Title

Influence of language experience on digit recognition by English and Chinese listeners

Journal

The Journal of the Acoustical Society of America, 138(3)

ISSN

0001-4966

Authors

Zhou, Xiaoqing
Yuan, Wei
Galvin, John J
et al.

Publication Date

2015-09-01

DOI

10.1121/1.4929617

Zhou *et al.*: JASA Express Letters [http://dx.doi.org/10.1121/1.4929617] Published Online 22 September 2015

Influence of language experience on digit recognition by English and Chinese listeners

Xiaoqing Zhou and Wei Yuan[a)]
Department of Otolaryngology, Southwest Hospital, Third Military Medical University, Gao Tan Yan Street, Shaping Ba District, Chongqing, 400038, China
tracyonly0603@sina.com, weiyuan175@sina.com

John J. Galvin and Qian-Jie Fu
Department of Head and Neck Surgery, David Geffen School of Medicine, University of California Los Angeles, Los Angeles, California 90095, USA
jgalvin@mednet.ucla.edu, qfu@mednet.ucla.edu

Ying Zhang
Department of Otolaryngology, First Affiliated Hospital of Kunming Medical University, Xichang road no. 295, Kunming, Yunnan, 650032, China
zhangy627@126.com

Abstract: Digit recognition was measured in quiet and in two noise conditions by English-native (EN) and Chinese-native (CN) listeners. EN listeners were tested using English digits and CN listeners were tested using both English and Chinese digits. In quiet, forward digit span recall worsened for both groups as the number of digits was increased. Significant effects of language experience were observed with five or more digits. Language experience had a significant effect on digit recognition in babble but not in steady noise. These results suggest that understanding of a nonnative language can be influenced by both cognitive load and listening environment.

[CC]
Date Received: February 25, 2015 **Date Accepted:** August 11, 2015

1. Introduction

Although bilingual communication is quite common, speech understanding with the nonnative, second language (L2) is often poorer than that with the native, first language (L1). For L2 listeners, everyday communication can be difficult, due to challenging acoustic environments (e.g., noise, interfering talkers, reverberation, etc.) and unfamiliarity with common expressions. Thus, non-native listeners in the "real" world must overcome both "imperfect" signals and knowledge of L2. Under adverse listening conditions, L2 listeners are more susceptible to interfering noise than are L1 listeners, regardless of speech stimulus type (phoneme, word, or sentence) (Rogers *et al.*, 2006; Cooke *et al.*, 2008; Shi, 2009; Broersma and Scharenborg, 2010; Garcia Lecumberri *et al.*, 2010; Jin and Liu, 2012). Noise can interfere with speech by overlapping the target speech spectrum ("energetic masking") and/or by presenting temporal information that is similar to the target temporal envelope ("informational masking"). Informational masking may occur even when the target and masker are spectrally remote (i.e., minimal energetic masking). Competing speech may contain both energetic and informational masking, depending on the voice characteristics of the talkers, the coincidence of temporal envelope information, and the linguistic content of the competing speech. While both L1 and L2 listeners may be similarly affected by energetic masking, L2 listeners may more adversely affected by informational masking, due to less familiarity with the language (Cooke *et al.*, 2008).

In previous studies, recognition of native and non-native speech is not only affected by the listening environment, but also by the speech stimulus used for testing (e.g., phonemes, words, sentences, etc.) (Garcia Lecumberri *et al.*, 2010). Digits are an important component of language as many everyday transactions involve numbers and are familiar to both native and nonnative listeners. Digits are also highly intelligible compared with other speech materials (Oba *et al.*, 2011, 2013). As such, digits may be very useful stimuli to test L1 and L2 speech perception testing under difficult listening conditions.

[a)] Author to whom correspondence should be addressed.

Mandarin Chinese and English are very different languages, as Mandarin Chinese is a tonal language and English is not. We hypothesized that, for bilingual listeners, digit recognition under difficult listening conditions may be affected by listeners' native language. To test this hypothesis, we measured forward digit span recall in quiet, as well as recognition of a fixed number of digits in two types of noise: (1) steady, speech-shaped noise (SSN) and (2) multi-talker speech babble. Chinese native (CN) subjects were tested while listening to L1 and L2 digits; as a control condition, English native (EN) subjects were tested while listening to L1 digits only. In quiet, we hypothesized that the cognitive load associated with short-term memory would be influenced by listeners' native language. In noise, we hypothesized that native language would have a greater influence with the babble noise due to the greater informational masking.

2. Method

2.1 Subjects

Ten EN and ten CN subjects participated in the study. All subjects were ≥18 years of age and all had normal hearing with pure-tone thresholds ≤15 dB hearing level at octave intervals between 250 and 8000 Hz. All subjects were adults who had reached at least the undergraduate college education level. All CN subjects had resided in the United States for less than three years and all had Test of English as a Foreign Language (TOEFL) scores of at least 213. All CN subjects reported speaking, reading, and writing English with excellent proficiency in terms of daily communication. Exclusion criteria included organic brain diseases and other physical or mental illness that could lead to cognitive impairment.

2.2 Digit stimuli

Ten English and Mandarin Chinese digits (0, 1, 2, 3, 4, 5, 6, 7, 8, and 9) were used as speech materials. The corresponding IPA symbols for the 10 Mandarin Chinese digits are shown as follows: liŋˊ(0), iˉ(1), ərˋ(2), sanˉ(3), sɿˋ(4), uˇ(5), liouˋ(6), ʨʻiˉ(7), pʌˉ(8), ʨiouˇ(9). English digits were produced by 1 English-speaking male talker and Chinese digits were produced by 1 Chinese-speaking male talker.

2.3 Procedure

All testing was conducted in sound field; subjects were seated in a double-walled sound-treated booth directly facing a single loudspeaker (Tannoy Reveal) 1 m away. In quiet and in noise, digit stimuli were presented at 65 dBA. All testing was conducted using custom software developed at House Research Institute.

For forward digit span in quiet, percent correct recognition of two, three, five, seven, or nine digit sequences was measured. Thus, a two-digit sequence might be 2-5, a five-digit sequence might be 4-2-17-4, and a nine-digit sequence might be 3-5-2-4-1-0-7-9-8-6. Performance for all digit sequences (two, three, five, seven, or nine digits) was measured within the same test block, with 20 presentations of each sequence condition randomly ordered during testing. One to three blocks were tested for each subject, depending on the subject's availability. During each trial, digits were randomly selected for each position in the sequence. After playback, the subject clicked on response buttons (labeled 0, 1, 2, 3, 4, 5, 6, 7, 8, and 9) in the order of the digits heard. No feedback was presented. EN subjects were tested using EN digits only. CN subjects were tested using both CN and EN digits; EN and CN digit recognition was measured separately. Before formal testing, subjects were provided with a brief training session using digits in their native language to familiarize them with the test procedure.

For digit recognition in noise, an adaptive procedure was used in which the signal-to-noise ratio (SNR) was adjusted from trial to trial according to the correctness of response, converging on the digit recognition threshold (DRT), defined as the SNR that produced 50% correct recognition of digits in noise. The target digit sequence consisted of 3 digits. DRTs were measured in two types of noise: (1) steady, speech-shaped noise or (2) six-talker speech babble. The steady noise was expected to produce energetic masking, while babble was expected to produce a combination of energetic and informational masking which might interact with subjects' native language. During each trial of the test, digits were randomly selected for each position in the sequence. The digits were presented at 65 dBA and the noise was adjusted in terms of long-term RMS according to the target SNR. After playback, the subject clicked on response buttons (labeled 0, 1, 2, 3, 4, 5, 6, 7, 8, and 9) in the order of the digits heard. If the subject answered correctly, the SNR was reduced by 2 dB; if the subject answered incorrectly, the SNR was increased by 2 dB. No feedback was provided. For each target digit sequence and noise condition, the DRT was estimated as the SNR that produced 50% correct digit recognition over a 25-trial block. One to three blocks were tested for each subject, depending on the subject's availability.

3. Results

Because subjects completed 1 – 3 test blocks depending on their availability, the data were analyzed for subjects and conditions in which multiple test blocks were completed. For EN subjects listening to L1 who completed three test blocks, a one-way repeated measures analysis of variance (RM ANOVA) was performed on the data, with test run (1, 2, or 3) as the factor and subject/condition as the repeated measure. Results showed no significant effect of test run [$F(15,30) = 0.202$; $p = 0.818$]. Because of a non-normal distribution, a one-way RM ANOVA was performed on ranked data for EN subjects who completed only two test blocks, with test run (1 or 2) as the factor and subject/condition as the repeated measure. Results showed no significant effect of test run (Chi-square = 0.348; $p = 0.555$). For CN subjects listening to L1 who completed three test blocks, a one-way RM ANOVA showed no significant effect of test run [$F(6,12) = 0.649$; $p = 0.540$]. Similarly, for CN subjects listening to L2 who completed three test blocks, a one-way RM ANOVA showed no significant effect of test run [$F(4,8) = 2.209$; $p = 0.172$]. For CN subjects listening to L1 who completed only two test blocks, a one-way RM ANOVA showed no significant effect of test run [$F(6,12) = 0.649$; $p = 0.540$]. Similarly, for CN subjects listening to L2 who completed three test blocks, a one-way RM ANOVA showed no significant effect of test run [$F(4,8) = 2.209$; $p = 0.172$]. Because of a non-normal distribution, a one-way RM ANOVA was performed on ranked data for CN subjects listening to L1 who completed only two test blocks. Results showed no significant effect of test run (Chi-square = 1.125; $p = 0.289$). Similarly, one-way RM ANOVA was performed on ranked data for CN subjects listening to L2 who completed only two test blocks. Results showed no significant effect of test run (Chi-square = 0.444; $p = 0.505$). Because there were no learning effects observed across test runs, data for each subject were averaged across runs and these mean data were used for subsequent analysis.

3.1 Forward digit span in quiet

Figure 1 shows mean percent correct digit recognition scores for the two-, three-, five-, seven-, and nine-digit sequences. For both EN and CN subjects, performance worsened as the number of digits in the sequence increased. For CN subjects, performance with L2 dropped more rapidly as a function of the number of digits, compared to performance with L1. Threshold was derived from sigmoid fits to the data, as was defined as the number of digits needed to produce 50% correct performance. For EN subjects listening to L1, the threshold was 7.8 digits. For CN subjects listening to L1, the threshold was 8.5 digits. For CN subjects listening to L2, the threshold was 6.3 digits. To analyze differences across listener groups and across sequence conditions, a split-plot RM ANOVA was performed on the L1 data, with native language (EN or CN) as the between-subject factor and the number of digits in the sequence (2, 3, 5, 7, or 9) as the within-subject factor. Results showed a significant effect for the number of digits in the sequence [$F(4,72) = 115.1$, $p < 0.001$], but no significant effect for native language [$F(1,18) = 3.5$, $p = 0.430$]; there was no significant interaction [$F(4,72) = 2.1$, $p = 0.094$].

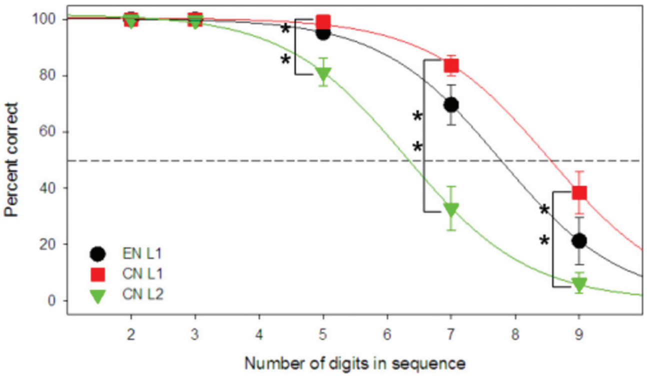

Fig. 1. Mean percent correct digit recognition as a function of the number of digits in the sequence for EN subjects listening to L1 (circles) and CN subjects listening to L1 (squares) or L2 (triangles). The lines show sigmoid fits to the data. The dashed line shows 50% correct. The error bars show the standard error. Asterisks indicate a statistically significant difference between L1 and L2 ($p < 0.05$).

To analyze native language effects across sequence conditions, a two-way RM ANOVA was performed on the CN data, with language (L1 or L2) and the number of digits in the sequence as factors. Results showed significant effects for language [$F(1,36) = 103.9$, $p < 0.001$] and the number of digits in the sequence [$F(1,36) = 160.0$, $p < 0.001$]; there was a significant interaction [$F(4,36) = 23.7$, $p < 0.001$]. *Post hoc* Bonferroni pairwise comparisons showed significant differences between L1 and L2 only for the five-, seven-, and nine-digit sequences ($p < 0.05$ in all cases). *Post hoc* Bonferroni pairwise comparisons also showed that, for L1, performance was significantly poorer for the nine-digit sequence compared with all other sequences ($p < 0.05$ in all cases); for L2, performance was significantly poorer for the seven- and nine-digit sequences compared with all other sequences, and poorer for the five-digit sequence compared with the two- and three-digit sequences ($p < 0.05$ in all cases).

Figure 2 shows mean digit recognition thresholds (DRTs) in noise as a function of listener group (left panel) or noise type (right panel). In general, there were only small differences in mean DRTs with SSN among listening groups (EN L1: −14.7; CN L1: −13.0; CN L2: −14.0). With babble, mean DRTs were poorer for CN L2 (−11.2) than for CN L1 (14.6) or EN L1 (−13.9). To analyze differences across listener groups and across noise types, a split-plot RM ANOVA was performed on the L1 data, with native language (EN or CN) as the between-subject factor and noise type (SSN or babble). Results showed no significant effects for noise type [$F(1,18) = 1.4$, $p = 0.257$] or native language [$F(1,18) = 0.6$, $p = 0.461$], but there was a significant interaction with native language [$F(1,18) = 11.1$, $p = 0.004$]. To analyze native language effects across noise conditions, a two-way RM ANVOA was performed on the CN data, with language (L1 or L2) and noise type as factors. Results showed no significant effects for language [$F(1,9) = 4.6$, $p = 0.061$] or noise type [$F(1,9) = 2.4$, $p = 0.158$]; there was a significant interaction [$F(1,9) = 20.7$, $p = 0.001$]. *Post hoc* Bonferroni pairwise comparisons showed significant differences between L1 and L2 only for speech babble ($p < 0.05$), and between SSN and speech babble only for L2 ($p < 0.05$).

4. Discussion

The present results suggest that cognitive load and listening environment can significantly affect listeners' understanding of a nonnative language. Under less challenging listening conditions (i.e., recognition of 2−3 digits in quiet, recognition of 3 digits in SSN), there was little difference in performance between L1 and L2 for the present CN subjects. This result is consistent with previous bilingual speech perception studies in quiet that report similar L1 and L2 performance (Rogers *et al.*, 2006; Cooke *et al.*, 2008; Stuart *et al.*, 2010); When the cognitive load was increased (i.e., recognition of 5−9 digits in quiet) or when the listening environment became more complex (recognition of three digits in babble), strong differences in performance between L1 and L2 were observed. This result was consistent with our hypothesis that challenging listening conditions would elicit differences between L1 and L2.

The present CN subjects had been living in the USA for several years and were skillful at everyday English communication according to self-reports. The present

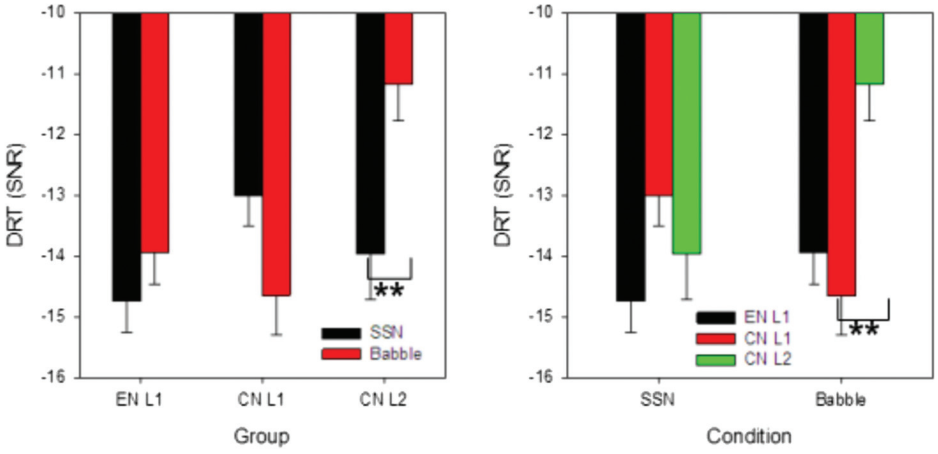

Fig. 2. Left panel: Mean DRTs in noise (in dB SNR) for SSN (black bars) and babble (red bars), as a function of listening group and target language. The error bars show the standard error. Asterisks indicate a statistically significant difference between SSN and Babble ($p < 0.05$). Right panel: the same data in left panel, but plotted as a function of noise type for EN L1 (black bars), CN L1 (red bars), and CN L2 (green bars). Asterisks indicate a statistically significant difference between L1 and L2 ($p < 0.05$).

study measured digit recognition using very simple stimuli. With more complex materials (e.g., difficult sentences), understanding of L2 most likely would be similarly difficult as observed with the present challenging conditions. Indeed, digit recognition in noise has been significantly correlated with sentence recognition in noise for EN subjects (Oba *et al.*, 2011).

There was no significant difference between the present EN and CN groups in terms of L1 perception, suggesting that the present digit recognition tasks were appropriate and not L1-dependent. Different from previous studies (Rogers *et al.*, 2006; Cooke *et al.*, 2008; Shi, 2009; Broersma and Scharenborg, 2010; Garcia Lecumberri *et al.*, 2010; Jin and Liu, 2012; Jin and Liu, 2014), we found a significant difference between CN L1 and L2 perception with only with interfering babble, not with SSN. Consistent with previous studies (Rogers *et al.*, 2006; Cooke *et al.*, 2008; Shi, 2009; Broersma and Scharenborg, 2010; Garcia Lecumberri *et al.*, 2010; Jin and Liu, 2012), the difference in performance between L1 and L2 was greater for babble than for SSN. Interestingly, a slight release from masking (babble DRT − SSN DRT) was only observed for L1 in CN subjects; L1 performance with babble slightly worsened for EN subjects, relative to SSN. For L2, the CN mean DRT worsened by almost 3 dB relative to SSN. Given the somewhat short duration of the digit stimuli, one might expect less release from masking with dynamic noise than found for longer sentence materials. In Mandarin Chinese, each digit corresponds to just one syllable while English digits may contain one or two syllables. It is possible that the CN subjects applied the rhythmic segmentation rule (each syllable corresponds to one digit) of their native language (Mandarin) to process English language (Cutler *et al.*, 1992; Rogers *et al.*, 2006; Mi *et al.*, 2013). For L1, CN subjects may have been able to use this segmentation and better attend to the target digits; for L2, attention may have been distributed more equally to the target and babble, given less certainty regarding L2 speech patterns. As such, the present CN subjects experienced more interference, rather than a release from masking with the dynamic babble noise for L2.

Acknowledgments

This study was partly supported by the Department of Otolaryngology in Southwest Hospital at Third Military Medical University in China (W.Y.) and NIDCD-R01-DC-004993 (Q.-J.F.).

References and links

Broersma, M., and Scharenborg, O. (**2010**). "Native and non-native listeners' perception of English consonants in different types of noise," Speech Comm. **52**, 980–995.

Cooke, M., Garcia Lecumberri, M. L., and Barker, J. (**2008**). "The foreign language cocktail party problem: Energetic and informational masking effects in non-native speech perception," J. Acoust. Soc. Am. **123**, 414–427.

Cutler, A., Mehler, J., Norris, D., and Segui, J. (**1992**). "The monolingual nature of speech segmentation by bilinguals," Cognit. Psychol. **24**, 381–410.

Garcia Lecumberri, M. L., Cooke, M., and Cutler, A. (**2010**). "Non-native speech perception in adverse conditions," Speech Commun. **52**, 864–886.

Jin, S. H., and Liu, C. (**2012**). "English sentence recognition in speech-shaped noise and multi-talker babble for English-, Chinese-, and Korean-native listeners," J. Acoust. Soc. Am. **132**, EL391–EL397.

Jin, S. H., and Liu, C. (**2014**). "English vowel identification in quiet and noise: Effects of listeners' native language background," Front Neurosci. **8**, 305.

Mi, L., Tao, S., Wang, W. J., and Dong, Q. (**2013**). "English vowel identification in long-term speech-shaped noise and multi-talker babble for English and Chinese listeners," J. Acoust. Soc. Am. **133**, EL391–EL397.

Oba, S. I., Fu. Q.-J., and Galvin, J. J. III (**2011**). "Digit training in noise can improve cochlear implant users' speech understanding in noise," Ear Hear. **32**, 573–581.

Oba, S. I., Fu. Q.-J., and Galvin, J. J. III (**2013**). "Minimal effects of visual memory training on auditory performance of adult cochlear implant users," J. Rehabil. Res. Dev. **50**, 99–110.

Rogers, C. L., Lister, J. J., Febo, D. M., Besing, J. M., and Abrams, H. B. (**2006**). "Effects of bilingualism, noise, and reverberation on speech perception by listeners with normal hearing," Appl. Psycholinguist. **27**, 465–485.

Shi, L. F. (**2009**). "Normal-hearing English-as-a-second-language listeners' recognition of English words in competing signals," Int. J. Audiol. **48**, 260–270.

Stuart, A., Zhang, J. L., and Swink, S. (**2010**). "Reception thresholds for sentences in quiet and noise for monolingual English and bilingual Mandarin-English listeners," J. Am. Acad. Audiol. **21**, 239–248.