The influence of spatial separation on divided listening[a)]

Virginia Best, Frederick J. Gallun, Antje Ihlefeld, and Barbara G. Shinn-Cunningham[b)]
Hearing Research Center, Boston University, Boston, Massachusetts 02215

(Received 12 January 2006; revised 21 June 2006; accepted 29 June 2006)

If spatial attention acts like a "spotlight," focusing on one location and excluding others, it may be advantageous to have all targets of interest within the same spatial region. This hypothesis was explored using a task where listeners reported keywords from two simultaneous talkers. In Experiment 1, the two talkers were placed symmetrically about the frontal midline with various angular separations. While there was a small performance improvement for moderate separations, the improvement decreased for larger separations. However, the dependency of the relative talker intensities on spatial configuration accounted for these effects. Experiment 2 tested whether spatial separation improved the intelligibility of each source, an effect that could counteract any degradation in performance as sources fell outside the spatial spotlight of attention. In this experiment, intelligibility of individual sources was equalized across configurations by adding masking noise. Under these conditions, the cost of divided listening (the drop in performance when reporting both messages compared to reporting just one) was smaller when the spatial separation was small. These results suggest that spatial separation enhances the intelligibility of individual sources in a competing pair but increases the cost associated with having to process both sources simultaneously, consistent with the attentional spotlight hypothesis. © *2006 Acoustical Society of America.* [DOI: 10.1121/1.2234849]

PACS number(s): 43.66.Dc, 43.66.Pn, 43.66.Qp [AJO] Pages: 1506–1516

I. INTRODUCTION

In the 1950s, Broadbent described auditory attention using the Filter Theory, in which some stimuli are filtered out and others are admitted on the basis of basic characteristics such as frequency and spatial location (Broadbent, 1958). This idea has since been developed most extensively for spatial attention in the visual perception literature, where it is known as the "spotlight" model. In the visual system a representation of space is available directly at the periphery, and is largely maintained at higher levels of the system. The spatial spotlight of attention is thought to operate directly on this representation to modulate competition between simultaneous objects. However, full development of the spotlight model in auditory spatial attention has proven to be a challenge for researchers. Although there is evidence that listeners can orient their attention spatially to enhance the detection and identification of simple targets (Spence and Driver, 1994; Mondor and Zatorre, 1995; Quinlan and Bailey, 1995), the role of spatial attention in the case of *simultaneous* sound sources is less clear. One difficulty lies in the fact that auditory location is computed in the auditory system rather than being represented in the sensory epithelium. As the peripheral representation is a frequency map, sounds coming from different locations often share the same set of peripheral receptors, and auditory source location must be computed from the mixture of signals reaching the left and right ears. This means that the ability to admit an acoustic source at one location and filter out a source at another location will be limited by the ability to separate the acoustic energy coming from different locations in addition to any constraints on the ability to distribute spatial attention.

When a listener must extract the content of one source (a "target") in the presence of competing sources ("maskers"), spatial separation of the target and masker is generally beneficial to performance (Bronkhorst, 2000). When the masker reduces the audibility of components of the target ("energetic masking"), there are two ways in which spatial separation offers an advantage. First, the relative energy of the target and masker at the ears changes with target and masker location, such that spatial separation increases the target audibility in each frequency band at one of the ears (the "better ear"). Second, binaural processing allows listeners to detect the presence of target energy at a particular time and frequency band if the target and masker contain different interaural time and/or level differences (Zurek, 1993; Bronkhorst, 2000). When competing sources do not have significant frequency overlap and reduced audibility is not the primary source of interference, a masker with similar spectrotemporal characteristics can still interfere with the perception of the target (so-called "informational masking"). One important source of informational masking is trial-to-trial variability in the target and/or masker, which leads to listener uncertainty as to how to classify a given spectrotemporal pattern. This kind of interference is reduced when the target and masker are distinguished in a way that reduces confusions between them. For example, differences in perceived spatial location have been shown to reduce informational masking by allowing listeners to selectively attend to the target at the location of interest (Freyman *et al.*, 1999; Freyman *et al.*, 2001; Best

[a)]Portions of this work were presented at the 2005 International Conference on Auditory Display and the 2006 mid-winter meeting of the Association for Research in Otolaryngology.

[b)]Author to whom all correspondence should be addressed. Electronic mail: shinn@cns.bu.edu *et al.*, 2005; Kidd *et al.*, 2005; Shinn-Cunningham *et al.*, 2005). In these situations, the spotlight of attention may play a role.

This study examines the effect of spatial separation when a listener must attend to *two* sustained sound sources simultaneously. In the example used here, keywords had to be extracted from each of two talkers in a competing pair. Previous studies of divided listening with speech have typically used dichotic signals in which each ear receives only one of the two competing sources (Broadbent, 1954; Massaro, 1976) and have not considered spatial factors in detail. One recent study (Shinn-Cunningham and Ihlefeld, 2004) examined the effect of spatially separating two competing talkers (by 90°) on the ability of listeners to report both messages. In that study, in which the talkers were presented at different relative intensities, the louder talker could always be recalled with relative ease. As a result, listeners appeared to allocate attention primarily to the quieter talker, a strategy similar to that employed in a selective attention task. Spatial separation improved performance, probably for the same reasons it improves performance in a true selective listening task (discussed above). In contrast, in the present study the two talkers were presented with equal intensity and were separated symmetrically about the midline. Thus, the two talkers are equally difficult to hear, and processing resources should be more equally allocated between the two competing talkers (i.e., the listening strategy is more likely to engage truly "divided" listening). In this case, it is not clear what the effect of spatial separation of the two targets might be. It is reasonable to expect that spatial separation would be advantageous in that it would enhance the audibility of the two sources as well as reducing confusion between them, as described above. However, if one considers the putative spotlight of spatial attention, spatial separation could be detrimental in a divided listening task. If the spotlight is focused at a given moment on one source, then the other is likely to be excluded if it is distant from the first, and simultaneous processing will be impaired.

In Experiment 1, the effect of spatial separation on the ability of listeners to report keywords from two simultaneous talkers was examined. Results suggest that there is little effect of spatial separation overall, except for some modulation of performance due to changes in energy at the two ears. In Experiment 2, an attempt was made to untangle two potentially opposing effects: (1) a benefit of spatial separation for segregating competing messages; and (2) a disadvantage of spatial separation when dividing spatial attention.

II. EXPERIMENT 1

A. Methods

1. Subjects

Eight paid subjects (ages 20–30) participated in the experiment. Four subjects had previous experience in psychophysical studies of a similar nature. All subjects participated in Experiment 1A, and six of the subjects went on to participate in Experiment 1B.

2. Speech materials

Speech materials were spoken sentences taken from the publicly available Coordinate Response Measure speech corpus (Bolia *et al.*, 2000). These sentences all contain seven words, three of which are keywords that vary from utterance to utterance. The form of the sentences is "Ready *call-sign* go to *color number* now," where the italicized words indicate keywords. In the corpus there are eight possible call-signs ("arrow," "baron," "charlie," "eagle," "hopper," "laker," "ringo," "tiger"), four possible colors ("blue," "green," "red," "white"), and eight possible numbers (1–8). All combinations of these words produce 256 phrases, which are each spoken by eight talkers (four male, four female), giving a total of 2048 sentences. The sentences are time aligned such that the word "ready" always starts at the same time, but some variations in overall rhythm occur between different sentences so that the keywords in different utterances are not exactly aligned.

3. Stimuli

For each trial, two sentences spoken by the same talker were chosen randomly from the corpus with the restriction that all keywords differed in the two sentences. In order to reduce the energetic interference between the two sentences, they were processed to produce intelligible speechlike signals that had little spectral overlap (Shinn-Cunningham *et al.*, 2005; see also Arbogast *et al.*, 2002; Brungart *et al.*, 2005, for similar approaches). The signals were bandpass filtered into eight nonoverlapping frequency bands of 1/3 octave width, with center frequencies spaced evenly on a logarithmic scale from 175 to 4400 Hz. Four bands were randomly chosen for the first sentence (two from the four lower bands and two from the four higher bands). The Hilbert envelope of each band was then used to modulate a sinusoidal carrier at the center frequency of that band, and the sentence was reconstructed by summing the four modulated sinusoids. For the second sentence, the remaining four frequency bands were chosen and the same procedure was followed. The two reconstructed sentences were root-mean-square (rms) normalized to result in a relative level of 0 dB (see Fig. 1 for example spectra).

The stimuli were processed to create binaural signals containing realistic spatial cues and presented over headphones. In Experiment 1A, a full set of spatial cues was used in the simulation. Binaural stimuli were created by convolving the speech signal with the appropriate anechoic left and right head-related transfer functions (HRTFs) measured on a KEMAR manikin at a distance of 1 m (Brungart and Rabinowitz, 1999). The two binaural stimuli were then added to simulate the two speech sources at their desired locations in external space. In Experiment 1B, level differences that were present in the HRTF simulation were removed in order to eliminate location-dependent changes in the relative level of the two sentences at the ears; thus only one spatial cue (the interaural time difference, ITD) was used in these simulations. A single, frequency-independent ITD was extracted from each HRTF pair by finding the time delay of the peak in the broadband interaural cross-correlation function. These

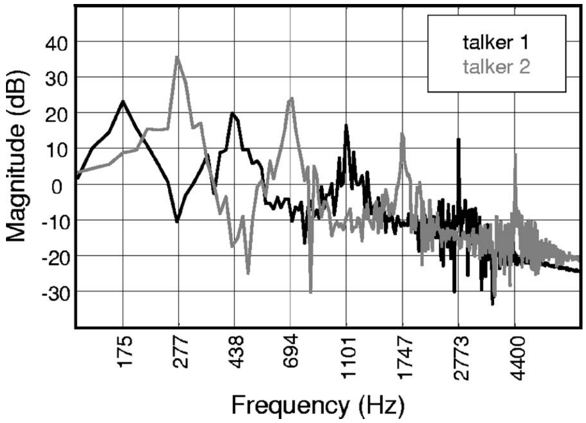

FIG. 1. Frequency spectra of two example sentences after processing. The two signals were processed to minimize their spectral overlap. Sentences were divided into eight 1/3 octave bands with center frequencies between 175 and 4400 Hz. Four different bands were chosen for each sentence and their envelopes used to modulate sinusoids at the center frequency of each band. Intelligible speech signals were reconstructed by summing the four modulated sinusoids.

ITDs were used to delay the left and right ear signals relative to one another to control the perceived lateral locations of the competing sources.

4. Procedure

Subjects were seated in a sound-treated booth in front of a personal computer (PC) terminal. Presentation of the stimuli was controlled by a PC, which selected the stimulus to play on a given trial. Digital stimuli were sent to Tucker-Davis Technologies hardware for digital-analog conversion and attenuation before presentation over insert headphones (Etymotic Research ER-2). Following each presentation, subjects indicated their responses by clicking on a graphical user interface displayed on the PC monitor. The interface consisted of an eight-by-four grid, with the color of the square and the number printed within it representing one of the 32 possible color/number pairs.

In each trial, two simultaneous sentences were presented and subjects were required to respond with *two* color/number pairs (in either order). No feedback was provided. A response was considered correct only if both color/number pairs were reported correctly. Note that chance performance, achieved by randomly guessing the two color/number pairs, is 0.3% for this task.

Stimulus locations were all on the horizontal plane passing through the ears (0° elevation). Performance was measured with sources separated symmetrically about the midline with separations of 0°, 30°, 60°, 90°, 120°, 150°, or 180°. The seven configurations were presented five times in a random order in each run. Each subject completed ten such runs for each experiment, for a total of 50 responses for each configuration. The 20 runs (ten each for Experiments 1A and 1B) were carried out over four to five sessions. Subjects did no more than one hour of testing per day.

5. Training

Before the start of the experiment, subjects participated in a short series of training runs designed to familiarize them with the stimuli and task. In a training test, subjects were presented with stimuli containing a single sentence in quiet, and were required to indicate the color/number pair they perceived. After each trial, correct-answer feedback was provided by a written message on the screen. A training run consisted of 130 trials. Subjects completed as many runs as required to bring their proportion of correct responses to at least 95%. All subjects reached this level within three training runs.

B. Results—Experiment 1A

Individual subjects differed in their absolute level of performance, but overall trends were similar. Mean percent correct scores across subjects (and standard errors) are shown in Fig. 2(a). Spatial separation had a modest effect on performance; for a given subject, performance did not vary by more than 30 percentage points across all spatial configurations. However, there were consistent patterns in the data: performance tended to first increase and then decrease with increasing source separation, peaking at 90°–120° separation.

In order to factor out overall differences in subject performance and concentrate on the effect of spatial separation, percent correct scores for each subject were normalized by subtracting the percent correct in the colocated (separation 0°) configuration. The resulting normalized values summarize how performance changed with source separation. Figure 2(b) shows the normalized data pooled across the eight subjects (means and standard errors). The trends described for the raw data are reinforced when individual differences are factored out in this way: increasing the spatial separation tended to first improve and then degrade performance. A repeated measures analysis of variance (ANOVA) confirmed that there was a significant effect of spatial separation on performance $[F(6,42)=7.1,\ p<0.001]$. Post-hoc analyses (pairwise comparisons with a Bonferroni correction) indicated that separations of 60°, 90°, 120°, and 150° were all significantly different from the colocated configuration (no other comparisons reached significance).

Although the two targets were nominally presented with equal intensity, variations in the HRTFs with source location caused variations in the level of each target at each ear. This is especially true for a target placed to the side, where the acoustic shadow cast by the head greatly attenuates the level received at the far ear, particularly at high frequencies (above about 2 kHz). Indeed, for a given spatial configuration, each of the two sources has a different ear in which their level (relative to that of the other source) is greater. Moreover, the magnitude of this better ear "level ratio" varies as a function of the spatial configuration. Note, however, that as the stimuli were composed of nonoverlapping frequency bands, the level ratio does not correspond to signal-to-noise ratio in the traditional sense (i.e., it is not the "within frequency band" signal-to-noise ratio and thus is not a direct measure of energetic masking). It may be better described as representing an overall loudness ratio of one target relative to the other.

An acoustic analysis was performed to examine whether

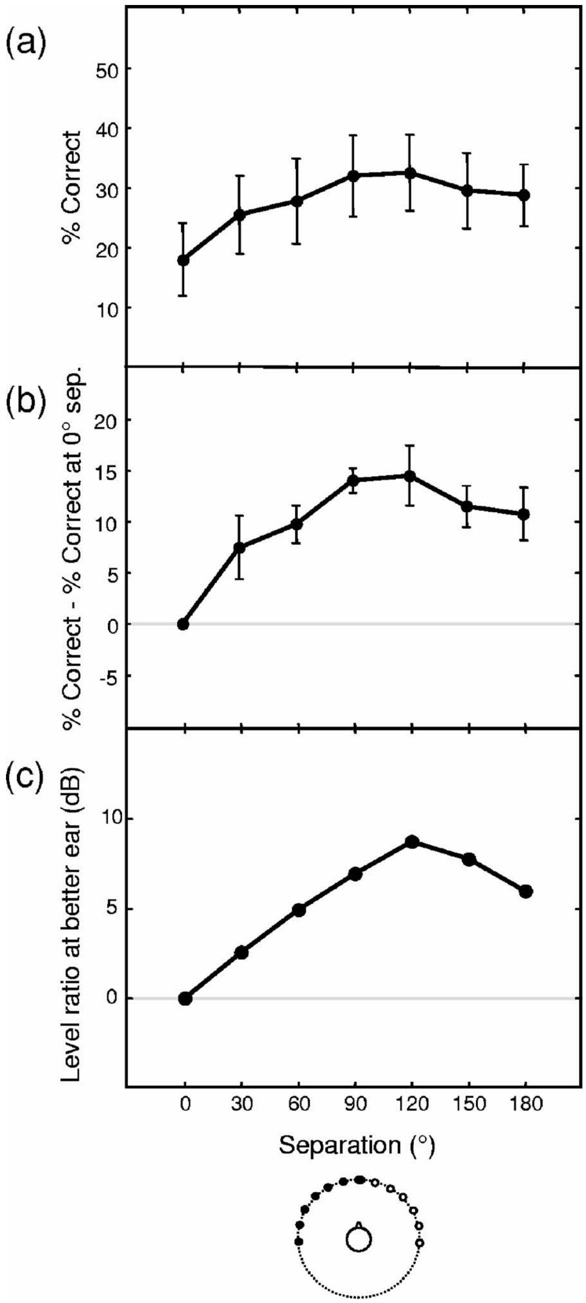

FIG. 2. (a) Mean percent correct scores for the different spatial configurations in Experiment 1A. Results are pooled across the eight subjects and error bars indicate standard error of the mean. (b) Normalized percent correct scores for Experiment 1A. Normalization was carried out for each individual by subtracting the score for the colocated configuration. Results are pooled across the eight subjects and error bars indicate standard error of the mean. (c) Level ratios for the different spatial configurations. Level ratios describe the level of a source in its "better ear" relative to the level of the other source. These ratios were calculated for 50 example stimuli and the means across the two sources (at their respective better ears) are shown.

the relative level of the competing sources at the ears might help to explain the trends seen in the behavioral data. For each spatial configuration, 50 speech pairs were generated and the level ratio (LR) for each source was calculated using the broadband rms level of each source after HRTF filtering for each ear. The changes in better-ear LR as a function of spatial separation (averaged across the two sources and their respective better ears) are shown in Fig. 2(c). Note that by definition, the LR for the colocated pair is 0 dB. It can be seen that the LR increases as separation grows to 120°, but then decreases with further separation. This analysis suggests that the relative overall loudness of the two talkers at each ear can at least partially account for the behavioral results. Performance was positively correlated with the mean of the computed LRs across the two better ears ($r^2=0.90$, $p=0.001$).

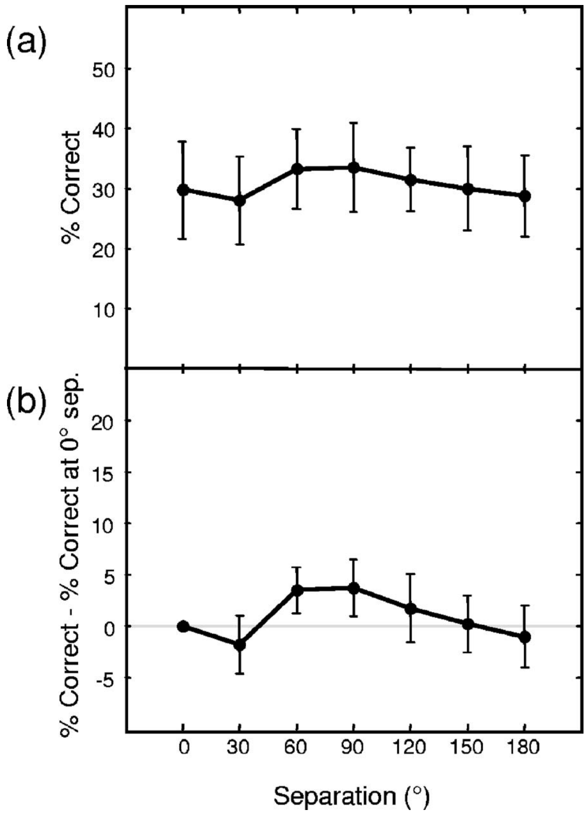

FIG. 3. (a) Mean percent correct scores for the different spatial configurations in Experiment 1B. Results are pooled across the six subjects and error bars indicate standard error of the mean. (b) Normalized percent correct scores for Experiment 1B. Normalization was carried out for each individual by subtracting the score for the colocated configuration. Results are pooled across the six subjects and error bars indicate standard error of the mean.

Experiment 1B was designed to eliminate energy effects in order to confirm their role in the results of Experiment 1A and to determine whether there is any residual influence of perceived spatial separation of the two sources in a divided attention task. By using only ITDs in the spatial simulation, the level variations induced by the HRTF processing in Experiment 1A were removed (in essence, the LRs for these stimuli are fixed at 0 dB).

C. Results—Experiment 1B

The mean percent correct scores (and standard errors) across the six subjects for the different configurations are shown in Fig. 3(a). The curve is noticeably flatter than that obtained in Experiment 1A. Interestingly, overall performance is better (by approximately five percentage points) in Experiment 1B than in Experiment 1A. However, it is important to keep in mind that all subjects in Experiment 1B

TABLE I. Distribution of error types for incorrect trials in Experiment 1A (left column) and Experiment 1B (right column). Results are pooled across subjects and across the seven spatial separations.

Error type	Experiment 1A (% trials)	Experiment 1B (% trials)
Drop 1	46.1	43.8
Drop 2 (same)	11.0	8.0
Drop 2 (different)	8.4	5.3
Switch	5.7	6.7
Drop 3	1.2	0.7
Drop 4	0.2	0.0

completed Experiment 1A first. Thus, overall improvements in performance may reflect the fact that subjects were more experienced in the task in Experiment 1B.

As in Experiment 1A, percent correct scores for each subject at each reference location were normalized by subtracting the score in the colocated (separation 0°) configuration. In Fig. 3(b), the normalized data pooled across the six subjects are shown. There was no consistent change in performance across subjects with increasing separation (a repeated measures ANOVA found no significant effect [$F(6,30)=1.1$, $p>0.05$]).

D. Error analysis

To examine the different kinds of errors that listeners made in incorrect trials, errors were classified as one of six types: (i) "drop 1" errors where one keyword was incorrect; (ii) "drop 2 (same)" errors where both keywords from one of the two stimuli were incorrect; (iii) "drop 2 (different)" errors where one keyword from each source was incorrect; (iv) "switch" errors where the four keywords were reported but in incorrect pairs; (v) "drop 3" errors where three of the four keywords were not reported; and (vi) "drop 4" errors where none of the keywords were included in the response. An example of each kind of error is given in the Appendix.

Table I shows the average error rates pooled across subjects and spatial separations, for Experiments 1A (left column) and 1B (right column). In both experiments, no influence of the side of the talker (left versus right) was found for error types involving more error for one talker than the other (sign test, $p>0.05$). Thus performance was roughly equally affected by errors for the sources in the left and right hemifields. Table I shows that the majority of errors in both experiments were "drop 1" followed by smaller numbers of "drop 2 (same)," "drop 2 (different)," and "switch" errors.

To investigate which of the different error types were responsible for the changes in performance observed with increasing spatial separation in Experiment 1A, error rates were normalized by subtracting the individual error rates at 0° separation. Figure 4(a) shows these individually normalized error rates for the different error types, although "drop 3" and "drop 4" errors are not shown due to their extremely low occurrence. Error bars have been omitted for the sake of clarity. The error type that changes most dramatically with spatial separation is the "drop 2 (same)" error. Repeated measures ANOVAs on the different error rates confirmed that this was the only type to change significantly with spatial separation [$F(6,42)=5.3$, $p<0.001$]. This suggests that the main influence of spatial separation in Experiment 1A was to alter the probability that a listener would miss one of the two sources completely.

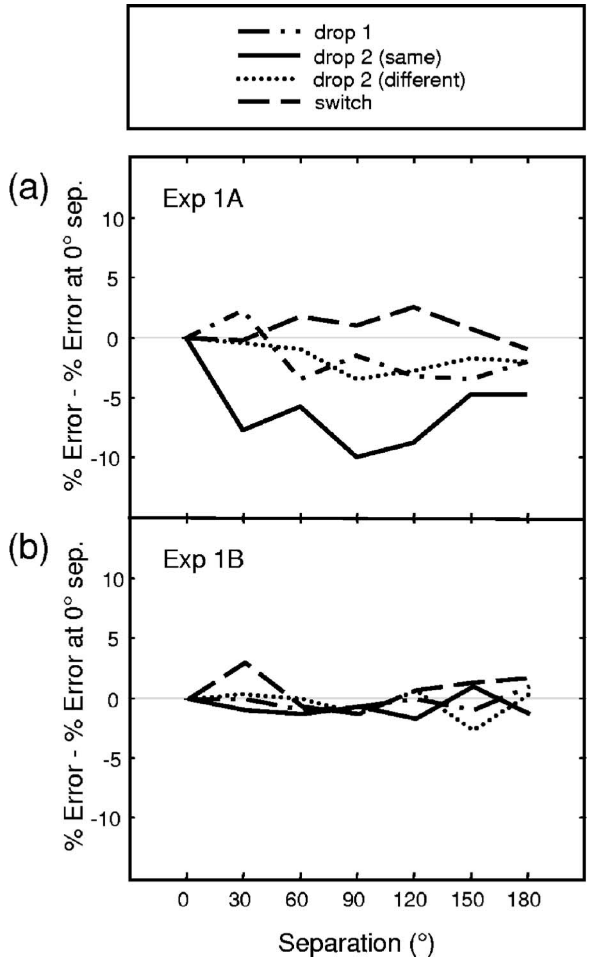

FIG. 4. Normalized error rates for incorrect trials in (a) Experiment 1A and (b) Experiment 1B. Normalization was carried out for each individual and each error type by subtracting the rate for the colocated configuration. Results are pooled across subjects. Error bars have been omitted for the sake of clarity, but statistical tests are described in the text.

Even though there was no significant change in overall performance with spatial separation in Experiment 1B, it is possible that the types of errors made depended on the separation. To investigate this possibility, error rates were normalized by subtracting the individual error rates at 0° separation as for Experiment 1A. Figure 4(b) shows these normalized error rates, although again "drop 3" and "drop 4" errors are not shown due to their extremely low occurrence. It appears that spatial separation has no consistent effect on the general pattern of errors, a conclusion confirmed using repeated measures ANOVAs on the different error types.

E. Discussion

These experiments examined the ability of listeners to track two simultaneous speech sources with minimal spectral overlap. When full HRTFs were used, performance improved with moderate separations (best performance for separations in the range 90°–120°) but then decreased with further separation. This trend was eliminated in Experiment 1B, when spatial locations were simulated using only ITDs. Furthermore, the pattern of responses in Experiment 1A was correlated with the relative levels of the two sources at their acoustically better ears. These results strongly suggest that variations in the relative level of the two sources at the ears modulated the difficulty of the task in Experiment 1A and, ultimately, the accuracy of responses. Indeed, in the best spatial configuration (120° separation), the mean level ratio was 9 dB, meaning that each target source was 9 dB more intense in its better ear than the competing source. Interestingly, if listeners are able to use these monaural level ratios it suggests that they use the information at the two ears *independently* when tracking two sources in different hemifields.

A surprising finding in the error analysis was that there was no change in the frequency of "switch" errors as spatial separation was varied. Previous experiments using the Coordinate Response Measure corpus have found that this kind of error is quite common, especially when the talkers are of the same sex and in close spatial proximity (Brungart *et al.*, 2001). It would thus be expected that this type of error would be high in the 0° separation condition and would decrease with spatial separation. The fact that these error rates were relatively low reflects a robust ability to properly stream the two talkers in this task, i.e., to keep the two talkers as separate perceptual objects even when they are at the same spatial location. In addition, it is likely that other kinds of errors (particularly "drop 1" errors) limited performance to such an extent that the influence of "switch" errors was masked.

There was no evidence in this experiment of differential processing of sources in the left and right hemifields. This is in contrast to a large body of literature demonstrating that speech stimuli coming from the right are preferentially processed. When two speech sources are presented simultaneously to the left and right sides, a right-side advantage has been demonstrated when spatial location is determined by ear of presentation, external loudspeaker location, or even ITD only (see Darwin *et al.*, 1978 and Morais, 1978 for reviews). It is not clear why a left/right-side difference was not observed in the present study, although the differences reported in the literature are often quite small (especially when the sources are mixed at the two ears such as in our simulation rather than presented to different ears) and may be revealed only in larger data sets. Furthermore, it may be that these kinds of speech-specific effects occur less robustly for highly degraded speech signals such as those used in this experiment.

The results of Experiment 1B suggest that changes in the lateral positions of the two sound sources (produced by differences in ITD) did not influence performance on this task. One possibility is that perceived location estimates based on ITD alone are "diffuse," and that the two broad images associated with the two sources were not very clearly defined in this experiment. However, this seems unlikely based on recent data showing that when energetic masking is not the primary factor limiting performance, changes in perceived location give rise to similar improvements in selective listening tasks, regardless of which spatial cues produce the differences in perceived location (Shinn-Cunningham *et al.*, 2005). Given this, the results of Experiment 1 can be interpreted as evidence against the idea of a spotlight of auditory attention. In this task there appears to be no increased difficulty in following sources that are widely separated compared to those in close proximity. However, there is another potential explanation for this result. Despite efforts to minimize spectral overlap between the two sources, there was undoubtedly some remaining spectral overlap between the talkers, which may have been reduced by spatial separation. Thus, spatial separation may have rendered *each of the two sources* more intelligible (Gallun *et al.*, 2005; Shinn-Cunningham *et al.*, 2005). Moreover, spatial separation may improve segregation of the talkers. Both of these effects would work in opposition to any degradation in performance as sources fall outside the spatial spotlight of attention.

III. EXPERIMENT 2

A second experiment was performed to disentangle the possible opposing effects discussed above. The experiment was similar to Experiment 1, but different levels of noise were added to the speech signals in order to equate the difficulty of understanding the two talkers in the different configurations. In effect, the interference between the talkers was matched in all spatial configurations, allowing the efficiency of dividing attention to be compared directly. As there was no longer a need to reduce energetic masking, unprocessed speech was used in this experiment. Furthermore, in an effort to increase the "naturalness" of the setup, the two speech samples were chosen to have different voices and were presented from a pair of loudspeakers in a classroom. Furthermore, the locations were fixed in blocks rather than changing from trial to trial.

A. Methods

1. Subjects

Five paid subjects (ages 20–30) participated in Experiment 2. All had previous experience in auditory psychophysical studies, and one participated in Experiment 1.

2. Stimuli

Speech materials were taken from the same corpus used in Experiment 1 (described in Sec. II A 2). For each trial, two sentences spoken by different male talkers were chosen randomly from the corpus with the restriction that all keywords differed in the two sentences. No processing of the signals was done other than rms normalization, which set the relative levels of the two signals to 0 dB.

3. Procedure

Subjects were seated on a chair fitted with a headrest in a quiet, empty, carpeted classroom. The two sentences were presented from two matched Bose cube loudspeakers positioned 1 m from the listener at ear level. Stimuli were generated by PC-controlled Tucker-Davis Technologies hardware, amplified by a Crown amplifier, and sent to the loudspeakers via an eight-relay output module (KITSRUS

K108). Following each presentation, subjects indicated their responses by pressing the appropriate letter/number keys on a hand-held keypad.

Testing was done at three spatial separations about the midline: 10° (close), 90° (intermediate), and 180° (far). For each spatial configuration, subjects completed a noise calibration test as well as tests measuring single-task and dual-task performance.

The noise calibration test was designed to find the appropriate level of broadband white noise to add to the two loudspeakers such that when selectively attending to one of the two talkers, each individual listener could report keywords from the attended talker with an accuracy of 85%. Pairs of sentences were presented in noise at a variety of levels. Subjects were asked to report the color/number from the left talker only. No feedback was provided. Six noise levels were tested, with each noise level presented 25 times in a random order for a total of 150 trials in the test. The range of noise levels tested was fixed across listeners but was different for each spatial configuration. The levels, stated in decibels relative to the level of each of the speech signals, were in 6 dB steps between −20 and 10 dB (close), −14 and 16 dB, (intermediate), and −16 and 14 dB (far). For each configuration, a logistic function was fit to the raw data and the noise level corresponding to 85% performance was estimated.

In the experimental sessions that followed the noise calibration test, subjects were presented with one sentence plus noise (at the appropriate level) from each loudspeaker. Performance was measured in both single-task and dual-task conditions. Importantly, the stimulus set was identical in these two situations; only the task of the listener changed. In the single-task condition, subjects were asked to report keywords from either the left talker (as in the noise calibration test) or from the right talker (the talker to be attended was fixed within a test). Based on the noise calibration test, it was expected that performance in these tests would be approximately 85%. In the dual-task condition, listeners were asked to follow *both* talkers, and were required to respond with *two* color/number pairs (as in Experiment 1). Verbal instructions indicated that listeners should enter their response to the left talker followed by their response to the right talker (an instruction that was not given in Experiment 1), and responses were considered correct if both color/number pairs were reported correctly and in the correct order (left then right). No feedback was provided during testing.

Two tests of 100 trials each were completed for each of the three tasks conditions (left single task, right single task, dual task) in each of the three configurations. The resulting 18 tests were completed in a counterbalanced fashion (where the random ordering of the first nine tests was reversed in the second nine tests) to eliminate biases due to learning. All testing (noise calibration and single-/dual-task) was completed over six to seven sessions per subject.

4. Training

Before performing Experiment 2, subjects participated in three short training runs designed to familiarize them with the stimuli and task. In a training test, subjects were presented with pairs of sentences using one of the spatial configurations and were instructed to attend to the left talker, the right talker, or to both talkers. The combinations of configuration and task were randomly chosen for each listener, however over the course of the three training runs, each listener was exposed to each task (left, right, dual) and each spatial configuration (close, intermediate, far). Subjects indicated the color/number pair(s) they perceived and received correct-answer feedback via a written message on the hand-held response unit. A training run consisted of 100 trials.

TABLE II. Noise levels determined from the noise calibration tests for individual subjects as well as the mean across subjects. Levels are in decibels relative to the level of the speech signals, and correspond to 85% performance on the single task for each spatial configuration.

Subject	Close	Intermediate	Far
S1	−15.8	1.3	1.9
S2	−13.9	−1.7	−0.3
S3	−25.9	−3.0	−5.3
S4	−11.9	0.8	−2.1
S5	−22.3	−5.5	−4.8
mean	−16.9	−1.6	−2.1

B. Results

Table II lists the noise levels (in decibels relative to the level of the speech signals) determined from the noise calibration test for each individual in each configuration. The mean noise levels across subjects are also shown. As predicted, the task of following one of the talkers was most difficult for the close configuration so that a lower level of noise had to be added to reduce performance to 85% than for the other configurations. Note that the increase in noise level with spatial separation in these single-task trials is one estimate of the amount of spatial unmasking (in decibels) for these separations.

Mean percent correct scores across subjects are shown in Fig. 5. The upper lines represent performance in the "attend right" (dashed line) and "attend left" (dotted line) trials, respectively. These data confirm that subjects performed the single task with approximately 85% accuracy in all three configurations, although performance in the "attend right" condition was better than in the "attend left" on average (significant using a sign test, $p < 0.05$). This may simply be an artifact of small differences in loudspeaker characteristics, or it may represent a "right-side advantage" as discussed in Sec. II E. Although no such asymmetry was found in Experiment 1, this may be because this subtle speech-specific bias is more prominent for clear unprocessed speech signals than for the processed stimuli used in Experiment 1. To confirm that subjects were able to distinguish the left from right loudspeaker successfully, the errors that resulted from subjects reporting the keywords from the wrong talker (i.e., reporting the right talker first or the left talker last) were counted. These errors were extremely rare (occurring in 0.9% of trials in total), confirming that listeners had no difficulty in judging the relative locations of the two talkers, even in the close configuration.

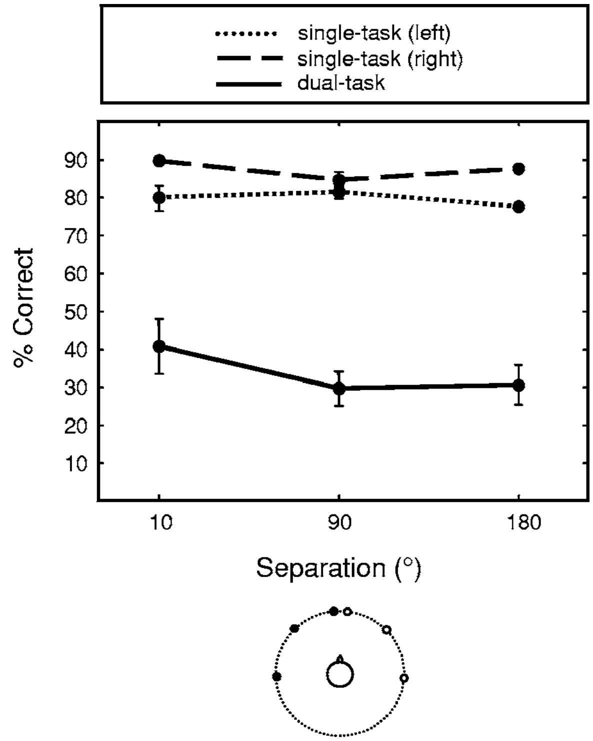

FIG. 5. Mean percent correct scores for the different spatial separations in Experiment 2 in both single-task (dashed and dotted lines) and dual-task (solid lines) trials. Single-task scores are shown separately for "attend left" and "attend right" conditions. Dual-task scores are based on trials in which both sources were correctly reported. Results are pooled across the five subjects and error bars indicate standard error of the mean.

The solid line represents performance in the dual-task trials. Performance was near 40% for close targets but near 30% for intermediate and far targets. This is slightly higher than the average performance in Experiment 1, probably due to the use of more natural and robust unprocessed speech signals.

To calculate a single measure of the "cost" associated with divided listening in the different spatial configurations, performance in the dual task was subtracted from average performance in the single task for each subject. These cost values are plotted for each subject in Fig. 6. A repeated measures ANOVA confirmed that there was a significant effect of spatial separation $[F(2,8)=9.6, p<0.05]$. Although post-hoc tests (pairwise comparisons with a Bonferroni correction) failed to reach significance, visual inspection of Fig. 6 makes it clear that the effect was due primarily to the fact that cost values were consistently smallest for the close configuration.

C. Error analysis

To further examine the different kinds of errors that listeners made in incorrect dual-task trials, errors were classified into different error types. The same error classification described in Experiment 1 was used, with the exception that drop errors associated with the left and right talkers were analyzed separately. Infrequent confusion between the left and right talkers, and combinations of a drop error and a left/right confusion, were pooled and classified as "other" errors.

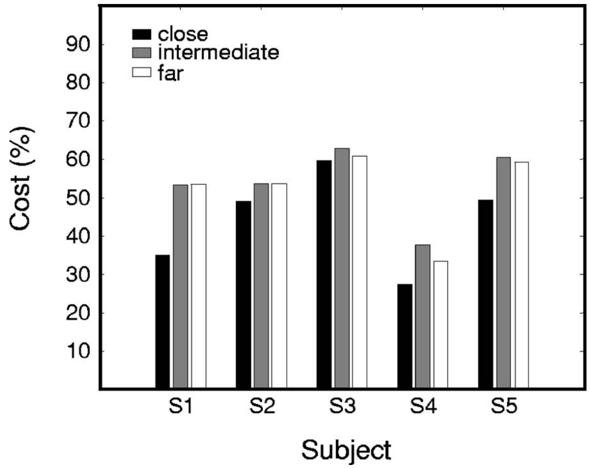

FIG. 6. Performance cost associated with responding to both targets (difference in percent correct between single-task and dual-task trials). Cost is shown for each subject in each of the three spatial configurations.

Table III shows the average error rates pooled across subjects and spatial separations. The majority of errors were of the kind "drop 1," as was the case in Experiment 1. For this experiment, however, a larger proportion of the "drop 1" errors were due to an error in reporting the right talker than the left. A smaller number of "drop 2" errors also occurred, with most of these involving errors in reporting both keywords from the right talker or one keyword from each talker. A small number of "switch" errors occurred, and these were approximately as frequent as they were in Experiment 1.

The fact that the error rates were much higher for the right source than for the left source—reversing any right-talker processing advantage observed in the single task—is most likely a result of the instructions given in this experiment. It seems that asking listeners to report the left talker first encouraged them to give *higher priority* to the left talker. In contrast, listeners in Experiment 1 were simply instructed to respond to both sources and no such asymmetry in performance was observed, presumably because attention was allocated more equally between the left and right talkers.

To investigate which of the error types drove the changes in performance observed with increasing spatial separation, error rates for the most common error types were normalized by subtracting the individual error rates at 10° separation (Fig. 7). Error bars have been omitted for the sake of clarity. This figure shows that most error types increased when spatial separation was increased, which explains the

TABLE III. Distribution of error types for incorrect trials in Experiment 2. Results are pooled across the five subjects and the three spatial separations.

Error type	% trials	
	Left	Right
Drop 1	15.6	25.8
Drop 2 (same)	2.9	6.1
Drop 2 (different)	5.3	
Switch	4.1	
Drop 3	0.8	
Drop 4	0.0	
Other	6.0	

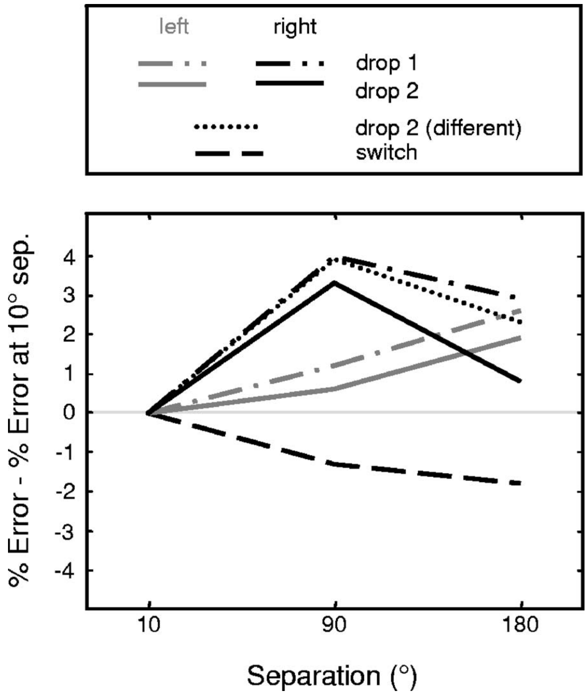

FIG. 7. Normalized error rates for incorrect dual-task trials in Experiment 2. Normalization was carried out for each individual and each error type by subtracting the rate for the "close" configuration. Results are pooled across the five subjects. Error bars have been omitted for the sake of clarity, but statistical tests are described in the text.

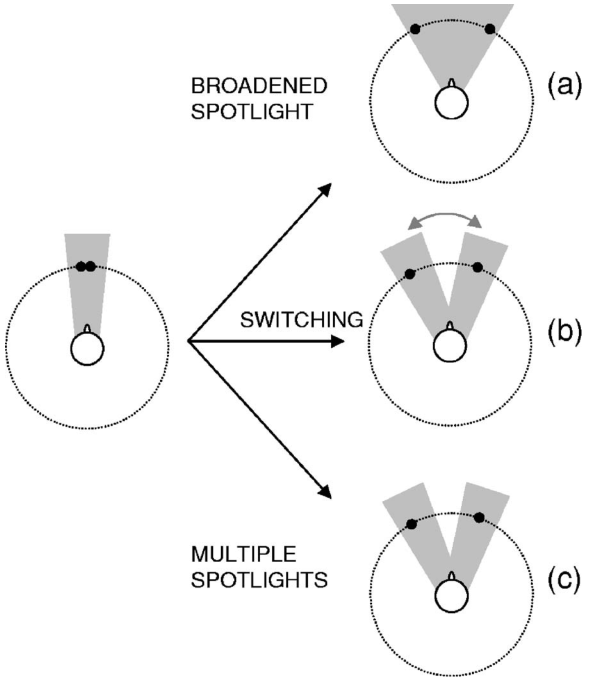

FIG. 8. If auditory spatial attention acts as a "spotlight" to enhance the perception of a relevant source and exclude others, it may be advantageous to have any sources of interest within a restricted region. Spatial separation of two targets of interest will require either (a) a broadening of the spotlight, (b) a strategy where attention is switched between the two sources, or (c) a splitting of the spotlight to form multiple spotlights.

overall decrease in performance at these larger separations. Interestingly, errors involving the right talker tended to be maximal at the 90° separation, and errors involving the left talker tended to be maximal at the 180° separation. Six repeated measures ANOVAs (one for each error type) were performed, and the effect of spatial separation reached significance only for the "drop 2 (right)" errors $[F(2,8)=6.1, p<0.05]$ (solid black line) and "drop 2 (different)" errors $[F(2,8)=9.8, p<0.05]$ (dotted black line). This suggests that the primary effect of spatial separation was to increase the probability that a listener would misreport the right talker completely or drop one keyword from each of the talkers. Note that "switch" errors *decreased* with spatial separation, consistent with spatial separation reducing confusion between elements of the two sources. Importantly, if these were not counted as errors (i.e., if the important criterion for "correctness" was simply how many of the four keywords were reported) the increase in dual-task cost as a function of spatial separation would be enhanced.

D. Discussion

These results show that there is a substantial performance cost associated with responding to two sentences compared to responding to one. This cost ranged between 30 and 60 %, depending on the subject and spatial configuration. This result is consistent with recent work (Gallun *et al.*, submitted) showing that performing two of the same kind of task (such as identifying two phrases simultaneously) results in a performance cost. Importantly, a number of factors may contribute to this cost. Not only are the attentional demands increased in the dual-task trials, but subjects must report four items instead of two and the memory load is increased. However, the main factor of interest in this study was not the magnitude or origin of the cost, but the impact of spatial configuration on the cost. It can be assumed that the memory load and general response demands were constant across spatial configurations. Furthermore, the design of the experiment meant that the accuracy with which each individual talker could be followed was constant across spatial configurations. Thus, we can be reasonably confident that changes in performance as a function of spatial separation are largely attributable to the distribution of attention between the competing talkers.

The main finding of the current experiment was that the cost associated with divided listening was smallest when the targets were in close spatial proximity. This finding is consistent with the idea that auditory spatial attention operates like a spotlight. As discussed in the Introduction, a spatially restricted attentional spotlight predicts that the simultaneous processing of two distant targets will be impaired. Several models have been put forward to describe this effect (see Fig. 8). It may be that the spotlight has to be broadened to simultaneously encompass two distant targets (the "zoom lens" model; see Eriksen and St. James, 1986) or switched rapidly between them, and in either case a decrease in processing efficiency may result. Alternatively, it may be that "multiple spotlights" of attention can be deployed (McMains and Somers, 2004; 2005), in which case there may be little or no cost in performance with spatial separation. Although the present experiments were not designed to address which of these mechanisms might apply in the auditory system, the error patterns give some preliminary indications. The fact that the majority of errors involved one source or the other suggests that attention was focused on only one source at a time. This is further supported by the fact that when subjects gave one talker priority on the basis of response requirements (the left talker in Experiment 2), they tended to make many more errors on the other talker. These errors tended to increase with spatial separation, consistent with the spotlight being spatially restricted. On the other hand, there were also a reasonable number of errors involving both talkers, and their occurrence increased with spatial separation. This is more consistent with a "broadening" of the spotlight; listeners attended to both sources but with reduced processing efficiency as they stretched limited attentional resources to cover a greater area (Eriksen and St. James, 1986). The occurrence of some "switch" errors (although small in number) at all spatial separations lends some weight to this possibility, as including both sources within the same attentional channel could disrupt perceptual segregation of the two sources. Importantly, these patterns are also consistent with listeners employing a restricted spotlight of attention that is switched between the sources, resulting in information being missed from one source during instances when attention was on the other source. A final possibility is that listeners are able to attend to two distinct locations in space (using "multiple spotlights"), but that there is an overhead associated with splitting the attentional spotlight in this way. This kind of model has been described in vision, where subjects are able to allocate attention to two noncontiguous zones of the visual field simultaneously (Awh and Pashler, 2000; Müller *et al.*, 2003; McMains and Somers, 2004). Furthermore, a similar division of attention has been shown for frequency detection in the auditory system, where listeners primed with two different frequencies are better able to detect probe tones occurring at those frequencies and not at frequencies in between (Macmillan and Schwartz, 1975; Scharf, 1998).

Further experiments examining spatial configurations in which a distracting, nontarget stimulus is located in between two targets could tease apart whether spatial attention is split or broadened. In experiments on visual spatial attention it is assumed that intervening distracters degrade performance in cases where the spotlight is broadened because they are obligatorily attended, but not in cases where the spotlight is effectively split (McMains and Somers, 2005). Importantly, however, adding intervening distractors in an auditory task has more complex implications than it does in vision. In particular, acoustic interactions between targets and the distractor will affect the peripheral representation of the target stimuli in addition to influencing higher processes such as the spatial distribution of attention. If experiments determine that the auditory spotlight can be split, further experiments will be required to distinguish between a sustained splitting of attention and a rapid switching of attention between locations. This distinction is impossible to make with long speech stimuli such as those used here, where accuracy is based on information accumulated over tens to hundreds of milliseconds. Experiments involving much briefer stimuli may be required to address this issue (Miller and Bonnel, 1994).

One final point regarding divided listening deserves consideration. It was noted in Experiment 2 that error patterns differed for the left (presumably "higher-priority") and right (presumably "lower-priority") talkers. It is possible that these differences expose different mechanisms involved in reporting the two sources. Errors involving the left talker were less frequent than those involving the right talker, but increased steadily with spatial separation. This pattern is highly consistent with subjects focusing a spatial spotlight of attention on the left source, and attempting to share it between two locations to simultaneously process the right talker. However errors involving the right talker peaked at the intermediate configuration and dropped again for the far configuration. It may be that processing of the lower-priority source depends more on temporary storage in working memory than on the focusing of spatial attention. The increase in errors involving the right source peaks at 90° separation, which is the configuration that (on average) required the addition of the highest level of noise in this experiment (see Table II). Thus it is possible that the addition of noise to the speech signals degraded the sensory trace storing the lower-priority right talker, and/or influenced recall of the lower-priority keywords. Future experiments will be aimed at determining the influence of cognitive factors in multiple-talker listening, in particular how working memory storage and recall are affected by the spatial arrangement, signal-to-noise ratio, and task-mediated prioritization of simultaneous sources.

IV. SUMMARY AND CONCLUSIONS

These experiments examined the effect of spatial separation of two competing speech sources on the ability of listeners to report keywords from them both. In particular, the "spotlight" model of spatial attention was examined by testing the hypothesis that attending to both talkers would be more efficient if they occupied the same spatial region.

The results of Experiment 1A, which systematically examined the effect of spatial separation, show that spatial separation modulates performance predominantly due to changes in relative energy levels at the two ears. Importantly, it seems that energy ratios at the ears can be independently utilized when listening to two simultaneous talkers. When energy variations were eliminated (Experiment 1B), performance was relatively stable across different spatial configurations, suggesting that spatial attention may not act like a spotlight. Experiment 2 was performed to examine whether this lack of modulation of performance with spatial separation was the result of two opposing effects: a poorer ability to attend to both sources working in opposition to an improved separability and/or intelligibility of the competing signals. To do this, the intelligibility of the two sources was equalized across configurations by adding broadband masking noise and the "cost" of divided listening was measured. Results suggest that small separations result in a smaller performance cost when two talkers must be processed simultaneously, and are largely consistent with the spotlight model of spatial attention.

ACKNOWLEDGMENTS

This work was funded by the Office of Naval Research Grant No. N00014-04-1-0131. Frederick Gallun was supported by F32 DC006526 from NIDCD. Prity Bengani assisted with data collection. Two anonymous reviewers gave important feedback that greatly strengthened the paper.

APPENDIX: TABLE A1

TABLE A1. Examples of the different error types for one example pair of sentences. Incorrect words are indicated in bold, and substitutions between the two sentences are indicated in italics.

Error type	Example
Correct	Ready Baron go to blue one now
	Ready Charlie go to red two now
Drop 1	Ready Baron go to **green** one now
	Ready Charlie go to red two now
Drop 2 (same)	Ready Baron go to **green three** now
	Ready Charlie go to red two now
Drop 2 (different)	Ready Baron go to **green** one now
	Ready Charlie go to red **four** now
Switch	Ready Baron go to blue *two* now
	Ready Charlie go to red *one* now
Drop 3	Ready Baron go to **green three** now
	Ready Charlie go to **white** two now
Drop 4	Ready Baron go to **green three** now
	Ready Charlie go to **white four** now

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.

Awh, E., and Pashler, H. (**2000**). "Evidence for split attentional foci," J. Exp. Psychol. Hum. Percept. Perform. **26**, 834–846.

Best, V., Ozmeral, E., Gallun, F. J., Sen, K., and Shinn-Cunningham, B. G. (**2005**). "Spatial unmasking of birdsong in human listeners: Energetic and informational factors," J. Acoust. Soc. Am. **118**, 3766–3773.

Bolia, R. S., Nelson, W. T., Ericson, M. A., and Simpson, B. D. (**2000**). "A speech corpus for multitalker communications research," J. Acoust. Soc. Am. **107**, 1065–1066.

Broadbent, D. E. (**1954**). "The role of auditory localization in attention and memory span," J. Exp. Psychol. **47**, 191–196.

Broadbent, D. E. (**1958**). *Perception and Communication* (Pergamon Press, London).

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

Brungart, D. S., and Rabinowitz, W. R. (**1999**). "Auditory localization of nearby sources. Head-related transfer functions," J. Acoust. Soc. Am. **106**, 1465–1479.

Brungart, D. S., Simpson, B. D., Darwin, C. J., Arbogast, T. L., and Kidd, G. Jr. (**2005**). "Across-ear interference from parametrically degraded synthetic speech signals in a dichotic cocktail-party listening task," J. Acoust. Soc. Am. **117**, 292–304.

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

Darwin, C. J., Howell, P., and Brady, S. A. (**1978**). "Laterality and localization: A "right ear advantage" for speech heard on the left," in: *Attention and Performance*, Vol. **VII**, edited by J. Renquin (Lawrence Erlbaum, Hillsdale, New Jersey), pp. 261–278.

Eriksen, C. W., and St. James, J. D. (**1986**). "Visual attention within and around the field of focal attention: A zoom lens model," Percept. Psychophys. **40**, 225–240.

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.

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

Gallun, F. J., Mason, C. R., and Kidd, Jr., G. (submitted). "Task-dependent costs in processing two simultaneous auditory stimuli," Percept. Psychophys.

Kidd, G., Mason, C. R., Brughera, A., and Hartmann, W. M. (**2005**). "The role of reverberation in release from masking due to spatial separation of sources for speech identification," Acust. Acta Acust. 114, 526–536.

Macmillan, N. A., and Schwartz, M. (**1975**). "A probe-signal investigation of uncertain-frequency detection," J. Acoust. Soc. Am. **58**, 1051–1058.

Massaro, D. W. (**1976**). "Perceptual processing in dichotic listening," J. Exp. Psychol. Hum. Percept. Perform. **2**, 331–339.

McMains, S. A., and Somers, D. C. (**2004**). "Multiple spotlights of attentional selection in human visual cortex," Neuron **42**, 677–686.

McMains, S. A., and Somers, D. C. (**2005**). "Processing efficiency of divided spatial attention mechanisms in human visual cortex," J. Neurosci. **25**, 9444–9448.

Miller, J., and Bonnel, A. M. (**1994**). "Switching or sharing in dual-task line-length discrimination?," Percept. Psychophys. **56**, 431–446.

Mondor, T. A., and Zatorre, R. J. (**1995**). "Shifting and focusing auditory spatial attention," J. Exp. Psychol. Hum. Percept. Perform. **21**, 387–409.

Morais, J. (**1978**). "Spatial constraints on attention to speech," in: *Attention and Performance*, Vol. **VII**, edited by J. Renquin (Lawrence Erlbaum, Hillsdale, New Jersey), pp. 245–260.

Müller, M. M., Malinowski, P., Gruber, T., and Hillyard, S. A. (**2003**). "Sustained division of the attentional spotlight," Nature (London) **424**, 309–312.

Quinlan, P. T., and Bailey, P. J. (**1995**). "An examination of attentional control in the auditory modality: Further evidence for auditory orienting," Percept. Psychophys. **57**, 614–628.

Scharf, B. (**1998**). "Auditory Attention: The Psychoacoustical Approach," in: *Attention*, edited by H. Pashler (Psychology Press, London), pp. 75–117.

Shinn-Cunningham, B., and Ihlefeld, A. (**2004**). "Selective and divided attention: Extracting information from simultaneous sound sources," *Proc. Int. Conf. Auditory Display*, International Community for Auditory Display, Sydney, Austrailia.

Shinn-Cunningham, B. G., Ihlefeld, A., Satyavarta, and Larson, E. (**2005**). "Bottom-up and top-down influences on spatial unmasking," Acta. Acust. Acust. **91**, 967–979.

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.

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.