Masker location uncertainty reveals evidence for suppression of maskers in two-talker contexts

Kachina Allen[a] and David Alais
School of Medical Sciences, University of Sydney, New South Wales, Australia 2106

Barbara Shinn-Cunningham
Department of Cognitive and Neural Systems, Boston University, Boston, Massachusetts 02215

Simon Carlile
School of Medical Sciences, University of Sydney, New South Wales, Australia 2106

(Received 11 February 2010; revised 5 August 2011; accepted 8 August 2011)

In many natural settings, spatial release from masking aids speech intelligibility, especially when there are competing talkers. This paper describes a series of three experiments that investigate the role of prior knowledge of masker location on phoneme identification and spatial release from masking. In contrast to previous work, these experiments use initial stop-consonant identification as a test of target intelligibility to ensure that listeners had little time to switch the focus of spatial attention during the task. The first experiment shows that target phoneme identification was worse when a masker played from an unexpected location (increasing the consonant identification threshold by 2.6 dB) compared to when an energetically very similar and symmetrically located masker came from an expected location. In the second and third experiments, target phoneme identification was worse (increasing target threshold levels by 2.0 and 2.6 dB, respectively) when the target was played unexpectedly on the side from which the masker was expected compared to when the target came from an unexpected, symmetrical location in the hemifield opposite the expected location of the masker. These results support the idea that listeners modulate spatial attention by both focusing resources on the expected target location and withdrawing attentional resources from expected locations of interfering sources. [DOI: 10.1121/1.3631666]

PACS number(s): 43.66.Dc [RYL] Pages: 2043–2053

I. INTRODUCTION

In everyday listening situations, a signal of relevance to the listener may often arrive at the ears simultaneously with a variety of signals from other sources. This creates a mixture of sounds at the ears and poses a challenge for the auditory system; it must isolate the signal of interest from other sources that must be ignored. Despite the complexity of everyday sound mixtures, the human auditory system is remarkably good at perceptually segregating sound sources of interest (such as speech) from interfering sources and noise. How is this problem solved? When Cherry (1953) first described this problem (which he dubbed the "cocktail party problem"), he suggested that differences in the spatial locations of competing sound sources (such as multiple talkers) could provide an important cue for segregating sounds. Sounds from different locations differ at the ears in terms of interaural time and level differences and in spectral content (Persson *et al.*, 2001; Hawley *et al.*, 2004; Edmonds and Culling, 2005). These spatial cues, together with other segregation cues such as pitch, voice differences and level (for a recent review, see Darwin, 2008), collectively provide a way for the auditory system to associate different sounds with different spatial locations. Numerous studies support Cherry's original conjecture; that spatially separating a target sound from interfering maskers improves the intelligibility of the speech target compared to when the target and maskers are co-located. The increase in intelligibility with spatial separation has come to be known as "spatial release from masking" (SRM; e.g., see Freyman *et al.*, 2001; Arbogast *et al.*, 2002; Brungart and Simpson, 2002; Noble and Perrett, 2002; Ebata, 2003; Litovsky, 2005).

SRM is thought to occur because spatially separating the target and masker leads to reductions in two kinds of masking. One is energetic masking: interference produced when the target and masking sounds overlap in the spectro-temporal domain, so that the signal is not robustly represented in auditory periphery. The second is informational masking: masking arising from more central interference, such as occurs when target and maskers are similar, easily confusable, and/or difficult to perceptually segregate from one another (e.g., see Kidd *et al.*, 1998; Freyman *et al.*, 1999; Brungart, 2001). Another factor that may contribute to informational masking and that is addressed in the experiments presented here is the uncertainty that occurs when there are multiple possible masker locations. This creates uncertainty about where to allocate attentional resources, and uncertainty itself may contribute to informational masking (Watson, Kelly, and Wroton, 1976; Watson, 2005). When there is significant informational masking between sources, spatially separating them can result in a large release from masking, much greater than when the main source of interference is energetic (Kidd *et al.*, 1998;

[a]Author to whom correspondence should be addressed. Electronic mail: kachinaa@princeton.edu

Brungart, 2001; Brungart *et al*., 2001; Freyman *et al*., 2001; Hall *et al*., 2005; Kidd *et al*., 2005b; Rhebergen *et al*., 2005). These results support the idea that differences in perceived source location provide one basis by which a listener can direct attention to a particular sound, selecting it from a mixture of competing sounds.

Several studies have reported that spatially directed attention helps listeners understand sound sources in multi-talker environments. Cuing the location of an auditory stimulus can decrease the response times to non-speech stimuli (Rhodes, 1987; Mondor and Zatorre, 1995; Sach *et al*., 2000) and improve the comprehension of speech embedded in similar maskers (Kidd *et al*., 2005a). Studies using recorded event related potentials (ERPs) show stronger responses to non-speech stimuli at attended locations compared to non-attended locations (Rhodes, 1987; Teder and Näätänen, 1994; Mondor and Zatorre, 1995; Teder-Sälejärvi and Hillyard, 1998; Teder-Sälejärvi *et al*., 1999; Widmann and Schröger, 1999; Sach *et al*., 2000), consistent with the idea that directed attention increases neural responses to stimuli at expected locations (see also Winkowski and Knudsen, 2006). While these studies all showed that prior knowledge of the location of a target influences performance, less is known about how prior knowledge of masker locations may affect target understanding and SRM. While some recent studies have varied both target and masker locations (Kidd *et al*., 2005a; Brungart and Simpson, 2007), the systematic manipulation of masker locations has seldom been examined.

One of the few studies to systematically manipulate prior knowledge about masker locations in a multi-talker experiment found that it had no effect on speech intelligibility (Jones and Litovsky, 2008). Based on other studies of auditory spatial attention, this was a counter-intuitive result. For example, reducing the probability of an auditory target being played from one location reduces the allocation of attentional resources to that location (Sonnadara *et al*., 2006), and a study of tone stimuli by Teder-Sälejärvi *et al*. (1999) indicated that the focus of attention is affected by the spacing of interferers. Together, these studies suggest that listeners can flexibly allocate attention given knowledge of both target and masker locations. Prior knowledge about masker location in Jones and Litovsky's experiment should therefore have allowed a listener to shift attentional resources away from the expected masker location, thus freeing up these resources for processing the desired target. However, the failure to demonstrate better performances in the expected masker condition may be due to the fact that spondees were used as target stimuli, and these have a relatively long duration (on the order of half to one second). This is important because switching the spatial focus of attention from one location to another may take only 80 to 200 ms (Teder-Sälejärvi and Hillyard, 1998). Thus, it may be that listeners in Jones and Litovsky's study were able to redirect attention to a new, unexpected spatial configuration of masker and target within the duration of the target stimulus.

In the current study, we manipulate prior knowledge about masker locations in an environment with competing speech maskers. Importantly, very brief stimuli are presented (single syllables, with listeners identifying the initial consonant), which prevents listeners from re-orienting spatial attention within trials. We compare performance for a masker coming from an unexpected location with performance for a masker coming from an expected location, for both co-located and separated target and masker. By testing whether prior knowledge of masker location affects performance, the results will shed light on an interesting theoretical proposal. Durlach *et al*. (2003a) noted that knowing the locations of targets and distractors could potentially allow two different, distinct strategies that could facilitate target identification. According to the "max" strategy, attention directed at the known target location could enhance sensitivity to sources at that location. Complementing this, a "min" strategy could involve suppression of signals from the known masker locations. While there is evidence consistent with the max strategy (knowing the target location improves identification of both speech and non-speech targets; Arbogast and Kidd, 2000; Ericson, Brungart and Simpson, 2004; Kidd *et al*., 2005a), whether spatially directed attention also leads to suppression of known distractors is less clear (Jones and Litovsky, 2008; Brungart and Simpson, 2007). The present study, by using stimuli too brief to permit re-orienting of spatial attention, is intended to explore whether there is suppression of distractors when their location is known.

II. GENERAL METHODS

A. Setup

Subjects were seated, facing forward, in a sound attenuated audiometric booth (size $= 3.5 \times 4.6 \times 2.4$ m) lined with 7.5 cm acoustic foam. An array of Tannoy active loudspeakers were placed 1.3 m away on the subject's audiovisual horizon at 20° intervals (e.g., $-20°$, 0°, 20°, 40). Subjects were instructed to remain facing directly ahead at all times, though their heads were not restrained. A laptop was provided on which subjects were asked to type responses. The laptop was placed at waist height to prevent any acoustic occlusion of the soundfield.

B. Stimuli

Broadband stimuli have been shown to be more accurately localized than band-limited signals for both speech (Best *et al*., 2005) and non-speech sounds (Carlile *et al*., 1999; Langendijk and Bronkhorst, 2002). Our corpus of nonsense syllables was recorded by an American female talker using a broad bandwidth (0–22.5 kHz) to ensure that the tokens produced robust spatial percepts. To reduce the possibility of slight differences in articulation or recording assisting in identification of the syllables, five tokens of each target word were recorded, from which tokens were randomly chosen on each trial. The same female talker was used for all target and masker words to minimize voice cue differences and maximize the similarity between target and masker (e.g., Brungart, 2001; Noble and Perrett, 2002). This corpus was previously used in experiments on spatial attention in Allen *et al*., 2009.

The ability to recognize a word depends on its frequency in everyday usage (Luce and Pisoni, 1998; Connine, 2004). To prevent frequency bias in our results, a corpus of nonsense syllables, rather than commonplace words, was used in the current study. The corpus consisted of target syllables that differed only in their initial, unvoiced stop consonants ("targ," "parg," and "karg") and maskers that started with voiced stop consonants and that had a different vowel ("boog," "doog," "goog," "borg," "dorg," and "gorg"). All stimuli were normalized to have the same RMS energy. The recordings were ramped with a 10-ms cosine windows at onset and offset to prevent clicks (a manipulation that did not significantly affect the recorded speech waveforms, but ramped the brief quiet portions of the recordings before and after the speech). They were played using MATLAB software (Mathworks, release 14.1) through a Hammerfall multiface soundcard (RME, Ltd.) at a sampling rate of 44.1 kHz and a sound pressure level of 57 dB.

C. Procedure

Subjects were seated in the audiometric booth and instructed to face straight ahead and attend at all times to the central speaker located directly in front of them. Evidence suggests that allocation of attentional resources depends on the probability of the target coming from a particular location (Kidd *et al.*, 2005a; Sonnadara *et al.*, 2006). Lowering the probability of the target being played from one location appears to decrease performance at the expected location while increasing performance at other locations. Thus, to maximize attention to the expected target location while permitting testing of alternative locations, the "expected configuration" was presented on a majority (80%) of trials in each block, with "unexpected configurations" occurring on the remaining 20% of trials.[1] Each unexpected trial was always followed by a trial with the expected configuration. To give subjects time to acclimatize to each condition, a group of six training trials was played at the beginning of each block (always using the expected configuration; results not recorded). Each trial was initiated by the subject's response to the preceding stimulus, so that presentation speed was self-paced, varying with the speed of response.

The target stimulus on a given trial consisted of one nonsense syllable randomly selected from the corpus list of "parg," "karg," and "targ" (each of which was represented by five different recorded tokens). The subjects were asked to identify the initial unvoiced consonant of the target (either "p," "k," or "t") and respond by typing it on a laptop computer. The masker was randomly selected from the maskers ("borg", "boog", "dorg", "doog," "gorg," and "goog"). The target and masker were temporally aligned to ensure that the unvoiced target phoneme was played during the initial masker phoneme.

Within each block of 200 trials, 40 trials were presented at each of five evenly spaced signal-to-noise ratio (SNR) levels, spanning a total range of 20 dB. The maximum and minimum SNR levels were varied slightly from block to block, depending on the performance of the specific subject, with the overall range across all subjects being between −30 and +10 dB SNR. To maximize data collection around the threshold of interest, most trials were presented between −20 and −0 dB SNR.

The target syllables could only be identified by the initial phoneme (the scoring letter), which lasted no longer than 80 ms for any syllable. This is important because switching attention between an expected and unexpected target location can take as little as 80 to 200 ms (Teder-Sälejärvi and Hillyard, 1998). Thus, listeners were unlikely to have been able to switch attention to any new location in cases where the target was played from an unexpected location.

Two conditions were tested in each experiment: (i) Co-located (target and masker were played from the same loudspeaker) and (ii) separated (target and masker were played from separate loudspeakers separated by 20° in azimuth). The initial stop consonant identification threshold (ISCIT) was measured for each condition, providing a baseline from which the spatial release from masking (SRM) could then be calculated. ISCIT was defined as the SNR at which 67% of the target phonemes were identified correctly. The ISCIT was estimated from the cumulative Gaussian psychometric function fitted using a maximum likelihood procedure (Watson, 1979). Each psychometric function was then resampled 1000 times using a bootstrapping procedure (Efron and Tibshirani, 1993). This produced a distribution of ISCIT thresholds from the 1 000 resampled functions; the standard deviation of the distribution was used for *post hoc* t-tests between conditions (see Alais and Carlile, 2005) with the type-I error rate set to $\alpha = 0.05$. SRM was calculated as

$$\mathrm{SRM}(\mathrm{condition}) = \mathrm{ISCIT}(\mathrm{co\text{-}located}) - \mathrm{ISCIT}(\mathrm{condition}). \quad (1)$$

D. Participants

All subjects recruited gave informed written consent, had normal hearing (tested using a standard pure-tone audiogram), and spoke English as their first and main language (except where indicated). Subjects were not remunerated for their participation.

III. EXPERIMENT 1: TARGET IDENTIFICATION WITH AN UNCERTAIN MASKER LOCATION

Experiment 1 examines whether prior knowledge of a masker location contributes to spatial release from masking. If knowing the location of a masker does contribute, then a masker from an unexpected location should provide more masking than an interferer of equal magnitude from an expected location. Thus, the ISCIT for a target paired with a masker at an expected location should be lower (i.e., performance should be better) than for the same target paired with a masker of equivalent energy at an unexpected location.

A. Participants

Four volunteers (three female, one male; mean age 32 ± 2.7 yr), including the first author, were recruited. All

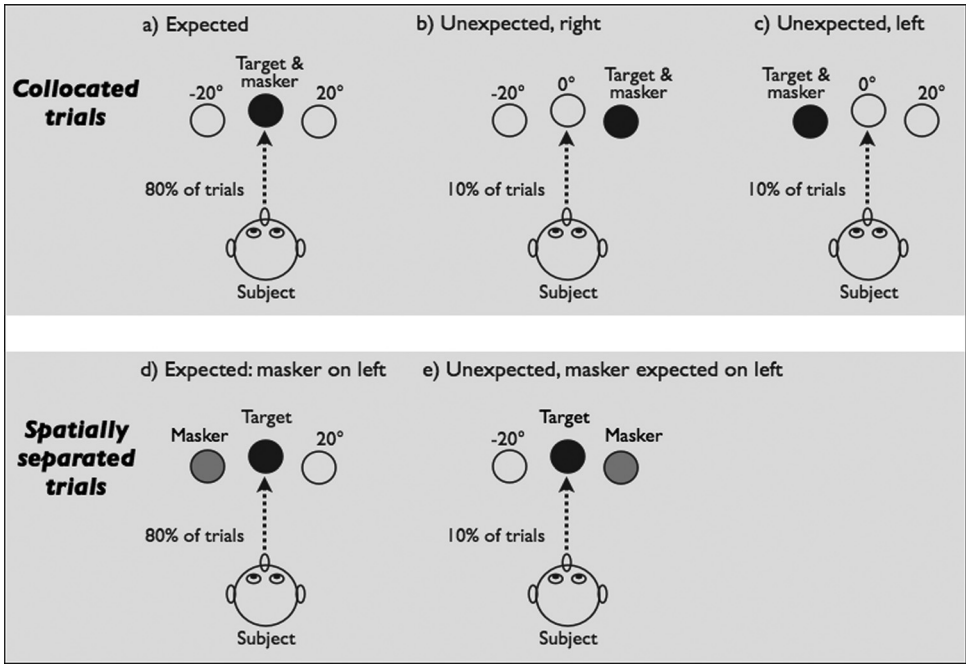

FIG. 1. Loudspeaker layout for experiment 1. Subject faces and attends to the central speaker directly ahead at all times. (a)–(c) co-located condition showing expected and unexpected configurations. (d) Separated condition with the masker expected on the left. (e) Separated condition, unexpected configuration with the masker expected on the left.

subjects had previous experience with auditory psychophysical experiments. Subject S2 spoke English fluently but did not learn it until the age of 18.

B. Procedure

Subjects faced three loudspeakers; one was directly ahead (0° azimuth), the other two flanked the center speaker, and were located 20° to the left and 20° to the right of the center, as depicted in Fig. 1. Subjects were instructed to face the central loudspeaker at all times and were told that the target would always be played from the central position. In the "co-located" condition, the target and single masker were always played from the same, central speaker (0° azimuth). A minimum of 400 trials (2 blocks) was carried out for co-located trials. In the "separated" condition, the masker was played from one of the flanking loudspeakers, either at +20° or −20° azimuth from the target loudspeaker. Within a each block, the masker was played either on the expected side (left or right) for 80% of trials ("masker expected" trials) and on the symmetrically opposite side for the remaining 20% of trials ("masker unexpected" trials). In this separated condition, which side (left or right) was the expected masker side varied randomly from block to block, and subjects were informed prior to each block on which side to expect the masker. Each subject completed four blocks (800 trials): two blocks for the "expect left" configuration and two for "expect right." The data were binned by expected versus unexpected masker location (pooling left and right locations, which will remove any hemispheric bias) and then ISCITs were calculated as described in Sec. II.

C. Results and discussion

ISCITs for four subjects are shown in Fig. 2(a), together with the group mean. Black columns show results for the "co-located" condition. The two gray columns show results for the "separated" condition, with the "masker expected" trials shown in light gray and "masker unexpected" shown in dark gray. Friedman's test ($\alpha = 0.05$) with Bonferroni-corrected *post hoc* *t*-tests revealed significantly higher ISCITs (poorer performance) when the masker was at the unexpected location compared to when the masker was at the expected location. Analysis of the group means using a paired *t*-test found that ISCITs were significantly lower (performance was better) when the masker was at the expected location (-18.2 ± 1.5 dB) compared to when the masker came unexpectedly from the symmetrically opposite side (-15.6 ± 1.4 dB). Similarly, SRM was significant when the masker came from the expected location (4.1 ± 0.6 dB), but not when it came from the unexpected location [1.5 ± 0.8 dB; Fig. 2(b)].

These results show that prior knowledge of the likely masker location improves performance. It is possible, however, that this result arises not due to a shift of top-down attentional focus, but to a different cause. A novel auditory stimulus is likely to elicit an exogenous shift of attention to the location of the novel stimulus (Spence and Driver, 1994; Kanai *et al.*, 2005). Such a shift may pull attentional resources from the target location and thus reduce target identification performance. In experiment 1, in the minority of trials where the masker is played from an unexpected location, the masker is effectively a novel stimulus and it may therefore draw exogenous attention, explaining the drop in performance at the target location. The aim of experiments 2 and 3 is to see whether expectations about masker location influence performance when the difference in novelty between the "masker expected" and "masker unexpected" conditions is reduced. These two experiments thus test whether an exogenous shift of attention toward an unexpected masker contributed to the results of the first experiment.

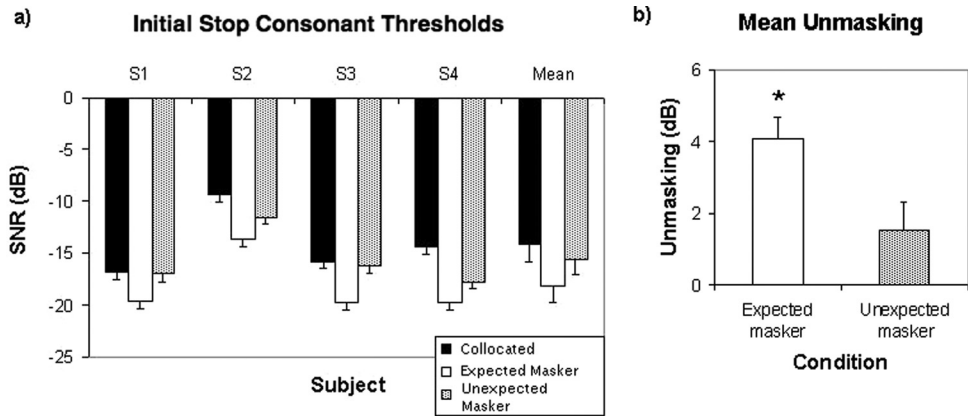

FIG. 2. (a) Initial stop consonant identification thresholds (ISCITs) for Experiment calculated from the 67% intelligibility level of a target and single masker co-located or with the masker separated by 20° from the target. In the separated conditions, the masker was played either on the expected side (80% of trials) or at the symmetrically opposite, unexpected location (20% trials; see Fig. 1). Target ISCITs were lower (performance better) when the masker was at the expected location than when the masker was at the unexpected location. Error bars for individual data are standard deviations calculated from 1 000 repetitions of a bootstrap technique. Error bars for mean data are standard errors. (b) Spatial release from masking for experiment 1 across subjects from the 67% intelligibility level of a target and single masker separated by 20° as compared to a co-located target and masker at the expected location. Conditions as described in Fig. 1. Error bars are standard errors. Unmasking was significantly higher with the expected masker as opposed to the unexpected masker location (paired t-test, $\alpha < 0.05$). Significant masking is marked with an asterisk. Error bars are standard errors.

IV. EXPERIMENT 2: TARGET IDENTIFICATION WITH UNEXPECTED TARGET AND MASKER LOCATIONS

Experiment 1 indicated that presenting a masker from an unexpected location leads to a reduction in target performance and in SRM. This result could potentially arise because the novelty of a masker from an unexpected location captures attention exogenously, leaving fewer resources available for processing the target. Experiment 2 balances novelty between different unexpected spatial configurations by shifting both target and masker to reduce the influence of any exogenous shifts of attention.

A. Design

In the co-located condition, the target and masker talkers were always played from a single loudspeaker. In 80% of co-located trials, target and masker were played from the central loudspeaker [Fig. 3(a)], while in 10% of trials they came from the loudspeaker 40° to the right [Fig. 3(b)] and in the remaining 10% of trials they were played from the loudspeaker 40° to the left [Fig. 3(c)].

In the separated conditions, the masker was always played from a position 20° more lateral than the target location. In 80% of trials, the target was played from the central

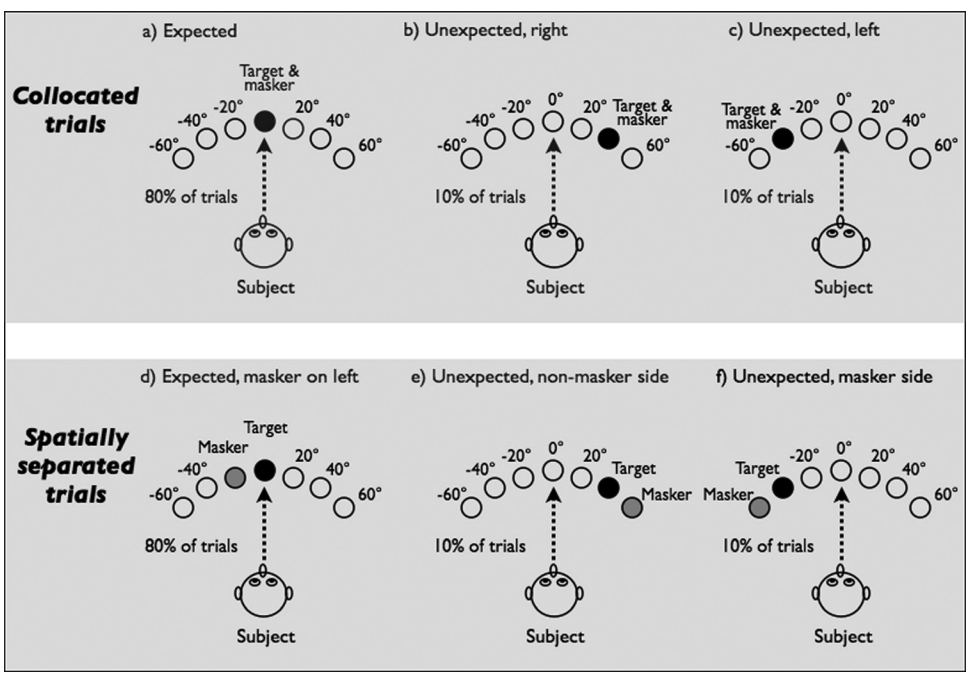

FIG. 3. Loudspeaker layout for experiment 2. Subject faces and attends to the central speaker directly ahead at all times. (a)–(c) Co-located condition showing expected and unexpected configurations. (d) Separated condition, expected configuration with the masker expected on the left. (e) Separated condition, unexpected configuration, non-masker side with the masker expected on the left. (f) Separated condition, unexpected configuration, masker side with the masker expected on the left.

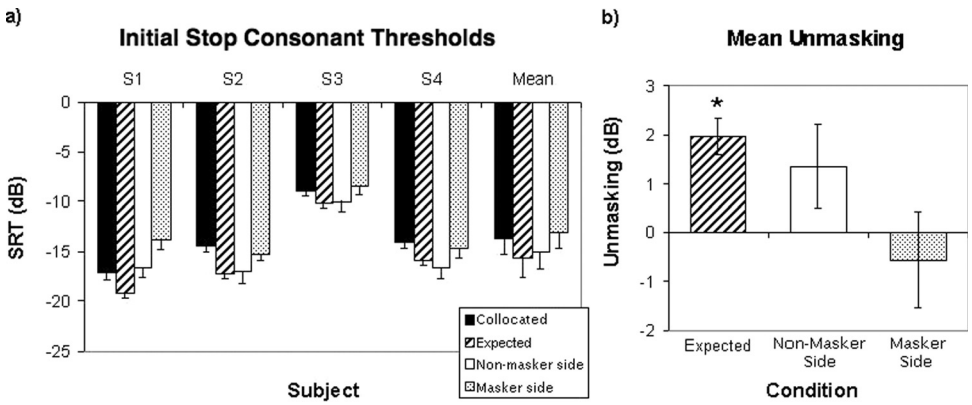

FIG. 4. (a) ISCITs for experiment 2 for trials where the target and single masker were co-located and when the masker was offset 20° to one side (separated). For separated conditions, trials were grouped by when the target was at the central, expected location (80% of trials), when the target was played from a distance of 40° from the expected location on the side with no masker (non-masker-side, 10% of trials) and when the target was played from a distance of 40° from the expected location on the side with a masker expected at 20° (masker-side, 10% of trials; see Fig. 3). Error bars as described in Fig. 2. There were significant differences between conditions with ISCTs on the non-masker side lower (performance better) than on the masker side location (Friedman's test with *post hoc* t-tests with Bonferroni corrections). (b) Mean spatial release from masking calculated by comparing ISCITs in separated conditions (single masker 20° from the target) described in Fig. 2, to the ISCIT where the masker was co-located with the target. Significant masking is marked with an asterisk. Unmasking was not significantly different between conditions (paired *t*-test). Error bars are standard errors.

loudspeaker [expected, Fig. 3(d)], with the masker 20° to the expected side for that block. In 10% of trials, the target was played from the loudspeaker 40° to the side opposite the expected side of the masker, with the masker 20° more lateral [Fig. 3(e)]; in the remaining 10% of trials, the target was played from 40° toward the expected masker side, with the masker 20° more lateral [Fig. 3(f)]. This ensured that in 10% of the trials, the target and masker were played from the same side as the expected masker (masker-side configuration) and 10% from the opposite side (non-masker-side configuration). However, in the masker-side condition, the target was displaced 20° laterally from the expected masker location to ensure that the locations of target and masker were similarly novel in the masker-side and non-masker side presentations.

All subjects carried out 200 unrecorded practice trials of co-located and separated conditions (see below) and a minimum of 1 000 experimental trials. For the separated condition, blocks of 200 trials were randomly selected to have the expected masker location on either the left and right side, and subjects were told before each block the side (left or right) from which the masker would usually be presented. Results were pooled across hemispheres and binned into trials where the target played from: 1—the central loudspeaker (expected condition); 2—the same hemisphere as the expected masker location (unexpected, masker-side condition); and 3—the hemisphere opposite the expected masker location (unexpected, non-masker-side condition).

B. Participants

Four female volunteers (mean age 35.8 ± 7.3 yr), including the first author, were recruited. All had previously participated in multi-talker studies.

C. Results and discussion

ISCIT data for the four individual subjects are shown in Fig. 4(a), together with the group means, shown on the far right. In the co-located condition, where the target and masker were played from the same loudspeaker, no significant differences (Friedman's test, $\alpha = 0.05$) were found between performance with target and masker at the expected location (80% of trials) and at the non-masker-side locations, which were $\pm 40°$ from the expected location (10% of trials to each side). As there were no significant differences in ISCIT between expected and unexpected target locations in the co-located condition, co-located results were pooled across the different absolute locations for all subjects; these pooled results are shown as black columns in Fig. 4(a).

For the separated conditions, ISCITs differed significantly between the non-masker-side and masker-side configurations [see Figs. 3(e) and 3(f)], the two conditions in which the spatial locations of the target and the masker were both unexpected. Friedman's test and Bonferroni-corrected *post hoc* *t*-tests demonstrated significantly lower ISCITs (better performance) when the target was played from the non-masker-side side than when it was played from the side where the masker was expected. Paired *t*-tests ($\alpha = 0.05$) showed higher ISCITs (poorer performance) when the target was played from the expected masker side (masker-side; -13.1 ± 1.6 dB) than when it was played from either the expected (-15.6 dB ± 1.9 dB) or the non-masker-side (-15.1 ± 1.7 dB) locations.

This pattern of data shows that performance was significantly poorer when the target appeared on the side of the expected masker than when the target and masker were both in the hemifield opposite the expected masker location. This is consistent with the idea that listeners divert attentional resources away from an expected masker location.

In experiment 1, the masker (but not the target) changed location in unexpected trials and may therefore have caused an exogenous shift in attention toward the novel masker location, degrading target performance. The masker and target in experiment 2 both moved from their expected locations in the masker-side and non-masker-side configurations. The purpose of this was to compensate for exogenous shifts

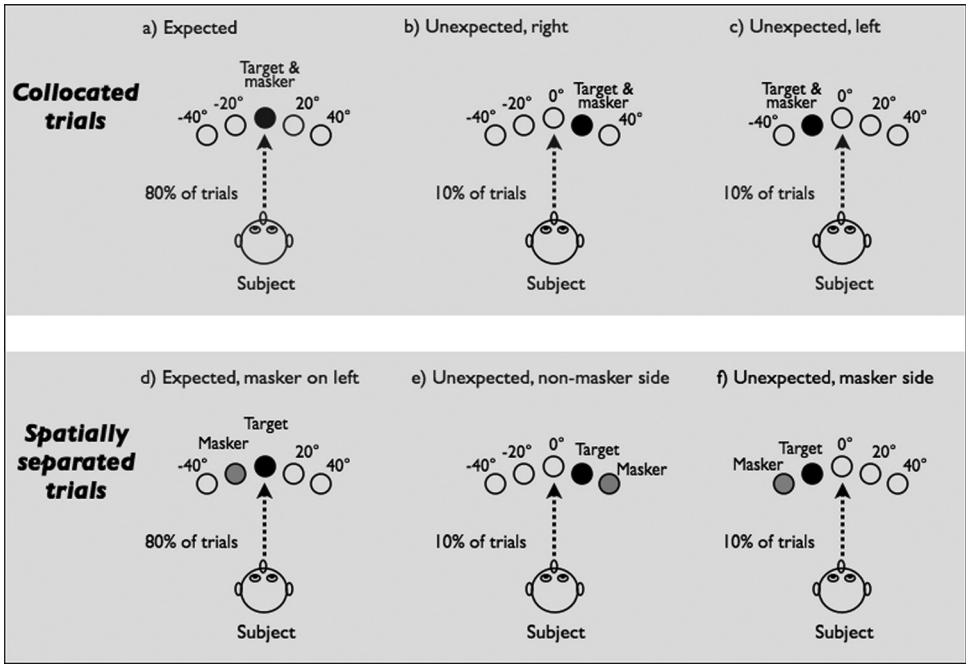

FIG. 5. Loudspeaker layout for experiment 3. Subject faces and attends to the central speaker directly ahead at all times. (a)–(c) Co-located condition showing expected and unexpected configurations. (d) Separated condition, expected configuration with the masker expected on the left. (e) Separated condition, unexpected configuration, non-masker side with the masker expected on the left. (f) Separated condition, unexpected configuration, masker location with the masker expected on the left.

of attention toward the novel masker in the two "unexpected" conditions. These results demonstrate a masker/non-masker-side hemispheric asymmetry in performance, and are thus not consistent with the hypothesis that exogenous attentional shifts fully explain the results from experiment 1.

The SRM data estimated for each target configuration is shown in Fig. 4(b). When masker and target were at the expected locations, 2.0 ± 0.4 dB of SRM was found (striped bar, z-test, $\alpha = 0.05$). For the masker-side and non-masker-side configurations, SRM was not significantly different from zero. This aligns with data from Experiment 1, which showed no significant SRM when a masker was played from an unexpected location. The results of this experiment, however, do not distinguish between the effects of masker and target uncertainty.

V. EXPERIMENT 3: TARGET IDENTIFICATION WITH THE TARGET AT AN EXPECTED MASKER LOCATION

Experiment 1 indicated that presenting a masker from an unexpected location leads to a reduction in target performance. Experiment 2 showed that this could not be explained by exogenous attentional shifts. Experiment 3 employs the same design as experiment 2 but specifically tests performance when the target comes from the expected masker location.

A. Design

In the co-located condition, the target and masker talkers were always played from a single loudspeaker. In 80% of co-located trials, they were played from the central loudspeaker [Fig. 5(a)], while in 10% of trials target and masker were played from the loudspeaker 20° to the right [Fig. 5(b)], and in the remaining 10% of trials from the loudspeaker 20° to the left [Fig. 5(c)].

In the separated condition, the target and masker were always separated by 20°. In 80% of separated trials, the target was played from the central loudspeaker [expected condition; Fig. 5(d)], with the masker 20° to the expected masker side. In 10% of trials, the target was played from the loudspeaker 20° to the side opposite the expected masker location with the masker also on the same side, but 20° more lateral than the target [non-masker-side condition; Fig. 5(e)]. In the remaining 10% of trials, the target was played from the expected masker location, with the masker 20° more lateral than the target [masker-side condition; Fig. 5(f)].

All subjects carried out 200 unrecorded practice trials of co-located and separated conditions (see below) and a minimum of 1 000 experimental trials of each condition, including a minimum of 1 000 trials with the masker expected on the left and a minimum of 1 000 trials with the masker expected on the right. In the separated condition, blocks of 200 trials were played with the expected masker either on the left or on the right, with subjects told before each block the side (left or right) from which the masker would usually be presented. Results were pooled across hemispheres and binned depending on where the target played from: (1) the central loudspeaker (expected condition); (2) the expected masker location (masker-side condition); and (3) the location symmetrically opposite the expected masker location (non-masker-side condition).

B. Participants

Four female volunteers (mean age 35.8 ± 7.3 yr), including the first author, were recruited. Subject S4 had not previously participated in tests of auditory perception; all other subjects had previous experience.

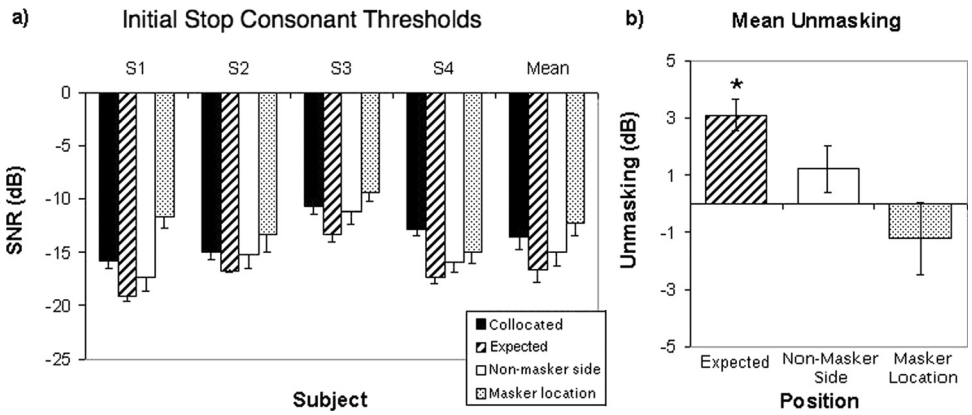

FIG. 6. (a) ISCITs by subject in experiment 3 calculated from the 67% intelligibility level of a target and single masker co-located or with the masker 20° to one side. For separated conditions, trials were grouped by when the target was at the central, expected location (80% of trials), when the target was played from a angle of 20° from the expected location on the side with no masker (non-masker side, 10% of trials) and when the target was played from the expected target location (masker location, 10% of trials; see Fig. 5). Results were collected from both the right and left side and were pooled between hemispheres. Error bars as described in Fig. 2. There were significant differences between the conditions with ISCITs lower (performance better) at the non-masker side than at the masker location (Friedman's test with *post hoc* t-tests with Bonferroni corrections). (b) Mean spatial release from masking in experiment 2 across subjects from the 67% intelligibility level of a target and single masker separated by 20° as compared to a co-located target and masker at the expected location. Conditions as described in Fig. 5. Error bars are standard errors. Significant unmasking is marked with an asterisk (z-test, $\alpha = 0.05$).

C. Results and discussion

ISCITs for the four individual subjects are shown in Fig. 6(a), together with the group means. In the co-located condition, where the target and masker were from the same loudspeaker, no significant differences were found between the 80% of trials played from the expected loudspeaker located directly ahead at 0° and the remaining 20% of trials where both sounds were played from one of the two side loudspeakers located at ± 20° (Friedman's test, $\alpha = 0.05$). As the co-located data did not differ with location, they were pooled; the group mean is plotted on the left of each data cluster in Fig. 6(a) (black column).

Results are similar to those from experiment 2. For the separated conditions, ISCITs differed between the unexpected non-masker and masker-side configurations [see Figs. 5(e) and 5(f)]. Performance was better for the non-masker-side than the masker-side conditions (ISCITs were significantly lower when the target was played from the non-masker side rather than from the expected masker location). Friedman's test ($\alpha = 0.05$) with Bonferroni-corrected *post hoc* paired *t*-tests found that ISCITs were higher (performance was poorer) when the target was played from the expected masker location (masker-side condition; −12.3 ± 1.2 dB) than when it was from the expected (−15.0 dB ± 1.3 dB) and non-masker-side (−16.6 ± 1.2 dB) locations. The difference of 2.6 dB between ISCITs for non-masker-side and masker-side conditions is similar to the 2.0 dB difference found in the corresponding conditions in experiment 2.

Figure 6(b) shows SRM calculated for each target configuration. SRM in the expected separated configuration was significantly greater than zero (z-test $p < 0.05$; 3.11 ± 0.6 dB) and was not significantly different from the SRM measured for the equivalent expected spatial configuration in experiment 1. In the unexpected configurations, no significant SRM was found for either the non-masker (1.2 ± 0.8 dB) or masker-side (−1.2 ± 1.3 dB) configurations. Overall, the results of experiment 3 support the interpretation of experiments 1 and 2, that prior knowledge of masker and target locations influences target performance and SRM.

VI. GENERAL DISCUSSION

In the classic cocktail party environment sounds from multiple talkers and noise sources arrive at the ears together. One of the ways the auditory system can isolate the talker of interest is to direct attention to the location of that talker. Directing spatial attention to stimuli at a particular location is associated with improved performance for auditory tasks involving both speech (Kidd *et al.*, 2005a; Allen *et al.*, 2009) and non-speech sounds (Rhodes, 1987; Teder and Näätänen, 1994; Mondor and Zatorre, 1995; Teder-Sälejärvi and Hillyard, 1998; Teder-Sälejärvi *et al.*, 1999; Widmann and Schröger, 1999; Sach *et al.*, 2000). In brain imaging studies, there is evidence that the expectation of a stimulus in either the auditory or visual sensory modality increases neural response to the stimulus (Foxe *et al.*, 2005). Widmann and Schröger (1999) also reported increased ERP amplitudes in response to an auditory stimulus from an attended location compared to when the stimulus was from an unattended location. These and other observations indicate that directing attention to a particular stimulus (based on different features, including location) increases perceptual sensitivity to the attended stimulus, both improving judgments about the stimulus and shortening reaction times to the stimulus (for a review, see Knudsen, 2007).

In the visual modality, improvements in performance due to directed attention come about both from a facilitation of neural excitation in response to the source at the attended location and neural suppression of stimuli from unexpected locations (Smith *et al.*, 2000; Hopf *et al.*, 2006; Kelly *et al.*, 2006), However, in the auditory modality, there is little evidence, either from physiological or behavioral studies, for suppression of stimuli that are not at expected target locations.

In the current study, experiment 1 showed increased masking (2.6 dB) when the target remained at the expected location, but the masker was played from an unexpected location. Experiment 2 indicated that exogenous attentional shifts toward the novel masker location were unlikely to be the cause of the performance loss, as experiment 2 compared performance in two different unexpected conditions. In both of the unexpected conditions in experiment 2, the target and masker were either played from novel locations on the side of the expected masker location or from the symmetrically equivalent locations in the opposite hemisphere. Even though the two unexpected conditions both presented target and masker from novel locations (so that both should have caused similar exogenous reorientation of attention), performance was worse (thresholds changed by 2.0 dB) when the target and masker were presented from the expected side of the masker rather than from the opposite configuration. In experiment 3 there was 2.6 dB more masking when the target was played at the expected location of the masker compared to when the target came from a symmetrically equivalent, unexpected location.

The current results indicate that prior knowledge of a masker location aids in speech recognition (at least at the phoneme identification level) in multi-talker environments. We found that performance is worse when the masker comes from an unexpected location. In contrast with our results, one previous study (Jones and Litovsky, 2008) found no difference when the masker came from an unexpected location compared to when it was at the expected location. This apparent contradiction may be related to the time course of the trials used in the Jones and Litovsky study as they used spondees, which have a relatively long duration. Thus, it is possible that when listeners heard a masker from an unexpected location, they had sufficient time to redirect their spatial attention to the new configuration within the time course of a single trial, thereby explaining why no effect of masker location was found. In the current study, such a strategy would be much less effective, as the acoustic information needed to identify the initial consonant of the target were concentrated at the start of the target tokens. The auditory system can use many different cues such as pitch, voice characteristics, sound level and location (for a review, see Darwin, 2008) to help comprehend a talker of interest in a noisy environment. The cues used vary with the specifics of the environment and required task. Our results hint that one of the cues that may be utilized is prior knowledge of the masker location. Further experiments will be needed to investigate under what circumstances this cue may be useful and what practical role it can play in a naturalistic multi-talker environment.

The results from the current study do not clearly identify how expectations about masker location can influence performance. Our results are consistent with active suppression of responses to auditory sources coming from the expected masker location (as in the "listener-min" strategy of Durlach *et al.*, 2003b). While such active suppression has been seen in past visual experiments, there is little clear-cut evidence for such effects in the auditory literature. Alternatively, since only a single masker was used, it is also possible that expectation of a masker location could cause attention to be focused on the known target location, but in an asymmetrical way such that it avoids the expected masker location (a slight variation on the "listener-max" strategy of Durlach *et al.*, 2003b). In other words, the degradation of performance when the target comes from the expected side of the masker may be a consequence either of attentional focus being directed away from the masker location or of active suppression directed toward the masker location. A combination of target maximization and masker minimization may both contribute to the current findings.

The vision attention literature shows that multiple forms of active suppression can occur, depending on the task. Past visual studies show suppression of sources in an unexpected hemisphere (Kelly *et al.*, 2006), from all unexpected locations (Smith *et al.*, 2000), and in a narrow spatial band surrounding the attended location (Hopf *et al.*, 2006). If active suppression of expected maskers is involved in audition, the current experiments may help to identify the nature of the suppression. In the current auditory experiments, performance is negatively affected when the target appears at the location of or the side of the expected masker location. Thus, the current results suggest that any suppression is directed toward locations from which a masker is expected, rather than affecting all locations from which the target is not expected. However, performance for a target coming from the expected masker location (experiment 3) is not very different from performance when the target is 20° offset from the expected masker location (experiment 2). This suggests that if there is a gradient in auditory spatial suppression, it is not a very steep one. It may also be the case that, as in vision, the processes involved vary with the task.

Performance tends to decrease as the distance of the target from an attended location increases (Allen *et al.*, 2009). Thus, performance for targets played from the unexpected, non-masker side locations in experiments 2 and 3 might have been expected to be poorer than for targets at the expected location. Yet, no significant differences in ISCIT were found (*post hoc t*-tests with Bonferroni correction) between the expected and non-masker-side target locations in either experiments 2 or 3 for either the co-located or separated condition. There is a slight trend in individual data in experiment 3 for performance for targets from the non-masker-side location to be worse than performance for targets from the expected location [Fig. 6(a)]; however, this difference is not statistically significant (*post hoc t*-tests with Bonferroni correction). With pure-tone stimuli, spatial attention gradients appear to depend on the experimental task and the physical spacing between the target and maskers (Teder-Sälejärvi *et al.*, 1999). Thus, a lack of any significant differences between the expected and non-masker-side locations in the current experiments may be due to the nature of the task itself. For instance, Teder-Sälejärvi *et al.*, 1999, proposed that closely flanking maskers promote a narrow focus of spatial attention. In the co-located conditions of all of the current experiments, no differences were found between trials with target and masker an expected locations and those played up to 40° in azimuth displaced from the expected locations. In co-located conditions, masker and target were played from the same loudspeaker; therefore, the attentional

focus could be unconstrained and broad enough so that there is no significant difference for sources coming from locations spanning an angle of 40°. Alternatively, spatial attention may only play a large role when there are competing sources from different locations, thus explaining why co-located sources from an unexpected direction yielded performance like that for co-located sources at the expected location. Regardless, these results show that the expected masker location alters how selective spatial attention is deployed.

Spatial release from masking in all experiments was significant only where the target and masker came from expected locations. In the co-located condition, when the target was played from an unexpected loudspeaker, there was only a single location from which sounds were played. In the separated condition, there was one masker and one target sound source. To perform well, subjects had to identify which of the sound sources was the target, which may have led to some confusion about the target location. The effect of location uncertainty is consistent with this idea. That is, when a listener is aware of three possible target locations, performance is much worse than if the listener knows *a priori* where the target will come from (Kidd *et al.*, 2005a). Similarly, assigning a target randomly to one of two locations significantly reduces SRM (Allen *et al.*, 2009). In the current experiments, it is likely that a similar effect of uncertainty arises, and that uncertainty about the target location creates masking that counters any SRM when the target is presented from a non-masker-side location.

VII. CONCLUSION

These findings are consistent with past literature showing that prior knowledge of target location is important in spatial selective auditory attention. Importantly, these results show that prior knowledge about masker location also affects performance. Using short stimuli, masking increases if a masker is played from an unexpected location, compared to when the masker is presented from an unexpected location (which gives rise to similar, but symmetrical acoustic cues). Spatial release from masking is reduced when the spatial configuration of target and masker is unexpected, possibly as a result of confusion about target location, which reduces identification performance.

[1]This approach is known in visual attention as the "Posner paradigm." It has been shown that the 80% likelihood of a target at the expected location is sufficiently high that subjects commit most of their attentional resources to that location, even though this is an inefficient strategy for 20% of trials (Posner, 1980). This approach was also utilized in the masker uncertainty experiments carried out by Jones and Litovsky (2008).

Alais, D., and Carlile, S. (**2005**). "Synchronizing to real events: Subjective audiovisual alignment scales with perceived auditory depth and speed of sound," Proc. Natl. Acad. Sci. U.S.A. **102**, 2244–2247.

Allen, K., Alais, D., and Carlile, S. (**2009**). "Speech intelligibility reduces over distance from an attended location: Evidence for an auditory spatial gradient of attention," Atten. Percept. Psychophys. **71**, 164–173.

Arbogast, T. L., and Kidd, G. (**2000**). "Evidence for spatial tuning in informational masking using the probe-signal method," J. Acoust. Soc. Am. **108**, 1803–1810.

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

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.

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

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.

Carlile, S., Delaney, S., and Corderoy, A. (**1999**). "The localisation of spectrally restricted sounds by human listeners," Hear. Res. **128**, 175–189.

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

Connine, C. M. (**2004**). "It's not what you hear but how often you hear it: on the neglected role of phonological variant frequency in auditory word recognition," Psychon Bull Rev **11**, 1084–1089.

Darwin, C. J. (**2008**). "Listening to speech in the presence of other sounds," Philos. Trans. R. Soc. London, Ser. B **363**, 1011–1021.

Durlach, N. I., Mason, C. R., Kidd, Jr., G., Arbogast, T. L., Colburn, H. S., and Shinn-Cunningham, B.G. (**2003a**). "Note on informational masking," J. Acoust. Soc. Am. **113**, 2984–2987.

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

Ebata, M. (**2003**). "Spatial unmasking and attention related to the cocktail party problem," Acoust. Sci. Tech. **24**, 208–219.

Edmonds, B. A., and Culling, J. F. (**2005**). "The spatial unmasking of speech: evidence for within-channel processing of interaural time delay," J. Acoust. Soc. Am. **117**, 3069–3078.

Efron, B., and Tibshirani, R. (**1993**). *An Introduction to the Bootstrap* (Chapman-Hall, New York), pp. 1–456.

Ericson, M. A., Brungart, D. S., and Simpson, B. D. (**2004**). "Factors that influence intelligibility in multitalker speech displays," Int. J. Aviation Psych. **14**, 313–334.

Foxe, J. J., Simpson, G. V., Ahlfors, S. P., and Saron, C. D. (**2005**). "Biasing the brain's attentional set: I. cue driven deployments of intersensory selective attention," Exp. Brain Res. **166**, 370–392.

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

Freyman, R., Helfer, K., McCall, D., and Clifton, R. (**1999**). "The role of perceived spatial separation in the unmasking of speech," J. Acoust. Soc. Am. **106**, 3578–3588.

Hall, J. W., 3rd, Buss, E., and Grose, J. H. (**2005**). "Informational masking release in children and adults," J. Acoust. Soc. Am. **118**, 1605–1613.

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

Hopf, J. M., Boehler, C. N., Luck, S. J., Tsotsos, J. K., Heinze, H. J., and Schoenfeld, M. A. (**2006**). "Direct neurophysiological evidence for spatial suppression surrounding the focus of attention in vision," Proc. Natl. Acad. Sci. U.S.A. **103**(4), 1053–1058.

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

Kanai, K., Ikeda, K., and Tayama, T. (**2005**). "The effect of exogenous spatial attention on auditory information processing," Psychol. Res. **71**(4), 418–426.

Kelly, S. P., Lalor, E. C., Reilly, R. B., and Foxe, J. J. (**2006**). "Increases in alpha oscillatory power reflect an active retinotopic mechanism for distracter suppression during sustained visuospatial attention," J. Neurophysiol. **95**(6), 3844–3851.

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

Kidd, G., Jr., Mason, C. R., and Gallun, F. J. (**2005b**). "Combining energetic and informational masking for speech identification," J. Acoust. Soc. Am. **118**, 982–992.
Kidd, G., Jr., 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.
Knudsen, E. I. (**2007**). "Fundamental components of attention," Annu. Rev. Neurosci. **30**, 57–78.
Langendijk, E. H., and Bronkhorst, A. W. (**2002**). "Contribution of spectral cues to human sound localization," J. Acoust. Soc. Am. **112**, 1583–1596.
Litovsky, R. Y. (**2005**). "Speech intelligibility and spatial release from masking in young children," J. Acoust. Soc. Am. **117**, 3091–3099.
Luce, P. A., and Pisoni, D. B. (**1998**). "Recognizing spoken words: the neighborhood activation model," Ear. Hear. **19**, 1–36.
Mondor, T., and Zatorre, R. (**1995**). "Shifting and focusing auditory spatial attention," J. Exp. Psychol. Hum. Percept. Perform. **21**, 387–409.
Noble, W., and Perrett, S. (**2002**). "Hearing speech against spatially separate competing speech versus competing noise," Percept. Psychophys. **64**, 1325–1336.
Persson, P., Harder, H., Arlinger, S., and Magnuson, B. (**2001**). "Speech recognition in background noise: monaural versus binaural listening conditions in normal-hearing patients," Otol. Neurotol. **22**, 625–630.
Posner, M. I., (**1980**). "Orienting of attention," Q. J. Exp. Psychol. **32**, 3–25.
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.
Rhodes, G. (**1987**). "Auditory attention and the representation of spatial information," Percept. Psychophys. **42**, 1–14.
Sach, A., Hill, N., and Bailey, P. (**2000**). "Auditory Spatial Attention Using Interaural Time Differences," J. Exp. Psychol. Hum. Percept. Perform. **26**, 717–729.
Smith, A. T., Singh, K. D., and Greenlee, M. W. (**2000**). "Attentional suppression of activity in the human visual cortex," Neuroreport. **11**(2), 271–277.
Sonnadara, R. R., Alain, C., and Trainor, L. J. (**2006**). "Effects of spatial separation and stimulus probability on the event-related potentials elicited by occasional changes in sound location," Brain Res. **1071**, 175–185.
Spence, C., and Driver, J. (**1994**). "Covert Spatial Orientation in Audition: Exogeneous and Endogenous Mechanisms," J. Exp. Psychol. Hum. Percept. Perform. **20**, 555–574.
Teder-Sälejärvi, W., and Hillyard, S. (**1998**). "The gradient of spatial auditory attention in free field: An event-related potential study," Percept. Psychophys. **60**, 1228–1242.
Teder-Sälejärvi, W., Hillyard, S., Roder, B., and Neville, H. (**1999**). "Spatial attention to central and peripheral auditory stimuli as indexed by event-related potentials," Brain Res. Cogn. Brain Res. **8**, 213–227.
Teder, W., and Näätänen, R. (**1994**). "Event-related potentials demonstrate a narrow focus of auditory spatial attention," NeuroReport **5**, 709–711.
Watson, A. B. (**1979**). "Probability summation over time," Vision Res. **19**, 515–522.
Watson, C. S. (**2005**). "Some comments on informational masking," Acta. Acust. Acust. **91**, 502–512.
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**, 1176–1186.
Widmann, A., and Schröger, E. (**1999**). "ERP indications for sustained and transient auditory attention with different lateralization cues," in *Psychophysics, Physiology, and Models of Hearing*, edited by V. H. B. K. T. Dau (World Scientific, Singapore), pp. 47–50.
Winkowski, D. E., and Knudsen, E. I. (**2006**). "Top-down gain control of the auditory space map by gaze control circuitry in the barn owl," Nature **439**, 336–339.
Wu, X., Wang, C., Chen, J., Qu, H., Li, W., Wu, Y., Schneider, B. A., and Li, L. (**2005**). "The effect of perceived spatial separation on informational masking of Chinese speech," Hear. Res. **199**, 1–10.