# Speech localization in a multitalker mixture[a)]

Norbert Kopčo[b)]
*Department Of Cybernetics and AI, Technical University of Košice, Košice 04001, Slovakia*

Virginia Best and Simon Carlile
*School of Medical Sciences, University of Sydney, Sydney, New South Wales 2006, Australia*

An experiment was performed that measured, for the frontal audio-visual horizon, how accurately listeners could localize a female-voice target amidst four spatially distributed male-voice maskers. To examine whether listeners can make use of *a priori* knowledge about the configuration of the sources, performance was examined in two conditions: either the masker locations were fixed (in one of five known patterns) or the locations varied from trial to trial. The presence of maskers disrupted speech localization, even after accounting for reduced target detectability. Averaged across all target locations, the rms error in responses decreased by 20% when *a priori* knowledge about masker locations was available. The effect was even stronger for the target locations that did not coincide with the maskers (error reduction of 36%), while no change in errors was observed for targets coinciding with maskers. The benefits were reduced when the target-to-masker intensity ratio was increased or when the maskers were in a pattern that made it difficult to make use of the *a priori* information. The results confirm that localization in speech mixtures is modified by the listener's expectations about the spatial arrangement of the sources.
© *2010 Acoustical Society of America.* [DOI: 10.1121/1.3290996]

## I. INTRODUCTION

It is known that spatial factors play a role in the ability of listeners to understand speech in noisy or complex listening environments. For example, acoustical advantages are provided by the spatial separation of speech from competing noise (Zurek, 1993; Bronkhorst, 2000). In the case of multiple competing talkers, it is believed that spatial differences also enable the correct "sorting" of the acoustic mixture into different sources and enable listeners to direct attention selectively to one source to enhance its processing (Yost *et al.*, 1996; Freyman *et al.*, 1999; Brungart, 2001; Shinn-Cunningham, 2008). In addition, the ability to rapidly locate a talker of interest in order to focus on them visually is clearly an important aspect of human communication. Despite the various roles of spatial hearing in dealing with competing speech sources, we know surprisingly little about how accurately listeners can localize in speech mixtures.

Under ideal conditions, humans can localize single broadband sounds within a few degrees of accuracy (Mills, 1958; Wightman and Kistler, 1989; Carlile *et al.*, 1997; Best *et al.*, 2005). When presented against a background of noise, localization does not suffer until the signal-to-noise ratio is negative, due in part to the fact that the *detection* of the target becomes compromised (Good and Gilkey, 1996; Good *et al.*, 1997; Abouchacra *et al.*, 1998; Lorenzi *et al.*, 1999). Few studies have examined more complex situations involving multiple talkers (for a summary see Faller and Merimaa, 2004). Hawley *et al.* (1999) measured localization of a known target sentence in the presence of one to three unknown sentences. Their task was a 1-of-7 loudspeaker identification (30° spacing), and they found that performance was relatively accurate (around 70%) and not affected significantly by the number or configuration of the maskers. In a similar paradigm with one to four distractors, Drullman and Bronkhorst (2000) found poorer performance (around 50%) for a 1-of-5 loudspeaker identification task (45° spacing). Finally, Simpson *et al.* (2006) required listeners to detect and localize a known word in a mixture of one to five synchronous words. The number and configuration of the maskers were varied from trial to trial. The authors found that localization errors increased systematically with the number of maskers even when miss trials (where the subject could not detect the target) were excluded from the analysis.

Few previous studies asked whether providing *a priori* information that can be used to direct automatic or strategic attention can improve sound localization (Spence and Driver, 1994; Sach *et al.*, 2000; Kopčo *et al.*, 2001). In these studies, the effect of cuing the target location was small: improvements in reaction times were observed (Spence and Driver, 1994), but little (Sach *et al.*, 2000) or no (Kopčo *et al.*, 2001) improvement in localization accuracy was found. A possible explanation for these weak effects is that the target was presented in isolation, and thus a lot of redundant information about its location was available. More complex scenes, in which sources compete for attention, may be more likely to reveal effects of *a priori* information.

The current study aimed to examine, in a realistic, complex listening situation containing competing sound sources, whether *a priori* knowledge of the location of maskers might

---
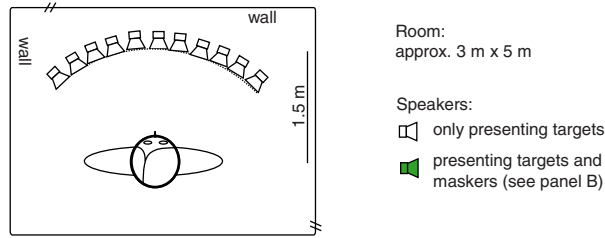
FIG. 1. (Color online) (A) Eleven loudspeakers evenly spaced in front of the listener (10° separation) were used to present stimuli. (B) Maskers were presented from loudspeaker locations arranged into one of five masker patterns. Each pattern had four maskers, presented concurrently with the target.

modify how they affect target localization performance. To this end, we measured the accuracy with which listeners can localize a speech target presented from a random location in the presence of four speech maskers. While all maskers were spatially separated from each other, targets could occur anywhere (including at masker locations). The maskers were presented in one of five masker patterns (see Fig. 1). In separate blocks, we either varied the masker pattern randomly from trial to trial (Mixed condition) or kept one of the masker patterns fixed throughout a block, providing the subject with information about the maskers' locations at the beginning of the block (Fixed condition). We expected that subjects would be able to use the masker location information in the Fixed condition to improve their localization performance (with regard to Mixed condition). For example, by actively suppressing the masker locations or actively attending away from masker locations, performance might improve for targets at nonmasker locations (but perhaps worsen for targets colocated with maskers).

The second goal of this study was to examine whether the extent to which *a priori* information can be utilized is influenced by the complexity of the masker distribution. Therefore, we included two simple masker configurations (patterns 1 and 2 in Fig. 1, in which all masker locations and all nonmasker locations were clustered together), two intermediate configurations (pattern 3, in which the masker locations were in one cluster while the nonmasker locations were in two clusters, and pattern 4, with the opposite arrangement), and one complex configuration (pattern 5, in which both the masker and the nonmasker locations were approximately evenly distributed). We expected the benefit of *a priori* knowledge to be larger for the simpler configurations, based on the assumption that spatial attention or masker suppression can be more efficiently applied to a single region than to multiple regions.

## II. METHODS

### A. Subjects

Seven subjects participated, one female and six males between the ages of 18 and 50 years. All had normal hearing

by self-report and gave informed consent as required by the University of Sydney's Human Research Ethics Committee.

### B. Environment, stimuli, and setup

The experiment took place in an empty office of dimensions $3 \times 5 \times 2.5$ m (width $\times$ length $\times$ height). The room had carpet on the floor, a concrete ceiling, and plasterboard on the lateral and back walls. The front wall was exposed brick and contained a large window that was filled with thick wool matting to reduce reflections and block out most of the incoming light; as a result visibility of the loudspeakers was minimal. Eleven loudspeakers on stands were positioned with a spacing of 10° to form a horizontal arc of radius 1.5 m [Fig. 1(A)] at the level of an average listener's ears when standing (1.6 m). The target was presented from 1 of the 11 loudspeakers at random. On masker trials, four simultaneous maskers were arranged in one of five configurations [Fig. 1(B)]. Listeners were aware that targets could fall on masker locations.

Speech materials were taken from a corpus of monosyllabic words recorded at Boston University's Hearing Research Center (Kidd *et al.*, 2008a). The target was the word "two" spoken by one of the female voices in the corpus. Maskers were nondigit words spoken by the eight male talkers in the corpus, and included names (e.g., "Jane"), verbs (e.g., "found"), adjectives (e.g., "red"), and nouns (e.g., "toys"). The four masker words were drawn randomly from this set with the constraints that they were four different words spoken by four different male voices. On catch trials (see below), the target was replaced by another randomly chosen masker word.

The maskers were all longer in duration than the target, and because they were spoken by male talkers they generally had a broader spectrum than the target. Thus, they did not provide substantial spectral or temporal gaps in which the listener could have a good "glimpse" at the target (as would be the case if the target and maskers were sentences containing natural silent breaks). On the other hand, keeping the target word fixed made the task of identifying the target easier than in natural situations in which the word spoken by the target speaker varies continuously (as do acoustic, phonetic, prosodic, and other characteristics of the utterance). This as well as the difference in gender between the target and the maskers enabled listeners to focus on localizing the target accurately without any ambiguity about which sound to localize.

Target words were presented at a level of approximately 60 dB SPL(A). Maskers were all equal in level, but presented at one of two levels relative to the target in order to vary the difficulty of the task. In the easier task, each masker was equal in level to the target [target-to-masker ratio (TMR) of 0 dB]; in the more difficult task, each masker was 5 dB louder (TMR of −5 dB).

The experiment was run in MATLAB on a PC-compatible control computer. On each trial, the appropriate stimuli were loaded from files stored on the computer hard disk (at a sampling rate of 48 kHz) and sent via a multichannel soundcard (RME Fireface 400), D/A converter (Apogee DA-16x),

J. Acoust. Soc. Am., Vol. 127, No. 3, March 2010

Kopčo *et al.*: Speech localization in a mixture    1451

and amplifier (Ashley Powerflex 6250), to Tannoy V6 loud-speakers.

Subjects indicated their responses by pointing their head in the perceived direction of the target and pressing a hand-held response button. A headtracker (Intersense IC3) mounted on a plastic headband was used to measure the orientation of the head at the time of response.

## C. Procedures

Before a session, the experimenter positioned the subject such that he/she was in the center of the loudspeaker array with his/her head pointing to 0° azimuth, and this location was recorded by the headtracker as the reference position. Before the stimulus was played on each trial, the subject was required to orient their head to this position and feedback was given by way of a small light-emitting diode display positioned above and behind the speaker array.

In runs containing maskers, it was expected that there would be a number of trials in which the listener would not be able to detect the target, and localization responses would either be to one of the masker locations or to some random location. To avoid these trials affecting the localization data, listeners were instructed to give a specific response if they did not detect the female target (miss trials). This response was to point to a location directly above the head—a response that was easily distinguished from regular localization responses that all had an elevation component on or near 0° (on the audiovisual horizon). To ensure that listeners were following this instruction, a number of catch trials were included in which the target was replaced by another random male masker, and thus false alarm rates could be monitored.

Control runs consisted of 55 trials (5 trials per target location). In masker runs, five catch trials were also included and thus these runs were 60 trials long. Each session consisted of 12 runs. The first and last of these were control runs with no maskers present. In five runs the masker pattern was kept fixed (Fixed) for the duration of the run. Each run used one of the five masker patterns, and the pattern was indicated at the start of the run by presenting a recording of the phrase "fixed maskers" sequentially at each of the four masker locations. In the remaining five runs the masker pattern was randomly chosen on each trial (Mixed) and the run was preceded by a presentation of the phrase "mixed maskers." The Fixed and Mixed runs were interleaved. Each subject completed four sessions, two at each TMR.

## III. RESULTS

## A. Control data

Figure 2 plots rms errors relative to the mean response to a single unmasked target as a function of location. Separate lines are shown for the control runs of the 0 dB TMR sessions and of the −5 dB TMR sessions (although the target level did not change). rms errors were consistent across the two masker sessions, growing with target laterality from about 2° to 5°. Also, there was a slight asymmetry in errors; on the left-hand side, rms errors grew approximately linearly with target eccentricity, whereas on the right-hand side the growth was initially steep (resulting in noticeably larger errors for the +30° and +40° targets compared to the −30° and −40° targets) and then leveled off (at +50°). This asymmetry could be related to minor asymmetries in the experimental setup or the room acoustics (even though the room was mostly left-right symmetric) or it could suggest that there is some perceptual asymmetry in speech localization.

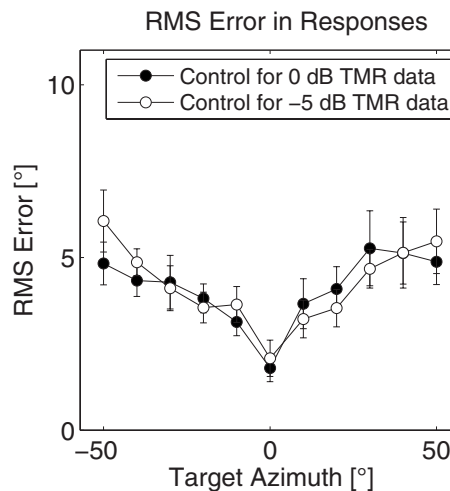

FIG. 2. Localization performance in the control condition with no maskers. Plotted are across-subject averages (±1 SEM) of the rms error as a function of the target location. Results are plotted separately for the two TMR sessions even though the target level was identical.

## B. Miss rates and false alarms

Table I shows mean miss rates and false alarm rates for the two masker conditions (Fixed and Mixed). Miss rates and false alarm rates were larger at −5 dB TMR than at 0 dB TMR but were relatively low overall. It was especially important that the false alarm rate was low, as this was our indicator that subjects were reliable at indicating that they did not hear the target (and thus that they were unlikely to give random localization responses). Note that only five catch trials were included in each experimental run, and thus the false alarm rate of 20% observed at the −5 dB TMR corresponds to just one false alarm per run. It is hard to estimate the impact this nonzero false alarm rate has on performance. However, importantly, miss rates and false alarm rates were similar for Fixed and Mixed conditions, meaning that differences between these conditions were unlikely to be attributable to differences in detection criteria.

TABLE I. Detection performance averaged across subjects, masker patterns and masker locations. Miss rate shows the percentage of trials on which the target was presented but not heard (out of 55 trials per run). False alarm rate shows the percentage of catch trials (out of five trials per run) on which no target was presented but the subject gave a localization response indicating that he/she heard a target.

| Masker condition | Miss rate (%) | | False alarm rate (%) | |
|---|---|---|---|---|
| | TMR 0 dB | TMR −5 dB | TMR 0 dB | TMR −5 dB |
| Fixed | 1 | 8 | 5 | 21 |
| Mixed | 1 | 9 | 8 | 19 |

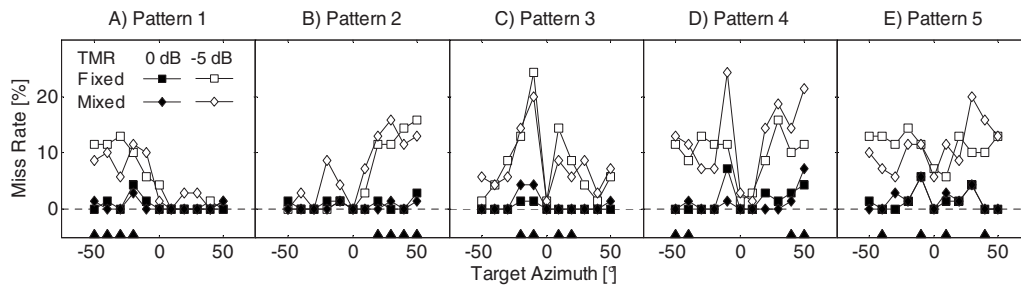Kopčo et al.: Speech localization in a mixture

FIG. 3. Across-subject average of the miss rate (percentage of trials on which target was presented but not heard) as a function of target location. Each panel shows data for one masker pattern (masker locations indicated by the filled triangles along the abscissa), separately for all combinations of the Fixed and Mixed conditions and of the 0 and −5 dB TMR. Mean miss rates collapsed across target locations are shown in Table I.

Figure 3 shows the miss rate as a function of target location for TMRs of 0 dB (filled symbols) and −5 dB (open symbols). The five panels show data for the five different masker patterns. Misses were more common for targets falling on masker locations (or within one loudspeaker from a masker) and very rare elsewhere. This effect, and its exaggeration at the lower TMR, is likely to be an effect of energetic masking, where colocated maskers simply reduce the audibility of the target.

## C. Effects of masking and *a priori* information on rms errors

Figure 4 shows the effect of maskers on rms errors for each target location. For each subject, rms errors in the control condition (see Fig. 2) were subtracted from rms errors in the different masker conditions[1] and plotted are the across-subject means of these differences. The two rows show data for the two TMRs, and the five columns show data for the five different masker patterns.

The effect of masking on rms errors depended in a complex way on all four parameters manipulated in this study, resulting in increases as large as 15° [patterns 4 and 5, Figs. 4(I) and 4(J)]. Overall, the presence of maskers always resulted in an increase in error, even at 0 dB TMR (all data points are positive in Fig. 4). It appears that the rms errors tended to increase most at target locations that corresponded

to masker locations. Moreover, the largest increases occurred for masker patterns 4 and 5, the patterns in which the maskers were distributed so that they did not form one group. Lowering the TMR resulted in an approximately constant increase in rms error across all patterns.

Increases in rms error were not perfectly left-right symmetric for the symmetrical patterns 3–5 or for patterns 1 vs 2 which are mirrored versions of one another. The asymmetry appears to parallel that seen in the control data (Fig. 2), where errors are slightly higher on average for targets on the right. It is difficult to determine from these results whether such asymmetries are perceptual or reflect asymmetries in the setup and the room. The main effects of interest here (like the differences between the Fixed and Mixed performance) did not appear to be influenced by this asymmetry.

Figure 4 shows that the effect of maskers on rms errors also depended on whether the masker locations were fixed or mixed within a run. For example, for pattern 1 at the poorer TMR [Fig. 4(F)], rms errors were larger in the Fixed condition when the target was presented from the left, but were larger in the Mixed condition when the target was presented from the right. In general there was a tendency for a transition such as this to occur at or near the boundary between masker regions and nonmasker regions.

Figure 5(A) provides a summary by showing the increase in the rms error in localization responses (relative to
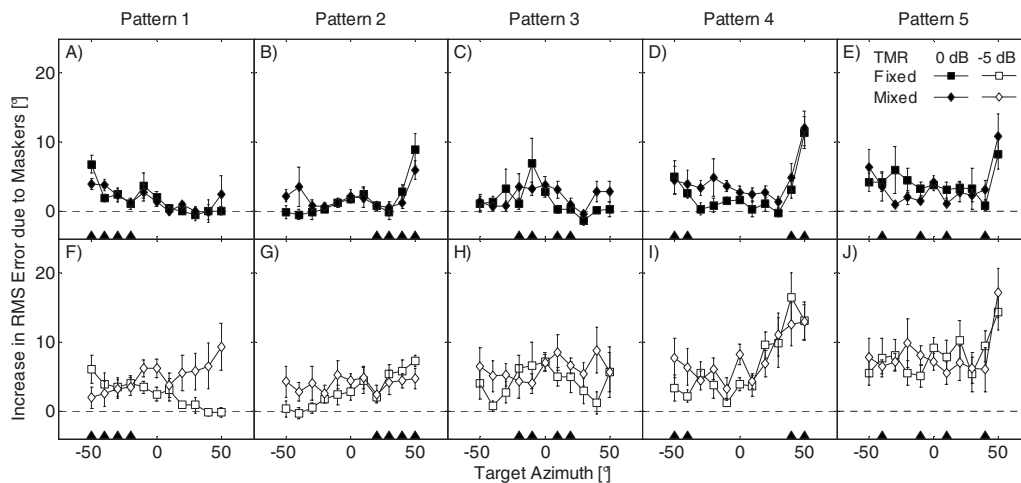


FIG. 4. Across-subject average (±1 SEM) of the increases in response rms errors (*re*. the control condition) as a function of the target location. Each column of panels shows the Fixed and Mixed condition data for one masker pattern (masker locations indicated by the filled triangles along the abscissa) and for the TMR of 0 dB (upper panels) and −5 dB (lower panels).
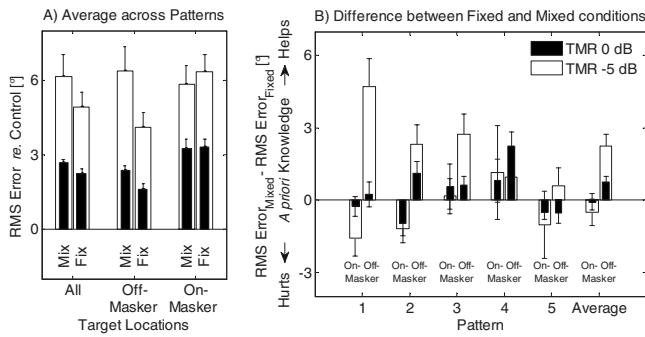
FIG. 5. (A) Effect of the maskers on localization accuracy shown as the increase in rms error in the responses to masked targets (*re.* rms error in the no-masker control condition) averaged across all patterns and across either all target locations (all), across the target locations from which no masker was presented for a given masker pattern (off-masker), or across the target locations from which a masker was presented for a given pattern (on-masker). Data are plotted separately for the Fixed and Mixed conditions and the two TMRs. (B) Effect of *a priori* knowledge on localization accuracy. The difference between rms errors in the Fixed and Mixed conditions is plotted separately for the on-and off-masker locations of each pattern and for the across-pattern average. All bars show the across-subject average ($\pm 1$ SEM).

the rms error in the control condition) averaged across masker patterns. Data are averaged across all target locations (All), across the locations at which the target was not colocated with a masker (off-masker), and across the locations at which the target was presented with a colocated masker (on-masker). Separate bars represent the Fixed and Mixed conditions and the two different TMRs.

Averaged across all target locations, the reductions in the rms error in the Fixed condition with regard to the Mixed condition were 15% at 0 dB TMR and 20% at −5 dB TMR [see the filled and open "All" bars in Fig. 5(A)]. When only the off-masker target locations were considered, the effect of *a priori* knowledge was even larger, reducing the rms errors by approximately 31% at 0 dB TMR and by approximately 35% at −5 dB TMR ["Off-Masker" bars in Fig. 5(A)]. On the other hand, the availability of *a priori* information had a modest effect on the on-masker targets, increasing the rms

errors by approximately 2% at 0 dB TMR and by approximately 9% at −5 dB TMR ["On-Masker" bars in Fig. 5(A)].

Figure 5(B) evaluates the effect of *a priori* knowledge directly by showing the difference between the Mixed and Fixed condition rms errors as a function of the masker pattern (including the across-pattern average), separately for the on- and off-masker locations. At the off-masker locations, *a priori* information provided a benefit for all masking patterns at −5 dB TMR and a smaller and less consistent benefit at 0 dB TMR. At −5 dB TMR, the largest benefit was observed with patterns 1–3 and at 0 dB TMR the largest benefit was for pattern 4. At both TMRs, the smallest off-masker benefit of *a priori* information was observed for masker pattern 5. At the on-masker locations, no consistent effect of *a priori* knowledge was observed.

These results were confirmed by submitting the data from Fig. 5(B) to a three-way repeated measures analysis of variance (ANOVA), with the factors of target location (on-masker vs off-masker), masker pattern (1–5) and TMR (0 and −5 dB). This ANOVA found a significant main effect of the target location and of the TMR, as well as a significant three-way interaction between the factors (Table IIA). Additional two-way ANOVAs were performed separately on the on-masker and off-masker data (Table IIB). No significant main effect or interaction was found for the on-masker data, while all main effects and interactions were significant for the off-masker data, confirming the trends shown in Fig. 5(B).

## IV. DISCUSSION

### A. Impact of maskers on detection and localization

One effect of presenting a speech target in the presence of four concurrent speech maskers was a reduction in the detectability of the target, as evidenced by the presence of miss trials (Table I), particularly at the poorer TMR. This was not surprising, given that concurrent speech maskers are well known to cause energetic and informational masking, both of which impede detection. We saw more misses for

TABLE II. (A) Three-way repeated measures ANOVA on the differences between the rms errors in the Mixed vs Fixed conditions (location × pattern × TMR). (B) Two two-way ANOVAs performed on the same data, but separately for the on-masker and off-masker locations (pattern × TMR).

| (A) Main factor/interaction | d. f. | F | Signif.[a] |
|---|---|---|---|
| Location (on vs off) | 1, 6 | 20.46 | *** |
| Pattern | 4, 24 | 1.98 | |
| TMR | 1, 6 | 4.34 | |
| Location × pattern | 4, 24 | 2.11 | |
| Location × TMR | 1, 6 | 3.51 | |
| Pattern × TMR | 4, 24 | 0.89 | |
| Location × pattern × TMR | 40, 240 | 3.12 | * |

| (B) | | On-masker | | Off-masker | |
|---|---|---|---|---|---|
| Main factor/interaction | d. f. | F | Signif.[a] | F | Signif.[a] |
| Pattern | 4, 24 | 1.39 | | 3.90 | * |
| TMR | 1, 6 | 0.42 | | 9.70 | * |
| Pattern × TMR | 4, 24 | 0.27 | | 3.79 | * |

[a]Significance levels: * $p < 0.005$, ** $p < 0.05$, *** $p < 0.01$

Kopčo *et al.*: Speech localization in a mixture

on-masker locations (Fig. 3) consistent with previous studies showing greater masking for colocated stimuli whether the task is detection (Simpson *et al.*, 2006; Balakrishnan and Freyman, 2008) or intelligibility (Bronkhorst, 2000; Brungart, 2001; Freyman *et al.*, 2001; Arbogast *et al.*, 2002).

In most previous studies that examined localization in the presence of maskers, disruptions to localization could largely be explained in terms of such reductions in detectability (Good and Gilkey, 1996; Good *et al.*, 1997; Lorenzi *et al.*, 1999). In the present study, however, localization performance was only measured for trials in which listeners reported to have heard the target (see also Simpson *et al.*, 2006). Using this approach, we still found that target localization was strongly degraded by the presence of the speech maskers. The maskers increased rms errors, depending on the configuration of the target and maskers and the TMR. The presence of false alarms on the catch trials (Table I) suggests that some of these effects may be a result of the listeners' tendency to respond and indicate a (probably random) target location even on trials on which they did not hear the target. However, these false alarms were very rare in the 0 dB TMR condition, and this condition gave rise to qualitatively similar patterns of results to the −5 dB TMR condition. Thus we are confident that random responses had only a minor impact on the results. Another possibility is that listeners would on occasion erroneously attribute the location of a masker to the target. It is likely that this "feature-binding" confusion would be more of a problem at the lower TMR where the target voice is more poorly segregated from the mixture and hence less distinct.

We compared the overall effect of maskers on speech localization in our study to that reported by Simpson *et al.* (2006). On average, our listeners showed rms errors of 7° in the control condition and 13° in the Mixed condition at 0 dB TMR. In a similar condition, with four different-sex maskers at 0 dB TMR, Simpson *et al.* (2006) reported rms errors (in the left-right dimension) of 8° in the control condition and 14° in the mixture case. This good correspondence in performance between our subjects and those of Simpson *et al.* (2006) suggests that the fact that our study was conducted in a reverberant office (rather than in an anechoic environment) did not increase errors in quiet or in a mixture. This was somewhat surprising given that reverberation has been shown to affect the segregation of competing speech sounds (Lavandier and Culling, 2007) as well as the localization of one sound in the presence of another (Braasch and Hartung, 2002; Kopčo *et al.*, 2007) in other studies using simpler stimuli.

## B. Effects of *a priori* information on localization

Many previous studies have demonstrated detrimental effects of stimulus uncertainty on target detection and identification in complex auditory mixtures (for review, see Kidd *et al.*, 2008b). Most of these studies have examined spectral or temporal uncertainty in the target or the masker(s). A handful of studies that examined uncertainty in the spatial domain found that within- or across-trial variability in the target location can disrupt intelligibility (Kidd *et al.*, 2005;

Brungart and Simpson, 2007; Best *et al.*, 2008). Variability or uncertainty in masker location, on the other hand, appears to have only a minor effect on target detection for the case of simultaneous noises (Fan *et al.*, 2008) and no effect at all on target intelligibility in the case of competing sentences (Jones and Litovsky, 2008). Furthermore, previous studies of spatial cuing for simple sound localization tasks have not found robust benefits (Spence and Driver, 1994; Sach *et al.*, 2000; Kopčo *et al.*, 2001).

The main hypothesis examined in this study was that the disruptive effect of maskers on speech localization in a multitalker mixture would be mitigated by providing the listener with *a priori* information about the masker locations. The rationale was that attentional processing may become important for localization in complex environments in which there is strong competition for processing resources. Consistent with this hypothesis, we found that the rms error was reduced by approximately 0.5° (or 15%) at 0 dB TMR and by 1° (or 20%) at −5 dB TMR.

We also predicted that the benefit of *a priori* information would be more pronounced for the off-masker target locations. In fact, we found that all of the benefit was restricted to the off-masker locations, at which the error was reduced by approximately 1° (or 31%) at 0 dB TMR and by approximately 2° (or 35%) at the −5 dB TMR. The improvement in performance for the off-masker targets is likely due to reassignment of processing resources to the off-masker locations (away from the known masker locations). However, the observed recovery from masking was not complete, showing that factors outside listener's strategic control also limited performance. These factors likely include direct acoustic interference of the competing sounds with the target sound and limitations in the binaural system's abilities to extract the relevant acoustic cues.

Another hypothesis was that the listeners would benefit more from the *a priori* information about the masker locations if the masker distribution was simple. This hypothesis was confirmed only partially. As expected, the smallest benefit was observed for pattern 5, where maskers were most distributed, thus making it difficult to strategically allocate attention to (or away from) a particular region. Interestingly, for the remaining patterns, the size of the benefit of *a priori* information depended on the TMR. At 0 dB TMR, the largest benefit was observed for pattern 4, whereas at −5 dB TMR the largest benefit was provided for patterns 1–3. While this result makes it difficult to identify the strategy employed by the listeners, it suggests that the strategy might change as the difficulty of the task increases.

## C. Potential mechanisms

As a final note, it is worth considering the mechanisms that might underlie sound localization in a complex speech mixture, with a view to understanding both the impact of maskers on accuracy and the moderating influence of prior knowledge about the their spatial arrangement.

Faller and Merimaa (2004) proposed a model that showed that the robust localization of sounds in the presence of distractors reported in previous studies (e.g., Hawley

J. Acoust. Soc. Am., Vol. 127, No. 3, March 2010

Kopčo *et al.*: Speech localization in a mixture    1455

*et al.*, 1999) could be predicted by considering binaural processing only at points in time when interaural coherence is higher than a (relatively high) critical threshold. Their model was successful at extracting independent binaural parameters for five simultaneous speech stimuli (in their case positioned at 0°, ±30°, and ±80° azimuth). This model is unlikely to handle stimuli like ours because it assumes that there are gaps in the masker profile that allow the listener to have a clear glimpse at the target. Such gaps were minimal in our stimuli because of the brief and synchronized nature of the five utterances. However, it is possible that if the threshold coherence criterion was sufficiently lowered, the model would give reasonable outputs and might even predict the increases in localization error in the presence of maskers observed here.

In order to describe the effects of masker spatial uncertainty on localization accuracy, it seems necessary to invoke more central mechanisms such as endogenous orientation (Posner and Petersen, 1990) whereby responses are modulated by attention and/or expectation. One challenge for any simple model of orienting, however, is to explain why the benefit of *a priori* information depended on the TMR and the specific masker pattern.

## V. SUMMARY

Localization of a monosyllabic speech target is degraded by the presence of concurrent speech maskers, particularly for poorer TMRs, even when reduced detectability is accounted for. The impact of the concurrent maskers depends in a complex way on the target/masker configuration and whether or not the target is spatially coincident with a masker. Listeners can use *a priori* information about the location of maskers to mitigate their adverse effects on target localization, in particular, for targets that do not coincide with maskers.

## ACKNOWLEDGMENTS

[1]The rms errors reported here were computed relative to the mean control responses. When rms errors were defined with respect to the actual target locations, the results were similar. We chose to use the rms errors relative to the control responses since they reflect the perceived locations and account for possible inaccuracies in the placement of the target speakers and in the response measurement system.

Abouchacra, K. S., Emanuel, D. C., Blood, I. M., and Letowski, T. R. (**1998**). "Spatial perception of speech in various signal to noise ratios," Ear Hear. **19**, 298–309.

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

Balakrishnan, U., and Freyman, R. L. (**2008**). "Speech detection in spatial and nonspatial speech maskers," J. Acoust. Soc. Am. **123**, 2680–2691.

Best, V., Carlile, S., Jin, C., and van Schaik, A. (**2005**). "The role of high frequencies in speech localization," J. Acoust. Soc. Am. **118**, 353–363.

Best, V., Ozmeral, E. J., Kopčo, N., and Shinn-Cunningham, B. G. (**2008**). "Object continuity enhances selective auditory attention," Proc. Natl. Acad. Sci. U.S.A. **105**, 13174–13178.

Braasch, J., and Hartung, K. (**2002**). "Localization in the presence of a distracter and reverberation in the frontal horizontal plane: I. Psychoacoustical data," Acust. Acta Acust. **88**, 942–955.

Bronkhorst, A. W. (**2000**). "The cocktail party phenomenon: A review of research on speech intelligibility in multiple-talker conditions," Acta. Acust. Acust. **86**, 117–128.

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., and Simpson, B. D. (**2007**). "Cocktail party listening in a dynamic multitalker environment," Percept. Psychophys. **69**, 79–91.

Carlile, S., Leong, P., and Hyams, S. (**1997**). "The nature and distribution of errors in sound localization by human listeners," Hear. Res. **114**, 179–196.

Drullman, R., and Bronkhorst, A. W. (**2000**). "Multichannel speech intelligibility and talker recognition using monaural, binaural, and three-dimensional auditory presentation," J. Acoust. Soc. Am. **107**, 2224–2235.

Faller, C., and Merimaa, J. (**2004**). "Source localization in complex listening situations: Selection of binaural cues based on interaural coherence," J. Acoust. Soc. Am. **116**, 3075–3089.

Fan, W. L., Streeter, T. M., and Durlach, N. I. (**2008**). "Effect of spatial uncertainty of masker on masked detection for nonspeech stimuli," J. Acoust. Soc. Am. **124**, 36–39.

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., 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.

Good, M. D., and Gilkey, R. H. (**1996**). "Sound localization in noise: The effect of signal-to-noise ratio," J. Acoust. Soc. Am. **99**, 1108–1117.

Good, M. D., Gilkey, R. H., and Ball, J. M. (**1997**). "The relation between detection in noise and localization in noise in the free field," in *Binaural and Spatial Hearing in Real and Virtual Environments*, edited by R. H. Gilkey and T. R. Anderson (Erlbaum, Hillsdale, NJ), pp. 349–376.

Hawley, M. L., Litovsky, R. Y., and Colburn, H. S. (**1999**). "Speech intelligibility and localization in a multi-source environment," J. Acoust. Soc. Am. **105**, 3436–3448.

Jones, G. L., and Litovsky, R. Y. (**2008**). "Role of masker predictability in the cocktail party problem," J. Acoust. Soc. Am. **124**, 3818–3830.

Kidd, G., Jr., 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, G., Jr., Best, V., and Mason, C. R. (**2008a**). "Listening to every other word: Examining the strength of linkage variables in forming streams of speech," J. Acoust. Soc. Am. **124**, 3793–3802.

Kidd, G., Jr., Mason, C. R., Richards, V. M., Gallun, F. J., and Durlach, N. I. (**2008b**). "Informational masking," in *Auditory Perception of Sound Sources*, edited by W. A. Yost, A. N. Popper, and R. R. Fay (Springer Handbook of Auditory Research, New York), pp. 143–190.

Kopčo, N., Best, V., and Shinn-Cunningham, B. G. (**2007**). "Sound localization with a preceding distractor," J. Acoust. Soc. Am. **121**, 420–432.

Kopčo, N., Ler, A., and Shinn-Cunningham, B. G. (**2001**). "Effect of auditory cuing on azimuthal localization accuracy," J. Acoust. Soc. Am. **109**, 2377.

Lavandier, M., and Culling, J. F. (**2007**). "Speech segregation in rooms: Effects of reverberation on both target and interferer," J. Acoust. Soc. Am. **122**, 1713–1723.

Lorenzi, C., Gatehouse, S., and Lever, C. (**1999**). "Sound localization in noise in normal-hearing listeners," J. Acoust. Soc. Am. **105**, 1810–1820.

Mills, A. W. (**1958**). "On the minimum audible angle," J. Acoust. Soc. Am. **30**, 237–246.

Posner, M. I., and Petersen, S. E. (**1990**). "The attention system of the human brain," Annu. Rev. Neurosci. **13**, 25–42.

Sach, A., Hill, N., and Bailey, P. (**2000**). "Auditory spatial attention using interaural time differences," J. Exp. Psychol. Hum. Percept. Perform. **26**, 717–729.

Shinn-Cunningham, B. G. (**2008**). "Object-based auditory and visual attention," Trends Cogn. Sci. **12**, 182–186.

Simpson, B. D., Brungart, D. S., Iyer, N., Gilkey, R. H., and Hamil, J. T. (**2006**). "Detection and localization of speech signals in the presence of competing speech signals," in Proceedings of the International Conference

Auditory Display, pp. 129–133.

Spence, C. J., and Driver, J. (**1994**). "Covert spatial orienting in audition: Exogenous and endogenous mechanisms," J. Exp. Psychol. Hum. Percept. Perform. **20**, 555–574.

Wightman, F. L., and Kistler, D. J. (**1989**). "Headphone simulation of free field listening II: Psychophysical validation," J. Acoust. Soc. Am. **85**, 868–878.

Yost, W. A., Dye, R. H., Jr., and Sheft, S. (**1996**). "A simulated "cocktail party" with up to three sound sources," Percept. Psychophys. **58**, 1026–1036.

Zurek, P. M. (**1993**). "Binaural advantages and directional effects in speech intelligibility," in *Acoustical Factors Affecting Hearing Aid Performance*, edited by G. A. Studebaker and I. Hochberg (Allyn and Bacon, Boston,), pp. 255–276.