CORE
Metadata, citation and similar papers at core.ac.uk
Provided by HKU Scholars Hub
The HKU Scholars Hub The University of Hong Kong 香港大學學術庫

Title	Individual differences in processing pitch contour and rise time in adults: A behavioral and electrophysiological study of Cantonese tone merging
Author(s)	OU, J; Law, SP
Citation	Journal of the Acoustical Society of America, 2016, v. 139 n. 6, p. 3226-3237
Issued Date	2016
URL	http://hdl.handle.net/10722/229535
Rights	Journal of the Acoustical Society of America. Copyright © Acoustical Society of America.; Copyright (year) Acoustical Society of America. This article may be downloaded for personal use only. Any other use requires prior permission of the author and the Acoustical Society of America. along with the following message: The following article appeared in (Journal of the Acoustical Society of America, 2016, v. 139 n. 6, p. 3226-3237) and may be found at (http://dx.doi.org/10.1121/1.4954252).; This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License.

CrossMark

Individual differences in processing pitch contour and rise time in adults: A behavioral and electrophysiological study of Cantonese tone merging

Jinghua Ou and Sam-Po Law[a)]
Division of Speech and Hearing Science, the University of Hong Kong, Hong Kong Special Administrative Region

(Received 16 July 2015; revised 18 April 2016; accepted 1 June 2016; published online 28 June 2016)

One way to understand the relationship between speech perception and production is to examine cases where the two dissociate. This study investigates the hypothesis that perceptual acuity reflected in event-related potentials (ERPs) to rise time of sound amplitude envelope and pitch contour [reflected in the mismatch negativity (MMN)] may associate with individual differences in production among speakers with otherwise comparable perceptual abilities. To test this hypothesis, advantage was taken of an on-going sound change–tone merging in Cantonese, and compared the ERPs between two groups of typically developed native speakers who could discriminate the high rising and low rising tones with equivalent accuracy but differed in the distinctiveness of their production of these tones. Using a passive oddball paradigm, early positive-going EEG components to rise time and MMN to pitch contour were elicited during perception of the two tones. Significant group differences were found in neural responses to rise time rather than pitch contour. More importantly, individual differences in efficiency of tone discrimination in response latency and magnitude of neural responses to rise time were correlated with acoustic measures of F0 offset and rise time differences in productions of the two rising tones. © *2016 Acoustical Society of America*. [http://dx.doi.org/10.1121/1.4954252]

[MAH] Pages: 3226–3237

I. INTRODUCTION

Theories of speech processing generally agree that speech production and perception interact in some manner. The motor theory (e.g., Liberman and Mattingly, 1985) proposes a strong link between perception and production in that a specialized phonetic module representing speech units in terms of articulatory gestures mediates both speech perception and production. Thus, the motor theory maintains that processes of speech motor planning are mandatory to speech perception, and predicts changes in production should modify perception. Contrary to the motor-centric view, other theorists have suggested that speech production relies on speech perception. For example, the Directions into Velocities of Articulators (DIVA) model proposes that auditory perceptual representations functions as acoustic templates to calibrate speech production (Guenther, 1995). Drawing on data from functional neuroimaging and aphasia, Hickok and Poeppel (2007) have also argued that auditory processing is critically involved in the production of speech. This perspective is also influential in models focusing on speech development, where the acoustic input to a prelingual child determines the speech patterns he or she acquires. Kuhl *et al.* (2008) have posited that the link between perception and production is forged based on perceptual experience and mapping between the two is learned during development. On the whole, perceptual systems are considered to have a stronger influence on production than motor systems have on perception (Lotto *et al.*, 2009).

One way of understanding the relationship between speech perception and production is to investigate cases where the two dissociate. Specific to dissociations between accurate perception but poor production, cases can be readily found in individuals learning a new language, or individuals with acquired language impairment as a result of brain damage, such as conduction aphasia, a syndrome characterized by good comprehension but frequent phonemic errors in production (Damasio and Damasio, 1980). In the present study, we took advantage of a unique opportunity in an on-going sound change in Hong Kong Cantonese (HKC)–tone merging (see below) to identify two groups of typically developed native speakers who show comparably accurate tonal distinction in perception but a difference in production. According to psycholinguistic speech production models (e.g., Levelt, 1999), the mismatch pattern of distinctive perception but non-distinctive production may not be difficult to reconcile as production involves motor programming of articulatory features subsequent to access to sensory/phonological representations. For instance, inaccurate production demonstrated by second language learners is perhaps due to the lack of practice of motor programming. Alternatively, to account for the dissociation exemplified by conduction aphasia, the duel-stream model (Hickok and Poeppel, 2007) proposes a disruption of the auditory-motor interface system, such that sensory representations can no longer provide online guidance for motor programming further leading to production errors. The presupposition inherent in these two accounts is

a)Electronic mail: splaw@hku.hk

 0001-4966/2016/139(6)/3226/12/$30.00 © 2016 Acoustical Society of America

RIGHTSLINK

that phonological representation underlying speech perception is accurate and/or remains intact. Nonetheless, neither account can satisfactorily explain the pattern of good perception and poor production among healthy normally developed native speakers. In this study, we investigated typically developed speakers differentiated by their speech production but not their speech perception abilities based on discrimination accuracy. We examined the neural responses measured with event-related potentials (ERPs) to rise time of sound amplitude envelope and the mismatch negativity (MMN) to pitch contour to see if the two groups differ at the brain level. If so, whether and how changes in production would be related to changes in neural processing of related features in the inputs. To the extent that different patterns of perception can be reflected in neural measures which are related to production, we gain a deeper understanding of the mechanism underlying the dissociation between speech perception and production, and hence the relationship between the two.

Perceptual studies generally agree that two aspects of the F0—the F0 level (high, middle, low) and the F0 contour (static, rising, falling)—are perceptual correlates in tone languages, including Mandarin Chinese (Gandour, 1983), Cantonese (Khouw and Ciocca, 2007), and Thai (Gandour *et al.*, 1994). The acoustic cues are language-dependent and, to a large extent, influenced by the composition of a tone system. For instance, in a contour tone system such as Mandarin Chinese, it has been demonstrated that Mandarin-speaking listeners attach more importance to pitch contour than pitch level (Chandrasekaran *et al.*, 2007). In contrast, in tone perception of Cantonese or Thai, where the tone system contains several level tones, the relative F0 levels play a more important role in distinguishing among the tones (Vance, 1976). Specific to Cantonese, Khouw and Ciocca reported that the F0 changes over the later part of the vocalic segment (i.e., F0 offset) were critical for distinguishing between the tones.

Besides the dominant role of spectral information, much attention has recently been paid to the importance of temporal information in parsing the acoustic signal into relevant segments for decoding during auditory/speech processing (Luo and Poeppel, 2012). Acoustic cues from the amplitude envelope have also been shown to successfully cue tone perception in Mandarin Chinese (Fu and Zeng, 2000; Kong and Zeng, 2006) as well as Cantonese (Zhou, 2012). The amplitude envelope of tone refers to the amplitude fluctuation in the waveform of a tone (Rosen, 1992), which reflects the overall rising, falling or steady trend of amplitude change throughout the production of a tone (Baken and Orlikoof, 2000). For instance, Fu and Zeng (2000) found that the amplitude envelope contributed significantly to the dipping and falling tone discrimination in Mandarin Chinese. Of the various cues of amplitude envelope, rise time, defined as the time taken for a sound to reach its maximum amplitude (Rosen, 1992), is proposed to be an important perceptual cue for the representation of amplitude envelope (Greenberg, 2006). The amplitude rise time has been found to be important in facilitating prosodic and syllable segmentation processes in children (Carpenter and Shahi, 2013), which are arguably critical for the formation of well-specified phonological representations (Goswami, 2011). Hence, one may question whether the rise time of sound amplitude envelope may likewise play a role in processing lexical tones. In other words, to process tones efficiently may entail the encoding of both spectral and temporal cues present in the speech signal to derive and access tone representations.

The dynamic process of tone perception may be characterized by individual variations in that listeners differ in their sensitivity to contrasts of different acoustic cues. In fact, sociolinguists have long recognized that individual differences exist not only in speech perception but also production among typically developed speakers of a linguistic community (e.g., Beddor, 2012; Johnson, 2006). In the case of HKC, the different behavioral patterns of tone perception and production can be captured in an on-going sound change–tone merging (Bauer *et al.*, 2003; Mok *et al.*, 2013). There are six contrastive tones for non-stopped syllables in HKC [see Fig. 1(a)], namely, T1 (high level tone [55]), T2 (high rising tone [25]), T3 (mid level tone [33]), T4 (low falling/extra low level tone [21]), T5 (low rising tone [23]), and T6 (low level tone [22]). The numbers in square brackets represent the relative starting and ending pitch levels of each tone, with 5 being the highest and 1 being the lowest pitch level (Chao, 1930). Previous observations have revealed three suspected tone mergers: T2 vs T5, T3 vs T6, and T4 vs T6 (Lee *et al.*, 2015; Mok *et al.*, 2013). Particularly, the high rising and low rising tones (T2/T5) have undergone an extensive merging in the community, and a significant number of Cantonese adult speakers can no longer distinguish the contrast between the two tones in perception and/or production (Bauer *et al.*, 2003; Mok *et al.*, 2013). While behavioral measures of accuracy and discrimination latency are traditionally and commonly used to identify perceptual patterns, studies of speech perception in recent years have also employed more sensitive methods to examine individual differences in speech perception at the neural level, including ERP (e.g., Díaz *et al.*, 2008). The ERP allows us to detect potential differences during on-line processing that are otherwise difficult, if not impossible, with behavioral tests.

To evaluate the neural correlates of acoustic and speech stimuli discrimination, researchers may measure participants' mismatch negativity (MMN) responses (Näätänen, 2001). The MMN is a fronto-central negative deflection peaking around 100–250 ms after change onset, and is usually elicited in an oddball paradigm, where a mismatch may be detected between a frequently repeated stimulus (the standard) and a stimulus deviating in at least one acoustic parameter (the deviant). Following the prevailing memory-based interpretation (Näätänen, 1990), MMN reflects the operation of a memory mechanism in which representations of the environment are used by a neural comparison process to detect auditory changes [see May and Tiitinen (2010) for an alternative adaptation/fresh-afferents account of MMN]. The MMN has been used extensively to examine the sensitivity of individuals to variations in speech sound contrasts, including place of articulation, voicing, and vowel (see Näätänen *et al.*, 2007 for a review). Its amplitude is often related to the magnitude of acoustic difference, and is thus considered a measure of individual sensitivity to auditory

RIGHTSLINK

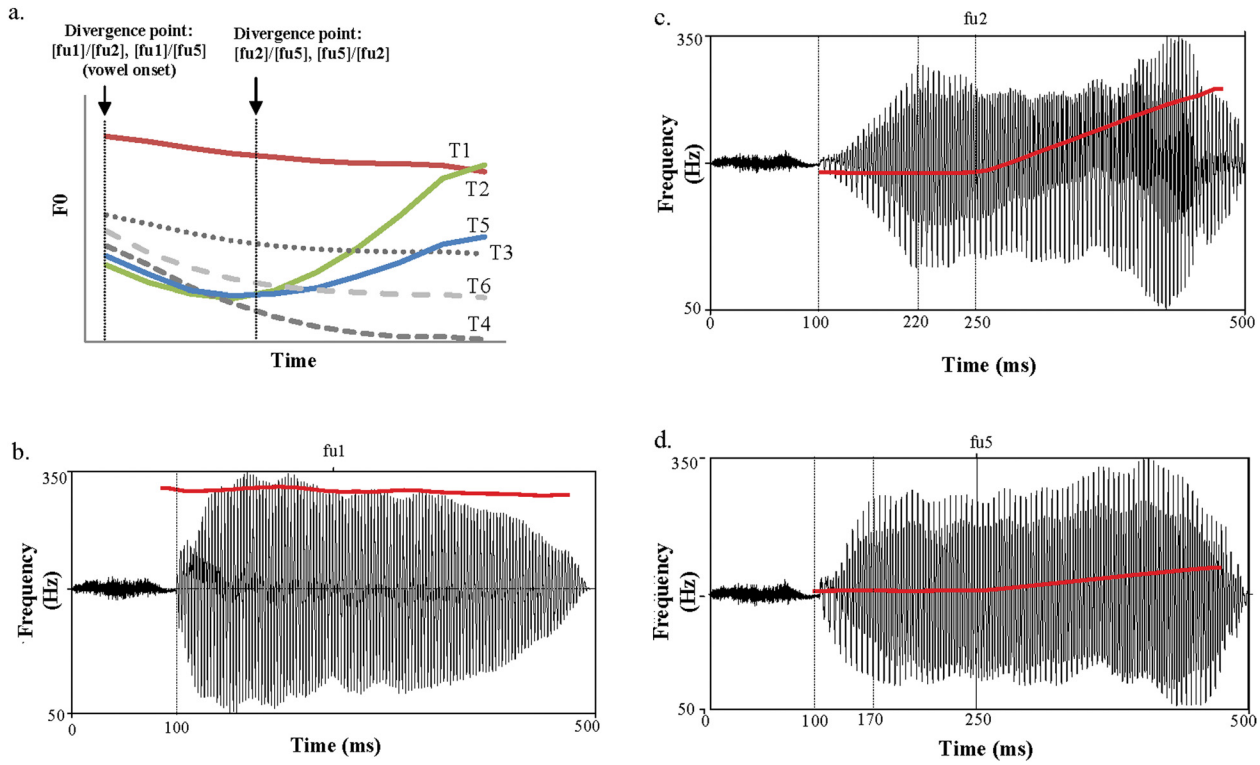

FIG. 1. (Color online) (a) F0 contours of the six contrastive tones in HKC with the three stimuli T1, T2, and T5 used in the EEG experiment indicated by the solid lines. Sound waveforms of /fu1/ (b), /fu2/ (c), and /fu5/ (d). The vowel onset is at 100 ms. The F0 divergence point for /fu2/ and /fu5/ is at 250 ms. The peak amplitude is at 120 ms for /fu2/ and 70 ms for /fu5/ post vowel onset. The solid line indicates F0 contour.

discrimination (Näätänen *et al.*, 2007). Moreover, the MMN can be elicited independently of participants' attention to the stimuli (i.e., passive oddball), and hence it is not assumed to be influenced by engagement of cognitive processes associated with task demands or strategies (Näätänen *et al.*, 2007). In addition, a positive-going ERP component typically found following the MMN is P3a, which is suggested to indicate the involuntary attention switching induced by the detection of deviant features in the passive oddball paradigm (Polich, 2003).

A few MMN studies using the passive oddball paradigm have examined tone discrimination in Cantonese (e.g., Law *et al.*, 2013; Tsang *et al.*, 2011). For instance, Tsang *et al.* demonstrated the size and latency of the MMN responses among Cantonese speakers were more sensitive to differences in pitch height than pitch contour, and the latency of P3a captured the presence of pitch contour change. Specifically, a smaller amplitude and longer latency of MMN was elicited by the high level vs high rising tone contrast (T1/T2) relative to the high level vs low level tone contrast (T1/T6), although the acoustic differences for T1/T6 and T1/T2 were comparable at the pitch onset. Moreover, the latency of P3a to T1/T2 was longer than T1/T6. A more recent study by Law *et al.* investigated the discrimination between the low falling and mid level tone contrast (T4/T6) in two groups of typically developed adult speakers of Cantonese whose difference represents the tone near-merger phenomenon. Behaviorally, both groups of participants could produce all six tones distinctively; they differed only in their perception of the T4/T6 contrast. As expected, the two participant groups showed differential responses to the T4/T6 contrast in the MMN. The significance of Law *et al.* lies not only in the use of ERP to study sound change phenomena traditionally in the realm of sociolinguistics, but also the potential of such phenomena to reveal the link between speech perception and production among typically developed speakers showing different behavioral patterns of perception and production.

The present study is the first examination of neural processes underlying the discrimination of the high rising and low rising tones T2/T5 in HKC from two groups of typically developed native speakers of HKC with comparable language and musical backgrounds. It is important to control for musical experience as it has repeatedly been shown to influence speech perception (e.g., Strait and Kraus, 2011). The participant groups represented, respectively, the pattern of good production and good perception of all Cantonese tones [+Pro+Per], and that of poor production of specifically the T2/T5 distinction but good perception of all tones [−Pro+Per]. We investigated the hypothesis that native speakers of HKC who have prototypical patterns of production and perception [+Pro+Per] would have different neural responses to those with [−Pro+Per]. The hypothesis was assessed using a passive oddball paradigm with ERP measures, which revealed the timing and strength of neural activities associated with the auditory stimuli unfolding over time. In addition, their productions were analyzed acoustically to identify the key acoustic features characterizing the differentiation between the two rising tones. We predicted that listeners would make use of information from F0 contour and potentially amplitude envelope to discriminate the

RIGHTSLINK

two highly similar contour tones. Last, we examined how differences in neural responses, if any, would relate to the acoustic differences in production, findings of which would provide significant insights into the relationship between perception and production.

II. METHODS

A. 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.

B. Participants

A total of 138 native speakers of Cantonese, all born and raised in Hong Kong, were recruited. No speaker reported a history of hearing abnormalities. They first participated in a tone perception and a tone production task.

1. Tone perception and production tasks

a. Stimuli. To control for syllable effects, only one CV root [fu] was used to derive the six tones. The six syllables were produced in single words, and recorded by a native female Cantonese speaker and served as the stimuli of the perception task. The stimuli were kept as natural as possible, but modifications were done in the vowel portion to standardized the syllable length to 500 ms using PRAAT (Boersma and Weenink, 2015), and the intensity to 70 dB sound pressure level (SPL) using AUDACITY (2015). The stimuli were delivered at this amplitude, as verified by measurement of a sound level meter (Brüel and Kjær 2250). For the production task, the six syllables were represented by six Chinese characters, i.e., 夫 fu1 "husband," 苦 fu2 "bitter," 褲 fu3 "trousers," 符 fu4 "symbol," 婦 fu5 "women," and 負 fu6 "negative." The characters were selected based on the results of a questionnaire completed by 85 undergraduate students, in which the students were asked to write down the first character they could think of associated with each of the six tones for the syllable [fu]. This was to ensure that the chosen characters are the ones most frequently linked to the respective syllables. Specifically, 夫 was chosen by 75% of the students to associate with T1, 苦 41% with T2, 褲 35% with T3, 符 41% with T4, 婦 80% with T5, 負 34% with T6.

b. Procedures. The perception and production tasks were conducted in a sound attenuated booth in the University of Hong Kong. The tasks were administered to the participants using the PRESENTATION® software (2015) running on an IBM laptop. Sounds were output through a Conexant 20672 SmartAudio HD sound card at a sampling rate of 44.1 kHz, and stimuli were presented diotically via Sennheiser headphones (HD-545) in the perception task. For the production task, speech outputs were recorded using an Audio-technica microphone (ATR2100), and sampled at 44.1 kHz digitized at 16 bits using AUDACITY. The production task was carried out before the perception task to eliminate any priming effect. It took approximately one hour to complete the two tasks.

The perception task was an AX discrimination test. Thirty-six tone pairs (6 AA pairs and 30 AB pairs counterbalanced in order of syllables) were repeated 10 times each, giving a total of 360 stimuli. A trial began with a fixation point on the screen for 300 ms, followed by the presentation of a tone pair with an inter-stimulus interval (ISI) of 500 ms. The participants had to indicate as soon as possible whether the tones presented were the same or different within 2 s. The inter-trial interval (ITI) jittered from 1 to 2 s. Both accuracy and reaction time were collected. Reaction time (RT) was measured from the divergence point in the second syllable. Additionally, participants' behavioral sensitivity (d′) to tone contrasts was computed based on hit (H) and false alarm (FA) rates, following the roving (differencing model) methods discussed in Macmillan and Creelman (1991) [pp. 147–152, Table V.4 provides values of true d′ for every (H, FA) pair].

In the production task, the six syllables were embedded in different positions of two sentence carriers: /ŋɔ^{13} ji$^{21/55}$ kɑ^{55} tʊk^{2} _ tsi^{22}/ "I am now reading the _ character" and /nei^{55} kɔ^{33} tsi^{22} hɐi^{22} _/ "This character is _." The stimuli were presented in written form. The 12 stimuli (6 syllables × 2 carriers) were repeated ten times each, thus generating 120 trials. For each trial, a fixation point appeared on the screen for 300 ms followed by the presentation of a sentence for 5 s. The participants were instructed to read aloud the sentence at a normal speech rate. The ITI was set to 1 s. The speech outputs were recorded digitally for phonetic transcription by a native Cantonese speaker with training in phonetics who was blind to the target stimuli, and 10% of the trials were randomly selected and transcribed by a second native Cantonese speaker also blind to the targets. The inter-rater reliability reached a 95% agreement.

Acoustic analysis of the production data was conducted to verify the auditory transcription analysis. By combining the auditory transcription and acoustic analyses, we ascertained that the production differences between the two groups were perceptually valid as well as acoustically significant. The F0 trajectory of the target syllable was analyzed using PRAAT. The start and end of the vocalic segment of the syllable were selected manually from the amplitude waveforms. The onset of the tone was marked by the start of vocalic modality, and the offset by the maximum point of the rising trajectory near the final vocalic portion. F0 (Hz) of the vocalic segment at ten equidistant time points was extracted via a PRAAT script. F0 values were then converted from Hz to a logarithm-based T value (Rose, 1987) to reduce cross-speaker variation. To identify the acoustic features that speakers used to separate the production of T2 and T5, different acoustic properties along the time course of the F0 trajectory were measured, including F0 onset, offset, onset-offset difference, and duration. Apart from the F0 information, the rise time of the amplitude envelope was also assessed, which was computed as the duration between vowel onset and amplitude peak during the vocalic segment (Tarr, 2013). For the other four tones, T1, T3, T4, and T6,

RIGHTSLINK

the mean F0 height was computed by averaging the T-values of the 10 points.

2. Participant selection

This section presents the descriptive statistics pertaining to the grouping of participants only, and additional findings arising from these two tasks are presented in the results section.

For tone perception, participants selected should score at least 95% correct in discrimination accuracy. As for tone production, a speaker had to score 100% correct in producing T2 and T5 according to the auditory transcription, to be classified as having good production. Poor production was defined by an accuracy of less than 60% for T2 and T5. For both groups, production accuracies of the other tones should be at least 90% correct. Based on the auditory transcription, 20 participants were classified as having good production and 21 for poor production.

Acoustic analyses were conducted to more objectively assess the production differences between the two groups. To ascertain the range of acoustic differences between prototypical productions of T2 and T5, different acoustic properties of T2 and T5 produced by participants of [+Pro+Per] were compared. Results revealed that the F0 duration and F0 onset did not differed significantly between the T2 and T5 productions in the [+Pro+Per] group (Bonferroni correction all $p > 0.01$, details see Sec. III A), whereas significant differences were found for the F0 offset and the F0 onset-offset difference, with T2 F0 offset higher than that of T5 [$t(19) = 33.95$, $p < 0.001$ (two-tailed unless specified otherwise), Cohen's $d = 11.02$], and T2 F0 onset-offset difference larger than that of T5 [$t(19) = 30.48$, $p < 0.001$, Cohen's $d = 5.06$]. As the significant differences of F0 onset-offset difference was driven by the differences in F0 offset, the difference between the T2 and T5 F0 offsets (T2 F0 offset minus T5 F0 offset) was then taken as an index of the degree of tonal differentiation (Barry and Blamey, 2004) demonstrated by a speaker. The T2-T5 F0 offset difference produced by participants who were classified as [+Pro+Per] based on auditory transcription were taken as a reference to verify the status of [−Pro+Per] participants. For a participant to be regarded as poor in distinguishing T2 and T5 in production, his or her F0 offset difference had to be at least 2.5 standard deviations (SDs) below the mean of the [+Pro+Per] group, in order to make sure the two groups represent statistically distinct distributions. Data from two participants were thus excluded from the final formation of the [−Pro+Per] group.

In all, on the basis of their performance on the perception and production tasks, 39 participants were selected and invited back to carry out a passive oddball task. The 39 participants were all right-handers according to the Edinburgh Handedness Inventory. Table I presents the characteristics of the two participant groups in terms of their performance on tone perception, tone production, and musical background. One group could distinctively perceive and produce all six Cantonese tones ([+Pro+Per], N = 20, female = 8); a second group could perceive all tones but fail to produce T2 and T5 distinctively ([−Pro+Per], N = 19, female = 11). The two groups were matched on the accuracy score [$t(37) = 1.11$, $p = 0.273$] and discrimination sensitivity index d′ [$t(37) = 1.05$, $p = 0.302$] of T2-T5 perception, age, years of formal education, and musical background in terms of onset and duration of training [all $t(37) < 0.84$, $p > 0.410$]. The F0 offset difference was confirmed to be significantly smaller for the [−Pro+Per] than the [+Pro+Per] group [$t(37) = 11.35$, Bonferroni correction $p < 0.001$, Cohen's $d = 3.71$], but not for the mean F0 heights of the other tones [all $t(37) < 0.67$, $p > 0.212$].

TABLE I. Background information on participants in [+Pro+Per] and [−Pro+Per] groups.

	[+Pro+Per] (N = 20)		[−Pro+Per] (N = 19)	
	M	SD	M	SD
Age	22.00	0.59	21.24	0.84
Years of education	17.10	0.67	17.00	0.12
Tone discrimination				
Distinguishing T2-T5				
Accuracy	0.99	0.01	0.98	0.01
Sensitivity (d′)	6.59	0.48	6.37	0.81
All other tone pairs				
Accuracy	0.99	0.01	0.97	0.01
Sensitivity (d′)	6.53	0.23	6.24	0.35
Tone production				
T2-T5 pitch offset difference[a]	2.95	0.38	1.11[b]	0.59
T1 mean pitch height	4.62	0.52	4.71	0.43
T3 mean pitch height	3.06	0.60	3.20	0.51
T4 mean pitch height	1.09	0.32	1.10	0.47
T6 mean pitch height	2.11	0.75	2.25	0.67
Musical background				
Onset	6.63	1.50	5.62	1.02
Duration	5.16	1.35	5.00	1.05

[a]The pitch is in normalized F0 values.
[b]The pitch offset of [−Pro+Per] group was significantly smaller than that of [+Pro+Per] group ($p < 0.001$).

C. EEG experiment

1. Stimuli

Three syllables /fu1/, /fu2/, and /fu5/ from the behavioral task were used in this experiment. The experiment consisted of four oddball conditions of different standard/deviant pairs, including T2/T5 (i.e., T2 as standard vs T5 as deviant) and T5/T2 as two experimental conditions, and two control conditions by pairing T2 and T5 with T1 as the common standard, i.e., T1/T2 and T1/T5. All three syllables were aligned to have the same vowel onset (100 ms) and vowel duration (400 ms). For the control conditions T1/T2 and T1/T5, the divergence point was at the vowel onset, where the F0 heights of the two stimuli began to deviate. The divergence point was different for the two experimental conditions, as T2 resembled T5 in the early part of the F0 contour and the two began to diverge at 250 post stimulus onset [Fig. 1(a)]. Additionally, in the period of 100–250 ms where the F0 contours of T2 and T5 fully overlapped, the

RIGHTSLINK

amplitude rise time differed between them. The rise time was 120 ms for T2 and 70 ms for T5 [Figs. 1(c) and 1(d)].

2. Procedure

The passive oddball task was administered to the participants using the PRESENTATION program running on a desktop. The participant was seated comfortably in front of a computer in a sound-attenuated electrically shielded booth. During the task, the participant was asked to watch a silent movie on a computer screen located at a distance of approximately 1 m away. Auditory stimuli were binaurally presented at 85 dB SPL through insert earphones simultaneously. The participant was told to concentrate on the movie while completely ignore all auditory stimuli.

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 in a quasi-random sequence with the constraint that there would be a minimum of five and a maximum of ten standards between consecutive deviants. Each trial consisted of the 500 ms syllable and an inter-stimulus interval (ISI) of 800 ms. The sequence of blocks were rotated across participants. The entire experiment lasted about 100 min.

3. Data recording and processing

The EEG was recorded on a SynAmps2 Neuroscan Inc. system (Compumedics Ltd., USA) from 64 Ag/AgCl electrodes (FPz, Fz, FCz, Cz, Pz, POz, Oz, FP1/2, F7/5/3/1/2/4/6/8, FT7/8, FC5/3/1/2/4/6, T7/8, C5/3/1/2/4/6, M1/2, TP7/8, CP5/3/1/2/4/6, P7/5/3/1/2/4/6/8, PO7/5/3/4/6/8, O1/2) arranged in an extended montage based on the international 10-20 system (using a Neuroscan 64-channel QuickCap, Compumedics Ltd., USA). M1 or M2 was selected as online reference electrode and ground was placed at AFz. Additional electrodes were placed on the supra- and infra-orbit ridges of the left eye and lateral to the outer canthus of both eyes to monitor vertical eye movements (VEOG) and horizontal eye movements (HEOG). Impedance for all electrodes was kept below 10 KΩ. Continuous data were digitized at a sampling rate of 500 Hz with a bandpass of 0.05 to 200 Hz.

The raw EEG data were preprocessed using the matlab toolbox FIELDTRIP (2015). The continuous data were first epoched with an 800 ms pre-stimulus interval and 1000 ms post-stimulus onset interval. Extreme trials with an amplitude larger than $\pm 300\,\mu$V were removed. Artifact reduction was performed using independent component analysis (ICA) to identify any components resembling eye blinks, horizontal eye movements, noisy channels, and other focal artefacts, which were then mathematically removed from the data. After ICA, the data were bandpass filtered between 1 and 20 Hz, baseline-corrected using the pre-stimulus interval (−200 to 0 ms) and re-referenced to average mastoids. Further artifact rejection was applied to reject trials exceeding 100 μV, or improbable data greater than 5 SDs. Thus, a total of 189 trials (or 0.44% of all trials) in the [+Pro+Per] group, and 228 trials (0.56%) in the [−Pro+Per] group were removed. For each condition, the remaining trials were categorized into three types: deviant, standard-before-a-deviant, and standard preceding standard-before-a-deviant (see Bishop *et al.*, 2011 for a similar analysis). The last set was then subtracted from all other trials, thus resulting in two types of difference waveforms, one true difference and one dummy difference set. These waveforms together represent the specific activity associated with mismatch after removing the ERP common to standards and deviants.

4. Non-parametric permutation analyses

Statistical differences between the true and dummy difference waves were evaluated by a non-parametric cluster-based random permutation approach (see Maris and Oostenveld, 2007 for details on the method, and see Law *et al.*, 2013 for a similar application to identify MMN), which was implemented in FIELDTRIP. The test first identifies sampling points (time-electrode) with t-statistic exceeding a critical threshold ($p < 0.05$, two-tailed). Clusters of adjacent (spatial-temporal) significant data points are computed, and for each cluster a cluster-level test statistic is calculated by taking the sum of all the individual t-statistics within the cluster. The maximum cluster-level test statistics were then computed to generate permutation distributions, one for positive clusters and one for negative clusters, based on 10 000 random partitions. The significance of a cluster was determined by whether it fell in the highest or the lowest 2.5th percentile of the corresponding distribution. The cluster-based permutation tests were carried out on each block for each participant group to identify significant ERP components reflecting responses to contrasts in pitch height/contour to different tone pairs and rise time between T2 and T5.

5. Conventional analyses at Fz and FCz

a. MMN and P3a to pitch height/contour. Conventional analyses were also performed to examine whether the two groups differed in the magnitude and latency of the ERP components that were identified in the cluster permutation test. Based on previous studies, data from the Fz and FCz electrodes were selected for statistical analyses, where the strongest mismatch effects were usually found (e.g., Chandrasekaran *et al.*, 2007; Tsang *et al.*, 2011). 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. For both components, the average amplitude was computed of a 100 ms time window centered on the MMN and P3a peaks, then averaged across the two selected electrodes. To verify the presence of components at the two selected electrodes, pair-wise comparisons were performed between the difference wave and the dummy wave for each component of interest in each participant group. Furthermore, separate t-tests were conducted for mean amplitude and peak latency for each component of interest in each condition to detect any differences between groups. Bonferroni correction was applied to control for the family-wise type I errors.

RIGHTSLINK

b. ERPs to rise time. The grand averaged ERPs for all occurrences (both standard and deviant) of T2 and T5 were computed, respectively, for each group, to assess whether there were 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 submitted to a two-way mixed-design analysis of variance (ANOVA) with condition (T2, T5) as a within-subject factor and group ([+Pro+Per], [−Pro+Per]) as a between-subject factor. *Post hoc* comparisons were conducted if a significant interaction was found (Bonferroni correction applied).

c. Relationship between perception and production of T2-T5. To examine the relationship between perception and production, Pearson product-moment correlation coefficients were computed between the production acoustic parameters, including T2-T5 F0 offset difference and T2-T5 rise time difference, and the perceptual responses, including (1) the discrimination response time to trials involving T2 and T5, (2) neural responses to rise time of T2, (3) neural responses to rise time of T5, (4) peak latencies and mean amplitudes of MMNs and/or P3a to T2/T5, (5) peak latencies and mean amplitudes of MMNs and/or P3a to T5/T2. To reduce the number of correlations, only those measures with significant differences between groups were entered in the analysis. The key correlations emerged were further subject to a partial correlation to examine the relationship between perception and production while controlling for effects of musical training. Bonferroni adjustment was applied to correct for multiple comparisons.

III. RESULTS

A. Behavioral results

1. Perception of tones

Results of the tone discrimination task showed that the [−Pro+Per] group had significantly longer RT of trials involving T2 and T5 than the [+Pro+Per] group, [$M_{[+Pro+Per]}$ = 1046.18 ms, SD = 80.19; $M_{[-Pro+Per]}$ = 1191.96 ms, SD = 155.03; $t(37) = -3.36$, $p = 0.002$, Cohen's $d = -1.18$], though both groups achieved high accuracies (above 98%) and maintained high discrimination sensitivities (d′ above 6.37).

2. Production of tones

Based on the auditory transcription analysis of tone production, a confusion matrix was constructed for the T2 and T5 produced by the [−Pro+Per] participants. Table II shows bi-directional confusions between the two tones; moreover, a considerable proportion of T2 and T5 productions were perceived as ambiguous forms in between the two. Results of a chi-square test revealed that the distributions of productions for the two tones were significantly different [χ^2 (2, N = 760) = 89.13, $p < 0.001$]. There is a stronger tendency of T5 stimuli being produced as T2 than T2 stimuli as T5

Table III presents the means and standard deviations of different acoustic properties, including F0 duration, onset, offset, and onset-offset difference for each participant group.

TABLE II. Confusion matrix of T2 and T5 produced by the [−Pro+Per] participants.

		Perceived			
		T2	T5	in-between T2/T5	Total
Target	T2	252 (66%)	61 (16%)	67 (18%)	380
	T5	138 (37%)	176 (46%)	66 (17%)	380

Significant group differences were found for the T5 F0 offset [$t(37) = -11.15$, $p < 0.001$, Cohen's $d = -3.56$] and the T5 onset-offset difference [$t(37) = -8.38$, $p < 0.001$, Cohen's $d = -2.68$], with higher T5 F0 offset and larger T5 onset-offset difference in [−Pro+Per] than in [+Pro+Per]. These findings are consistent with the tendency shown in the confusion matrix in that the T5 F0 onset-offset difference was close to that of T2 for [−Pro+Per]. The other measures, i.e., F0 duration and F0 onset, were not different between the two groups (Bonferroni correction $p > 0.005$). Apart from the difference in F0 acoustics, the two groups also differed in the acoustic measurement of amplitude rise time, with T5 rise time of the [+Pro+Per] group significantly shorter than that of the [−Pro+Per] group [$t(37) = -7.92$, $p < 0.001$, Cohen's $d = -2.58$], but not for T2 ($p = 0.47$).

As mentioned earlier, T2 and T5 F0 offset difference was computed to index the degree of production differentiation between the two rising tones, and was confirmed to be significantly different between the two groups ($p < 0.001$). Additionally, the difference in amplitude rise time between T2 and T5 (T2 rise time minus T5 rise time) was computed as another index for the T2-T5 production differentiation. Results showed that T2-T5 rise time difference of the [+Pro+Per] group was significantly larger than that of the [−Pro+Per] group [$M_{[+Pro+Per]}$ = 29.6 ms, SD = 18.7; $M_{[-Pro+Per]}$ = 1.6 ms, SD = 4.71; $t(37) = 7.23$, $p < 0.001$, Cohen's $d = 2.05$].

B. ERP results

1. Cluster-based permutation tests

The results of the cluster-level permutation test revealed several significant clusters in different conditions in the two

TABLE III. Means and standard deviations of different acoustic properties of T2 and T5 for both participant groups.

	[+Pro+Per]		[−Pro+Per]		
	Mean	SD	Mean	SD	p[a]
T2 duration (ms)	482.04	74.58	498.28	72.44	0.50
T5 duration (ms)	478.28	62.44	476.37	80.29	0.38
T2 F0 onset[b]	1.35	0.81	1.53	0.93	0.51
T5 F0 onset	1.47	0.85	2.27	1.12	0.02
T2 F0 offset	5.00	0.00	4.97	0.11	0.31
T5 F0 offset	2.05	0.39	3.86	0.61	0.001[c]
T2 F0 onset-offset difference	4.08	0.56	4.36	0.54	0.128
T5 F0 onset-offset difference	0.94	0.68	3.17	0.22	0.001[c]
T2 amplitude rise time (ms)	74.31	38.50	72.24	37.87	0.47
T5 amplitude rise time (ms)	44.83	13.27	74.02	19.28	0.001[c]

[a]Bonferroni adjusted significance level at $p < 0.005$.
[b]All F0 acoustics is in normalized F0 values.
[c]A significant group difference was found.

RIGHTSLINK

TABLE IV. Averaged amplitudes of true difference wave and dummy difference wave at Fz and FCz electrodes in the T1/T2, T1/T5, and T2/T5 conditions for both participant groups.

		[+Pro+Per]				[−Pro+Per]			
		True difference wave (μV)	Dummy difference wave	t	p	True difference wave	Dummy difference wave	t	p
T1/T2	MMN	−3.52	−0.20	−4.76	0.000	−3.68	−0.40	−4.16	0.001
	P3a	2.38	−0.34	3.29	0.004	1.26	−0.46	3.04	0.007
T1/T5	MMN	−3.33	0.21	−3.33	0.004	−3.10	−0.66	−3.39	0.003
T2/T5	MMN	−1.74	−0.76	−2.84	0.010	−1.01	0.25	−2.26	0.036

participant groups. 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 to 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 to 150 ms post vowel onset.

a. MMN. In the [+Pro+Per] 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 mainly distributed 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 ($p < 0.001$), between 100 and 166 ms for T1/T5 ($p = 0.006$), and between 150 and 200 ms for T2/T5 ($p = 0.015$). A significant negative cluster in the time window between 150 and 238 ms was observed in the T5/T2 condition ($p = 0.044$) but with a centro-parietal distribution, which was hence not considered as an MMN. In the [−Pro+Per] group, MMNs were also elicited in the T1/T2 (110–166 ms, $p = 0.006$), T1/T5 (104–154 ms, $p = 0.025$) and T2/T5 (150–200 ms, $p = 0.015$) conditions, but no significant negative cluster was observed in the T5/T2 condition.

b. P3a. The contrast between T1 and T2 elicited a significant positive cluster immediately following the MMN for both groups, in the time window of 300 to 400 ms for [+Pro+Per] ($p = 0.025$) and 342 to 404 ms for [−Pro+Per] ($p = 0.039$), which can be considered P3a. No significant positive clusters were found in the other conditions.

c. 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. Both participant groups exhibited an early positive-going cluster in the T2/T5 condition in the time window between 62 and 154 ms for [+Pro+Per] ($p = 0.015$) and between 64 and 144 ms for [−Pro+Per] ($p = 0.025$). For the T5/T2 condition, an early negative-going component was observed only in the [+Pro+Per] group in the time window of 36 to 176 ms ($p = 0.039$).

In summary, similar patterns of mismatch response were demonstrated in the T1/T2, T1/T5, and T2/T5 conditions for both participant groups. In the T5/T2 condition, no significant clusters were observed in the [−Pro+Per] group, whereas two negative-going components were obtained for the [+Pro+Per] group.

2. T-tests and ANOVAs of neural responses at Fz and FCz

a. MMN and P3a to pitch height/contour. The conventional analyses were restricted to the components that were identified with appropriate scalp distributions in the cluster permutation test, i.e., MMNs to T1/T2, T1/T5, and T2/T5, as well as P3a to T1/T2. Statistical analyses showed that the mean amplitudes of the true difference waves were more negative than those of dummy difference waves of the MMNs in the three conditions in both groups. The presence of P3a to T1/T2 was also verified in both participant groups (see Table IV). The mean amplitudes and peak latencies of the difference waves of each component of interest are shown in Table V. Results of group comparisons revealed that that none of the above measures significantly differed between the two groups at the adjusted p value of 0.006 (all $p > 0.011$).

b. Early neural responses to rise time of T2 and T5. The averaged ERPs to all occurrences of T2 and T5 for both participant groups are shown in Fig. 2. Results of a mixed ANOVA of the average amplitudes showed main effects of tone condition [$F = (1, 37) = 46.46$, $p < 0.001$, $\eta^2 = 0.57$] and group [$F = (1, 37) = 5.08$, $p = 0.030$, $\eta^2 = 0.121$], with T5 ($M = 2.56$, $SD = 0.27$) eliciting more positive responses than T2 ($M = 1.01$, $SD = 0.04$), and stronger responses from [+Pro+Per] ($M = 2.38$, $SD = 0.37$) than [−Pro+Per] ($M = 1.19$, $SD = 0.38$). No significant group by tone condition interaction effect was found ($p = 0.108$). However, pairwise comparisons between groups for the two tones found that the [+Pro+Per] group showed a significantly higher amplitude than the [−Pro+Per] group to T5 [$t(37) = 2.917$, $p = 0.006$,

TABLE V. Averaged amplitude and peak latency of the difference wave (true difference minus dummy difference waves) at the Fz and FCz electrodes in the T1/T2, T1/T5, and T2/T5 conditions for both participant groups.

			[+Pro+Per]	[−Pro+Per]	p
T1/T2	MMN	Mean amplitude (μV)	−3.32	−3.28	0.837
		Peak latency (ms)	131.20	145.05	0.011
	P3a	Mean amplitude	2.72	1.72	0.195
		Peak latency	395.30	408.73	0.482
T1/T5	MMN	Mean amplitude	−3.12	−2.44	0.749
		Peak latency	132.20	137.47	0.401
T2/T5	MMN	Mean amplitude	−0.98	−0.76	0.328
		Peak latency	154.63	154.29	0.971

RIGHTSLINK

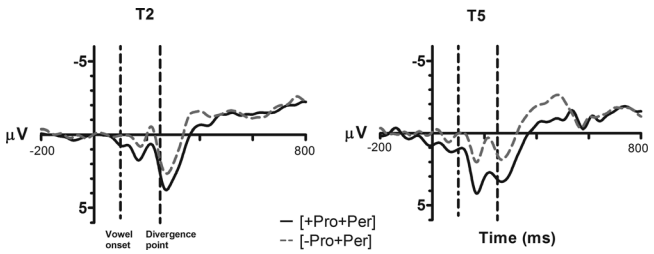

FIG. 2. Averaged ERPs to all occurrences (both standards and deviants) of T2 and T5 at Fz and FCz electrodes for the two participant groups.

Cohen's $d = 0.93$], but no significant difference for T2 ($p = 0.183$).

c. Relationships between production and perception. The T2-T5 production acoustic indices were correlated with three measures of tone perception, i.e., the RTs to trials involving T2 and T5 in the tone discrimination task, and the respective mean amplitude of neural responses to rise time of T2 and T5. The Bonferroni corrected p-value with six correlations was adjusted at 0.0083. As can be seen in Table VI, the T2-T5 F0 offset difference and T2-T5 rise time difference were significantly and negatively correlated with the discrimination RT ($p < 0.008$), with higher production distinction associated with shorter discrimination RTs. Furthermore, both production indices were positively correlated with the mean amplitude of the brain responses to rise time of T5 ($p < 0.008$), the higher the production distinction, the larger the response. The above correlations were further subject to a partial correlation whilst controlling for musical training. Significant negative correlations between RTs and the production indices (F0 offset difference: $r = -0.55$, $p < 0.001$; rise time difference: $r = -0.50$, $p = 0.001$) remained, so did the positive correlations between ERPs to T5 rise time and the production indices (F0 offset difference: $r = -0.55$, $p < 0.001$; rise time difference: $r = 50$, $p = 0.001$), indicating that musical training had little influence on the relationship between perception and production in this study.

IV. DISSCUSION

Utilizing the passive oddball paradigm, the present study compared the neural processes underlying the discrimination of the high rising and low rising tones in Cantonese between two groups of typically developed native speakers differing critically in their production of the rising tone contrast. The design allowed us to gain insights in the online processing of the two highly similar contour tones and to reveal the relationship between tone perception and production. The main findings are the longer discrimination latency demonstrated by the [−Pro+Per] compared with the [+Pro+Per] participants and the significant differences in neural responses between the two participant groups to the subtle acoustic cue of rise time in the amplitude envelope, rather than the pitch contour of the rising tones. More importantly, these perceptual differences are shown to be associated with acoustic differences in producing the two rising tones with respect to F0 offset and amplitude rise time.

TABLE VI. Correlations between T2-T5 production acoustic parameters and perceptual responses to T2 and T5 comparison across participants.

	T2-T5 production—F0 offset difference	T2-T5 production—rise time difference
Discrimination RTs to trials involving T2 and T5	−0.55[a]	−0.48[a]
T5 rise time mean ERP amplitude	0.55[a]	0.48[a]
T2 rise time mean ERP amplitude	0.29	0.34

[a]Significant correlation at Bonferroni-corrected level of 0.0083 (0.05/6).

The behavioral RT finding is convergent with the study by Mok *et al.* (2013), in which the authors showed that Cantonese speakers with reduced pitch differences between T2/T5, T3/T6, and T4/T6 in production were slower in all conditions of a tone discrimination task. The results suggest that the speakers with non-distinctive production demonstrated similar effectiveness of tone discrimination as the speakers with distinctive production but with significantly longer reaction times. Based on the results of auditory transcription of tone productions by participants of this study, the major confusion pattern among speakers with non-distinctive production was T5 produced as T2, which seemed to align with the acoustic analyses. The acoustic forms of T5 were found to approximate those of T2 in terms of F0 offset and amplitude rise time [see Fig. 2(a) and Table III]. For the ERP findings, early components were elicited in the T2/T5 condition from the permutation test for both groups, whereas an early component in the T5/T2 condition was observed only in the [+Pro+Per] group. These results are particularly interesting as temporal cues are generally recognized as secondary or concomitant to the primary cue of F0 in the literature of tone perception (Gandour, 1983; Khouw and Ciocca, 2007). In cognitive frameworks of auditory perception (Holt and Lotto, 2008, 2010), the speech signal encompasses a multitude of acoustic information unfolding over time. Any of the multiple cues may be informative for the identity of a phoneme, but the salience of each cue may vary depending on the context, speaker or other sources of variability. Speech perception is a highly dynamic process in which listeners have to make use of the most informative cues from the acoustic distributions to aid in deriving the sound representations (Holt and Lotto, 2008). In our ERP experiment, the acoustic waveforms of the T2 and T5 stimuli showed that during the segment from vowel onset to pitch divergence point [see Figs. 1(c) and 1(d)], F0 may not be as informative as other acoustic cues, i.e., rise time of the amplitude envelope. Consequently, listeners may extract these robust cues to facilitate the parsing of acoustic signal into tone categories. The present finding corroborates with reports of rise time perception being a significant predictor of phonological awareness of lexical tones in Mandarin Chinese (Goswami *et al.*, 2011).

The positive-going responses from 50 to 150 ms post stimulus onset shown in the present study may seem inconsistent with previous findings that cortical encoding of rise time changes is usually indexed by obligatory N1-P2 responses (or N1b, see Thomson *et al.*, 2009), with fronto-

RIGHTSLINK

central scalp distributions (Carpenter and Shahi, 2013). The N1 and P2 components have been shown to be modulated by varying rise times, specifically, a rise time change from short (30 ms) to long (300 ms) resulting in a latency increase and an amplitude decrease in the components of interest. Thomson *et al*. (2009) measured brain responses to rise time discrimination of synthetic pure tone stimuli (15 ms vs 185 ms) among healthy adults, and found that when serving as standards in the passive oddball task, both stimuli showed negative deflection during 100–150 ms post-stimulus onset, with the stimulus of short rise time consistently inducing greater amplitudes. The discrepancy in terms of the component polarity between the current study and previous work may be due to the use of natural speech stimuli in this study, whereas non-linguistic stimuli were usually employed in previous investigations (e.g., Thomson *et al*., 2009; Goswami *et al*., 2013). Besides the polarity difference, the positive-going component found in this study emerged around 50 ms earlier than those reported previously. The difference in rise time between the T2 and T5 stimuli was 50 ms in the present study, and this difference was already detectable by typically developing English-speaking 12-year-olds (Goswami *et al*., 2013). On the other hand, typically developing Chinese 10-year-olds require a difference around 87 ms (Goswami *et al*., 2011). The early latency observed in this study is perhaps due to more efficient speech processing among adults. Nonetheless, our findings are compatible with previous observations that sounds of shorter rise time (T5 in the present study) are associated with stronger brain responses relative to those of long rise time (T2).

Amplitude rise time is suggested to be an important cue for segmenting speech stream into syllables, onsets or rimes (Goswami *et al*., 2002; Scott, 1998). Difficulties in rise time perception have been proposed to underlie poor phonological processing in children and adults with dyslexia (Goswami, 2011). For instance, the N1 component has been found to show differences between individuals with and without dyslexia as a function of rise time (Hämäläinen *et al*., 2008). The "rise time" hypothesis of developmental dyslexia states that sensitivity to the rhythmic properties of speech, such as those cued by rise time change, may facilitate the development of well-specified phonological representations (Goswami, 2011), which are critical for learning letter-sound correspondences (Snowling, 1981). In the present study, significant group differences were found in brain responses associated with perception of short rise time (i.e., T5); particularly, stronger responses were found in individuals with distinctive T2 and T5 production than those without. Following the rise time hypothesis, higher acuity as reflected in stronger neural responses to rise time exhibited by the [+Pro+Per] participants may result in more distinctive representations in the perceptual space, rendering more efficient (or faster) perceptual discrimination on the one hand, and more distinctive production of these tones on the other hand.

Our findings of the correlation analyses between T2/T5 production distinction and specific perceptual measures lend support to the hypothesis of an association between higher perceptual discrimination of speech sounds and distinctive speech production. The acoustic distinctions of rise time and F0 offset between the two rising tones were associated with how fast the speakers discriminated the tones and the strength of their neural responses to rise time of T5 in particular. More importantly, the two production indices correlated negatively with discrimination latencies and positively with neural responses to rise time. In other words, speakers who are more efficient in discriminating similar speech sounds as well as extracting and encoding rise time information tend to produce the relevant sounds more distinctively. These observations shed light on the mechanism underlying the relationship between speech perception and production.

While the correlation between the acoustic differences in producing rise time and neural responses to the temporal cue indicates a link between speech perception and production, the null difference in the MMN amplitude to the T2/T5 contrast between the participant groups vis-à-vis the non-distinctive production (based on perceptual judgments of listeners) between the rising tones among the [−Pro+Per] participants might be taken to support to a dissociation between the two. However, in light of the stronger neural responses to the temporal cue in the [+Pro+Per] group, we propose that more distinctive perceptual representations resulting from better acuity to temporal information (as discussed above) may associate with more precise acoustic templates (the DIVA model, Guenther, 1995) or sensory targets (Hickok and Poeppel, 2007) of less within-phoneme variability and larger between-phoneme distance when computing a motor program. Thus with respect to the [−Pro+Per] participants, productions of the two rising tones concerning F0 offset, although statistically different from one another [$t(18) = -8.197$, $p < 0.001$], were not of a large enough between-toneme difference to result in distinction of the two tones on the part of the listener. On the whole, despite the apparent dissociation between perception and production exhibited by the [−Pro+Per] group at the behavioral accuracy level, our findings demonstrate that the link between perception and production can be at a more subtle level.

Besides the main findings, several interesting observations in the results deserve further consideration. In the control conditions, the P3a was elicited in T1/T2 but not T1/T5, while MMNs of comparable amplitudes and latencies were elicited. According to one acoustic view of lexical tone processing, the physical acoustic properties of lexical tones would dominate the early perceptual processes, and the linguistic nature of lexical tone might only exert effects at a later stage (Luo *et al*., 2006). In the initial stage, the "sensory-memory-mismatch" MMN detects the pitch height differences between T1 and the two contour tones, which are identical in the two conditions [see Fig. 1(a)]. As the tonal stimuli unfold over time, the linguistic features of the tones may modulate a later ERP component, i.e., P3a. This component is thought to be an index of involuntary orienting to a salient or novel auditory stimulus and it reflects attention switching (Escera and Corral, 2007). More interestingly, the P300, of which P3a is a subcomponent, has been suggested to also index phonological discrimination, with P300 amplitude being greater for deviants that are perceived as phonologically distinct from the standard (Frenck-Mestre *et al*., 2005). The pitch contour change in the high rising tone (T2)

RIGHTSLINK

might be more attention-capturing, thus inducing the P3a in the T1/T2 comparison, whereas the pitch contour feature is a lot less salient in the low rising tone (T5).

Another observation worth considering is the asymmetric pattern of MMN to the contrast between T2 and T5 as revealed by the cluster-based permutation test. Since the acoustic differences between the two stimuli are the same, why would the MMN be elicited in one condition (T2/T5) but not the other (T5/T2)? Previous studies have indeed compared the amplitude and peak latency of MMN elicited in one deviant/standard allocation with the reversed arrangement and obtained similar asymmetric patterns (e.g., German vowels: Eulitz and Lahiri, 2004; Japanese vowels: Ikeda *et al.*, 2002). The memory-based comparison process indexed by the MMN component compares incoming stimuli to representations generated from the repetitive sound sequence (Näätänen, 2001). To explain the occurrence of asymmetric MMN in some German vowels, Eulitz and Lahiri proposed that the standard stimulus in a passive oddball task accesses its underlying (or phonemic) representation, while the deviant corresponds to the surface (or phonetic) representation based on the acoustic signal. However, we are not sure if such an account is applicable to the present finding. The observation nonetheless adds to the extant literature of asymmetric MMN in lexical tone processing, and it would be important to see whether the observation would be replicated in future study.

In conclusion, the present study has investigated the processing of the two contour tones in Cantonese, high rising and low rising tones, and demonstrated that tone perception is highly dynamic and exploits different acoustic cues at different stages of processing—rise time at the sensory/perceptual level and pitch feature at the cognitive level. Moreover, our findings have revealed differential perceptual acuities between individuals with and without distinctive production of these tones as evidenced by the differences in discrimination latency and magnitude of the brain responses to short rise time. Specifically, higher perceptual acuity (as reflected in larger neural responses to rise time) is associated with more distinctive productions. Against the background of an on-going tone merging in HKC, the present investigation makes a good complement to the majority of sociolinguistic research by focusing on the internal perceptual factor, with results demonstrating that changes in neural responses to auditory inputs can be observed even if perception appears prototypical at the behavioral level.

ACKNOWLEDGMENTS

We would like to thank three anonymous reviewers, whose comments and suggestions have been extremely helpful in revising the manuscript. 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"].

AUDACITY (**2015**). http://audacity.sourceforge.net (Last viewed 7/19/2015).

Baken, R. J., and Orlikoff, R. F. (**2000**). *Clinical Measurement of Speech and Voice*, 2nd ed. (Singular Thomson Learning, San Diego), pp. 322–324.

Barry, J. G., and Blamey, P. J. (**2004**). "The acoustic analysis of tone differentiation as a means for assessing tone production in speakers of Cantonese," J. Acoust. Soc. Am. **116**(3), 1739–1748.

Bauer, R. S., Cheung., K. H., and Cheung, P. M. (**2003**). "Variation and merger of the rising tones in Hong Kong Cantonese," Lang. Var. Change **15**(2), 211–225.

Beddor, P. S. (**2012**). "Perception grammars and sound change," in *The Initiation of Sound Change: Perception, Production, and Social Factors*, edited by M. J. Solé and D. Recasens (Benjamins, Amsterdam), pp. 37–55.

Bishop, D. V., Hardiman, M. J., and Barry, J. G. (**2011**). "Is auditory discrimination mature by middle childhood? A study using time-frequency analysis of mismatch responses from 7 years to adulthood," Developmental Sci. **14**(2), 402–416.

Boersma, P., and Weenink, D. (**2015**). "Praat: Doing phonetics by computer," [Computer program], http://www.praat.org/ (Last viewed 7/19/2015).

Carpenter, A. L., and Shahin, A. J. (**2013**). "Development of the N1–P2 auditory evoked response to amplitude rise time and rate of formant transition of speech sounds," Neurosci. Lett. **544**, 56–61.

Chandrasekaran, B., Krishnan, A., and Gandour, J. T. (**2007**). "Mismatch negativity to pitch contours is influenced by language experience," Brain Res. **1128**, 148–156.

Chao, Y. R. (**1930**). "A system of tone letters," Le Maitre Phonet. **45**, 24–27.

Damasio, H., and Damasio, A. R. (**1980**). "The anatomical basis of conduction aphasia," Brain **103**, 337–350.

Díaz, B., Baus, C., Escera, C., Costa, A., and Sebastián-Gallés, N. (**2008**). "Brain potentials to native phoneme discrimination reveal the origin of individual differences in learning the sounds of a second language," Proc. Natl. Acad. Sci. U.S.A. **42**, 16083–16088.

Escera, C., and Corral, M. J. (**2007**). "Role of mismatch negativity and novelty-P3 in involuntary auditory attention," J. Psychophysiol. **21**(3–4), 251–264.

Eulitz, C., and Lahiri, A. (**2004**). "Neurobiological evidence for abstract phonological representations in the mental lexicon during speech recognition," J. Cogn. Neurosci. **16**(4), 577–583.

FIELDTRIP (**2015**). http://fieldtrip.fcdonders.nl/ (Last viewed 7/19/2015).

Frenck-Mestre, C., Meunier, C., Espesser, R., Daffner, K., and Holcomb, P. (**2005**). "Perceiving nonnative vowels: The effect of context on perception as evidenced by event-related brain potentials," J. Speech Lang. Hear. Res. **48**(6), 1496–1510.

Fu, Q. J., and Zeng, F. G. (**2000**). "Identification of temporal envelope cues in Chinese tone recognition," Asia Pacific J. Speech Lang. Hear. **5**(1), 45–57.

Gandour, J. (**1983**). "Tone perception in far eastern-languages," J. Phonetics **11**(2), 149–175.

Gandour, J., Potisuk, S., and Dechongkit, S. (**1994**). "Tonal coarticulation in Thai," J. Phonetics **22**(4), 477–492.

Goswami, U. (**2011**). "A temporal sampling framework for developmental dyslexia," Trends Cogn. Sci. **15**(1), 3–10.

Goswami, U., Huss, M., Mead, N., Fosker, T., and Verney, J. P. (**2013**). "Perception of patterns of musical beat distribution in phonological developmental dyslexia: Significant longitudinal relations with word reading and reading comprehension," Cortex **49**(5), 1363–1376.

Goswami, U., Thomson, J., Richardson, U., Stainthorp, R., Hughes, D., Rosen, S., and Scott, S. K. (**2002**). "Amplitude envelope onsets and developmental dyslexia: A new hypothesis," P. Natl. Acad. Sci. U.S.A. **99**(16), 10911–10916.

Goswami, U., Wang, H. L. S., Cruz, A., Fosker, T., Mead, N., and Huss, M. (**2011**). "Language- universal sensory deficits in developmental dyslexia: English, Spanish, and Chinese," J. Cogn. Neurosci. **23**(2), 325–337.

Greenberg, S. (**2006**). "A multi-tier framework for understanding spoken language," in *Understanding Speech: An Auditory Perspective*, edited by S. Greenberg and W. Ainsworth (LEA, Mahweh, NJ), pp. 411–433.

Guenther, F. H. (**1995**). "Speech sound acquisition, coarticulation, and rate effects in a neural network model of speech production," Psychol. Rev. **102**(3), 594–621.

Hämäläinen, J. A., Leppänen, P. H. T., Guttorm, T. K., and Lyytinen, H. (**2008**). "Event-related potentials to pitch and rise time change in children with reading disabilities and typically reading children," Clin. Neurophysiol. **119**(1), 100–115.

Hickok, G., and Poeppel, D. (**2007**). "The cortical organization of speech processing," Nat. Rev. Neurosci. **8**(5), 393–402.

RIGHTSLINK

Holt, L. L., and Lotto, A. J. (**2008**). "Speech perception within an auditory cognitive science framework," Curr. Dir. Psychol. Sci. **17**(1), 42–46.
Holt, L. L., and Lotto, A. J. (**2010**). "Speech perception as categorization," Atten. Percept. Psycho. **72**(5), 1218–1227.
Ikeda, K., Hayashi, A., Hashimoto, S., Otomo, K., and Kanno, A. (**2002**). "Asymmetrical mismatch negativity in humans as determined by phonetic but not physical difference," Neurosci. Lett. **321**(3), 133–136.
Johnson, K. (**2006**). "Resonance in an exemplar-based lexicon: The emergence of social identity and phonology," J. Phonetics **34**(4), 485–499.
Khouw, E., and Ciocca, V. (**2007**). "Perceptual correlates of Cantonese tones," J. Phonetics **35**(1), 104–117.
Kong, Y. Y., and Zeng, F. G. (**2006**). "Temporal and spectral cues in Mandarin tone recognition," J. Acoust. Soc. Am. **120**(5), 2830–2840.
Kuhl, P. K., Conboy, B. T., Coffey-Corina, S., Padden, D., Rivera-Gaxiola, M., and Nelson, T. (**2008**). "Phonetic learning as a pathway to language: New data and native language magnet theory expanded (NLM-e)," Philos. Trans. R. Soc. London B **363**(1493), 979–1000.
Law, S. P., Fung, R., and 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., and Tong, M. C. (**2015**). "Lexical tone perception in native speakers of Cantonese," Intl. J. Speech Lang. Pathol. **17**(1), 53–62.
Levelt, W. J. (**1999**). "Models of word production," Trends Cogn. Sci. **3**(6), 223–232.
Liberman, A., and Mattingly, I. (**1985**). "The motor theory of speech perception revised," Cognition **21**, 1–36.
Lotto, A. J., Hickok, G. S., and Holt, L. L. (**2009**). "Reflections on mirror neurons and speech perception," Trends Cogn. Sci. **13**(3), 110–114.
Luo, H., Ni, J. T., Li, Z. H., Li, X. O., Zhang, D. R., Zeng, F. G., and Chen, L. (**2006**). "Opposite patterns of hemisphere dominance for early auditory processing of lexical tones and consonants," P. Natl. Acad. Sci. **103**(51), 19558–19563.
Luo, H., and Poeppel, D. (**2012**). "Cortical oscillations in auditory perception and speech: Evidence for two temporal windows in human auditory cortex," Front. Psychol. **3**, 170.
Macmillan, N. A., and Creelman, C. D. (**1991**). *Detection Theory: A User's Guide* (Cambridge University Press, New York), pp. 147–152.
Maris, E., and Oostenveld, R. (**2007**). "Nonparametric statistical testing of EEG- and MEG-data," J. Neurosci. Meth. **164**(1), 177–190.
May, P. J., and Tiitinen, H. (**2010**). "Mismatch negativity (MMN), the deviance-elicited auditory deflection, explained," Psychophysiology **47**(1), 66–122.
Mok, P. P., Zuo, D., and Wong, P. W. (**2013**). "Production and perception of a sound change in progress: Tone merging in Hong Kong Cantonese," Lang. Var. Change **25**(03), 341–370.
Näätänen, R. (**1990**). "The role of attention in auditory information processing as revealed by event-related potentials and other brain measures of cognitive function," Behav. Brain Sci. **13**(02), 201–233.
Näätänen, R. (**2001**). "The perception of speech sounds by the human brain as reflected by the mismatch negativity (MMN) and its magnetic equivalent (MMNm)," Psychophysiology **38**, 1–21.
Näätänen, R., Paavilainen, P., Rinne, T., and Alho, K. (**2007**). "The mismatch negativity (MMN) in basic research of central auditory processing: A review," Clin. Neurophysiol. **118**(12), 2544–2590.
Polich, J. (**2003**). "Overview of P3a and P3b," in *Detection of Change: Event- Related Potential and fMRI Findings*, edited by J. Polich (Kluwer, Boston, MA), pp. 83–98.
presentation (**2015**). www.neurobs.com (Last viewed 7/19/2015).
Rose, P. (**1987**). "Some considerations in the normalization of the fundamental frequency of linguistic tone," Speech Commun. **6**(4), 343–352.
Rosen, S. (**1992**). "Temporal information in speech: Acoustic, auditory and linguistic aspects," Philos. Trans. R. Soc. London B **336**(1278), 367–373.
Scott, S. K. (**1998**). "The point of P-centres," Psycho. Res. **61**(1), 4–11.
Snowling, M. J. (**1981**). "Phonemic deficits in developmental dyslexia," Psychol. Res **43**(2), 219–234.
Strait, D. L., and Kraus, N. (**2011**). "Can you hear me now? Musical training shapes functional brain networks for selective auditory attention and hearing speech in noise," Front. Psychol. **2**, 113.
Tarr, E. (**2013**). "Processing perceptually important temporal and spectral characteristics of speech," Doctoral dissertation, The Ohio State University, Columbus, Ohio.
Thomson, J. M., Goswami, U., and Baldeweg, T. (**2009**). "The ERP signature of sound rise time changes," Brain Res. **1254**, 74–83.
Tsang, Y. K., Jia, S., Huang, J., and Chen, H. C. (**2011**). "ERP correlates of pre-attentive processing of Cantonese lexical tones: The effects of pitch contour and pitch height," Neurosci. Lett. **487**(3), 268–272.
Vance, T. J. (**1976**). "An experimental investigation of tone and intonation in Cantonese," Phonetica **33**(5), 368–392.
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, New York, NY.

RIGHTSLINK