Hearing Research 331 (2016) 109–118

Contents lists available at ScienceDirect

Hearing Research

journal homepage: www.elsevier.com/locate/heares

Research paper

Mandarin Chinese vowel-plus-tone identification in noise: Effects of language experience

Mingshuang Li [a, b], Wenjing Wang [a, b, **], Sha Tao [a, b], Qi Dong [a, b], Jingjing Guan [c], Chang Liu [c, *]

[a] State Key Laboratory of Cognitive Neuroscience and Learning, IDG/McGovern Institute for Brain Research, Beijing Normal University, Beijing, 100875, China
[b] National Innovation Center for Assessment of Basic Education Quality, Beijing Normal University, Beijing, 100875, China
[c] Department of Communication Sciences and Disorders, University of Texas at Austin, 1 University Station A1100, Austin, TX, 78712, USA

ARTICLE INFO

Article history:
Received 16 July 2015
Received in revised form
9 November 2015
Accepted 10 November 2015
Available online 30 November 2015

Keywords:
Mandarin vowel-plus-tone identification
Native Chinese listeners
Native English experience
Noise processing

ABSTRACT

Several studies found better English vowel identification in English multi-talker babble (MTB) and temporally-modulated (TM) noise, but not in quiet condition for native Chinese listeners in the US (CNU) with the US residency of 1–3 years than native Chinese listeners in China (CNC) with no residency history in English speaking countries. Two possible explanations were proposed: (1) CNU listeners used temporal dips of noise more efficiently than CNC listeners; and (2) CNU listeners had less informational masking of MTB than their CNC peers. The current study explored whether the difference in noise processing between CNU and CNC listeners was also presented for their native speech perception. Chinese vowel-plus-tone identifications were measured for CNU and CNC in quiet, stationary and TM noise, babble-modulated noise, and MTB. The identification scores of CNU listeners were significantly higher than CNC listeners in most noisy backgrounds, whereas both groups had the same performance in quiet. Moreover, compared with CNC listeners, CNU listeners gained greater masking releases from the temporal modulation in TM noise at low SNRs, whereas no significant difference was found in informational masking between the two groups. In conclusion, the native English experience may improve native Chinese listeners' capacity to use temporal glimpses in TM noise, possibly depending on the modulation frequency and depth, while it may not improve their ability to resist the informational masking of babble when perceiving their native speech.

1. Introduction

The effect of the second language (L2) on the native language (L1) as a transfer effect has become a hotly discussed issue in the field of second-language acquisition. Many studies have demonstrated this effect in the level of phonology, morphosyntax, lexis, semantics, pragmatics, rhetoric, and conceptual representations (for reviews, see Pavlenko, 2000; Cook, 2003; Kecskes, 2008); however, there is a lack of research investigations on how L2 learning affects speech perception in L1, particularly under noise conditions. Although a number of studies have extensively examined speech perception for non-native listeners in a variety of listening conditions (Cutler et al., 2004; Cooke et al., 2008; Broersma and Scharenborg, 2010; Stuart et al., 2010; Mi et al., 2013; Guan et al., 2015; for a review, see Garcia Lecumberri et al., 2010), few studies have focused on the effect of L2 learning experience on the perception of native speech. Given that native English experiences changed native Chinese listeners' English speech perception in noise (Mi et al., 2013; Guan et al., 2015), one primary goal of the present study was to explore whether native English experience also influenced native Chinese listeners' Mandarin Chinese speech recognition in quiet and noise conditions.

Previous work in our laboratories found that one to three years of native English experience for native Chinese listeners might significantly enhance their English vowel identification in English multi-talker babble (MTB) (Mi et al., 2013). Native Chinese listeners in the US (CNU) and China (CNC) had comparable performance of English vowel identification in quiet and long-term speech-shaped noise, while the CNU group outperformed the CNC group in English

* Corresponding author.
** Corresponding author.
E-mail address: changliu@utexas.edu (C. Liu).

http://dx.doi.org/10.1016/j.heares.2015.11.007
0378-5955/

MTB, especially at low signal-to-noise ratios (Mi et al., 2013). Two possible explanations were proposed: first, the ability to use temporal dips might be improved for CNU listeners who had regular and extensive English exposures. In general, temporal dips of dynamic noise are used in auditory processing to gain lower noise masking and a higher signal-to-noise ratio (SNR), thereby resulting in better speech recognition compared with stationary noise (Peters et al., 1998; Cooke, 2006). This effect was named as masking release from temporal dips of noise (Miller, 1947; Miler and Licklider, 1950; Festen and Plomp, 1990). The masking release was found not only in native listeners (Festen and Plomp, 1990; Festen, 1993; Bacon et al., 1998; Peters et al., 1998), but also in non-native listeners (Broersma and Scharenborg, 2010; Stuart et al., 2010; Guan et al., 2015). Stuart et al. (2010) reported that when both groups of listeners perceived their native speech, native American English listeners took better advantage of temporal dips in noise than native Chinese listeners. Another recent study by Guan et al. (2015) suggested that CNU listeners with one to three years of the US residency showed significantly greater masking release from the temporal dips of noise for the task of English vowel identification than CNC listeners who had no residency history in English-speaking countries. Such differences between CNU and CNC listeners in the masking release from the temporal dips of noise may be associated with the difference in the temporal structures of English and Chinese speech, e.g., Mandarin Chinese speech had less temporal modulations (e.g., smaller temporal dips) than English speech (Calandruccio et al., 2010).

The second explanation is that native English speech experience might help CNU listeners have less informational masking than CNC listeners. Informational masking was commonly equated with non-energetic and central masking, which occurs beyond the energetic masking at the cochlear level (Durlach et al., 2003). It has been demonstrated that the amount of informational masking is influenced by the similarity of the target and the masker (Brungart, 2001; Van Engen, 2010) and by the listeners' familiarity of the multi-talker babble (Van Engen, 2010). For example, the informational masking was higher when the target and masking babble were in the same language than when they were in different languages (Van Engen and Bradlow, 2007; Calandruccio et al., 2010; Van Engen, 2010; Gautreau et al., 2013; Calandruccio and Zhou, 2014). In addition, many studies found a higher native advantage for native listeners in speech recognition under babble noise than under quiet condition, likely because non-native listeners suffered more from the informational masking of babble (Mayo et al., 1997; Cutler et al., 2004; Garcia Lecumberri and Cooke, 2006; Cooke et al., 2008; Mi et al., 2013). In addition, Mi et al. (2013) suggested that the native English experience might acclimate CNU listeners to English speech environments such that they distinguished English speech stimuli from English babble noise more easily, resulting in less informational masking than CNC listeners.

The studies above on the L2 speech perception in noise suggested that the native English experience appeared to result in CNU listeners' advantages of noise processing, i.e., the better use of temporal dips in noise and greater resistance to informational masking of babble, rather than improved their English phonetic processing (Mi et al., 2013; Guan et al., 2015). It is therefore hypothesized in the present study that the benefits of noise processing may be also presented in L1 speech perception. Thus, another goal of this study was then to investigate whether the two improvements of noise processing could be observed in Chinese speech identification.

Mandarin Chinese is a tonal language with four lexical tones based on F0 contours: level (tone 1), rising (tone 2), dipping (tone 3), and falling (tone 4; Chao, 1968). For example, the syllable [ma] means “mother” with tone 1, “hemp” with tone 2, “horse” with tone 3, and “scold” with tone 4. On the other hand, Mandarin Chinese has a sparse vowel system with six monophthongs: three high vowels, /i, y, u/, two mid-vowels, /ɤ, ə/, and one low vowel, /ɑ/ (Duanmu, 2008). Different from speech perception in non-tone languages, speech recognition in tone languages such as Mandarin Chinese relies on not only phonemic perception, but also tone recognition. In the present study, vowel-plus-tone perception was measured for native Mandarin Chinese listeners.

Two experiments were conducted for CNU and CNC listeners: (1) Chinese vowel-plus-tone identification in stationary and temporally modulated (TM) noise, and (2) Chinese vowel-plus-tone identification in English and Chinese multi-talker babble (MTB) and the corresponding babble-modulated long-term speech-shaped noise (LTSSN). For the first experiment, the masking release was computed as the difference in the identification scores between stationary and TM noise for the two groups of native Chinese listeners. In the second experiment, the amount of informational masking was calculated as the difference of the identification score between babble-modulated LTSSN (e.g., English LTSSN and Chinese LTSSN) and MTB (e.g., English multi-talker babble and Chinese multi-talker babble), respectively. Based on the hypothesis that the 1–3 year residency in the US might help CNU listeners have better noise processing such as a more efficient use of temporal dips in noise and greater resistance to the informational masking of MTB, it was expected that CNU listeners would have a greater masking release from the temporal dips of noise and show less informational masking of MTB than CNC listeners.

2. Experiment 1: vowel-plus-tone identification in quiet, stationary, and temporally modulated noise

2.1. Methods

2.1.1. Participants

Two groups of native Mandarin Chinese listeners (CNC and CNU) participated in the study. Each group comprised 14 participants aged 18–28 years old. All listeners had normal hearing with pure-tone thresholds $\leq$15 dB HL at octave intervals between 250 and 8000 Hz (ANSI, 2010). Both groups of listeners started their school-based English education from 9 to 13 years old. The CNU group consisted of graduate or undergraduate students at the University of Texas at Austin (UT) with the TOEFL score higher than 80, while participants in the CNC group were undergraduate or graduate students at Beijing Normal University (BNU) whose College English Test Band 4 (CET-4) authorized by the Ministry of Education of China were between 470 and 560. The CET-4 is required for undergraduate students to receive a baccalaureate degree in most universities at China. Listeners in both groups were from northern China, speaking fluent Mandarin Chinese – putonghua. CNU listeners entered the US at 18–26 years old and resided in the US for one to three years, while CNC listeners had no residency history in English-speaking countries.

2.1.2. Speech stimuli and noise

Five mandarin Chinese vowels, /ɑ, ɤ, i, u, y/ combined with the four tones (i.e., a total of 20 stimuli) served as speech signals. They were originally produced by a young female native speaker of Mandarin Chinese and the duration of the vowels plus tones varied from 236 ms to 772 ms.

Two types of noise were used as the masker: stationary and temporally modulated (TM) noise. The stationary noise was generated by applying a filter with an average spectrum of six-talker Mandarin babble on Gaussian noise. The TM noise was created by applying a temporally modulated envelope on the stationary noise. The modulation phase was randomized between