Brain & Language 173 (2017) 67–75

Contents lists available at ScienceDirect

Brain & Language

journal homepage: www.elsevier.com/locate/b&l

Phonological experience modulates voice discrimination: Evidence from functional brain networks analysis

Xueping Hu [a,1], Xiangpeng Wang [a,1], Yan Gu [a], Pei Luo [a], Shouhang Yin [a], Lijun Wang [a], Chao Fu [c], Lei Qiao [a], Yi Du [b,*], Antao Chen [a,*]

[a] Key Laboratory of Cognition and Personality of Ministry of Education, Faculty of Psychology, Southwest University, Chongqing 400715, China
[b] CAS Key Laboratory of Behavioral Science, Institute of Psychology, Chinese Academy of Sciences, Beijing 100101, China
[c] School of Psychology, Northwest Normal University, Lanzhou 730070, China

ARTICLE INFO

Article history:
Received 14 October 2016
Revised 19 April 2017
Accepted 3 June 2017
Available online 27 June 2017

Keywords:
Voice discrimination
Language familiarity effect
Phonological experience
Cognitive control
Large-scale brain network

ABSTRACT

Numerous behavioral studies have found a modulation effect of phonological experience on voice discrimination. However, the neural substrates underpinning this phenomenon are poorly understood. Here we manipulated language familiarity to test the hypothesis that phonological experience affects voice discrimination via mediating the engagement of multiple perceptual and cognitive resources. The results showed that during voice discrimination, the activation of several prefrontal regions was modulated by language familiarity. More importantly, the same effect was observed concerning the functional connectivity from the fronto-parietal network to the voice-identity network (VIN), and from the default mode network to the VIN. Our findings indicate that phonological experience could bias the recruitment of cognitive control and information retrieval/comparison processes during voice discrimination. Therefore, the study unravels the neural substrates subserving the modulation effect of phonological experience on voice discrimination, and provides new insights into studying voice discrimination from the perspective of network interactions.

1. Introduction

Human voice discrimination is defined as the ability to judge whether two voices are similar or different based on matching of basic auditory parameters (Van Lancker & Kreiman, 1987). Developed early in newborns, the ability is essential for human communication and social interactions (Belin, Fecteau, & Bedard, 2004; Belin, Zatorre, Lafaille, Ahad, & Pike, 2000; Latinus, McAleer, Bestelmeyer, & Belin, 2013). As a voice-based process, voice discrimination is influenced by speaker variability, and in addition, language familiarity (Fleming, Giordano, Caldara, & Belin, 2014). In particular, listeners are better at identifying voices in their familiar language than in an unfamiliar one. This phenomenon is captured by the language familiarity effect (LFE) (Fleming et al., 2014; Thompson, 1987; Winters, Levi, & Pisoni, 2008). Since firstly described by Thompson (1987), the LFE has attracted a large amount of attention and been confirmed in bilinguals (Fleming et al., 2014; Thompson, 1987; Winters et al., 2008), infants (Johnson, Westrek, Nazzi, & Cutler, 2011), and even in dyslexic, receptive aphasic and phonagnosia patients (Perrachione, Del Tufo, & Gabrieli, 2011; Van Lancker, Cummings, Kreiman, & Dobkin, 1988).

Concerning the LFE, a main point comes into controversy is whether the effect is based on comprehension of linguistic information or familiarity with the phonological structure (Fleming et al., 2014). While no conclusion has been made, the study by Fleming et al. (2014) indicated that the phonological aspects of language ability only could influence voice discrimination. In particular, the authors employed time-reversed speech that excludes the influence of linguistic comprehension but retains acoustic cues important for voice identification (Johnson et al., 2011; Van Lancker, Kreiman, & Emmorey, 1985). They found that both English and Chinese groups were more sensitive to voice in their native language than in their non-native language. The observation is obviously interesting; however, since behavioral data alone could do little to provide further evidence, the mechanism subserving

* Corresponding authors at: CAS Key Laboratory of Behavioral Science, Institute of Psychology, Chinese Academy of Sciences, 16 Lincui Road, Chaoyang District, Beijing, China, (Y. Du). Key Laboratory of Cognition and Personality (Ministry of Education), Faculty of Psychology, Southwest University, No. 2, Tiansheng Road, Beibei, Chongqing, China (A. Chen).

E-mail addresses: duyi@psych.ac.cn (Y. Du), xscat@swu.edu.cn (A. Chen).

[1] X. P. Hu and X. P. Wang contributed equally to this work.

http://dx.doi.org/10.1016/j.bandl.2017.06.001
0093-934X/

the modulation of phonological experience on voice discrimination remains unclear.

In a typical voice discrimination process, listeners should shift attention from irrelevant information to goal-directed information and map the acoustic patterns onto phonetic idiosyncrasies of the talker (Francis, Baldwin, & Nusbaum, 2000; Magnuson & Nusbaum, 2007; Nusbaum & Magnuson, 1997). Then listeners get accesses to the "acoustical voice space" to track the acoustical-phonetic imprint when hearing the first voice stimuli (Andics et al., 2010; Latinus, Crabbe, & Belin, 2011). After that, by detecting the differences between the first and second voices, listeners make a judgement about whether the two voices are produced by the same speaker or not (Van Lancker & Kreiman, 1987). Therefore, in addition to acoustic processing involved in voice discrimination, various cognitive processes (e.g., working memory, selective attention) should also be recruited.

In the current study, we sought to explore the modulation of phonological experience on voice discrimination from the perspective of neural substrates. The materials we used were time-reversed speech sentences similar to those used in Fleming et al. (2014). Meanwhile, phonological experience in the current study was indexed by language familiarity, and we employed three different languages: native language (Mandarin Chinese) and non-native languages (English, second language; Korean, unfamiliar language) to construct a gradient change. During fMRI scanning, participants were asked to perform a same/different voice discrimination task across different language contexts. The current study first examined the regional activity in response to the task. In addition, since voice discrimination is based on the processing of multi-source information, we hypothesized that network-based neural interactions are key elements subserving the phenomenon. In particular, we examined several networks that might be recruited in voice discrimination in an exploratory way. To specify, there were a voice acoustic network and a voice identity network (Andics et al., 2010; Blank, Wieland, & von Kriegstein, 2014; Latinus et al., 2011) sensitive to a general voice-identity processing; a default mode network associated with person-identity information (Arnott, Heywood, Kentridge, & Goodale, 2008; Blank et al., 2014; Simmons, Reddish, Bellgowan, & Martin, 2010); a salience network engaged in detecting salience stimuli (Menon, 2011); and a fronto-parietal network responsible for attentional control (Menon, 2011; Power et al., 2011). Neural interactions in the current study were calculated using a multi-regional psychophysiological interaction (PPI) modeling analysis. And this method has been demonstrated as effective in investigating dynamic functional connectivity between regions/networks (Cocchi, Zalesky, Fornito, & Mattingley, 2013; Cocchi et al., 2014; Gerchen, Bernal-Casas, & Kirsch, 2014). We predicted that typical brain regions of voice processing could be activated, along with control-related systems. Importantly, regional activity and neural interactions could operate as a function of phonological experience.

2. Methods

2.1. Participants

Thirty-six native Mandarin Chinese volunteers from Southwest University participated in the current study (four males; M = 21.11 ys, SD = 0.23). All participants are seniors of English majors without any experience of Korean, and have learned English for more than 10 years (M = 10.71, SD = 0.292). All participants were right-handed with normal hearing, and reported no neurological or psychiatric disorders. Written informed consents were obtained from all participants before scanning. To evaluate the language proficiency of English, all subjects finished an English language proficiency test (the Transparent Language Proficiency Test: http://www.transparent.com/) and the Bilingual Switching Questionnaire test (BSWQ) (Rodriguez-Fornells, Krämer, Lorenzo-Seva, Festman, & Münte, 2012). The mean score of language proficiency test was 84.92 ± 1.27 (% of correct responses), and that of the BSWQ was 9.91 ± 0.24 for L1S (switch to Chinese), 7.71 ± 0.29 for L2S (switch to English), 9.03 ± 0.31 for CS (contextual switch), 8.46 ± 0.31 for US (unintended switch) and 35.11 ± 0.80 for OS (overall switch), respectively. More information about the BSWQ could be find in the research of Rodriguez-Fornells et al. (2012). These scores suggest that the L2 proficiency of volunteers is above average. The study was approved by the Human Ethics Committee from Southwest University, China.

2.2. Stimuli

The testing stimuli were selected from two speech corpuses and then edited. The Mandarin Chinese and Korean stimuli were selected from OSCAAR (The Online Speech/Corpora Archive and Analysis Resource, https://oscaar.ci.northwestern.edu/index.html), the English speech stimuli were selected from the PN/NC corpus (McCloy et al., 2013). In order to avoid influence from paralinguistic information of voices, such as gender (Belin et al., 2004; Pisanski, Cartei, McGettigan, Raine, & Reby, 2016), the speech stimuli included six sentences from five native male speakers in each language respectively. Recordings were sampled at 16 bit and 22.05 kHz, time-reversed (Fig. 1B), and normalized for root mean square (RMS) amplitude to 70 dB SPL. The average duration of speech recordings is 1777.47 ms (SD = 115.79 ms) for Mandarin Chinese, 1856.17 ms (SD = 106.86 ms) for English and 1760.60 ms (SD = 123.98 ms) for Korean, respectively. All stimuli were edited using Adobe Audition 3.0 and Praat (Boersma, 2001; Boersma & Weenink, 2015). The detailed acoustic features of stimuli are listed in Table S1.

2.3. Experimental procedures

For each language, the recording of the six sentences by five native male speakers resulted in 30 time-reversed speech stimuli. For each language, the same time-reversed sentence by the same speaker was repeated twice and formed the same condition (SC) of 30 speech pairs; while for the different condition (DC), the same time-reversed sentence by two distinct speakers was presented consecutively, resulting 60 kinds of speech pairs. To balance the number of the stimuli, the final materials included 60 pairs of "same" stimuli and 60 pairs of "different" stimuli for each language across three runs. Each run contained 40 pairs of stimuli with an equal proportion of the "same" and "different" trials. In total, the experiment consisted of 360 trials. The order of language runs was counterbalanced across subjects. To avoid the consecutive presentation of the same speaker's speeches, the sequence of the stimuli within each language was pseudo-randomized.

Stimuli were presented binaurally at a comfortable intensity level using MR-compatible headphones. Each trial started with a yellow speaker icon lasted for 2000 ms (during which the first speech was presented), followed by a 500 ms blank screen. After that, a second yellow speaker icon was displayed for 4000 ms, during which the second speech was presented and subjects were told to discriminate whether the two voices were the same or different and make responses by pressing the "left" key for "same" trials and the "right" key for "different" trials on a two-button pad using in the scanner. The inter-trial-interval (ITI) was jittered from 1 s to 3 s (2 s in average).