Available online 13 November 2004

e ct of perceived spatial separation, induced by p nce from noi speech masking was inves-
g esearch 199 1–10

et al., 1999, 2001; Arbogast et al., 2002; Brungart, 2001; Brungart and Simpson, 2002; Kidd et al., 1994, 1998).

However, when both the signal source and masking source are speech, the speech masker may interfere with the processing of the speech target because both may activate linguistic and semantic systems involved in speech recognition and language comprehension. Hence a speech masker can interfere with the perception and recognition of the targeted speech at both peripheral and central (cognitive) levels. In the literature, any central level interference resulting from stimulus (speech or non-speech sound) uncertainty is referred to as informational masking (Arbogast et al., 2002; Brungart, 2001; Brungart and Simpson, 2002; Durlach et al., 2003; Freyman et al., 1999, 2001; Kidd et al., 1994, 1998).

It is di cult, however, to assess the relative contribution of these two types of masking. Theoretically, if one could equate a speech masker to a non-speech masker with respect to all peripherally-significant acoustic properties, then any di erences in target recognition between these two types of maskers would reflect the contribution of informational masking. Recently, Freyman et al. (1999) appear to have accomplished this by show-

spectrum noises. Because the acoustics at each ear do not change substantially with a switch in the perceived location of the masker (see Freyman et al., 1999 for a discussion of this issue), the larger advantage of perceived spatial separation when masking stimuli are nonsense sentences is presumably associated with higher level processes.

1.3. Energetic and informational masking in Mandarin Chinese

In the present paper, we attempted to replicate and expand on Freyman et al.'s (1999) results using Mandarin-speaking Chinese listeners. Chinese is one of the most popular languages in the world. To date, however, there is little literature available on whether there is a similar advantage of perceived spatial separation for recognition of Chinese speech, or the extent to which release from informational masking is modulated by the characteristics of the language in which the information is presented. Indeed there are at least two reasons to suspect that the extent of the release from informational masking due to perceived spatial separation may di er between English and Mandarin Chinese. First, there is some evidence that the pattern and extent of energetic masking di ers substantially between English and Chinese. Second, it is possible that the tonal nature of Mandarin Chinese may modulate the degree of release from informational masking due to perceived spatial separation.

not provide any contextual support for recognition of key words. These sentences were recorded digitally onto a computer disk, sampled at 22.05 kHz and saved as 16-bit PCM wave files. The digital waveforms were examined on a computer monitor for artifacts such as excessive noise and/or peak clipping that would require replacement of the sentence. The sentences were arbitrarily divided into 24 lists of 13 sentences.

Target sentences were presented by both the right and the left loudspeakers with the right speaker leading the left speaker by 3 ms. Thus participants perceived the target sentence images as coming from the right side.

There were two types of masking stimuli: noise and speech. To obtain a noise whose spectrum was representative of young female Chinese talkers, 5000 speech samples from 10 young female Chinese talkers (20–26 years old, 500 for each talker) were mixed using Matlab software at the sampling rate of 22.05 kHz with 16 bit quantization. The resulting 0.66-s noise sample was then continuously repeated (without a pause between segments) to provide a stream of Chinese speech spectrum noise. Fig. 1 shows the long-term average spectrum of the noise sound used in this study. Because the sample was repeated, the Chinese speech spectrum noise had a periodicity of 0.66 s, which is approximately the length of three Chinese words. The speech masker was a continuous recording of numerous Chinese nonsense sentences simultaneously spoken by two other young female talkers (Talkers B and C). Nonsense sentences in the masker were similar in linguistic structure to the target nonsense sentences but di ered in their content. Also, each of the masking sentences spoken by Talkers B and C was di erent.

Targets and maskers were calibrated using a B&K

was fit to each individual's data, using the Levenberg–Marquardt method (Wolfram, 1991), where y is the probability of correct identification of keywords, x is the SNR corresponding to

e di erent left/right onset delays,
d a masker image as coming from
left, respectively. The perceptual re-
the precedence e ect can be induced
ıg speech or noise (Freyman et al.,
ven in a nonanechoic testing chamber as
e.

veraged across participants, percent correct
ntification increased monotonically with SN
ı of the six masking conditions (2
× 3 Perceived Locations), without
ıs or dip as reported in previ
01; Freyman et al., 1999

vement in tl
of the psycl
for the condi
Here, however,
) was larger (Fig.

ense sentences or speech-s
d by the two spatially separat

NOISE MASKING

100
90

plateaus were observed when both the target sentences and the speech masker were perceived to be emanating from the same location. The absence of nonmonotonicity in our data is in agreement with the results reported by Arbogast et al. (2002).

The present study used Chinese nonsense sentences as speech signals and obtained results that are comparable to those reported by Freyman et al. (1999). When the masker was noise, the improvement of recognition of nonsense Chinese speech was minor (1 dB), even though a large perceived spatial separation (45° or 90°) was in-

Freyman et al. (1999) ect.y similarly a tion when th

spatial separation provides a cue that facilitates perceptual segregation of target speech from informational maskers, and strengthens the connection of the relevant elements in target speech across time. However, perceived spatial separation only slightly releases the target from energetic masking.

Interestingly, no di erences in the amount of release from masking were observed for the conditions in which the masker was perceived to be located frontally and when the masker was perceived to be in the opposite hemifield. These results indicate that the 45° perceived separation is su ciently large and that further increases in perceived separation do not provide an additional benefit. In both speech and noise masking situations, the masking stimuli from the two loudspeakers were correlated. Thus the monaural spectral profiles of masking stimuli were di erent between the perceived 90° separation and 45° separation, because of the e ect of the time lag on the spectrum of the sum of the two correlated sounds (comb filtering). Also, repetition of the noise-masker segment (1.52 Hz) might have modified the monaural spectral profiles. However, the lack of any difference between the perceived 90° separation and 45° separation suggests that the di erence in monaural spectral profiles produced by consistent phase-linked e ects (comb filtering) may be diminished in the non-anechoic condition and/or the spectral cue did not contribute to the perceived spatial advantage at all. Finally, the present data raise the issue of whether or not there is special advantage if the masker and target are perceived to be in di erent hemifields (di erent sides of the head). Boehnke and Phillips (1999) have argued that there might be two central spatial channels, one for the left hemifield and one for the right hemifield, with the two channels overlapping in the center. If these di erent channels are accessed by perceived location rather than by actual physical location, we might expect di erences in the degree of release from masking when the midline was crossed. However, no such e ect was observed.

For Chinese speech, recognition of initial consonants is critical to recognition of the associated words. Since there are more voiceless consonants, Chinese words would be more vulnerable to energetic masking than English words (Kang, 1998). Also, perception of tones of syllable in Chinese is closely linked to lexical meaning, which may provide listeners with additional cues to connect syllables in target speech across time. In spite of these characteristics of Chinese speech, results of the present study indicate that the advantage of perceived separation in unmasking speech is not limited to English but also extends to tonal Chinese. At this moment it is not clear why under speech masking conditions the perceived-spatial-separation advantage obtained for Chinese is smaller than reported by Freyman et al. (1999) for English. In the future, the e ect of perceived spatial separation on cross-language informational masking should be investigated with further refined controls of target/masker similarities.

A

The authors would like to thank Jane W. Carey and Chenfei Ma for their assistance in data acquisition and illusion construction. The authors would also like to thank Dr. Steve Colburn, Dr. Brian C. J. Moore, and one anonymous reviewer for helpful critiques. This work was supported by the China Natural Science Foundation (No. 60172055, 69635020), the China National High-Tech R&D Project (863 Project, No. 2001AA114181), a grant from the Ministry of Science and Technology of China (No. 2002CCA01000), and a ‘‘985’’ grant from Peking University. It was also supported by the Natural Sciences and Engineering Research Council of Canada and the Canadian Institutes of Health Research.

R

Arbogast, T.L., Mason, C.R., Kidd, G., 2002. The e ect of spatial separation on informational and energetic masking of speech. J. Acoust. Soc. Am. 112, 2086–2098.

Blauert, J., 1997. Spatial Hearing. MIT, Cambridge, MA.

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

Bronkhorst, A.W., Plomp, R., 19par9-1uetic,c04205i.7(Ahce)-224g2-219.6.8(of)23

Kang, J., 1998. Comparison of speech intelligibility between English and Chinese. J. Acoust. Soc. Am. 103, 1213–1216.
Koehnke, J., Besing, J.M., 1996. A procedure for testing speech intelligibility in a virtual listening environment. Ear. Hear. 17, 211–217.