Speech detection in spatial and nonspatial speech maskers

Uma Balakrishnan[a] and Richard L. Freyman
Department of Communication Disorders, University of Massachusetts, 358 N. Pleasant Street, Amherst, Massachusetts 01003, USA

(Received 22 March 2007; revised 28 February 2008; accepted 3 March 2008)

The effect of perceived spatial differences on masking release was examined using a 4AFC speech detection paradigm. Targets were 20 words produced by a female talker. Maskers were recordings of continuous streams of nonsense sentences spoken by two female talkers and mixed into each of two channels (two talker, and the same masker time reversed). Two masker spatial conditions were employed: "RF" with a 4 ms time lead to the loudspeaker 60° horizontally to the right, and "FR" with the time lead to the front (0°) loudspeaker. The reference nonspatial "F" masker was presented from the front loudspeaker only. Target presentation was always from the front loudspeaker. In Experiment 1, target detection threshold for both natural and time-reversed spatial maskers was 17–20 dB lower than that for the nonspatial masker, suggesting that significant release from informational masking occurs with spatial speech maskers regardless of masker understandability. In Experiment 2, the effectiveness of the FR and RF maskers was evaluated as the right loudspeaker output was attenuated until the two-source maskers were indistinguishable from the F masker, as measured independently in a discrimination task. Results indicated that spatial release from masking can be observed with barely noticeable target-masker spatial differences.
 [DOI: 10.1121/1.2902176]

PACS number(s): 43.66.Dc, 43.66.Pn, 43.66.Qp [RYL] Pages: 2680–2691

I. INTRODUCTION

The term informational masking has been used in a wide variety of contexts involving both speech and nonspeech signals. The use of this common terminology implies a common basis for this type of masking across speech and nonspeech stimuli. However, some key differences in methodology preclude the development of a global understanding of the concept of informational masking across both classes of signals. Not surprisingly, the majority of work with speech targets and maskers has involved *recognition* tasks (e.g., Carhart *et al.*, 1969; Freyman *et al.* 1999; Brungart, 2001; Arbogast *et al.*, 2005; Rakerd *et al.*, 2006). Some nonspeech studies have employed pattern identification or similar tasks (e.g., Watson *et al.*, 1976; Kidd *et al.*, 1995, 1998) and shown evidence of substantial informational masking. However, most of the work with nonspeech stimuli has involved *detection* experiments where the threshold of a tone is measured in the presence of a spectrally and/or temporally complex masker (e.g., Neff and Green, 1987; Wright and Saberi, 1999; Lutfi, 2003; Richards and Neff, 2004; Durlach *et al.*, 2005).

Only one study that we are aware of (Helfer and Freyman, 2005) examined detection of speech in conditions in which the focus was on informational masking. In that study, detection thresholds of topic-based sentences in the presence of masking sentences were obtained as part of an investigation of audiovisual speech perception. Sentence detection thresholds were obtained for both audiovisual and audio alone conditions for two spatial conditions. In the nonspatial condition the target sentence and the masker were presented from a single loudspeaker directly in front of the listener (the F-F condition). The spatial condition was the same except that an additional copy of the masker was presented 60° to the right and a 4 ms time advance was imposed on the right loudspeaker (the F-RF condition). In this condition, the masker was perceived to be well to the right of the target due to the precedence effect. Detection thresholds improved through the addition of visual cues or target-masker spatial separation only when the masker was speech (two talkers of the same sex as the target talker), not when it was noise. The fact that the F-RF condition produced no improvement for the noise masker was interpreted, as in previous recognition studies (e.g., Freyman *et al.*, 1999) to be evidence that this spatial configuration produced no release from energetic masking. Therefore, the observation of improvement in the F-RF condition with the speech masker was ascribed to release from informational masking. Monaural control experiments conducted by Freyman *et al.* (2001) had shown that, at least with recognition, such improvements were realized only with binaural listening, and presumed to result from perceived spatial differences between target and masker.

Because of the common use of detection for studying informational masking with nonspeech stimuli, its application to speech stimuli may offer improved opportunities to compare informational masking across speech and nonspeech signals. An additional important advantage of studying detection, relative to recognition, is that it would be expected to lower the signal-to-noise (S-N) ratio at which a threshold level of performance can be measured. The classic literature shows the speech detection threshold to be about 9 dB below the speech reception threshold (Thurlow *et al.*, 1948; Chaiklin, 1959). This difference could be useful when release from informational masking is measured in degraded

[a]Author to whom correspondence should be addressed. Electronic mail: umab@vermontel.net.

 0001-4966/2008/123(5)/2680/12/$23.00

speech or in hearing-impaired populations. For example, Arbogast *et al.* (2005) showed that release from informational masking, and possibly informational masking itself, was reduced in a group of hearing-impaired subjects relative to normal-hearing listeners. However, the authors found that the base line nonspatial condition was so difficult for the hearing-impaired listeners that speech reception thresholds were often obtained at slightly positive S-N ratios. At positive S-N ratios, the target is louder than the masker, a condition that may not lend itself to informational masking. The vicinity of 0 dB S-N ratio may thus create a ceiling for informational masking, and the nonspatial threshold may have been truncated. Arbogast *et al.* (2005) suggested that because of this truncation it was not possible to determine conclusively whether informational masking differed between normal-hearing and hearing-impaired listeners. A speech detection paradigm could potentially avoid this issue because of the lower S-N ratios at which thresholds can be obtained.

The first experiment in the current paper employed a refined version of the detection paradigm used by Helfer and Freyman (2005) to study spatial release from informational masking. The sentence detection task used in that earlier study was most sensibly limited to single-interval trials because of the duration of the sentence stimuli. In the present study, we used words excised from the sentences and sought to determine whether a 4AFC paradigm could be employed successfully. The shortening of the stimuli from sentences to words also offered better control of stimulus variables and eliminated the potential use of semantic or syntactic connections between words available in sentence-level targets as cues, increasing the potential of making comparisons with nonspeech stimuli. Using this 4AFC paradigm, we investigated release from informational masking by creating target-masker spatial differences that do not produce unmasking with purely energetic maskers. These included the F-RF condition, described above, as well as an F-FR condition, which was identical to F-RF except that the front masker led the right masker by 4 ms. Both natural and time-reversed speech maskers were used.

In the course of Experiment 1 we (1) evaluated the extent to which a word detection paradigm lowered threshold S-N ratios, thereby mitigating the ceiling issues raised by Arbogast *et al.* (2005); (2) compared the amount of the informational masking that this detection paradigm produces relative to existing speech recognition data; (3) determined whether time-reversing the masker reduces spatial release from masking, a typical result in recognition studies; and (4) determined whether there is a difference in spatial release between the F-RF and F-FR conditions, where the former produces a large difference in perceived spatial location and the latter does not.

Experiment 2 explored in further detail the F-RF versus F-FR comparison and addressed the extent and type of spatial differences between target and masker that are essential for releasing informational masking. A classic study (Carhart *et al.*, 1969), as well as several recent studies (Freyman *et al.* 1999; Brungart *et al.*, 2005; Edmonds and Culling, 2005; and Rakerd *et al.*, 2006), have shown strong release from speech-on-speech masking even when spatial separation of target-masker images was not expected to create a clear separation in the centroid of target and masker auditory images. Freyman *et al.* (1999) found that spatial release from masking was almost as large for their F-FR condition as it was for the F-RF condition. The F-FR condition used the same loudspeaker configuration as the F-RF but with the time lead favoring the front loudspeaker. In this case the perceived horizontal location of target and masker was expected to be similar. Brungart *et al.* (2005) and Rakerd *et al.* (2006) explored the effect of masker delay over a wide range and found about the same masking release regardless of whether the time lead was imposed on the front or right loudspeaker. As the conditions where the masker time lead favors the front (target position) are not expected to create large horizontal shifts in the centroid of the masker image, their effectiveness may include or even be dominated by target-masker differences in other spatial properties, such as spatial width and shape. Blauert (1997) used the term "spaciousness" to describe this difference between two-source and single-source stimuli. Inasmuch as our two-loudspeaker spatial maskers simulate a source and a single reflection, the FR masker simulates only a reflection, whereas the RF masker simulates a reflection *and* also moves the masker source. It is not clear at this time whether the added spatial separation of masker source in the RF masker contributes to greater release from informational masking when the task is signal detection as opposed to recognition. In our second study we systematically attenuated the output of the right loudspeaker in both FR and RF spatial conditions and considered the difference in detection in these masking conditions in relation to their discriminability from the front only masker condition.

Finally, we investigated the effects of subject training on detection thresholds in the nonspatial (F-F) condition. This was motivated by the fact that in pilot listening, detection thresholds in the F-F condition showed considerable variability across runs and across listeners. Such variation was minimal in the spatial (F-RF and F-FR) conditions. Our purpose was to tease out any learning effects that might impact the overall amount of informational masking measured in the reference, nonspatial condition, so that we could make more accurate measurements of the release from informational masking provided by the spatial manipulations of the masker.

II. EXPERIMENT 1: RELEASE FROM MASKING IN DETECTION

A. Methods

1. Stimuli

Target stimuli were 20 consonant-vowel-consonant words excised from nonsense sentences recorded by a female talker. This was the same target talker used by Freyman *et al.* in previous studies (e.g., Freyman *et al.*, 1999). The 20 target words were chosen for clarity of production and ease of excision from the continuously produced sentence waveforms. Typically, the target word was the second "content" (meaning loaded) word of each utterance. In those instances where the second word in the sentence was not easy to extract from its surround, content words in other locations (first or third)

within the utterance were taken. The 20 target words were scaled to equate their root-mean-square (rms) amplitude and then padded with zeroes to match the duration of the longest word on the list (500 ms). The 20 words were concatenated to create a single file from which the experimental software randomly selected a single word and played it on each trial.

Two types of maskers were used: two-talker speech ("natural") and two-talker time-reversed speech ("reversed"). These were a 35 s duration mixture of the recording of two female talkers reciting nonsense sentences (see Freyman *et al.*, 2001). The choice of two-talker speech stream was determined by results shown by Freyman *et al.* (2004) that informational masking increased as number of talkers increased from one to two, and diminished thereafter as additional talkers were added. Other investigators have also shown that two to three-talker speech maskers are particularly effective in masking speech targets (Carhart *et al.*, 1975; Yost *et al.*, 1996; Brungart *et al.*, 2001; Hall *et al.*, 2002).

Each stereophonic masker was created by copying the two-talker masker signal into a second channel and imposing a 4 ms pad of zeroes at the beginning of one channel to delay it with respect to the other channel. Another 4 ms pad of zeroes was appended to the end of the second channel, to maintain equal duration of maskers in both channels. Both speech maskers were equated for rms as were the two channels of each masker. Two time-lead conditions were used for the natural masker: right-leading (RF) and front-leading (FR). Only the RF masker condition was used for the reversed speech masker. For comparison, a front-only (F) masker condition was created by turning off the right loudspeaker and playing the masker only from the loudspeaker that produced the target speech (the F-F condition). Thus for the front-only masker condition, the single-channel masker output was 3 dB less than that of the two-channel maskers. The target words always originated from the front loudspeaker. Therefore, for a given listening trial the target-masker configuration could be front-front (F-F), target front and masker right-front (F-RF) or target front and masker front-right (F-FR).

2. Apparatus

The experiments were conducted in an anechoic chamber measuring 4.9 m × 4.1 m × 3.12 m. The walls, floor, and ceiling are lined with 0.72 m foam wedges. Subjects were seated in the center of the room in front of a foam-covered semicircular arc on which two loudspeakers were positioned. The Front loudspeaker was at 0° horizontal azimuth; the Right loudspeaker was at 60° to the right. Both were 1.9 m from the approximate center of the subjects' head and were at ear height for the typical adult.

The target words were delivered via TDT System I instrumentation. The output of the 16 bit digital-to-analog converter (TDT DA1) running at 20 kHz was low-pass filtered at 8.5 kHz (TDT), attenuated (TDT PA3), and mixed with the masker before being delivered to a Crown D40 amplifier and a Realistic Minimus 7 loudspeaker. The masker was delivered from a second computer (Dell Dimension XPD 333) via audio software (Cool Edit Pro). The 35-s-long interference segment was played continuously in the loop mode over the duration of an adaptive track fed through PA4 attenuators (Tucker Davis System II). Calibration of target and maskers was by means of a 1 in. microphone (B&K 4145) fitted with a random incidence corrector and lowered to the position of the subjects' head with the subject absent. A sound level meter (B&K 2204) located outside the chamber measured the microphone output using the C-scale and Fast meter response. The target was calibrated to a sawtooth noise equated for average power to the target word stream. The maskers were calibrated for each channel using a speech-spectrum noise masker (Byrne *et al.*, 1994).

3. Subjects

Listeners were five young adult students. Four of the five subjects had hearing thresholds ≤20 dB hearing level (HL) in the tested frequency range 250–6000 Hz (ANSI S3.6, 1996). The fifth subject (coded as S1 in this paper) had hearing thresholds ≤20 dB HL in the frequency range 250–4000 Hz, and thresholds of 35 and 30 dB HL at 6000 Hz in the left and right ears, respectively, with recovery to within normal limits at 8000 Hz.

4. Procedures

For all conditions, masker level was fixed at 53 dBC in each masker channel while the target level was adapted. A four-alternative forced-choice (4AFC) paradigm with a 2-down 1-up stepping rule was employed to estimate the 70.7% criterion performance (Levitt, 1971). An individual adaptive track consisted of 10 reversals with the threshold computed as the arithmetic mean of the last six reversals. The initial step size for the adaptive track was 16 dB and the final, 2 dB. For all five subjects, data were collected for six listening conditions—two maskers (two-talker natural speech, two-talker reversed speech,) x three loudspeaker configurations (F-F, F-RF, F-FR). The masker was turned on prior to the initiation of each adaptive track and left running in loop play until the end of the track. Subjects responded using a button box with light-emitting diode (LED) lights that marked the four intervals, one of which contained the target. Feedback was provided via an LED, display that illuminated the target interval. For each condition, four adaptive tracks were obtained and threshold determined as the arithmetic mean of the four runs. All five subjects received the same order of listening conditions with two runs per condition before going on to the next. The sequence of presentation is shown in Table I.

At the beginning of the first listening session, subjects were verbally instructed, familiarized with the list of 20 target words (print and audio), and given practice runs in quiet and with the masker in F-F and F-RF loudspeaker conditions to familiarize them with the task. Listeners sat facing the front loudspeaker and were advised to not turn their heads; however, no head restraint was used. Subjects were given a brief break after the first six runs. A total of 12 runs completed the first listening session, with each subject returning for a second session to complete the second set of 12 adaptive tracks, again with a break after the first six runs of the

TABLE I. Order of presentation of listening conditions for Experiment 1. All subjects received the same sequence over two sessions. Maskers were two-talker speech natural and time reversed. Subjects received a final "Quiet" target track without any masker at the end of the second session.

Session 1	1, 2 Natural F-RF	3, 4 Natural F-F	5, 6 Natural F-FR	Break	7, 8 Reversed F-RF	9, 10 Reversed F-F	11, 12 Reversed F-FR	
Session 2	13, 14 Reversed F-FR	15, 16 Reversed F-F	17, 18 Reversed F-RF	Break	19, 20 Natural F-FR	21, 22 Natural F-F	23, 24 Natural F-RF	25 Quiet threshold

second session. Following the 12 adaptive tracks run with maskers, a "target only" track was run for each subject without the masker to ensure that target audibility in quiet was well below masker overall level. For all five subjects the quiet threshold was below 10 dBC.

B. Results

Figure 1 displays detection thresholds for the five subjects as well as the group mean for the two speech maskers. An additional set of four adaptive thresholds was obtained for S2 who showed high variability of F-F thresholds on the initial set (standard deviation of the mean=7.75). The second

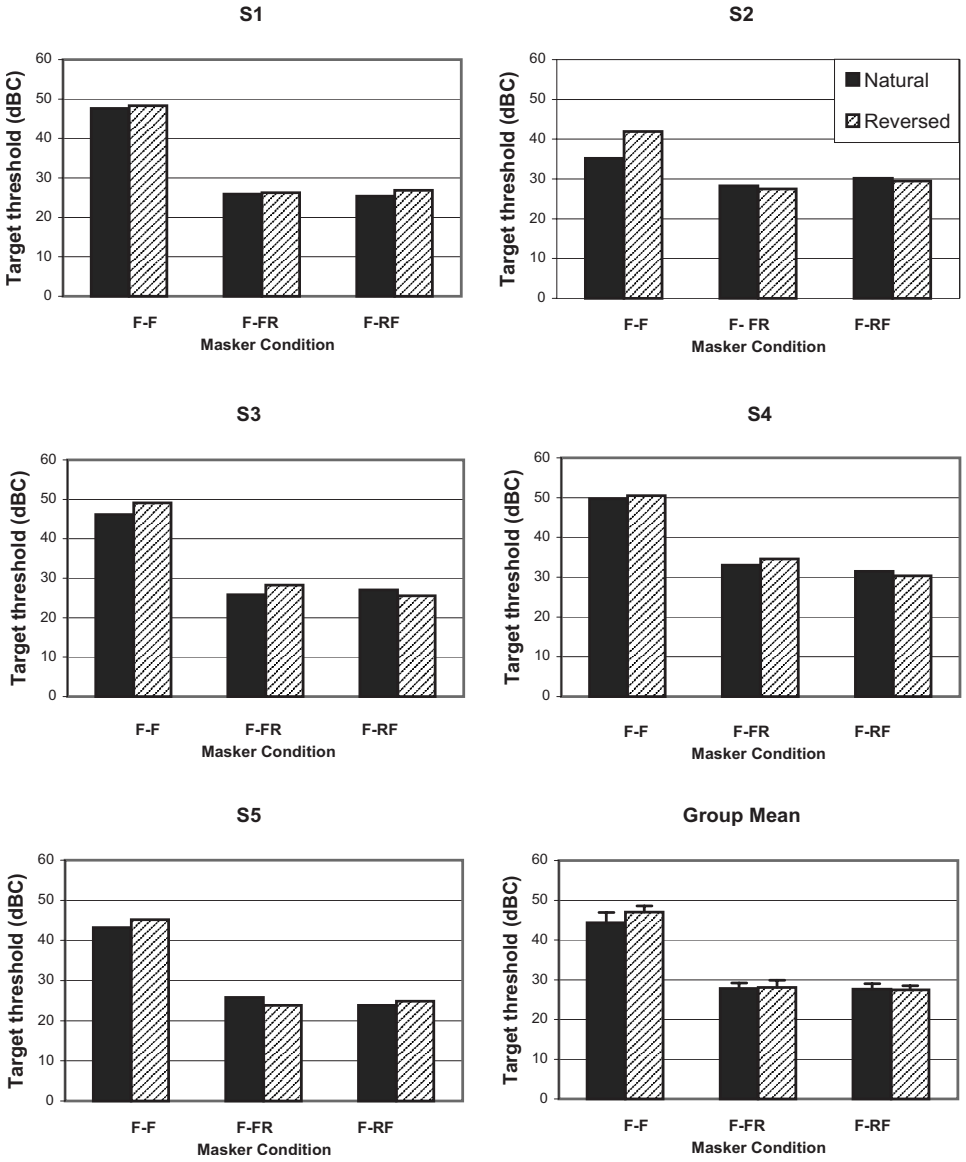

FIG. 1. Target thresholds for five subjects and group mean thresholds for the two-talker maskers in the F-F (nonspatial), F-FR (spatial, front-leading), and F-RF (spatial, right-leading) conditions. Solid bars represent the natural speech masker and the shaded bars represent the time-reversed speech masker. The group mean data (bottom left panel) also display one standard error of the mean.

set of four adaptive tracks also yielded a high standard deviation (6.37). Therefore, final threshold for the F-F condition for this subject was reported as the mean of all eight adaptive tracks.

As indicated earlier, masker output levels were fixed at 53 dBC for each loudspeaker. Threshold signal-to-noise (S-N) ratios are specified as the target level relative to 53 dBC. Across subjects target thresholds ranged from −2 dB S-N ratio for the nonspatial masker condition (S1) to nearly 30 dB below masker level for the RF spatial condition (S5). For individual subjects and the group, clear improvement in thresholds occurred for both spatial speech masker conditions (F-RF and F-FR) as compared to the F-F condition. Mean threshold improvement for the group (bottom right panel) in both spatial conditions relative to F-F was roughly 17 dB for the natural masker and 20 dB for the reversed masker. These values are substantially larger than those observed for nonsense sentence recognition using the same target talker and maskers (Freyman *et al.* 2001). It was also larger than the spatial advantages reported for detection by Helfer and Freyman (2005) for a different set of sentence stimuli using a single-interval task.

A noteworthy finding in the data displayed in Fig. 1 is the fact that both spatial speech masker conditions yielded substantial target threshold improvements, even though one (RF) was perceived well to the right of the target and the other (FR) was not. This is consistent with the findings reported in recognition studies (Freyman *et al.*, 1999; Brungart *et al.*, 2005; Rakerd *et al.*, 2006).

The improvement in the spatial conditions for the speech masker is interpreted here, as before, as indicating a release from informational masking. It is assumed that the high thresholds obtained in the F-F condition reflect approximately the same energetic masking as in the spatial conditions, with the difference of 17–20 dB reflecting informational masking. The fact that detection thresholds in the F-F condition were no better for the time-reversed speech than for the natural speech masker suggests that understandability of the masker is of virtually no importance in this detection experiment. If anything, the unintelligible time-reversed speech was a marginally more effective masker than the natural speech; the difference, however, was nonsignificant $t(4)=-1.962$; $p=0.121$. We suggest that this outcome is logical based on the brevity of the target signal used in this experiment (monosyllabic words) in conjunction with the nature of the detection task itself. That is, for our targets, the informational overlap between the target and masker occurs at the syllabic level, or, even more elementally, at the level of the phoneme. While natural and time-reversed speech streams are clearly differentiable from each other at the sentence level, they may be more similar to each other at the syllabic or phonemic level. For this reason, both types of speech maskers can be equally effective for the targets and the tasks we used. In contrast, when the same two maskers were used in a recognition experiment, the time-reversed masker was shown to be less effective than the natural masker in the F-F condition (Freyman *et al.*, 2001). Similar results have been reported for recognition by others (for example, Rhebergen *et al.*, 2005; Marrone *et al.*, 2007). These differences might suggest that the underlying psychometric functions created by time-reversed speech maskers might be steeper than those generated by natural speech, with the two functions converging at the lower end of the functions, near detection thresholds, and diverging at the higher end, where recognition performance is easier. In Freyman *et al.*, 2001 (Fig. 8, p. 2119) the backward and forward masker recognition performance functions for the F-F condition did converge at low S-N ratios (around −12 dB) while being divergent at high S-N ratios. However, because of the observed floor effects on performance, it is not possible to determine whether the convergence at low S-N ratios reflects an actual difference in the slopes of the two underlying functions.

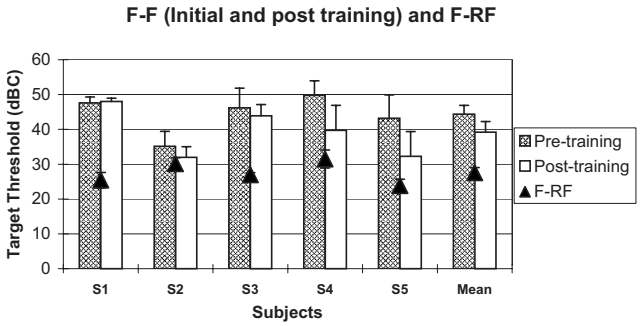

FIG. 2. Comparison of mean target thresholds for the initial four runs of Experiment 1 and the mean of the last four of ten training runs in the presence of the F-F (nonspatial) masker for five subjects and the group. Solid triangles are a replotting of the target thresholds obtained for the F-RF spatial masker condition for the natural speech masker shown in Fig. 1. Vertical bars show one standard deviation for individual subjects across the four pre- and post-training runs, and one standard error of the group mean data.

C. Learning effects in detection

In the F-F listening condition, four of the subjects showed higher variability in thresholds across the four adaptive tracks as compared to the F-RF condition. This raised the issue as to whether listeners could partially resolve the informational masking occurring in the F-F condition under some circumstances. To address this question, all five subjects were given ten successive adaptive tracks for the natural speech masker in the F-F condition. Results revealed variability across listeners, and while most listeners were able to improve performance on a given track, the gains did not carry over to successive tracks in a predictable or consistent manner.

Figure 2 treats the mean of the last four of the 10 runs as "post-training" and compares it to the mean of the initial four F-F runs of Experiment 1 for the five subjects and for the group. S1 shows no improvement, S2 and S3 show modest (less than 3 dB) gains, and S4 and S5 show large gains (about 10 dB). A paired-samples analysis of group mean data did not show a statistically significant difference between the pre- and post-training thresholds [$t(4)=2.316$; $p=.081$].

Also shown in Fig. 2 are the F-RF results for the natural speech masker reproduced from Fig. 1 (filled triangles). With the exception of S2, who did not show much spatial release from masking to begin with, subjects still showed substantial advantages in the F-RF condition even over the post-training

F-F results. Group mean differences between the F-F post-training and F-RF were statistically significant [$t(4)=3.182$; $p=.033$]. These results suggest that for most listeners, the release from masking provided by the spatial maskers was well beyond any improvement resulting from persistent listening and training. Inconsistent or limited responsiveness to extensive training has also been shown for nonspeech stimuli in informational masking tasks (for example, Neff and Callaghan, 1988; Neff and Dethlefs, 1995).

III. EXPERIMENT 2: REDUCING SPATIAL DIFFERENCES THROUGH RIGHT-CHANNEL ATTENUATION

In Experiment 1, listeners were able to benefit nearly equally from both the F-FR and F-RF spatial configurations, despite the fact that the spatial differences created are highly dissimilar. With the RF masker there is a substantial horizontal shift of the masker away from the target. As with any two-source sound, the RF masker is expected to have increased spaciousness relative to the F target, but the horizontal shift stands out as the most obvious difference. With the FR masker, the horizontal shift in position is expected to be fairly small and subjects may indeed use differences in spaciousness as a cue. Here, it may be worth noting an informal impression by the authors that listening to the F-FR condition does not require any special conscious or active effort to sort out how and where the target appears in relation to the masker. One listens for the target directly in front, and, at the relevant S-N ratios, the addition of the right loudspeaker causes it to stand out without any obvious increase in effort by the listener.

In the current experiment, target-masker spatial differences produced in the F-FR configuration were minimized by attenuating the right loudspeaker in graded steps. For comparison, we made the same right-loudspeaker attenuations for the RF masker, although in this case the time-intensity trade could create split or poorly defined spatial impressions. There were two parts to the experiment. In the first part, target words presented from the front loudspeaker were detected in the presence of the two-talker maskers as in Experiment 1, but now as a function of right-loudspeaker attenuation. In the second part, the spatial maskers were discriminated from the nonspatial (F) masker to determine the right-loudspeaker attenuation at which the addition of the right loudspeaker was undetectable. The goal of the combination of the two studies was to determine whether spatial release from masking could be observed with the front-right masker even when it was barely discriminable from the front masker.

A. Methods

Subjects were the same five listeners that completed Experiment 1. The speech target detection experiment used the same procedures as Experiment 1, except that the attenuation of the right loudspeaker in both the F-RF and F-FR conditions was increased from 0 dB (identical to the spatial conditions of Experiment 1) to 16 dB in 4 dB steps. Four adaptive tracks were obtained for each attenuation value. All subjects received the same stepwise order of presentation of conditions, starting with 0 dB and progressing to 16 dB of attenuation. One subject, (S3), received an additional attenuation step of 20 dB due to the fact that for this subject target threshold remained invariant as right loudspeaker attenuation was increased. The target words were the same as those used in Experiment 1 and the masker was the natural two-talker speech. Adaptive procedures and hardware setup were as described in Experiment 1.

For the discrimination task, 20 500 ms segments were randomly excised from a single channel version of the two-talker speech masker. Two kinds of average power relationships between segments were created. One set of 20 segments was allowed to retain the "natural" amplitude variation (N) (rms range=7.55 dB, standard deviation =1.93 dB), and in a second set the 20 segments were equated to each other for rms (Eq). The segments were converted into stereo waveforms by audio software (Cool Edit Pro), 20 ms linear rise/fall times were imposed, and 4 ms delays to one channel were added using zero padding as in Experiment 1. In each of the four intervals within a trial, an independent randomly selected segment was presented from among the 20 segments. In three of the intervals the stimulus was presented only from the front loudspeaker. The right loudspeaker was turned on during only one interval, to be detected by the listener, with feedback provided after each trial. Across trials, the right channel attenuation was adapted using the same tracking criteria and protocol employed in the other experiments described in this paper. Starting level for all stimuli was 40 dBC in both channels. Each subject received four adaptive tracks for each of four conditions in the following sequence: (1) FR-N–front-leading spatial masker with natural amplitude variation across segments; (2) RF-N–right-leading spatial masker with natural amplitude variation; (3) FR-Eq.–front-leading spatial masker rms equated; and (4) RF-Eq.–right-leading spatial masker rms equated. Subjects were the same five individuals who had participated in the previous experiments.

B. Results

The discrimination data are presented first (Fig. 3) so that they can be used to help interpret the target detection results later in Fig. 4. The abscissa marks the four spatial-amplitude combination conditions. The ordinate displays the relative level in the right loudspeaker required for discrimination from the front loudspeaker alone. Only small differences were noted between the results for the naturally roving level and the rms-equated stimuli. The purpose of the using the natural rove was to make the possible use of loudness cues available in the addition of the second loudspeaker less reliable. However, this appeared to have little effect. Because of the similarity across the "N" and "Eq." processing conditions, only the results for the natural rove will be discussed. There was a fairly wide variation across listeners, but the mean data show that the just-noticeable difference (jnd) for FR versus F alone occurred when the right loudspeaker was attenuated by 9 dB, whereas the jnd for the RF masker was obtained at a mean attenuation of approximately 16 dB.

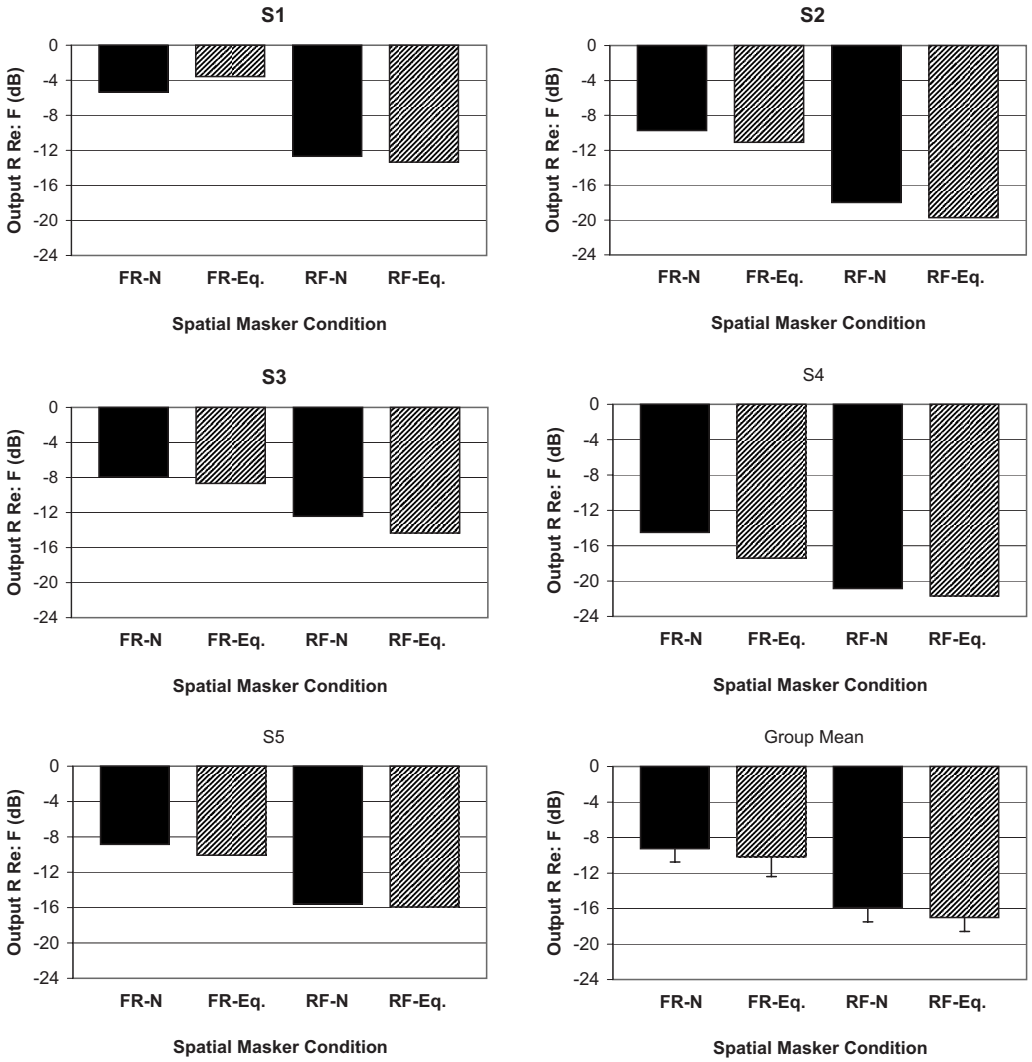

FIG. 3. Just-noticeable difference (jnd) in dB for discrimination of the spatial maskers from the nonspatial masker for five subjects and for the group. Jnd is expressed as right loudspeaker output relative to the fixed front loudspeaker output of 40 dBC. For each spatial masker, data are shown for the natural amplitude varying condition ("N") and the amplitude equated condition ("Eq."). Vertical bars for the group mean data display one standard error.

Paired samples *t*-test results showed that this was a statistically significant difference [$t(4)=10.6$; $p<0.001$]. The fact that the right loudspeaker output could be detected at a lower level for the RF masker may be attributable to a broadened or split spatial impression similar to that reported in lateralization studies in time-intensity trade conditions (Hafter and Jeffress, 1968; Hafter and Carrier, 1971). Rakerd and Hartmann (1985) suggested that the large variations in time-intensity trade data reported in the literature might be attributable to listeners expectations of the reliability and plausibility of conflicting directional cues.

Figure 4 displays target threshold for the two maskers as a function of right loudspeaker attenuation. Also shown for each subject is the F-F mean threshold for the last four of ten training runs (open square), which was assumed to be the value that the thresholds would approach as right loudspeaker output was attenuated and spatial release from informational masking was virtually eliminated. For all but S3, the functions for F-FR progressed from low thresholds averaging about 25 dB sound pressure level at 0 dB attenuation to higher thresholds approaching the F-F results at 16 dB attenuation. Thus, for the most part, 16 dB of attenuation effectively eliminated all effect of the right loudspeaker in producing release from masking. The progression varied among individual listeners. For example, S1's data show a large change between 12 and 16 dB while S4's data show the largest change with the first 4 dB of attenuation. For the other subject, S3, it appears that additional learning took place during the collection of the data such that performance actually improved as the experiment progressed and the attenuations increased. A 5 dB difference between F-FR with 0 dB attenuation (left-most data point) and the most attenuated condition occurred, but this was not nearly as large as the original F-FR versus F-F difference of 19 dB for that subject. The relationships between the F-RF and F-FR results are difficult to summarize on an individual subject basis, except to state that when there was a difference between the two thresholds the F-RF was usually lower. When thresholds were averaged across subjects (bottom-right panel), the RF masker provided a slightly but consistently greater release from masking relative to the FR masker for each right loudspeaker attenuation value (range of 0.55–2.93 dB). How-

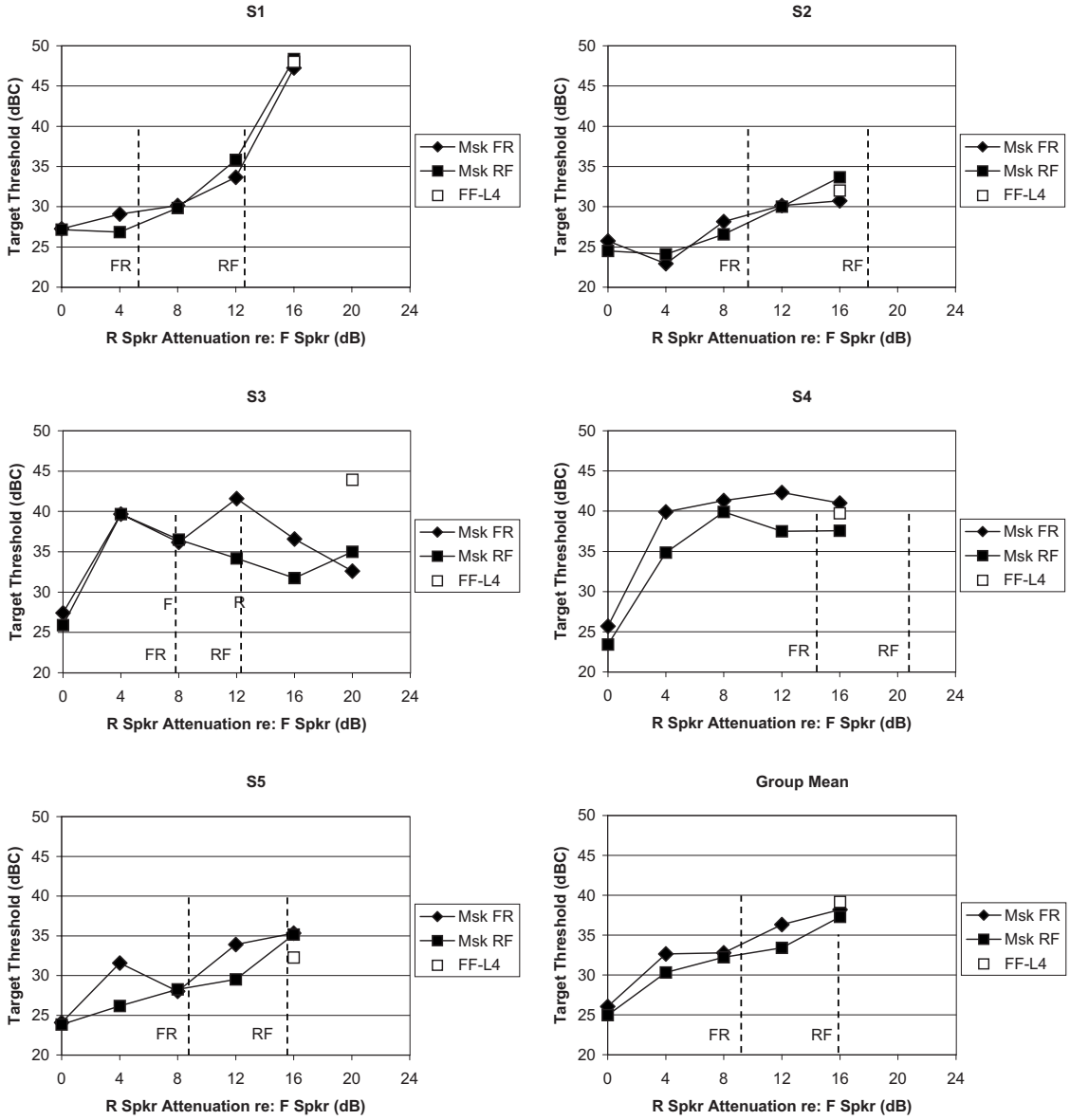

FIG. 4. Target thresholds for two spatial maskers for five subjects and for the group as a function of right loudspeaker attenuation. Dashed lines intersecting the abscissa replot the jnd values for each spatial masker from Fig. 3.

ever, analysis of variance results showed that these were not statistically significant [$F(4,1)=4.5$; $p=0.102$].

For comparison with the discrimination data, the natural-rove data from Fig. 3 are replotted in Fig. 4 as indicated by the dashed lines. These lines indicate the right loudspeaker attenuation values along the abscissa at which the spatial masker was just barely discriminable from the nonspatial. The comparison reveals the relationship between the threshold level of the right loudspeaker presentation and the effectiveness of the addition of the right loudspeaker in releasing masking for the front-only target words. Consider the group data first (bottom-right panel). The average just noticeable difference (jnd) for the F-FR configuration was approximately 9.3 dB of attenuation. Were the FR masker to be used with this right loudspeaker output level, the interpolated target detection threshold would be approximately 33.9 dBC (that is, the point at which the FR dashed line intersects the filled diamond trace depicting the FR threshold function). However, the nonspatial F-F threshold (unfilled square) was actually about 39.2 dBC. This means that 5.3 dB improvement in target detection for the FR masker took place at right loudspeaker attenuation values where the discrimination between FR and F was below 71% correct. In the case of one subject, S1, *most* of the improvement occurred at attenuations where FR and F configurations were less than one jnd apart. Thus, it must be concluded that spatial release from masking can occur with barely discriminable spatial differences between target and masker.

As had been shown in Fig. 3, detection of the signal from the right loudspeaker occurred at lower right loudspeaker output levels for the RF configuration than for FR (distance between the dashed lines along the abscissa). However, when the RF masker was used with the front-only target words, this RF-FR difference did not translate to correspondingly large or consistent changes in release from masking of the target words (difference between squares and

diamonds along the ordinate). This may have been because at low right-loudspeaker levels the RF masker likely produced split images, with one image remaining near the target and creating the same type of confusion caused by the front loudspeaker masking alone. If so, the lack of masking release would not be unlike what was observed by Brungart *et al.* (2005) and Rakerd *et al.* (2006). In those studies, masking release in the spatial maskers disappeared at long delays between the front and right loudspeakers where it was assumed that images were split.

IV. DISCUSSION

In these studies, a standard psychophysical multiple-interval forced-choice detection procedure using monosyllabic words revealed substantial evidence of spatial release from informational masking for speech stimuli. We believe that the general paradigm holds promise for helping to understand the commonalties and differences in what has been called informational masking for both speech and nonspeech stimuli.

In Experiment 1 we gathered basic information about informational masking and masking release using this word detection paradigm. We investigated (1) the S-N ratios at which base line performance could be obtained, (2) the size of the spatial release from informational masking that could be observed, (3) the dependence of informational masking on the understandability of the masker, and (4) the kinds of spatial differences that are required for spatial release from masking.

With regard to (1) above, the detection of words in the presence of two-talker masking was shown to produce thresholds of approximately −25 dB S-N ratio in spatial conditions, where informational masking was presumably minimal. By contrast, threshold *recognition* performance for sentence targets occurred around −10 dB S-N ratio in a recent study also using two-talker speech maskers (Freyman *et al.*, 2007). Others have found low threshold S-N ratios in recognition studies in spatial conditions using single-talker maskers (e.g., Hawley *et al.* 2004); however, less informational masking is expected than with two-talker maskers (e.g., Yost *et al.*, 1996; Freyman *et al.*, 2004). The advantage of the low threshold S-N ratio observed in the present study is that there is ample headroom to observe the effects of informational masking that may occur in nonspatial conditions without bumping into what might be a ceiling for informational masking at around 0 dB S-N ratio. The availability of this headroom could be useful for studying how a variety of stimulus manipulations affect informational masking, and could also make it easier to study informational masking in populations that have higher base line thresholds (see Arbogast *et al.*, 2005).

The second question concerned the size of the spatial release from masking that would be observed with these methods and stimuli. The answer was 17–20 dB for the natural and time-reversed speech maskers, respectively. All of the effect is presumed to be the result of release from informational masking. The basis for this assumption is that the spatial conditions used in this study have not been shown to produce masking release with purely energetic maskers (e.g., Freyman *et al.*, 1999; Brungart *et al.*, 2005; Rakerd *et al.*, 2006), even when using this very same detection paradigm (Freyman *et al.*, 2006). Under the further assumption that informational masking is approximately zero in the spatial conditions, a spatial release of 17–20 dB is indicative of informational masking of the same magnitude. This value is considerably larger than the approximately 5 dB of informational masking estimated with the same maskers for sentence-level target recognition (Helfer and Freyman, 2005; Freyman *et al.*, 2007). Helfer and Freyman (2005) also obtained detection thresholds for sentences against these maskers; the informational masking measured was about 6 dB, suggesting that the use of word level stimuli was a major contributing factor to the greater informational masking seen in the current experiment. A comparison of the word detection thresholds obtained in the current work with the sentence detection thresholds from Helfer and Freyman (2005) revealed that virtually all of the difference occurred in the nonspatial (F-F) condition (−17 dB S-N ratio in the earlier paper, and approximately −7 dB in the current study). This suggests that informational masking was greater for words than for sentences.

In restricting the task to word detection, our results more closely approximated the size of informational masking effects reported for brief nonspeech stimuli by other investigators. For instance, Kidd *et al.* (1998) obtained up to 15 dB of masking release at a spatial separation of 60° for pattern recognition in the presence of informational maskers. Using a detection paradigm, Oxenham *et al.* (2003) found on the average 10 and 25 dB of informational masking for a 1 kHz tone burst for musicians and nonmusicians, respectively. Oh and Lutfi (1999) estimated 11–12 dB informational masking for everyday sound maskers when they were easily recognized by subjects. It is also possible to observe large amounts of informational masking (on the order of 15 dB) in speech recognition studies (e.g., Arbogast *et al.*, 2005). However, these results have been reported for stimuli such as filtered coordinate response measure (CRM) sentences (Bolia *et al.*, 2000) and conditions specifically designed to maximize informational masking and minimize energetic masking.

The fact that spatial release is so large in our detection experiment may reflect a difference in the cues listeners are likely to use in the spatial and nonspatial conditions. In the two spatial conditions, F-RF and F-FR, the spatial image produced by the masker was different from that of the target, and the listener had only to detect the presence of any stimulus appearing to come from the front loudspeaker. By contrast, in the F-F condition both target and masking sounds appeared only from the front. In order to extract the target from the midst of an ongoing two-talker mixture of voices it was necessary to attend to other cues. At a minimum, subjects could have listened for the presence of the words with which they had been familiarized, abrupt amplitude changes, the voice characteristics of the target talker, or other linguistic features that separated the target speech from the masker speech. The threshold S-N ratios in the F-F condition (−7 dB average of natural and time-reversed maskers) suggest that none of these characteristics was easy to extract until the

target was within a few dB of the speech level of the individual talkers within the two-talker masking complex.

The third question considered in Experiment 1 was whether time reversing the masker would affect the amount of informational masking measured with this task. Results showed that masker time reversal produced no reduction in masking in either the spatial or nonspatial conditions. Indeed, a slight (though statistically insignificant) increase of about 2 dB was noted for the time-reversed speech masker. By contrast, sizable reduction of masking efficacy has been shown with masker time reversals for the recognition of short "everyday" sentences (Rhebergen *et al.*, 2005), nonsense sentences (Freyman *et al.*, 2001), and CRM stimuli (Marrone *et al.*, 2007). In recognition tasks, the absence of meaningful words in the masker could eliminate one of the main sources of confusion which, at least for CRM stimuli, is clearly supported by error patterns (Kidd *et al.*, 2005). By contrast, understandability could be less relevant in the detection paradigm used in this study, because any detectable portion of the target (such as a phoneme) can improve performance. The listener is able to use—and may even be dependent on—the phonological or linguistic information present in overlapping masker-target consonant and vowel segments rather than on word meaning and context cues. Thus, time reversing the speech masker does not reduce confusability between target and masker, because the confusion lies not at the word level but at the segmental level.

The absence of an effect of masker time reversal suggests that the speech-on-speech masking measured in the current study could be quite unlike the type of informational masking revealed when, for example, two or three CRM sentences compete with one another for attention (e.g., Ericson and McKinley, 1997; Arbogast *et al.*, 2002; Brungart and Simpson, 2002; Brungart *et al.*, 2005; Kidd *et al.*, 2005; Shinn-Cunningham *et al.*, 2005; Rakerd *et al.*, 2006). The difference may lead to the concern that a common terminology is being applied to different masking processes. However, although it would appear that in this study we measured the effect of a type of auditory confusion that is closer to what is seen in nonspeech informational masking, it would be premature to conclude that the speech-on-speech masking measured in this study had no linguistic basis. Even the time-reversed masker could be clearly recognized as speech. Listeners may have imposed the phonological and lexical rules of English to "extract" the target segment from the speech surround, in both natural and time-reversed speech maskers. It will likely require more work to identify the linguistic and nonlinguistic factors contributing to informational masking in this study.

A final question asked in Experiment 1 concerned the kinds of spatial differences between target and masker that are effective in releasing informational masking. The F-RF condition, where the masker from the right loudspeaker leads the presentation from the front loudspeaker, produces a dramatic target/masker spatial difference in which the masker is heard well to the right of the target. The F-FR condition, in which the front masker leads the right masker, still produces a noticeable spatial difference, but the change is more difficult to characterize. According to accounts from several papers (e.g., Shinn-Cunningham *et al.*, 1993; Litovsky and Macmillan, 1994; Chiang and Freyman, 1998) as well as classic texts on the spatial hearing (Blauert, 1997), the auditory image produced by a two-source masker such as the FR masker is expected to be less punctate than that of the single-source, front-only target, with the center of gravity shifted to the right by perhaps 5–10°. Despite the seemingly less obvious target-masker differences in the F-FR configuration, in *recognition* experiments release from masking relative to F-F was nearly the same for F-RF and F-FR configurations (Freyman *et al.*, 1999; Brungart *et al.*, 2005; Rakerd *et al.*, 2006). The current paper (Fig. 1) shows that this same result is obtained also for the word detection task. This suggests that small differences in spatial separation and/or spatial width cues are no less efficient in releasing informational masking than the dramatic spatial separation provided by the right-leading F-RF masker. The implication is that target-masker perceptual spatial differences other than spatial *separation* can be potent facilitators of release from masking.

In Experiment 2 we exploited the forced-choice word detection technique, and the large spatial release it produces, to explore further the subtlety of spatial difference required for masking release in the F-FR configuration. We titrated the more than 15 dB of difference in thresholds available between F and FR configurations by attenuating the output of the right loudspeaker in F-FR until the results were essentially no different than F-F. The functions were then related to the results of a second portion of the experiment, which determined the threshold of discrimination between F and FR maskers with systematic attenuation of the right loudspeaker in the FR condition. For comparison, all conditions were repeated with the RF masker, but still with attenuation of the right loudspeaker.

The 4AFC task with a two-down one-up stepping rule estimates a d' of 1.53 (Macmillan and Creelman, 1991). Using that threshold criterion, the mean just discriminable difference between FR and F alone occurred when the right loudspeaker was attenuated by 9 dB. With that same 9 dB of attenuation, an average of 5 dB of spatial release from masking still occurred in the F-FR configuration. Thus, it must be concluded that spatial release from masking can be observed with barely discernable target-masker spatial differences. This is consistent with other studies showing that small changes in masker-target spatial relationships can substantially affect target recognition (e.g., Brungart *et al.*, 2005; Gallun *et al.*, 2005).

There are also at least two observations in the literature where the opposite occurred, i.e., presumably large spatial differences between target and masker produced no release from masking (Brungart *et al.*, 2005; Rakerd *et al.*, 2006). In both of these cases release from masking disappeared when the delay in a two-source speech masker was increased from 32 to 64 ms. Presumably, the 64 ms delay produced split images for the masker, one near the front target and one near the lateral masker. Informational masking is attributed to the masker image located near the target. In the current Experiment 2, attenuations applied to the right loudspeaker were much more easily detected in the RF configuration than the FR configuration, but the release from masking created was

only marginally increased (Fig. 4) and did not reach statistical significance. It is possible that the severe attenuation applied to the leading (right) loudspeaker in the RF case produced a split image due to the time-intensity incompatibility and may have failed to produce increased release from masking for the same reasons presumed to occur in the Brungart *et al.* (2005) and Rakerd *et al.* (2006) studies. If this is indeed the explanation, then it is fascinating that listeners cannot use this additional synchronized separated image to inform them of the temporal characteristics of the masker in order better focus on the target. This kind of active study of the masker in order to learn what to ignore does not appear to be used or useful. Rather, barely discriminable differences between the spatial image produced by the target and the most similar image produced by the masker seem to facilitate improved attention on the target.

ACKNOWLEDGMENTS

The authors would like to thank Gail Brown, John Ackland Jones, Laurel Slongwhite, and Lauren Sullo for their assistance in data collection. We would also like to thank three anonymous reviewers and the associate editor for their thoughtful comments on earlier versions of this paper. This work was supported by a grant from the National Institute on Deafness and other Communicative Disorders (DC 01625).

ANSI (**1996**). ANSI S3.6-1996, "Specifications for audiometers" (American National Standards Institute, New York).

Arbogast, T. L., Mason, C. R., and Kidd Jr., G. (**2002**). "The effect of spatial separation on informational and energetic masking of speech," J. Acoust. Soc. Am. **112**, 2086–2098.

Arbogast, T. L., Mason, C. R., and Kidd Jr., G. (**2005**). "The effect of spatial separation on informational masking of speech in normal-hearing and hearing-impaired listeners," J. Acoust. Soc. Am. **117**, 2169–2180.

Blauert, J. (**1997**). *Spatial Hearing* (MIT, Cambridge, MA).

Bolia, R. S., Nelson, W. T., Ericson, M. A. and Simpson, B. D., (**2002**). "A speech corpus for multitalker communications research," J. Acoust. Soc. Am. **102**, 1065–1066.

Brungart, D. S., and Simpson, B. D. (**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., Ericson, M. A., and Scott, K. R. (**2001**). "Informational and energetic masking effects in the perception of multiple simultaneous talkers," J. Acoust. Soc. Am. **110**, 2527–2538.

Brungart, D. S., and 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.

Brungart, D. S., Simpson, B. D., and Freyman, R. L. (**2005**). "Precedence-based speech segregation in a virtual auditory environment," J. Acoust. Soc. Am. **118**, 3241–3251.

Byrne, D., Dillon, H., Tran, K., Arlinger, S., Wilbraham, K., Cox, R., Hagerman, B., Hetu, R., Kei, J., Lui, C., Kiessling, J., Nasser Kotby, M., Nasser, N. H. A., El Kholy, W. A. H., Nakanishi, Y., Oyer, H., Powell, R., Stephens, D., Meridith, R., Sirimanna, T., Tavarkiladze, G., Frolenkovgi, G. I., Westerman, S., and Ludvigsen, C. (**1994**). "An international comparison of long-term average speech spectra," J. Acoust. Soc. Am. **96**, 2108–2120.

Carhart, R., Johnson, C., and Goodman, J. (**1975**). "Perceptual masking in mutliple sound backgrounds," J. Acoust. Soc. Am. **45**, 694–703.

Carhart, R., Tillman, T. W., and Greetis, E. S. (**1969**). "Perceptual masking of spondees by combinations of talkers," J. Acoust. Soc. Am. **58**, 694–703.

Chaiklin, J. B. (**1959**). "The relation among three selected auditory speech thresholds," J. Speech Hear. Res. **2**, 237–243.

Chiang, Y. C., and Freyman, R. L. (**1998**). "The influence of broadband noise on the precedence effect," J. Acoust. Soc. Am. **104**, 3039–3047.

Durlach, N. I., Mason, C. R., Gallun, F. J., Shinn-Cunningham, B. G., Colburn, H. S., and Kidd, G. (**2005**). "Informational masking for simultaneous nonspeech stimuli: Psychometric functions for fixed and randomly mixed maskers," J. Acoust. Soc. Am. **118**, 2482–2497.

Edmonds, B. A., and Culling, J. F. (**2005**). "The role of head-related time and level cues in the unmasking of speech in noise and competing speech," Acta. Acust. Acust. **91**, 546–553.

Ericson, M. A., and McKinley, R. L. (**1997**). "The intelligibility of multiple talkers separated spatially in noise," in *Binaural and Spatial Hearing in Real and Virtual Environments*, edited by R. H. Gilkey, and T. B. Anderson (Lawrence Erlbaum, Hillsdale, N. J.)

Freyman, R. L., Helfer, K. S., McCall, D. D., and 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., and Helfer, K. S. (**2001**). "Spatial release from informational masking in speech recognition," J. Acoust. Soc. Am. **109**, 2112–2122.

Freyman, R. L., Balakrishnan, U., and Helfer, K. S. (**2004**). "Effect of number of masking talkers and auditory priming on informational masking and speech recognition," J. Acoust. Soc. Am. **115**, 2246–2256.

Freyman, R. L., Balakrishnan, U., and Helfer, K. S. (**2006**). "Exploration of spatial release from masking in a simulation of cochlear implant listening," paper presented at the *29th Mid-Winter Meeting of the Association for Research in Otolaryngology*, Baltimore, MD.

Freyman, R. L., Helfer, K. S., and Balakrishnan, U. (**2007**). "Variability and uncertainty in masking by competing speech," J. Acoust. Soc. Am. **121**, 1040–1046.

Gallun, F. J., Mason, C. R., and Kidd Jr., G. (**2005**). "Binaural release from informational masking in a speech recognition task," J. Acoust. Soc. Am. **118**, 1614–1625.

Hafter, E. R., and Jeffress, L. A. (**1968**). "Two-image lateralization of tones and clicks," J. Acoust. Soc. Am. **44**, 563–569.

Hafter, E. R., and Carrier, L. A. (**1972**). "Binaural interaction in low-frequency stimuli: the inability to trade time and intensity completely," J. Acoust. Soc. Am. **51**, 1852–1862.

Hall, J. W., Grose, J. H., Buss, E., and Dev, M. B. (**2002**). "Spondee recognition in a two-talker masker and a speech-shaped noise masker in adults and children," Ear Hear. **23**, 159–165.

Hawley, M. L., Litovsky, R. Y., and Culling, J. F. (**2004**). "The benefit of binaural hearing in a cocktail party: Effect of location and type of interferer," J. Acoust. Soc. Am. **115**, 833–843.

Helfer, K. S., and Freyman, R. L. (**2005**). "The role of visual speech cues in reducing energetic and informational masking," J. Acoust. Soc. Am. **117**, 842–849.

Kidd Jr., G., Arbogast, T. L., Mason, C. R., and Gallun, F. J. (**2005**). "The advantage of knowing where to listen," J. Acoust. Soc. Am. **118**, 3804–3815.

Kidd Jr., G., Mason, C. R., and Rohtla, T. L. (**1995**). "Binaural advantage for sound pattern identification," J. Acoust. Soc. Am. **98**, 1977–1986.

Kidd Jr., G., Mason, C. R., Rohtla, T. L., and 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.

Levitt, H. (**1971**). "Transformed up-down methods in psychoacoustics," J. Acoust. Soc. Am. **49**, 467–477.

Litovsky, R. Y., and Macmillan, N. A. (**1994**). "Sound localization precision under conditions of the precedence effect: Effects of azimuth and standard stimuli," J. Acoust. Soc. Am. **96**, 752–758.

Lutfi, R. A., Kistler, D. J., Callahan, M. R., and Wightman, F. L. (**2003**). "Psychometric functions for informational masking," J. Acoust. Soc. Am. **114**, 3273–3282.

Macmillan, N. A., and Creelman, C. D. (**1991**). *Detection Theory: A User's Guide.*, (Cambridge U. P., Cambridge, England).

Marrone, N. L., Mason, C. R., and Kidd Jr., G. (**2007**) "Spatial release from speech-on-speech masking with symmetrically placed maskers," paper presented at the *30th Mid-winter Meeting of the Association for Research in Otolaryngology*, Denver, CO.

Neff, D. L., and Callaghan, B. P. (**1988**). "Effective properties of multicomponent simultaneous maskers under conditions of uncertainty maskers," J. Acoust. Soc. Am. **83**, 1833–1838.

Neff, D. L., and Dethlefs, T. M. (**1995**). "Individual differences in simultaneous masking with random frequency, multicomponent maskers," J. Acoust. Soc. Am. **98**, 125–134.

Neff, D. L., and Green, D. M. (**1987**). "Masking produced by spectral uncertainty with multicomponent maskers," Percept. Psychophys. **41**, 409–415.

Oh, E. L., and Lutfi, R. A. (**1999**). "Informational masking by everyday

sounds," J. Acoust. Soc. Am. **106**, 3521–3528.
Oxenham, A. J., Fligor, B. J., Mason, C. R., and Kidd Jr., G. (**2003**). "Informational masking and musical training," J. Acoust. Soc. Am. **114**, 1543–1549.
Rakerd, B., and Hartmann, W. M. (**1985**). "Localization of sound in rooms: The effects of a single reflecting surface," J. Acoust. Soc. Am. **78**, 524–533.
Rakerd, B., Aaronson, N. L., and Hartmann, W. M. (**2006**). "Release from speech-on-speech masking by adding a delayed masker at a different location," J. Acoust. Soc. Am. **119**, 1597–1605.
Rhebergen, K. S., Versfeld, N. J., and Dreschler, W. A. (**2005**). "Release from informational masking by time reversal of native and non-native interfering speech," J. Acoust. Soc. Am. **118**, 1274–1277.
Richards, V. M., and Neff, D. L. (**2004**). "Cueing effects for informational masking," J. Acoust. Soc. Am. **115**, 289–300.
Shinn-Cunningham, B. G., Zurek, P. M., and Durlach, N. I. (**1993**). "Adjustment and discrimination measurements of the precedence effect," J. Acoust. Soc. Am. **93**, 2923–2932.
Shinn-Cunningham, B. G., Ihlefeld, A., Satyavarta, and Larson, E. (**2005**). "Bottom-up and top-down influences on spatial unmasking," Acta. Acust. Acust. **91**, 967–979.
Thurlow, W. R., Silverman, S. R., Davis, H., and Walsh, T. E. (**1948**). "A statistical study of auditory tests in relation to the fenestration operation," Laryngoscope **58**, 43–66.
Watson, C. S., Kelly, W. J., and Wroton, H. W. (**1976**). "Factors in the discrimination of tonal patterns II: Selective attention and learning under various levels of stimulus uncertainty," J. Acoust. Soc. Am. **60**, 1176–1185.
Wright, B. A., and Saberi, K. (**1999**). "Strategies to detect auditory signals in small sets of random maskers," J. Acoust. Soc. Am. **105**, 1765–1775.
Yost, W. A., Dye, R. H., and Sheft, S. (**1996**). "A simulated cocktail party with up to three sound sources," Percept. Psychophys. **58**, 1026–1036.