EFFECTS OF NOISE ON LEXICAL TONE PERCEPTION BY NATIVE AND NON-NATIVE LISTENERS

Tasha M. Dees[1], *Ann R. Bradlow*[2], *Sumitrajit Dhar*[1], *Patrick C.M. Wong*[1,3]

[1]The Roxelyn and Richard Pepper Department of Communication Sciences and Disorders, Northwestern University, U.S.A.
[2]Department of Linguistics, Northwestern University, U.S.A.
[3]Northwestern University Interdepartmental Neuroscience Program, Northwestern University, U.S.A.
pwong@northwestern.edu

ABSTRACT

Speech in ecological environments is often accompanied by noise, which often proposes challenges to listeners. In this study we examined differences in native Mandarin-speaking and native English-speaking adults' ability to perceive lexical tones and vowels in white noise and multi-talker babble noise at different signal-to-noise ratios. We found that although white noise impaired both listener groups' speech perception ability, Mandarin multi-talker babble noise impaired Mandarin listeners to a greater extent. These results provide evidence for the differential effects of energetic and informational masking on native and non-native speech perception and the perceptual consequence of the reorganization of the auditory system after native language learning.

1. INTRODUCTION

Speech is more difficult to discern in noise than in quiet, and differences in masking type (energetic and informational) influence the level of difficulty [3,9]. Recently, some studies have shown that competing speech in a language foreign to the listener does not mask as well as competing speech in the listener's native language and that foreign speech is more difficult to discern in difficult listening situations than native speech [3, 9].

However, these studies did not systematically investigate the influence of white noise, native babble, and non-native babble on perception of native and non-native phonemic contrasts, nor did they investigate the influence of noise on suprasegmental (lexical tone), such as those used in Mandarin Chinese (/ma/ spoken in a level or rising tone can mean "mother" or "hemp," respectively).

Considerations of different types of noise and their effects on native and non-native phonemic contrasts, both segmental and suprasegmental, will provide valuable information for developing a more comprehensive theory of speech perception in ecologically relevant listening situations, as well as for developing more robust automatic speech recognition systems.

The present study examined how white (energetic masking) and Mandarin multi-talker babble (informational masking) noise impair native Mandarin- and English-speaking listeners' ability to perceive lexical tones and vowels. We hypothesize that although energetic masking impairs speech perception in both groups, informational masking will result in additional listening difficulty for the Mandarin listeners, despite that fact that lexical tones were learned early in life.

2. EXPERIMENT 1

2.1. Method

2.1.1. Subjects

Twelve native English-speaking adults (20-25 years, mean=23) who had no previous experience with any tone language, and 12 native Mandarin-speaking adults (20-30 years, mean=26) participated in this study. Mandarin subjects were individuals born in a country where Mandarin Chinese is an official language, for whom Mandarin Chinese was the first language learned while growing up. None reported having begun to use English on a daily basis before 12 years of age.

2.1.2 Materials

Four native Mandarin speakers who did not serve as subjects (listeners) produced /a/, /i/, and /u/ (these vowels occur in both Mandarin and English). These originally produced tokens were then resynthesized to include three variants differing only in pitch patterns (fundamental frequency). The three pitch patterns, resembling Mandarin Tones 1, 2, and 4, were interpolated linearly through the voiced portion of each stimulus using Pitch-Synchronous Overlap and Add (PSOLA) from Praat [2]. The pitch values are based on Shih [7]. Similar resynthesis procedures were used in Wong & Perrachione [11].

All stimuli were embedded in two levels of white noise, generated in Audacity [6], resulting in signal-to-noise ratios (SNR) of -15 dB and -25 dB. The RMS amplitude of the signal and the noise were normalized using Level16 [8]. The duration of the white noise was identical to the stimuli.

2.1.3. Procedures

Each participant performed two three-alternative forced-choice identification (ID) tasks in the two levels of masking noise. In the tone identification tasks (Tone ID), all participants were shown a picture of a level, rising, and falling arrow; played an example of each tone; and instructed to choose the arrow corresponding to the tone they heard throughout the task. In the vowel identification task (Vowel ID), participants were shown a picture of the vowels "a," "i," and "u" with example words containing those vowels in their native language and instructed to choose the letter corresponding to the sound they heard throughout the task.

The order of the tasks (Vowel ID or Tone ID) was counterbalanced across subjects. For each task, the higher SNR condition (the less noisy condition) was always performed first.

2.2. Results and Discussion

Data were analyzed with a 2x2 (SNR x Group) repeated measures ANOVA. Accuracy results are summarized in Figures 1 and 2.

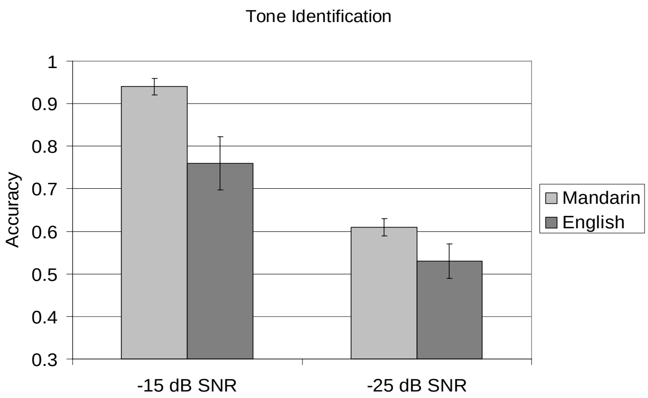

Figure 1. Accuracy of Tone ID by Mandarin and English listeners in white noise.

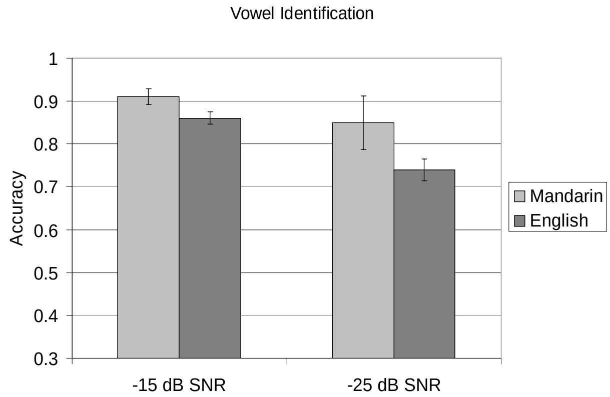

Figure 2. Accuracy of Vowel ID of English and Mandarin listeners in white noise.

In Tone ID, there was a main effect of noise level in accuracy [$F(1,22)=146.63$, $p<.001$], which showed that both groups were impaired by the lower SNR. There was a main effect of group [$F(1,22)=6.020$, $p=.023$], where native English speakers identified tones less accurately than Mandarin speakers, and a significant interaction effect [$F(1,22)=4.47$, $p=.046$]. Reaction time data in Tone ID showed no exchange of speed for accuracy. There was no main effect of noise level, though this value was approaching significance [F(1,22)=3.387, p=.079]. There was no main effect of group [F(1,22)=.249, p=.623], though there was a marginally significant interaction [F(1,22)=3.463, p=.076]. These data suggest that Mandarin speakers identified tones more quickly and accurately than English speakers due to their native language listening advantage.

In Vowel ID, there was a main effect of noise level in accuracy [$F(1,22)=10.243$, $p=.004$]. There was neither a main effect of group [$F(1,22)$, $p=.073$] nor a significant interaction [$F(1,22)=1.654$, $p=.212$], which indicated that when phonemes that two linguistic groups perceive as native were imbedded in white noise, the white noise did not differentially affect perception ability. Reaction time data in Vowel ID showed no exchange of speed for accuracy. There was no main effect of noise level [F(1,22)=.086, p=.772] nor was there a significant interaction [F(1,22)=.401, p=.533], but there was a marginal main effect of group [F(1,22)=3.925, p=.060]. These data showed that, in general, Mandarin speakers identified vowels slightly more quickly and accurately than native English speakers due to the use of Mandarin-produced vowels.

While Experiment 1 examined the effects of energetic masking on lexical tone and vowel perception and showed impairments across groups, it did not consider how informational masking may affect the perception of native lexical tone contrasts, which will be addressed in Experiment 2.

3. EXPERIMENT 2

3.1. Methods

3.1.1. Participants

Twenty-four adults who did not participate in Experiment 1, including 12 native English speakers (18-24 years, mean=20) and 12 native Mandarin speakers (18-32 years, mean=24), participated in this experiment.

3.1.2. Materials and Procedures

The target vowels with pitch patterns used in this experiment were the same as those in Experiment 1. These stimuli were embedded in Mandarin Chinese 6-talker babble taken from Van Engen & Bradlow [9]. The noise was created by asking 6 (3 male and 3 female) native speakers of Mandarin Chinese to provide recordings in Chinese.

The duration of the noise was edited to match the target vowels. The noise was RMS amplitude normalized, as in Experiment 1, resulting in SNRs of -5 dB and -20 dB. The SNR was unlike that of Experiment 1 to avoid a performance ceiling effect. The procedures were identical to Experiment 1.

3.1. Results and Discussion

Data were analyzed with a 2x2 (SNR x Group) repeated measures ANOVA. Accuracy results are summarized in Figures 3 and 4.

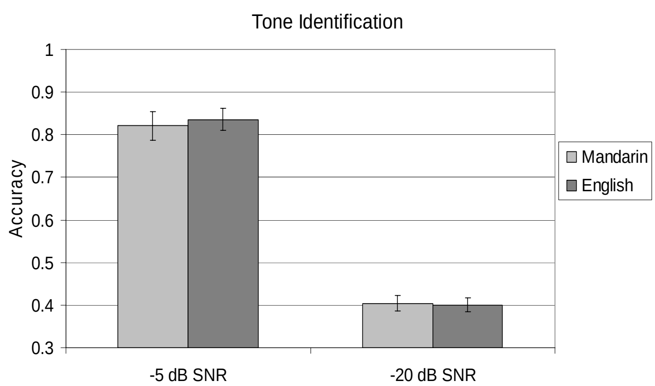

Figure 3. Accuracy of Tone ID by Mandarin and English listeners in Mandarin babble.

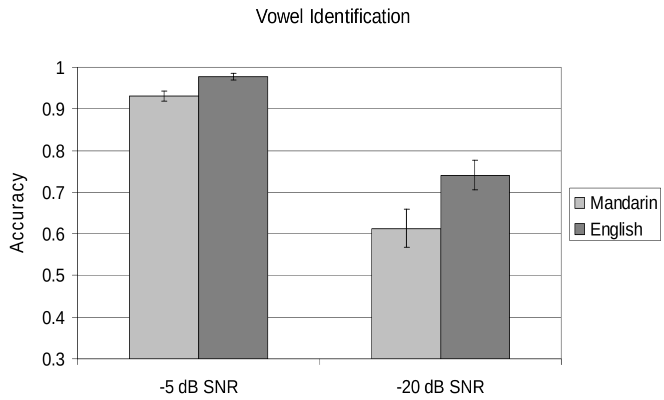

Figure 4. Accuracy of Vowel ID by Mandarin and English listeners in Mandarin babble.

In Tone ID, there was a main effect of noise level in accuracy [$F(1,22)=602.074$, $p<.001$], which showed that both linguistic groups were impaired by the lower SNR. There was not a main effect of group [$F(1,22)=.039$, $p=.846$] nor a significant interaction effect [$F(1,22)=.264$, $p=.612$], even though native Mandarin speakers have surpassed native English speakers in lexical tone identification in quiet [1] and in white noise, as is shown in Experiment 1. Reaction time data in Tone ID showed no exchange of speed for accuracy. There was a main effect of noise level [$F(1,22)=5.471$, $p=.029$], but no main effect of group [$F(1,22)=.094$, $p=.762$] and no interaction [$F(1,22)=.334$, $p=.569$]. These data suggest that Mandarin speakers were further impaired by Mandarin babble noise than English speakers even when the target phonemes were foreign to the English speakers.

In Vowel ID, there was a main effect of noise level [$F(1,22)=95.097$, $p<.001$], and a main effect of group [$F(1,22)=7.724$, $p=.011$] in accuracy. Native English speakers identified vowels with higher accuracy in both high and low SNR than Mandarin speakers. There was no significant interaction [$F(1,22)=2.030$, $p=.168$]. Reaction time data in Vowel ID showed no exchange of speed for accuracy. There was a main effect of noise level [$F(1,22)=34.135$, $p<.001$], but no significant interaction [$F(1,22)=1.378$, $p=.253$] or main effect of group [$F(1,22)=.575$, $p=.456$]. These data suggest that Mandarin speakers were further impaired by Mandarin babble noise than English speakers.

4. GENERAL DISCUSSION

Our results support and extend previous findings that listeners are less disrupted by competing speech in a foreign language than in their native language [2,5], providing evidence that native multi-talker babble differentially affects the perception of native and non-native phonetic contrasts. Notably, Van Engen and Bradlow [9] found that native and foreign 6-talker babble, do not differentially affect the perception of native speech. This discrepancy may be due to the fact that Van Engen and Bradlow's task provided the listener with more opportunity to perceive redundant cues, making the task easier.

Our findings may be explained by theories which suggest that learning the native language drives the reorganization of the auditory perceptual system [10]. The reorganization specifically involves tuning the brain in order to facilitate with more efficient processing of native speech sounds, which explain why adults, unlike infants, do not display the ability to identify foreign language phonemes with as much accuracy as a native speakers. This same perceptual bias has been shown in infants [5] and adults [4] with tones.

In the context of these studies, our results may show that, though an individual's auditory perception may exhibit enhanced processing under relatively good conditions, and though those enhancements are robust enough to persevere in many difficult listening situations, once a listener's auditory perceptual system is exposed to sufficient sensory and cognitive loads (as in the case of native babble) these processing advantages are diminished, or may disappear entirely. The listener is then left with the perceptual abilities of the non-native, un-tuned auditory system. This framework may explain the diminished performance by Mandarin speakers as compared to English speakers when identifying vowels in Mandarin babble. The Mandarin speakers' systems were inundated with high sensory and cognitive loads due to native babble and began to revert to un-tuned performance. The English speakers' systems were inundated with less sensory and cognitive loads and they retained their native efficiency tuning.

The discrepancy between our present study and the work by Garcia Lecumberri and Cooke [3] when examining non-native phonemes in babble may be due to the fact that the 0 dB SNR used in the Garcia Lecumberri and Cooke study did not expose the system of the native speakers to sufficient sensory and cognitive loads to cause un-tuned performance.

This framework may also be used to explain the interaction effect between SNR and linguistic group in tone identification in white noise. The low SNR, but not the high SNR, was disruptive enough to interfere with the efficient native phoneme processing of Mandarin speakers.

The same line of reasoning could also be used to explain why Mandarin speakers identified tones in Mandarin babble with the same level of accuracy as native English speakers. The additional interference of native babble may have added sufficient sensory and cognitive loads to the auditory systems of the Mandarin listeners to interfere with their native tuning.

Future research in this field should continue to catalog the differential effects of energetic and informational masking on varying levels of native and non-native speech. In light of the explanation proposed to account for the results of this study, future research should also attempt to discover neural correlates of cross-linguistic perception of speech in native and non-native background babble.

5. REFERENCES

[1] Bent, T., Bradlow, A., & Wright, B. (2006). The influence of linguistic experience on the cognitive processing of pitch in speech and nonspeech sounds. *J Exp Psychol: Human, 32*(1), 97-103.

[2] Boersma, P., & Weenink, D. (2004). Praat Version 4.126.

[3] Garcia Lecumberri, M., Cooke, M. (2006). Effect of masker type on native and non-native consonant perception in noise. *JASA, 119*, 2445-2454.

[4] Krishnan, A., Xu, Y., Gandour, J., & Caiani, P. (2005). Encoding of pitch in the human brainstem is sensitive to language experience. *Cognitive Brain Res, 25*, 161-168.

[5] Mattock, K. & Burnham, D. (2006). Chinese and English infants' tone perception: Evidence for perceptual reorganization. *Infancy, 10*(3), 241-265.

[6] Mazzoni, D. (2006). Audacity 1.2.6.

[7] Shih, C-L. (1988). Tone and intonation in Mandarin. *Working Papers of the Cornell Phonetics Laboratory, 3*, 83-109.

[8] Tice, R. & Carrel, T. (1998). Level16 (Version 2.0.3): University of Nebraska.

[9] Van Engen, K. & Bradlow, A. (2007). Sentence recognition in native- and foreignlanguage multi-talker background noise. *JASA, 121*(1), 519-526.

[10] Werker, J. & Tees, R. (2005). Speech Perception as a Window for Understanding Plasticity and Commitment in Language Systems of the Brain. *Dev Psychobiol, 46*(3), 233-251.

[11] Wong, P., & Perrachione, T. (in press). Learning pitch patterns in lexical identification by native English-speaking adults. *Appl Psycholinguist.*

This work is supported by the National Institutes of Health (U.S.A.) grants HD051827 & DC007468 awarded to Patrick Wong who is the corresponding author.