Available online at www.sciencedirect.com
SCIENCE DIRECT®
Hearing Research 199 (2005) 1–10
HEARING RESEARCH
www.elsevier.com/locate/heares

# The effect of perceived spatial separation on informational masking of Chinese speech

Xihong Wu [a,b,*], Chun Wang [a], Jing Chen [b], Hongwei Qu [b], Wenrui Li [a], Yanhong Wu [a], Bruce A. Schneider [c], Liang Li [a,b,c]

[a] Department of Psychology, Peking University, Beijing 100871, China
[b] National Key Laboratory on Machine Perception, Speech and Hearing Research Center, Peking University, Beijing 100871, China
[c] Department of Psychology, Centre for Research on Biological Communication Systems, University of Toronto at Mississauga, Mississauga, Ont., Canada L5L 1C6

Received 13 October 2003; accepted 5 March 2004
Available online 13 November 2004

**Abstract**

The effect of perceived spatial separation, induced by the precedence effect, on release from noise or speech masking was investigated. Listeners were asked to orally repeat Chinese nonsense sentences, which were spoken by a female talker and presented by both the left (−45°) and right (+45°) loudspeakers, when maskers, which were either speech-spectrum noise sounds or Chinese nonsense sentences spoken by two other female talkers, were presented by the same two loudspeakers. Delays between identical sounds presented over the two loudspeakers were used to control the perceived locations of the target (right only) and masker (right, center, or left). The results show that perceived 45° and 90° separations of target speech from masking speech led to equivalently marked improvement in speech recognition, even though the degree of improvement was smaller than that reported in [J. Acoust. Soc. Am. 106 (1999) 3578 (using English nonsense speech)]. When the masker was noise, however, perceived separation only marginally improved speech recognition. These results indicate that release from informational masking, due to perceived target/masker spatial separation induced by the precedence effect, also occurs for tonal Chinese speech. Compared to the 45° perceived within-hemifield separation, the 90° perceived cross-hemifield separation does not produce further unmasking.


*Keywords:* Chinese speech; Nonsense sentences; Precedence effect; Perceived spatial separation; Informational masking; Energetic masking

## 1. Introduction

### 1.1. Energetic versus informational masking

Listeners often find it difficult to comprehend and participate in conversations that take place in a noisy environment, especially when the "noise" is other people talking, such as in a "cocktail party" (Cherry, 1953) environment. This difficulty is due, in large part, to "energetic" masking of the speech signal by the speech masker. The presence of any acoustic source, other than the signal, will provide an energy floor masking the signal. If the level of the signal is too close to this energy floor, the peripheral neural activity elicited by the competing noise will overwhelm that of the signal. Thus, it is difficult to detect and comprehend a speech signal when other sound sources (speech or non-speech) are present. This sort of masking is referred to as "energetic" (Freyman

*Abbreviations:* SNR, signal-to-noise ratio

* Corresponding author. Address: National Key Laboratory on Machine Perception, Speech and Hearing Research Center, Peking University, Beijing 100871, China. Tel./fax: +86 10 62759989.
*E-mail address:* wxh@cis.pku.edu.cn (X. Wu).

0378-5955/$ - see front matter © 2004 Elsevier B.V. All rights reserved.
doi:10.1016/j.heares.2004.03.010

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 difficult, 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 differences 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 showing that the release from masking that occurs when the target and masker are perceived to be spatially separated is greater when the masker is informational than when it isn't (see below).

### *1.2. Using perceived spatial separation to compare energetic and informational masking*

It has been well documented that spatially separating the source of an auditory signal from a source of masking improves the recognition of the signal (for a review see Zurek, 1993). For example, when a noise masker is presented from a loudspeaker located in the lateral field, thresholds for detecting sound signals, which are presented from a loudspeaker located in the frontal field, are lower than when the noise masker is presented from the same frontal-field loudspeaker as the target (Arbogast et al., 2002; Dubno et al., 2002; Duquesnoy, 1983; Freyman et al., 1999; Gelfand et al., 1988). This physical separation can improve the signal-to-noise ratio (SNR). For example if the masker moves to the right while the target remains at the front, the SNR in the left ear will improve because the head shadow lowers the level of the masker in the left ear. In addition, moving the masker to the right makes the interaural time delay of the signal different from that of the masker, a change that is known to improve detectability (Bronkhorst and Plomp, 1988; Zurek, 1993). When both target and masker are speech and physically separated, the release from masking could be due to either acoustic cues (head shadow effects and binaural interaction) created by this physical separation or reduced difficulty of both perceptually segregating the target from the masker and inhibiting linguistic and semantic processing of the masker. Hence, we might expect to find a greater spatial separation effect when the masker is speech than when the masker is noise.

Freyman et al. (1999) negated the effectiveness of these head-shadow and binaural cues for unmasking signals by using the precedence effect to manipulate the perceived locations of target and masker (see below). If the release from masking produced by a difference in perceived spatial position is greater for a speech than for a noise masker, this indicates that informational masking occurs for the former.

In a reverberant environment, listeners not only receive the direct wavefront from a sound source but also numerous time-delayed reflections of the source. If the time delays between the arrival of the direct wave and each of the reflected waves are sufficiently short (1–10 ms or more, depending on the nature of the stimulus), listeners typically perceive a single "fused" image of the source located at or near the original site of the source. This phenomenon has been generally known as the precedence effect (Wallach et al., 1949; for reviews see Blauert, 1997; Li and Yue, 2002; Litovsky et al., 1999; Zurek, 1980). In the laboratory, the precedence effect is typically simulated by presenting the same stimulus over two spatially separated loudspeakers. Delaying presentation of a sound over one of the loudspeakers simulates the situation in which there is a single source located at or near the leading loudspeaker and a reflection of that source coming from the direction of the lagging loudspeaker. By changing which loudspeaker leads the other, one can switch the perceived location of the sound. Freyman et al. (1999) used the precedence effect to induce a perceived separation of images of target and masking stimuli. In two of their experimental conditions (FR–FR and FR–RF conditions), both a frontal loudspeaker and a lateral loudspeaker (in the right hemifield) delivered both target stimuli (nonsense sentences) and masking stimuli (nonsense sentences or speech spectrum noise). For target sentences, the frontal loudspeaker always led the right loudspeaker by 4 ms. Thus the perceived images of target sentences seemed to be from the frontal loudspeaker. For the masking stimuli, the frontal loudspeaker either led or lagged behind the right loudspeaker by 4 ms. Thus the perceived masker images were around either the frontal loudspeaker or the right loudspeaker. In other words, the perceived locations of the target and the masker could be manipulated as either spatially the same or separated, even though the masker and the target were presented physically from both loudspeakers. Freyman et al. found a large advantage (4–9 dB) of the perceived spatial separation in the recognition of nonsense sentences spoken by a female talker when masking stimuli were nonsense sentences spoken by another female talker, but a much smaller advantage (less than 1 dB) when the masking stimuli were speech-

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 differ between English and Mandarin Chinese. First, there is some evidence that the pattern and extent of energetic masking differs 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.

The structure of a Chinese syllable can be divided into two or three components: an initial consonant (only a small number of syllables have no initial consonants), followed by a vowel, which is sometimes followed by final consonants. Compared to English, Chinese syllables have more voiceless consonants and fewer voiced consonants. Voiceless consonants are more easily masked than voiced consonants because they have less energy. Thus masking noise might cause more perceptual confusion among Chinese consonants than in English. In other words, Chinese syllables might be more vulnerable to energetic masking. It has been reported that the intelligibility of Chinese speech is considerably worse than that of English speech under conditions of noise masking (Kang, 1998).

On the other hand, unlike European languages, the pitch contour of the vowel is phonemic. For example, changing the pitch glide in the syllable "ma" from flat, to rising, or to rising and falling, or to falling, changes the meaning of the word. Thus there might be some distinct patterns of informational masking for Chinese speech. When Mandarin listeners are attending to a target Mandarin talker, they have to use the pitch contours in that talker's vowels in order to correctly identify the phoneme and therefore the word. Because pitch contour information is phonemic, changes in pitch contours are likely to initiate activity in the language (non-auditory) pathways. It is possible that the degree to which there is release from informational masking depends on the types of informational confusion between the target and speech maskers. Thus, the amount of release could differ across language groups.

In the present study, we used the precedence effect to induce perceived spatial separation of target Chinese nonsense sentences from either informational or energetic maskers. In addition, we also investigated if the size of the release of speech from masking depended on whether the perceived location of the masker was in the same or opposite hemifield relative to the perceived location of the target.

## 2. Materials and methods

### 2.1. Participants

Twelve young university students (mean age = 21.1 years old) with normal and balanced (less than 15 dB difference between the two ears) hearing thresholds, confirmed by audiometry, participated in the study. Their first language was Mandarin Chinese.

### 2.2. Apparatus and materials

Participants were seated in a chair at the center of a sound-attenuating chamber, which was 192 cm in length, 181 cm in width, and 196 cm in height (EMI Shielded Audiometric Examination Acoustic Suite). All acoustic signals were digitized at the sampling rate of 22.05 kHz using the 24-bit Creative Extigy sound blaster (with a built-in antialaising filter) and audio editing software (Cooledit), under the control of a computer with a Pentium IV processor. The analog outputs were delivered from two loudspeakers (Creative Inspire 4.1), which were in the frontal azimuthal plane at the left and the right 45° positions symmetrical with respect to the median plane. The loudspeaker height was approximately ear level for a seated listener with average body height, and the distance from each of the two loudspeakers to the center of the participants' head was 1.5 m.

Target speech stimuli were Chinese "nonsense" sentences spoken by a young female talker, the author CW (Talker A). The direct English translations of these sentences are similar but not identical to the English nonsense sentences that were developed by Helfer (1997) and also used in studies by Freyman et al. (1999, 2001). These sentences are syntactically correct but not meaningful. In each of the target sentences, for example, "One appreciation could retire his ocean", there are three key words (as indicated by the words underlined in the example) that are scored during speech recognition testing. Note that the sentence frame does

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 differed in their content. Also, each of the masking sentences spoken by Talkers B and C was different.

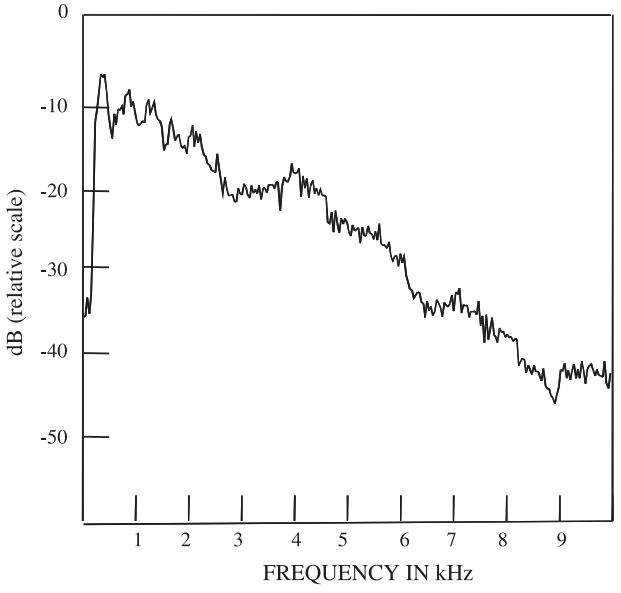


Fig. 1. The relative spectrum of a 0.66 s segment of Chinese speech-spectrum noise. In constructing this plot the power was averaged over

Targets and maskers were calibrated using a B&K sound level meter (Type 2230) whose microphone was placed at the central location of the participants' head when the listener was absent using a "Fast"/"RMS" meter response. Measurements were conducted separately for each loudspeaker. During a session, the target sentences were presented at a level such that each loudspeaker, playing alone, would produce an average sound pressure of 54 dBA at the location corresponding to the center of the listener's head. The sound pressure level of the target remained constant throughout the experiment. The sound pressure levels of the masker were adjusted to produce four SNRs: −12, −8, −4, and 0 dB.

### 2.3. Procedure

Six of the twelve participants heard the target sentences against the noise background before hearing different target sentences against the speech background. The remaining six participants were tested in the opposite order.

As indicated in Fig. 2, the masker was presented over the two loudspeakers using one of the three delay times: (1) right leading left by 3 ms (R − L = +3 ms); (2) no lag between the loudspeakers (R − L = 0 ms); and (3) right lagging behind left by 3 ms (R − L = −3 ms). For R–L delays of +3, 0, and −3 ms, all the participants heard the masker as originating from right, center, and left, respectively.

Twenty-four blocks of 13 sentences each were created for all possible combinations of the three masker delays, four SNRs, and two types of maskers (speech-spectrum noise and nonsense sentences). Hence, the type of masker, its level and perceived location remained constant during each 13-trial block. The order of presentation of the different perceived locations of the masker was completely counterbalanced across participants with each participant experiencing the four SNRs in a different random order.

In each trial, the listener pressed the central button of the response box to start the masking sound. About 1 s later, a single target sentence was presented. The masker was gated off with the target. Participants were instructed to repeat the target sentence as best they could immediately after the sentence was completed. The experimenters indicated on a marking sheet which of the key words had been identified correctly. The number of correctly identified words was tallied later.

## 3. Results

A logistic psychometric function,

$$y = 1/(1 + e^{-\sigma(x-\mu)})$$

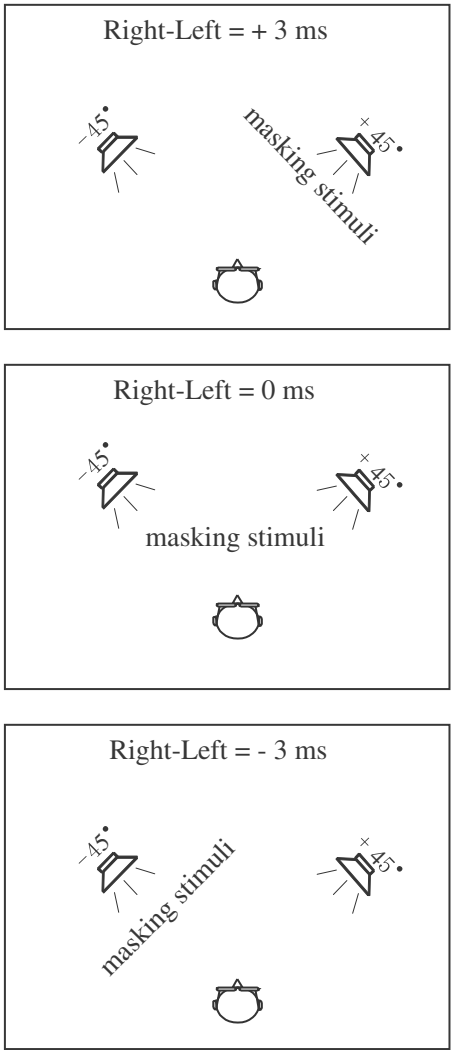


Fig. 2. Diagrams showing the perceived locations of masking stimuli at the three different left/right sound delays: (1) The right loudspeaker led the left loudspeaker by 3 ms (right leading time: right – left = +3 ms). Under this delay, the listener perceived masking sounds as coming from the right side. (2) The right loudspeaker started simultaneously with the left loudspeaker (right leading time: right – left = 0 ms). Under this delay, the listener perceived masking sounds as coming from the front. (3) The right loudspeaker lagged behind the left loudspeaker by 3 ms (right leading time: right – left = −3 ms). Under this delay, the listener perceived masking sounds as coming from the left side.

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 $y$, $\mu$ is the SNR corresponding to 50% correct identification (the threshold ratio), and $\sigma$ determines the slope of the psychometric function.

Fig. 3 shows percent-correct word identification as a function of SNR for each of the 12 subjects in the following six masking conditions: (1) noise masker perceived left (NL); (2) noise masker perceived centrally (NC); (3) noise masker perceived right (NR); (4) speech masker perceived left (SL); (5) speech masker perceived centrally (SC); (6) speech masker perceived right (SR). In general, the psychometric functions provide a good fit to the individual data.

The psychometric functions in Fig. 3 were used to determine mean thresholds (the SNRs corresponding to 50% correct identification) across participants. Mean thresholds for the six listening conditions are shown in Fig. 4. For both noise and speech maskers, thresholds were lower when the perceived location of the masker differed from that of the target (NL and NC versus NR for the noise masker, and SL and SC versus SR for the speech masker), indicating a perceived spatial location effect for both noise and speech maskers. However, the effect was much larger for speech masking than it was for noise masking. In addition, when the masker was perceived to originate from the same spatial location as the target (NR and SR), thresholds for the noise and speech maskers were about the same. This pattern of results was confirmed by a 2 (Masker) by 3 (Perceived Location) within-participant ANOVA which revealed a significant main effect of Masker, $F(1,11) = 13.719$, MSE = 2.359, $p = 0.003$, a significant main effect of Perceived Location, $F(2,22) = 21.984$, MSE = 1.801, $p < 0.001$, and a significant interaction between Masker and Perceived Location, $F(2,22) = 3.503$, MSE = 2.794, $p = 0.048$. To determine the locus of this interaction effect we conducted separate ANOVAs for the noise and speech maskers.

For the noise masker, the location effect on threshold was not quite significant, $F(2,22) = 3.430$, MSE = 1.898, $p = 0.051$. However, for the speech masker, the location effect on threshold was highly significant, $F(2,22) = 15.896$, MSE = 2.697, $p = 0.000$. Pairwise comparisons indicated that the perceived left and central locations of the speech maskers did not differ from one another ($p = 1.000$) but both left and center locations did differ significantly from the right location ($p < 0.001$, $p = 0.003$, respectively).

Fig. 5 shows how the slope parameter varied across the six listening conditions. In general, slopes are steeper for the noise masker than for the speech masker. Fig. 5 also suggests that slopes might be shallower for perceived location on the left. However, although the ANOVA on the slope parameter revealed a main effect due to Masker, $F(1,11) = 22.595$, MSE = 0.009, $p = 0.001$, neither the main effect of Perceived Location, $F(2,22) = 1.691$, MSE = 0.007, $p = 0.207$, nor the interaction between Masker and Perceived Location, $F(2,22) = 0.126$, $p = 0.883$, were significant.

Fig. 6(a) plots mean percent correct as a function of SNR for the noise masker. In accordance with the results from the ANOVA, a single psychometric function was fit to the left and central perceived locations and the slopes for the two functions (masker right and masker off-right) were constrained to be equal. Fig. 6(b) shows the equivalent data for the speech masker where the

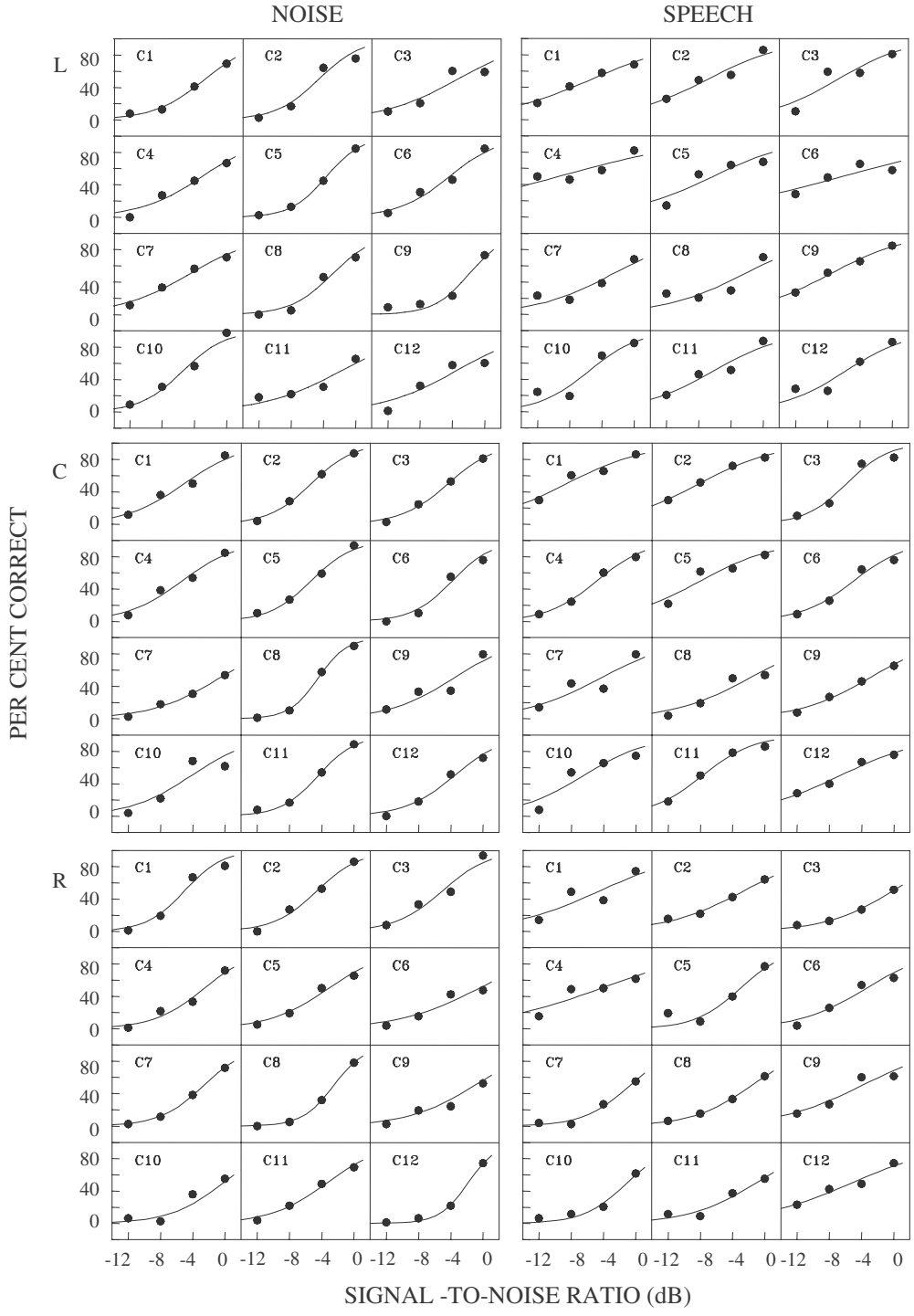


Fig. 3. Percent-correct word identification as a function of signal-to-noise ratio (SNR) for each of the 12 subjects in the following six masking conditions: (1) noise masking and left location of perceived masker image (NOISE, L); (2) noise masking and central location of perceived masker image (NOISE, C); (3) noise masking and right location of perceived masker image (NOISE, R); (4) speech masking and left location of perceived masker image (SPEECH, L); (5) speech masking and central location of perceived masker image (SPEECH, C); (6) speech masking and right location of perceived masker image (SPEECH, R).

psychometric functions were subject to the same constraints. Fig. 6(a) shows that the psychometric function that fits the data for the condition in which the perceived locations of masker and target were the same, when shifted by about 1 dB to the left, also provides a good description of the functions for conditions in which the masker was perceived to have a different location than that of the target. Hence, the effect of a perceived

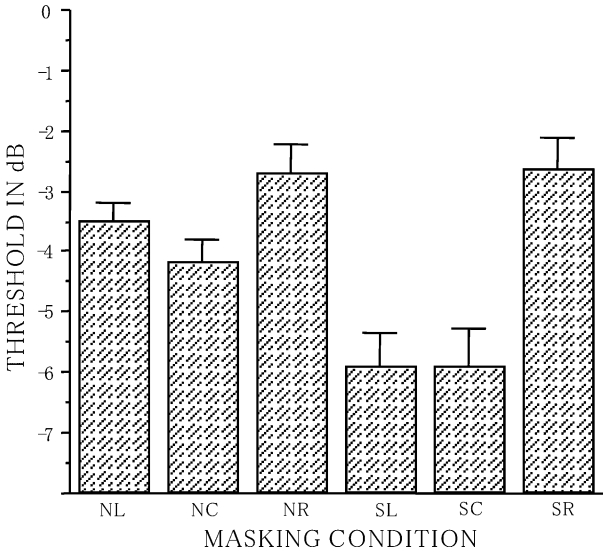


Fig. 4. Mean thresholds (average of 50% points on the psychometric functions in Fig. 3) for the six masking conditions: (1) noise masker on left, (NL); (2) noise masker at center (NC); (3) noise masker on right (NR); (4) speech masker on left (SL); (5) speech masker at center (SC); (6) speech masker on right (SR). The error bars indicate the standard errors of the mean.

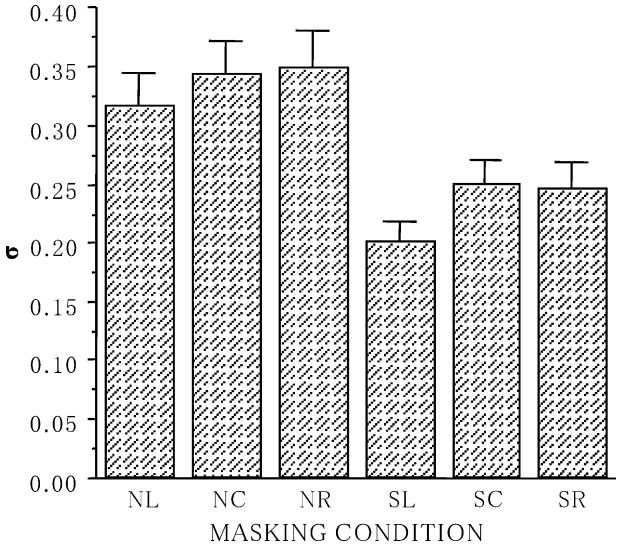


Fig. 5. Mean value of $\sigma$ for the six masking conditions. The slope of the psychometric function at the intensity level corresponding to 50% correct is $\sigma/4$. The error bars indicate the standard errors of the mean.

spatial separation was a small improvement in threshold but without a change in the slope of the psychometric function. A similar result holds for the conditions in which the masker was speech. Here, however, the shift in the function (about 3.3 dB) was larger (Fig. 6(b)).

## 4. Discussion

When either nonsense sentences or speech-spectrum noise were delivered by the two spatially separated loudspeakers, with the three different left/right onset delays, participants perceived a masker image as coming from the right, front, or left, respectively. The perceptual results confirm that the precedence effect can be induced with long-lasting speech or noise (Freyman et al., 1999, 2001), even in a nonanechoic testing chamber as we used here.

When averaged across participants, percent correct word identification increased monotonically with SNR in each of the six masking conditions (2 Masker types × 3 Perceived Locations), without displaying plateaus or dip as reported in previous studies (Brungart, 2001; Freyman et al., 1999). In particular, no dips or

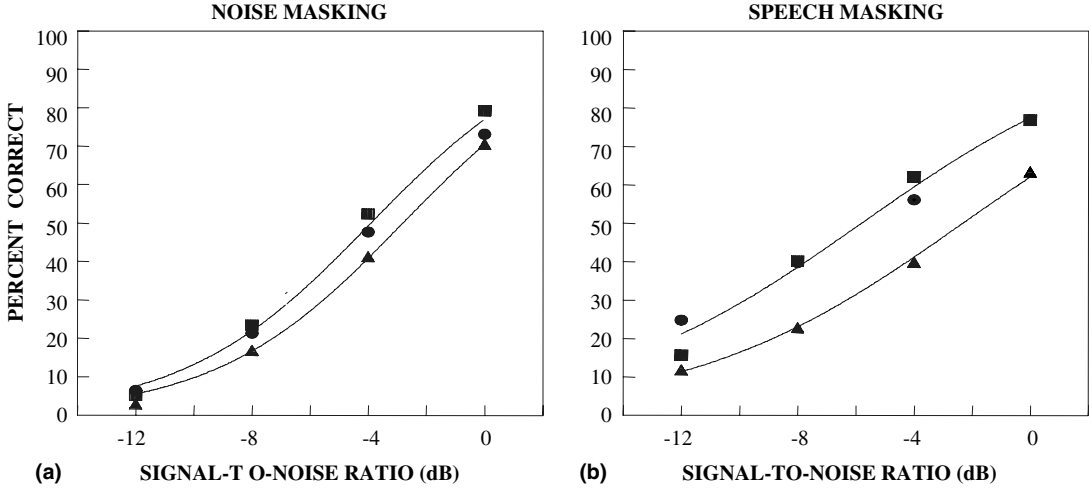


Fig. 6. Mean percent correct word identification as a function of SNR for the three different perceived locations of the masker. Panel (a): Noise masker. Panel (b): Speech masker. In each panel, the psychometric function on the right one is for the condition under which both the target and masker were perceived as coming from the "right". The other psychometric function (the left one) is for the condition under which the target was perceived as coming from the "right" but the masker was perceived as coming from "off-right". Symbols: triangles (masker right); squares (masker center); circles (masker left).

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 induced by the precedence effect. Freyman et al. (1999) found a similarly small improvement in word identification when the perceived location of the noise masker differed (60° separation) from that of the target. Hence, for a noise masker, the benefit due to perceived spatial separation between target and masker is very small. As mentioned in the Introduction, manipulating perceived location through use of the precedence effect minimizes head shadow effects and binaural cues. The small effect due to perceived separation in the precedence situation for the noise masker supports the view that the large effect observed for physically separated targets and noise maskers is mainly based on head shadow effects and binaural cues.

When the masker was nonsense speech, which produced both informational and energetic masking, the perceived spatial separation of the target speech from the speech masker markedly improved recognition of the target. The improvement in threshold (3.3 dB) observed for Chinese speech, however, was somewhat smaller than that (4–9 dB) reported by Freyman et al. (1999). It is possible that the larger energetic masking effect for Chinese speech (Kang, 1998) may be responsible, in part, for differences in the size of the effect in the two languages. However, the fact that a substantial effect was observed in both languages reinforces the argument that the release is not due to peripheral acoustic features (which differ substantially in these two languages) but rather to the operation of higher-order linguistic or semantic processes. Perceived spatial separation tends to reduce perceived target/masker spatial similarity, which may interact with other dimensions of target/masker similarities. Mandarin speech may have a different target/masker similarity pattern than English speech. Thus more adequately defined target/masker similarities and a general target/masker similarity metric (see discussion of this issue by Durlach et al., 2003) would be useful for both clarifying various dimensions of target/masker similarities and explaining the relatively smaller effects of perceived spatial separation on reducing informational masking of Chinese speech. Clearly, further research is needed with respect to this issue.

It is interesting to note that when both target and masker were perceived to originate from the same spatial location, thresholds for target recognition were identical for both noise and speech maskers. One might have expected a greater degree of masking by a speech masker than by a noise masker because a speech masker should give rise to both energetic and informational masking while a noise masker should produce energetic masking only. However, fluctuations in the envelope of the speech masker could have attenuated energetic masking effects relative to those observed with a noise masker (because it would be easier to recognize the target speech during a trough in the envelope of the speech masker), thereby lowering recognition thresholds. But the presence of competing information in the speech masker may have offset this reduction in the degree of energetic masking, leading to equivalent thresholds for both speech and noise maskers.

Consistent with previous results (e.g., Freyman et al., 1999; Brungart, 2001), Fig. 5 shows that under all conditions, the slopes of the psychometric function are steeper for noise maskers than they are for speech maskers. Because there is considerable variation in the energy envelope of a speech masker, there will be time periods in which the SNR is high (e.g., vowels in the target speech occurring when there is a pause or unvoiced consonants in the masking speech), and other time periods in which the SNR is low (unvoiced consonants in the target occurring when the energy in the masking stimulus is high). The effect of these fluctuations in local SNR would be to flatten the psychometric function for a speech masker (as compared to a broadband noise masker). This was the pattern that was found in this experiment.

It is also interesting to note that release from informational masking is observed even when the environment is not anechoic. Apparently, the surface reflections have a negligible effect on the perceived locations of target and masker. More importantly, these reflections make it even less likely that the source of the informational masking effect is peripheral in nature since reverberant environments reduce the effectiveness of peripheral acoustic cues (see discussions by Freyman et al., 1999; Koehnke and Besing, 1996).

Since the benefits of head-shadow and binaural-timing effects on unmasking signals are markedly reduced (Freyman et al., 1999; Koehnke and Besing, 1996) when perceived spatial location is manipulated using the precedence effect in a reverberant room, the release from masking that occurs with perceived spatial separation when both masker and target are speech cannot be explained by the acoustical cues in Zurek's model (Zurek, 1993). Rather, the release from masking produced by perceived spatial separation suggests that it is easier to suppress the linguistic and semantic interference from the masker when the masker is perceived as coming from a different location than that of the target. Our data support Freyman et al.'s (1999, 2001) notion that perceived

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 differences 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 sufficiently 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 different between the perceived 90° separation and 45° separation, because of the effect 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 difference in monaural spectral profiles produced by consistent phase-linked effects (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 different hemifields (different 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 different channels are accessed by perceived location rather than by actual physical location, we might expect differences in the degree of release from masking when the midline was crossed. However, no such effect 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 effect of perceived spatial separation on cross-language informational masking should be investigated with further refined controls of target/masker similarities.

## Acknowledgements

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.

## References

Arbogast, T.L., Mason, C.R., Kidd, G., 2002. The effect 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., 1988. The effect of head-induced interaural time and level differences on speech intelligibility in noise. J. Acoust. Soc. Am. 83, 1508–1516.

Brungart, D.S., 2001. Informational and energetic masking effects in the perception of two simultaneous talkers. J. Acoust. Soc. Am. 109, 1101–1109.

Brungart, D.S., Simpson, B.D., 2002. The effects of spatial separation in distance on the informational and energetic masking of a nearby speech signal. J. Acoust. Soc. Am. 112, 664–676.

Cherry, E.C., 1953. Some experiments on the recognition of speech with one and two ears. J. Acoust. Soc. Am. 25, 975–979.

Dubno, J.R., Ahlstrom, J.B., Horwitz, A.R., 2002. Spectral contributions to the benefit from spatial separation of speech and noise. J. Sp. Lan. Hear. Res. 45, 1297–1310.

Duquesnoy, A.J., 1983. Effect of a single interfering noise or speech source upon the binaural sentence intelligibility of aged persons. J. Acoust. Soc. Am. 74, 739–743.

Durlach, N.I., Mason, C.R., Shinn-Cunningham, B.G., Arbogast, T.L., Colburn, H.S., Kidd, G., 2003. Informational masking: counteracting the effects of stimulus uncertainty by decreasing target-masker similarity. J. Acoust. Soc. Am. 114, 368–379.

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.

Freyman, R.L., Balakrishnan, U., Helfer, K.S., 2001. Spatial release from informational masking in speech recognition. J. Acoust. Soc. Am. 109, 2112–2122.

Gelfand, S.A., Ross, L., Miller, S., 1988. Sentence reception in noise from one versus two sources: effects of aging and hearing loss. J. Acoust. Soc. Am. 83, 248–256.

Helfer, K.S., 1997. Auditory and auditory-visual perception of clear and conversational speech. J. Sp. Lan. Hear. Res. 40, 432–443.

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.

Kidd, G., Mason, C.R., Deliwala, P.S., Woods, W.S., Colburn, H.S., 1994. Reducing informational masking by sound segregation. J. Acoust. Soc. Am. 95, 3475–3480.

Kidd, G., Mason, C.R., Rohtla, T.L., Deliwala, P.S., 1998. Release from masking due to spatial separation of sources in the identification of nonspeech auditory patterns. J. Acoust. Soc. Am. 104, 422–431.

Li, L., Yue, Q., 2002. Auditory gating processes and binaural inhibition in the inferior colliculus. Hear. Res. 168, 113–124.

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

Wolfram, S., 1991. Mathematica: A System for Doing Mathematics by Computer. Addison-Welsey, New York.

Wallach, H., Newman, E.B., Rosenzweig, M.R., 1949. The precedence effect in sound localization. Am. J. Psychol. 62, 315–336.

Zurek, P.M., 1980. The precedence effect and its possible role in the avoidance of interaural ambiguities. J. Acoust. Soc. Am. 67, 952–964.

Zurek, P.M., 1993. Binaural advantages and directional effects in speech intelligibility. In: Studebaker, G.A., Hockberg, I. (Eds.), Acoustic Factors Affecting Hearing Aid Performance. Allyn and Bacon, Boston, MA.