speech perception: Evidence from mismatch negativity. *Journal of Experimental Psychology: Human Perception and Performance*, 39(3), 757.
- Cutler, A., & Clifton, C. (1999). Comprehending spoken language: A blueprint of the listener. In C. M. Brown & P. Hagoort (Eds.), *The neurocognition of language* (pp. 123–166). Oxford: Oxford University Press.
- Diaconescu, A. O., Alain, C., & McIntosh, A. R. (2011). The co-occurrence of multisensory facilitation and cross-modal conflict in the human brain. *Journal of Neurophysiology*, 106(6), 2896–2909.
- Díaz, B., Baus, C., Escera, C., Costa, A., & Sebastián-Gallés, N. (2008). Brain potentials to native phoneme discrimination reveal the origin of individual differences in learning the sounds of a second language. *Proceedings of the National Academy of Sciences of the United States of America*, 105(42), 16083–16088.
- Escera, C., Alho, K., Winkler, I., & Näätänen, R. (1998). Neural mechanisms of involuntary attention to acoustic novelty and change. *Journal of Cognitive Neuroscience*, 10(5), 590–604.
- Escudero, P., & Boersma, P. (2004). Bridging the gap between L2 speech perception research and phonological theory. *Studies in Second Language Acquisition*, 26(04), 551–585.
- Eulitz, C., & Lahiri, A. (2004). Neurobiological evidence for abstract phonological representations in the mental lexicon during speech recognition. *Journal of Cognitive Neuroscience*, 16(4), 577–583.
- Facoetti, A., Trussardi, A. N., Ruffino, M., Lorusso, M. L., Cattaneo, C., Galli, R., ... & Zorzi, M. (2010). Multisensory spatial attention deficits are predictive of phonological decoding skills in developmental dyslexia. *Journal of Cognitive Neuroscience*, 22(5), 1011–1025.
- Francis, A. L., & Nusbaum, H. C. (2002). Selective attention and the acquisition of new phonetic categories. *Journal of Experimental Psychology: Human Perception and Performance*, 28(2), 349.
- Francis, A. L., & Nusbaum, H. C. (2009). Effects of intelligibility on working memory demand for speech perception. *Attention, Perception, & Psychophysics*, 71(6), 1360–1374.
- Fu, Q. J., & Zeng, F. G. (2000). Identification of temporal envelope cues in Chinese tone recognition. *Asia Pacific Journal of Speech, Language and Hearing*, 5(1), 45–57.
- Fung, R., & Wong, C. (2010). *Mergers and near-mergers in Hong Kong Cantonese tones*. Presented at Tone and Intonation 4, Stockholm, Sweden.
- Fung, R., & Wong, C. (2011). Acoustic analysis of the new rising tone in Hong Kong Cantonese. In *Proceedings of the 17th International Congress of Phonetic Sciences* (pp. 716–718).
- Gandour, J. (1983). Tone perception in far eastern languages. *Journal of Phonetics*, 11(2), 149–175.
- George, E. M., & Coch, D. (2011). Music training and working memory: An ERP study. *Neuropsychologia*, 49(5), 1083–1094.
- Goldstone, R. L. (1994). Influences of categorization on perceptual discrimination. *Journal of Experimental Psychology: General*, 123(2), 178.
- Goldstone, R. L. (1998). Perceptual learning. *Annual Review of Psychology*, 49(1), 585–612.
- Gordon, P. C., Eberhardt, J. L., & Rueckl, J. G. (1993). Attentional modulation of the phonetic significance of acoustic cues. *Cognitive Psychology*, 25(1), 1–42.
- Goswami, U., Thomson, J., Richardson, U., Stainthorpe, R., Hughes, D., Rosen, S., & Scott, S. K. (2002). Amplitude envelope onsets and developmental dyslexia: A new hypothesis. *Proceedings of the National Academy of Sciences*, 99(16), 10911–10916.
- Guenther, F. H. (1995). Speech sound acquisition, coarticulation, and rate effects in a neural network model of speech production. *Psychological Review*, 102(3), 594.
- Hari, R., & Renvall, H. (2001). Impaired processing of rapid stimulus sequences in dyslexia. *Trends in Cognitive Sciences*, 5(12), 525–532.
- Jakoby, H., Goldstein, A., & Faust, M. (2011). Electrophysiological correlates of speech perception mechanisms and individual differences in second language attainment. *Psychophysiology*, 48(11), 1517–1531.
- Janse, E. (2012). A non-auditory measure of interference predicts distraction by competing speech in older adults. *Aging, Neuropsychology, and Cognition*, 19(6), 741–758.
- Janse, E., & Adank, P. (2012). Predicting foreign-accent adaptation in older adults. *The Quarterly Journal of Experimental Psychology*, 65(8), 1563–1585.
- Jesse, A., & Janse, E. (2012). Audiovisual benefit for recognition of speech presented with single-talker noise in older listeners. *Language and Cognitive Processes*, 27(7–8), 1167–1191.
- Jusczyk, P. W. (2002). How infants adapt speech-processing capacities to native-language structure. *Current Directions in Psychological Science*, 11(1), 15–18.
- Khouw, E., & Ciocca, V. (2007). Perceptual correlates of Cantonese tones. *Journal of Phonetics*, 35(1), 104–117.
- Knudsen, E. I. (2007). Fundamental components of attention. *Annual Review of Neuroscience*, 30, 57–78.
- Kong, E. J., & Edwards, J. (2011). Individual differences in speech perception: Evidence from visual analogue scaling and eye-tracking. In *Proceedings of the XVII International Congress of Phonetic Sciences* (pp. 1126–1129). Hong Kong: ICPhS.

- Kong, Y. Y., & Zeng, F. G. (2006). Temporal and spectral cues in Mandarin tone recognition. *The Journal of the Acoustical Society of America*, 120(5), 2830–2840.
- Kraus, N., Strait, D. L., & Parbery-Clark, A. (2012). Cognitive factors shape brain networks for auditory skills: Spotlight on auditory working memory. *Annals of the New York Academy of Sciences*, 1252(1), 100–107.
- Lallier, M., Tainturier, M. J., Dering, B., Donnadieu, S., Valdois, S., & Thierry, G. (2010). Behavioral and ERP evidence for amodal sluggish attentional shifting in developmental dyslexia. *Neuropsychologia*, 48(14), 4125–4135.
- Larson, E., & Lee, A. K. (2013). The cortical dynamics underlying effective switching of auditory spatial attention. *NeuroImage*, 64, 365–370.
- Law, S. P., Fung, R., & Kung, C. (2013). An ERP study of good production vis-à-vis poor perception of tones in Cantonese: Implications for top-down speech processing. *PLoS One*, 8(1), e54396.
- Lee, K. Y., Chan, K. T., Lam, J. H., van Hasselt, C. A., & Tong, M. C. (2015). Lexical tone perception in native speakers of Cantonese. *International Journal of Speech-Language Pathology*, 17(1), 53–62.
- Lee, A. K., Larson, E., Maddox, R. K., & Shinn-Cunningham, B. G. (2014). Using neuroimaging to understand the cortical mechanisms of auditory selective attention. *Hearing Research*, 307, 111–120.
- Lev-Ari, S., & Peperkamp, S. (2014). The influence of inhibitory skill on phonological representations in production and perception. *Journal of Phonetics*, 47, 36–46.
- Lotto, A. J., Hickok, G. S., & Holt, L. L. (2009). Reflections on mirror neurons and speech perception. *Trends in Cognitive Sciences*, 13(3), 110–114.
- Mattys, S. L., & Wiget, L. (2011). Effects of cognitive load on speech recognition. *Journal of Memory and Language*, 65(2), 145–160.
- McClelland, J. L., & Elman, J. L. (1986). The TRACE model of speech perception. *Cognitive Psychology*, 18(1), 1–86.
- McQueen, J. M., Norris, D., & Cutler, A. (1994). Competition in spoken word recognition: Spotting words in other words. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 20(3), 621.
- Mok, P., Zuo, D., & Wong, P. W. (2013). Production and perception of a sound change in progress: Tone merging in Hong Kong Cantonese. *Language Variation and Change*, 25(03), 341–370.
- Moore, D. R., Ferguson, M. A., Halliday, L. F., & Riley, A. (2008). Frequency discrimination in children: Perception, learning and attention. *Hearing Research*, 238(1), 147–154.
- Näätänen, R., Paavilainen, P., Rinne, T., & Alho, K. (2007). The mismatch negativity (MMN) in basic research of central auditory processing: A review. *Clinical Neurophysiology*, 118(12), 2544–2590.
- Newman, R. S. (2003). Using links between speech perception and speech production to evaluate different acoustic metrics: A preliminary report. *The Journal of the Acoustical Society of America*, 113(5), 2850–2860.
- Norris, D. (1994). Shortlist: A connectionist model of continuous speech recognition. *Cognition*, 52(3), 189–234.
- Nosofsky, R. M. (1986). Attention, similarity, and the identification–categorization relationship. *Journal of Experimental Psychology: General*, 115(1), 39.
- Oldfield, R. C. (1971). The assessment and analysis of handedness: The Edinburgh inventory. *Neuropsychologia*, 9(1), 97–113.
- Ou, J., & Law, S. P. (2016). Individual differences in processing pitch contour and rise time in adults: A behavioural and electrophysiological study of Cantonese tone merging. *Journal of the Acoustical Society of America*, 139(6), 3226–3237.
- Ou, J., Law, S. P., & Fung, R. (2015). Relationship between individual differences in speech processing and cognitive functions. *Psychonomic Bulletin & Review*, 22, 1725–1732.
- Perkell, J. S., Guenther, F. H., Lane, H., Matthies, M. L., Stockmann, E., Tiede, M., & Zandipour, M. (2004). The distinctness of speakers’ productions of vowel contrasts is related to their discrimination of the contrasts. *The Journal of the Acoustical Society of America*, 116(4), 2338–2344.
- Perkell, J. S., Lane, H., Ghosh, S., Matthies, M. L., Tiede, M., Guenther, F., & Ménard, L. (2008). Mechanisms of vowel production: Auditory goals and speaker acuity. In *Proceedings of the Eighth International Seminar on speech production, Strasbourg, France* (pp. 29–32).
- Polich, J. (2003). Overview of P3a and P3b. In J. Polich (Ed.), *Detection of change: Event-related potential and fMRI findings* (pp. 83–98). Boston: Kluwer.
- 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, 1547–1570.
- Renvall, H., & Hari, R. (2002). Auditory cortical responses to speech-like stimuli in dyslexic adults. *Journal of Cognitive Neuroscience*, 14(5), 757–768.
- Robertson, I. H., Ward, T., Ridgeway, V., & Nimmo-Smith, I. (1994). *The test of everyday attention: TEA*. Bury St. Edmunds: Thames Valley Test Company.
- Ruffino, M., Trussardi, A. N., Gori, S., Finzi, A., Giovagnoli, S., Menghini, D., ... Facoetti, A. (2010). Attentional engagement deficits in dyslexic children. *Neuropsychologia*, 48(13), 3793–3801.
- Schertz, J., Cho, T., Lotto, A., & Warner, N. (2015). Individual differences in phonetic cue use in production and perception of a non-native sound contrast. *Journal of Phonetics*, 52, 183–204.
- Smith, D. V., Davis, B., Niu, K., Healy, E. W., Bonilha, L., Fridriksson, J., ... Rorden, C. (2010). Spatial attention evokes similar activation patterns for visual and auditory stimuli. *Journal of Cognitive Neuroscience*, 22(2), 347–361.
- Stevens, M., & Reubold, U. (2014). Pre-aspiration, quantity, and sound change. *Laboratory Phonology*, 5(4), 455–488.
- Stewart, M. E., & Ota, M. (2008). Lexical effects on speech perception in individuals with “autistic” traits. *Cognition*, 109(1), 157–162.
- Tabachnick, B. G., & Fidell, L. S. (1996). *Using multivariate statistics* (3rd ed.). New York: Harper Collins.
- Thompson, B. (2000). Canonical correlation analysis. In L. Grimm & P. Yarnold (Eds.), *Reading and understanding more multivariate statistics* (pp. 207–226). Washington, DC: American Psychological Association.
- Thomson, J. M., Goswami, U., & Baldeweg, T. (2009). The ERP signature of sound rise time changes. *Brain Research*, 1254, 74–83.
- Tong, X., McBride, C., Lee, C. Y., Zhang, J., Shuai, L., Maurer, U., & Chung, K. K. (2014). Segmental and suprasegmental features in speech perception in Cantonese-speaking second graders: An ERP study. *Psychophysiology*, 51(11), 1158–1168.
- Tsang, Y. K., Jia, S., Huang, J., & Chen, H. C. (2011). ERP correlates of pre-attentive processing of Cantonese lexical tones: The effects of pitch contour and pitch height. *Neuroscience Letters*, 487(3), 268–272.
- Wechsler, D. (2010). *Wechsler Adult Intelligence Scale* (4th ed.). San Antonio, TX: Pearson/PsychCorp.
- Whalen, D. H., & Xu, Y. (1992). Information for Mandarin tones in the amplitude contour and in brief segments. *Phonetica*, 49(1), 25–47.
- Yip, M. (2002). *Tone*. Cambridge: Cambridge University Press.
- Yu, A. C. L. (2010). Perceptual compensation is correlated with individuals’ “autistic” traits: Implications for models of sound change. *PLoS ONE*, 5(8), e11950.
- Zhou, Y. V. (2012). *The role of amplitude envelope in lexical tone perception: Evidence from cantonese lexical tone discrimination in adults with normal hearing* (Doctoral dissertation). City University of New York, NY.