Perceived Spatial Separation Induced by the Precedence Effect Releases Chinese Speech From Informational Masking

Jing Chen[1], Chun Wang[1], Hongwei Qu[1], Wenrui Li[1], Yanhong Wu[1], Xihong Wu[1], Bruce A. Schneider[2], Liang Li[1,2]

[1]National Key Laboratory on Machine Perception, Speech and Hearing Research Center, Center for Brain and Cognitive Sciences, Department of Psychology, Peking University, Beijing, China, 100871. chenj@cis.pku.edu.cn
[2]Centre for Research on Biological Communication Systems, Department of Psychology, University of Toronto at Mississauga, Mississauga, Ontario, Canada L5L 1C6

INTRODUCTION

Spatial separation of the source of signal sound from the source of interference sound improves the recognition of the signal (Zurek et al., 1993). In reverberant environments, listeners usually "fuse" the direct sound wave from a source with its reflections and perceive a single event as originating from the direction of the source. This phenomenon is called the precedence effect (Litovsky, 1999). Using the perceptual consequence of the precedence effect (Freyman et al., 1999) induced a perceived separation of images of target and masking stimuli, and found a much larger advantage of the perceived spatial separation in English speech recognition when masking stimuli were informational than energetic.

Compared to English, Chinese syllables have more voiceless consonants and less voiced consonants. Chinese syllable initials would be more vulnerable to energetic masking. It has been reported that intelligibility of Chinese speech is considerably worse than that of English speech under conditions with noise masking (Kang, 1998). On the other hand, standard Chinese (also called Mandarin) is a type of tonal language and the pitch pattern of a Chinese single syllable is lexically meaningful. Fox and Unkefer (1985) reported that the perception of tones in Chinese is affected by the lexical status of the speech token. These features in Chinese syllables would induce a distinct pattern of informational masking that does not exit in English, when masking stimuli are Chinese speeches.

In the present study, we used the precedence effect to induce perceived spatial separation of target Chinese speech from either informational or energetic maskers. Instead of a single-sized separation within one semifield (within perceptual channel) as used in the study by Freyman et al., we also investigated the advantages of perceived spatial separation with two different sizes (45° and 90°).

MATERIALS AND METHODS

Target speech stimuli were Chinese "nonsense" sentences spoken by a young female talker. These sentences are syntactically correct but not meaningful. In each of the target sentences there are 3 key words that are scored during speech recognition testing. Target sentences were presented by both the right and the left loudspeakers with the right speaker leading the left speaker by 3 ms. Listeners perceived the target sentence images as coming from the right side. The single-source presentation level of the target sentences was fixed at 54 dB SPL.

There were two types of masking stimuli: noise and speech. The noise masker was steady speech-spectrum noise made from Chinese speech. The speech masker was a recording of continuous Chinese nonsense sentences spoken by other two young female talkers and the contexts were different from target speech. There were 3 perceived locations for the masking stimuli: right, central and right. There were 4 single-source intensity levels for both the speech masking and the noise masking stimuli: 66, 62, 58, and 54 dB SPL, which corresponded to the 4 signal/noise ratios: -12, -8, -4, and 0 dB, respectively.

RESULTS

The percent-correct speech rrecognition across 12 listeners for 2 masking conditions and 3 perceived locations of each masking condition are presented in Figure 1. Under noise masking conditions, the location effect on threshold was not quite significant, $F(2, 22) = 3.430$, $MSE = 1.898$, $p = 0.051$. Under speech masking conditions, the location effect on threshold was significant, $F(2, 22) = 15.896$, $MSE = 2.697$, $p = 0.00$. Pairwise comparisons indicated that the perceived left and central locations of speech maskers did not differ from one another ($p = 1.000$) but both locations differ significantly from the right location ($p = 0.000$, $p = 0.003$, respectively).

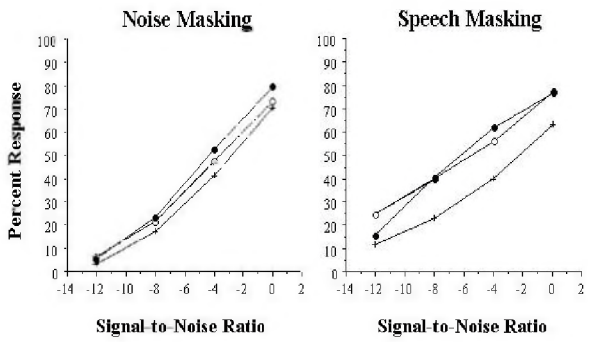

Figure 1. Symbols: crosses, perceived right location; filled circles, perceived center location; open circles, perceived left location.

DISCUSSION

The present study used Chinese nonsense sentences as speech signals and obtained results that are comparable to those reported from Freyman et al's study. When the masker was noise, the improvement of recognition of nonsense Chinese speech was minor (1dB), event though a large perceived spatial separation (45° or 90°) was induced by the precedence effect. When the masker was nonsense speech, the perceived spatial separation of target speeches from masking speeches markedly improved recognition of the target speeches. The analyses of psychometrical function reveal a 3.3-dB improvement. This improvement is somehow smaller than those (4-9 dB) reported by Freyman et al. For Chinese speeches, recognition of syllable initials is critical to recognition of the associated words. Since initials have broad spectrums like those of noises, Chinese words would be more vulnerable to energetic masking than English words (Kang, 1998). Also, perception of tones of syllable finals in Chinese is closely linked to lexical meaning, which may provide listeners with additional cues to connect syllables in target speeches across time. In spite of these characteristics of Chinese speeches, the results indicate that advantage of perceived separation in releasing speeches from masking can extend to tonal Chinese. Interestingly, the perceived separation across hemifields (perceived 90° separation) produced equivalent unmasking effect, compared to the perceived separation within hemifield (perceived 45° separation). These results are not consistent with the prediction of the two-channel model (Boehnke et al. 1999) since the perceived across-channel masking (perceived 90° separation) did not cause less masking than the perceived within-channel masking (perceived 45° separation). A prospective explanation of this disagreement is that the two spatial channels are more associated with processing for locations of sound sources, while perceived locations of "fused" image induced by the precedence effect are more related to higher-order processing.

At this moment it is not clear why under speech masking conditions the perceived-spatial-separation advantage obtained here is smaller than those reported by Freyman et al. In the future, at least two cross-language issues should be addressed: (1) whether Chinese and English speeches have different or similar vulnerabilities to noise masking, and (2) whether Chinese and English masking speeches have different or similar interference effects on recognition of either Chinese or English speeches.

REFERENCES

Zurek, P. M., 1993. Binaural advantages and directional effects in speech intelligibility. In: Studebaker, G.A.,

Litovsky, R. Y., Colburn, H. S., Yost, W. A., Guzman, S. J., 1999. The precedence effect. J. Acoust. Soc. Am. 106, 1633-1654.

Freyman, R.L., Helfer, K.S., McCall, D.D., Clifton, R.K.,1999. The role of perceived spatial separation in the unmasking of speech. J. Acoust. Soc. Am. 106, 3578-3588.

Kang, J., 1998. Comparison of speech intelligibility between English and Chinese. J. Acoust. Soc. Am. 103,1213-1216.

Fox, R.A, Unkefer, J., 1985. The effect of lexical status on the perception of tone + Chinese phonology. J. Chinese Linguist.13, 69-90.

Boehnke, S.E., Phillips, D.P., 1999. Azimuthal tuning of human perceptual channels for sound location. J. Acous. Soc. Am.106, 1948-1955.