Cross-modal selective attention: On the difficulty of ignoring sounds at the locus of visual attention

CHARLES SPENCE and JANE RANSON
University of Oxford, Oxford, England

and

JON DRIVER
University College London, London, England

In three experiments, we investigated whether the ease with which distracting sounds can be ignored depends on their distance from fixation and from attended visual events. In the first experiment, participants shadowed an auditory stream of words presented behind their heads, while simultaneously fixating visual lip-read information consistent with the relevant auditory stream, or meaningless "chewing" lip movements. An irrelevant auditory stream of words, which participants had to ignore, was presented either from the same side as the fixated visual stream or from the opposite side. Selective shadowing was less accurate in the former condition, implying that distracting sounds are harder to ignore when fixated. Furthermore, the impairment when fixating toward distractor sounds was greater when speaking lips were fixated than when chewing lips were fixated, suggesting that people find it particularly difficult to ignore sounds at locations that are actively attended for visual lipreading rather than merely passively fixated. Experiments 2 and 3 tested whether these results are specific to cross-modal links in speech perception by replacing the visual lip movements with a rapidly changing stream of meaningless visual shapes. The auditory task was again shadowing, but the active visual task was now monitoring for a specific visual shape at one location. A decrement in shadowing was again observed when participants passively fixated toward the irrelevant auditory stream. This decrement was larger when participants performed a difficult active visual task there versus fixating, but not for a less demanding visual task versus fixation. The implications for cross-modal links in spatial attention are discussed.

The majority of selective-attention research has focused on selection within just one sensory modality. In many everyday situations, however, we need to coordinate our selective attention across different sensory modalities. For example, when listening to a conversation at a noisy gathering (the archetypal "cocktail party" situation), one must not only attend selectively to the sound of the speaker's voice but also integrate this auditory information with related visual information regarding the speaker's lip movements and gestures, while ignoring irrelevant events in both modalities.

A number of audiovisual cross-modal links in spatial attention have now been demonstrated using variants of the spatial cuing methodology introduced by Posner (1978). For instance, Spence and Driver (1996) found that when people expected a visual target from one location, their per-

formance for occasional auditory targets was better at that location than on the other side, even though auditory probes were somewhat more likely on the other side. This suggests a tendency for voluntary (or *endogenous*) spatial attention to shift to corresponding locations across hearing and vision. Cross-modal links have also been demonstrated for more reflexive (or *exogenous*) aspects of attention within the spatial cuing paradigm (e.g., Spence & Driver, 1997a; Spence, Nicholls, Gillespie, & Driver, 1998; Ward, 1994).

However, these results have come from relatively sparse situations, with just a single isolated target on each trial, which is far removed from the classic selective-shadowing studies (e.g., Broadbent, 1958; Cherry, 1953; Treisman, 1969; Wood & Cowan, 1995), which instigated the study of our ability to ignore continuous distraction under multiple stimulation, as exemplified by the cocktail party. Here, we examined cross-modal issues for such situations. One robust finding to emerge from the classic shadowing research has been that people find it easier to select one stream and ignore another when the spatial separation between the relevant and irrelevant auditory streams increases (e.g., Broadbent, 1958; Cherry, 1953; Nelson, Bolia, Ericson, & McKinley, 1998; Speith, Curtis, & Webster, 1954; see also Darwin & Hukin, 1999). Similar findings have also been reported within vision and touch

(see Driver & Grossenbacher, 1996; Eriksen & Eriksen, 1974; Johnston & Dark, 1986), so the spatial separation principle seems quite general. In the present study, we investigated the possible implications of this spatial principle for cross-modal attention. In particular, we considered whether it becomes easier to ignore distracting sounds when they are placed farther away from the focus of visual attention. Thus, in this study, we examined the spatial separation of cross-modal sources rather than separation within just one modality, as in previous studies.

It has long been known that it can be difficult to ignore stimuli presented in one modality while attending to concurrent stimuli in another modality, even when the distracting modality is known to be utterly irrelevant to the prescribed task (e.g., Broadbent, 1956; Klein, 1977, Experiment 3; Simon & Craft, 1970). Recent demonstrations of this include cross-modal Stroop effects and negative priming effects from a stream of auditory distractors on responses to a concurrent stream of visual targets (e.g., Cowan, 1989; Cowan & Barron, 1987; Driver & Baylis, 1993; Elliott, Cowan, & Valle-Inclan, 1998; Greenwald, 1970, 1972; Morton, 1969; Shimada, 1990; though see Miles, Madden, & Jones, 1989; Thackray & Jones, 1971). A further example comes from the *unattended speech effect* (USE) in serial recall (e.g., Colle & Welsh, 1976; Jones, 1993, 1999; LeCompte, Neely, & Wilson, in press; Salamè & Baddeley, 1982), in which the immediate serial recall of a short series of nameable visual stimuli (e.g., digits or words) is significantly impaired by a concurrent stream of spoken auditory distractors, thus showing that auditory distractors can impair performance for concurrent visual targets. However, few studies have made any serious attempt to test whether such interference from auditory distractors is affected by their spatial proximity to the visual targets, despite the fundamental importance of this issue for models of cross-modal attention.

Jones and Hapeshi (1991, Experiment 4; cited in Jones, 1993) suggested that cross-modal Stroop effects may be modulated by the spatial separation between attended and irrelevant streams. They presented a visual stream of successive colored letter strings as targets. The distracting auditory stream comprised spoken words, presented either from a loudspeaker directly above the visual targets in front of participants or from a loudspeaker behind their heads. Reaction time (RT) to name ink colors was delayed when the auditory distractor was an incongruent color word. This cross-modal Stroop effect was reduced when the auditory distractors were behind the participant rather than in front near the visual stimuli. These results apparently suggest a spatial constraint on cross-modal attention, with greater interference from auditory distractors when at the same location as visual targets. However, an alternative explanation is that the sounds were simply more salient when in front, regardless of their position relative to the visual targets. People usually find it easier to hear speech from in front than from behind (the well-known frontal speech advantage; e.g., Hublet, Morais, & Bertelson, 1976, 1977; Morais & Bertelson, 1973;

Morais, Cary, Vanhaelen, & Bertelson, 1980). Moreover, any such frontal speech advantage may have been accentuated in Jones and Hapeshi's study, because the loudspeaker in front of participants was much closer to them than the one at the rear (0.6 m vs. 2.5 m), and so the auditory distractors were presumably louder when presented from the front.

Only a few studies of the unattended speech effect found in serial recall have considered the location of auditory distractors relative to visual targets. Colle (1980) presented irrelevant sounds over headphones to the right ear, the left ear, dichotically, or binaurally; he found that all produced equivalent interference on serial recall for visual targets presented directly in front of the participant. Jones, Miles, and Page (1990) reported that changes in the location of irrelevant speech (either moving between left and right ear or at a fixed position near the midline) had little effect on proofreading performance. This led them to claim that "spatial location does not appear to be an important feature of the distracting effect of irrelevant speech" (Jones et al., 1990, p. 99). However, since sounds were always presented over headphones in both of these studies, one might argue that the effective changes in auditory location were fairly small (since the sounds always came from a source very close to the head) and, hence, might not be expected to affect the observed cross-modal interference.

Recently we (Spence & Driver, 1999) investigated this further by conducting a standard study of the unattended speech effect while varying the relative location of the visual targets and auditory distractors substantially. Participants saw a list of eight successive visual digits at a location on their extreme right or left. Concurrent random auditory digits could be played from the same or opposite side. Serial recall was then made by typing the sequence of seen digits on a keyboard. The irrelevant auditory digits disrupted subsequent recall performance (as compared with a silent control condition), but did so equivalently regardless of whether they were presented at the same location as the visual targets or 60° away on the other side. Thus, spatial location does not seem to affect the well-studied effect of irrelevant speech on serial recall for nameable visual items. However, the retrospective nature of serial recall means that it may be a relatively insensitive measure for the ease of on-line distractor rejection.

The cross-modal Stroop effects (Cowan, 1989; Cowan & Barron, 1987; Driver & Baylis, 1993; Elliott et al., 1998; Greenwald, 1972; Morton, 1969; Shimada, 1990) and negative priming effects (Driver & Baylis, 1993) mentioned earlier might seem to offer better paradigms for studying any effects of cross-modal spatial separation on the ability to ignore distractors. However, these paradigms may also be inappropriate for isolating difficulties with perceptual selection in particular. They might instead primarily reflect the ease of response selection, when incongruent response tendencies are activated (in Stroop situations) or when the response required to the

current target has just been suppressed for a preceding target (in negative priming). Accordingly, in the present study, we did not utilize any of the established cross-modal interference effects. Instead, we chose to examine how selective shadowing would be affected when auditory distractors were close to or far from the focus of visual attention, reasoning that this should provide a more direct measure for the efficiency of perceptual selection than any of the other available paradigms.

Our tasks were adapted from a previous series of shadowing experiments reported by Driver and Spence (1994), in which participants shadowed an auditory stream presented from one loudspeaker at 30° to one side, while ignoring an irrelevant auditory stream presented from a loudspeaker on the other side. In this study, there were also two visual streams: one at the left loudspeaker, and the other at the right loudspeaker. In some blocks of trials, these two visual streams each consisted of lip-read information (videos of a person recording either of the auditory messages). These visual streams were sometimes placed on the same side as the corresponding auditory messages and at other times on the opposite side (e.g., the lip-read information for the relevant message might be on the right, while the sounds to be shadowed were on the left). Participants fixated the visual stream that matched the message they had to shadow.

More relevant words were correctly reported when the attended visual stream was on the same side as the relevant auditory stream, suggesting that participants found it easier to direct auditory and visual attention to the same position rather than in opposite directions (see also Spence & Driver, 1996). Merely fixating passively in one direction or another (in control conditions in which the visual stream now comprised meaningless "chewing" lip movements, rather than visual lip-read information) had less effect on shadowing performance than actively attending in either direction to process lip-read information.

Although this study suggests cross-modal links in spatial attention during selective shadowing, it does not resolve our specific question of whether it is particularly difficult to ignore irrelevant information in one modality (e.g., distracting sounds) at a location that is selected for another modality (i.e., for the actively attended visual stream). Here, we adapted Driver and Spence's (1994) paradigm, in order to investigate whether increasing just the spatial separation between an *irrelevant* auditory stream and a relevant visual stream would make it easier to ignore the auditory distractors. Participants had to shadow a relevant auditory stream of words presented from a loudspeaker directly behind their heads, while ignoring a concurrent irrelevant auditory stream of words presented from a loudspeaker in front of them, on one side or the other (see Figure 1 for a schematic of the experimental set-up, seen from above the participant's cartooned head and upper body). At the same time, participants also had to fixate a visual display, which could be placed either at the same location as the irrelevant auditory
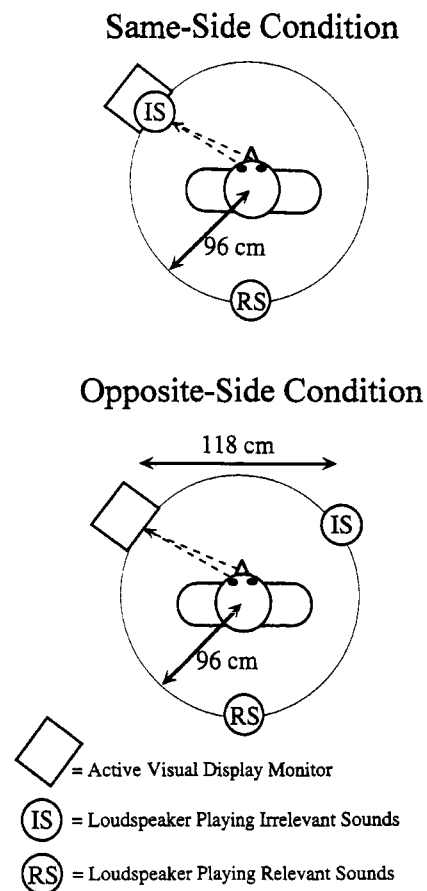

Figure 1. Schematic bird's-eye view of the experimental set-up, showing the position of the loudspeakers that could be used to present the relevant and irrelevant auditory streams and the monitors used to present the visual stream, around the participant's cartooned head and body (possible direction of fixation indicated by the dotted lines). The virtual circle on which the loudspeaker cones were placed is also shown. In the same-side condition, the visual stream and irrelevant sounds were either both on the left (as shown) or both on the right (equally likely). In the opposite-side condition, for half of the blocks, the visual stream was presented on the left and the irrelevant sounds on the right (as shown); for the remaining half of the blocks, the reverse arrangement held. Note that, in the experiment, there was always one monitor and one loudspeaker visible in front of the participant on each side; we simply manipulated which pieces of the fixed apparatus presented stimuli at any one time.

stream (same-side condition) or in the mirror-symmetric position on the other side (opposite-side condition).

Our question was whether the irrelevant sounds would be harder to ignore, thus leading to poorer selective shadowing, when presented at the location that was visually attended rather than on the opposite side (see Figure 1). Following Driver and Spence (1994), we also sought to disentangle any effects caused by actively processing visual events at one location from those due merely to fixating passively there. Accordingly, just as in Driver and Spence's study, the visual streams either consisted of mean-

ingful lip movements that matched the relevant auditory message (the speaking-lips condition, intended to induce active visual attention) or consisted of meaningless lip movements (chewing lips) that merely had to be passively fixated.

Note that the spatial arrangement of the present study held the separation between relevant and irrelevant sounds constant and also held the separation between relevant visual events and the relevant sounds constant, unlike Driver and Spence (1994). All that varied was whether the fixated visual stream was close to or far from the irrelevant auditory stream. This design was therefore expected to allow us to determine for the first time whether it is particularly hard to ignore auditory distractors when presented at the current spatial focus of visual attention. Moreover, if any such effect is stronger with active visual processing than with mere passive fixation, then it should be more pronounced in the speaking-lips conditions than in the chewing-lips conditions.

As regards the latter passive fixation conditions, several previous studies have suggested that the mere direction of fixation may sometimes affect selective listening. Gopher (1973), who monitored gaze while listeners shadowed one ear during dichotic presentation over headphones, found a tendency for the listeners to look toward the shadowed ear. Further studies reported better selective shadowing when listeners looked toward an external loudspeaker presenting relevant sounds rather than the loudspeaker presenting irrelevant sounds (Larmande, Elghozi, Sintes, Bigot, & Autret, 1983; Morais et al., 1980; also see Hiscock, Hampson, Wong, & Kinsbourne, 1985). This comparison cannot distinguish a possible cost when fixating distractor sounds from a possible benefit when fixating target sounds. Reisberg, Scheiber, and Potemken (1981), who manipulated whether listeners looked toward a relevant loudspeaker, a distracting loudspeaker, or directly in between, found only a small cost from fixating distractor sounds. However, their result has proved controversial and difficult to replicate (see Wolters & Schiano, 1989), perhaps because relatively small spatial separations were used. Note that the present study used a wider separation between the sounds and thus should be more sensitive to any effect from the direction of passive fixation. In addition, our study tested for the first time whether active visual processing might produce larger effects than mere fixation.

EXPERIMENT 1

Method

Participants. Sixteen participants (7 men and 9 women) were recruited by advertisement for this experiment and were each paid £3.50 (approximately U.S. $6) for a 50-min session. Their mean age was 26 years (range, 17–45 years). All participants but 1 were right-handed by self-report. The participants were naive as to the purpose of the experiment, and all were native English speakers.

Apparatus and Materials. The participant was seated at a table in a darkened room, facing straight ahead, with head movements precluded by an adjustable temple clamp and chinrest. Two video monitors (Panasonic TC-15M1R) were placed 96 cm from the participant, arranged symmetrically at 38° from the participant's midline on either side at eye level. Only one displayed a visual stream at any one time, and the participant always fixated toward this visual stream. Two videorecorders (Panasonic AG-7650-B and AG-7750-B) and a video-editing system (Panasonic AG-A770) were used to control the presentation of auditory and visual streams. Three loudspeakers (each 6-in. diameter, 15 ohm) were arranged on a virtual circle (radius 96 cm) centered on the participant's head, and each could be used to present an auditory stream of words (see Figure 1). The relevant auditory stream was always presented via the loudspeaker situated directly behind the participant's head.[1] The concurrent irrelevant auditory message was played from either of two loudspeakers situated in front of the participant on either side, in the same location as one or the other of the two video monitors. The visual stream of moving lips subtended about 1° of visual angle horizontally (the whole head subtended an angle of approximately 3° horizontally) and was presented 2° above the center of a loudspeaker. The apparatus was not moved between conditions; instead, switches were used to send the appropriate stimuli to a particular loudspeaker or monitor. Relevant and irrelevant auditory streams were each presented at approximately 68 dB (A) on average, peaking at 71 dB (A), as measured from the participant's ear position. In order to bring shadowing performance below ceiling and thus encourage the participant to engage in active lipreading when speaking lips were visible, white noise was presented continuously at 52 dB (A) from a loudspeaker strapped to the underside of the participant's chair throughout the experiment.

The materials comprised two videorecordings of the first author reading random triplets of two-syllable words, all with stress on the initial syllable, and with the two different recordings played back to the participant concurrently (e.g., the two simultaneous triplets might be "olive, notice, topic" and "jewel, reason, under"; see Driver & Spence, 1994, for more details). Concurrent words from the two different tapes were approximately matched for spoken length and for spoken frequency (Brown, 1984). During each recording, the author aimed to pronounce the words at a fairly level intensity and at a regular rate (by means of a series of computerized pacing clicks). These clicks were audible on the recordings and thus provided information to warn the participant of the imminent onset of a triplet in the experiment. Word triplets were read at a rate of 1.6 sec per triplet, with a 4-sec gap between triplets.

There were 16 experimental test blocks, each consisting of 13 triplets of words. These were created in the following manner: 8 blocks of 13 word triplets were recorded onto one videotape (Tape 1), and a further 8 blocks of 13 triplets of different words (approximately matched for word length and word frequency) were recorded onto a second videotape (Tape 2). The original 8 blocks recorded on Tape 2 were then copied onto to the end of Tape 1, and, likewise, the original words recorded onto Tape 1 were copied onto the end of Tape 2, to yield two 16-block tapes, with quality (original vs. copy) counterbalanced across tapes. The visual track of Tape 2 was then overdubbed with meaningless lip movements (the person who had recorded the messages now vigorously chewed gum with a closed mouth, instead of speaking, to produce visual chewing-lips stimuli). Only one of the two possible visual channels (speaking lips from Tape 1 or chewing lips from Tape 2) was shown at any time in the experiment, from just one monitor. However, the two auditory channels (from Tape 1 and Tape 2) were presented concurrently and always comprised different words at a given time. The two videotapes were synchronized every 4 blocks, using a framecounter on the videoeditor to set them to framecounts previously judged by the experimenters to provide maximum phenomenal simultaneity for reference click trains on the two tapes. For each participant, particular words appeared only once in the speaking-lips condition and once in the chewing-lips condition. The spatial condition in which particular words appeared (same side vs. opposite side) was perfectly

counterbalanced across participants. The experiment began with 8 blocks of practice trials (which were not analyzed), each comprising 8 triplets of words different from those used in the experimental blocks.

Design. The speaking-lips and chewing-lips conditions alternated between successive blocks. Which of the two video monitors displayed the visual stream (and hence whether the participant fixated leftward or rightward) was switched every 4 blocks. Hence, there were eight equiprobable conditions produced by crossing the three within-participants factors of lips (speaking vs. chewing), relative position of the visual stream with respect to the irrelevant auditory stream (same side vs. opposite side), and hemispace of irrelevant sounds (i.e., whether the irrelevant auditory stream was on the left or right).

The side of the visual stream stayed constant within every block of 13 successive triplets, whereas the side of the irrelevant auditory stream changed after the sixth triplet (i.e., within each block of trials). Switching the position of auditory distractors in this way ensured that the relative position factor changed within each block (and likewise for hemispace of irrelevant sounds), so that any variability due to long-term drifts in performance should have minimal impact on these factors. The switch also allowed us to examine any influence on performance when the distracting sound suddenly changed position (i.e., on the 7th triplet in each block). This might attract attention toward the distracting stream, in the automatic "exogenous" manner revealed by spatial cuing studies of attention with spatially uninformative peripheral cues (e.g., Spence & Driver, 1994, 1997a; Spence et al., 1998).

Of course, the main aim of our study was to investigate the cross-modal implications of directing deliberate (or "endogenous") attention toward a location, to test whether or not endogenously attending to a visual location makes auditory distractors at that location more potent. However, we were also interested in how any such endogenous effects might interact with exogenous attention capture by a sudden change in the location of distracting sounds. The spatial cuing literature has already highlighted a number of qualitative differences between exogenous (or reflexive) and endogenous (or deliberate) mechanisms of spatial attention. However, rather few studies have looked at the possible interplay between exogenous and endogenous factors (although see Theeuwes, 1994), and, to our knowledge, absolutely none have ever done so for a cross-modal situation. We expected performance to suffer on the 7th triplet in each block, because the sudden change in the location of the distracting auditory stream should attract attention toward it in an exogenous manner (cf. Spence & Driver, 1994, 1997a; Näätänen, 1992; Näätänen, Paavilainen, Tiitinen, Jiang, & Alho, 1993). The more intriguing question was whether the switch in distractor position

would modulate any endogenous effects due to where the participant was deliberately attending in vision; this might reveal how exogenous and endogenous factors interact in directing selective attention in a cross-modal situation.

Procedure. The participants were instructed to maintain fixation on the single monitor that displayed a visual stream, either on their far left or far right in each block, and adherence to this fixation instruction was monitored throughout by the experimenter. The participants had to repeat as many words as possible from each triplet in the relevant message (i.e., the sounds from behind them, with or without supporting lip movements on one side in front). They responded in the interval between successive triplets. The experimenter recorded the participants' responses, which were scored as correct, incorrect, or intrusions (i.e., reports of words from the irrelevant auditory stream) without regard to the order of report for words within each triplet. Close approximations (e.g., derivational errors) were scored as incorrect.

Results

We excluded the 7th triplet in each block from our initial scoring, since it may have been unusual due to the distractor sounds suddenly changing location; this might have attracted attention exogenously, as explained above. We will return to the analysis for any such exogenous effect later. The interparticipant mean percentages of correct responses (and of intrusion errors) after this exclusion are shown for each condition in Table 1. It can be seen that performance was better overall when speaking lips gave useful lip-read information to boost the relevant auditory message. More important, in support of our critical prediction, performance was worse in the same-side condition, in which the visual stream was at the same location as the irrelevant auditory stream that had to be ignored, as compared with the opposite-side condition. Moreover, this effect of relative position appeared larger for the speaking-lips condition than for the chewing-lips condition.

The percent correct data were analyzed using a within-participants analysis of variance (ANOVA), with three factors [lips (2) × relative position (2) × hemispace of irrelevant sound (2)]. This revealed that the participants did indeed report significantly more relevant words when fixating relevant lip-read information ($M = 68.9\%$) than

Table 1
Mean Percentages of Correct Responses and Intrusion Errors, With Their Standard Errors, as a Function of the Lips Condition and the Relative Position of the Irrelevant Auditory Stream and the Fixated Visual Stream (Same vs. Opposite Side) for Irrelevant Auditory Streams Presented to the Left and the Right and Averaged Over Side in Experiment 1

| Irrelevant Auditory Stream | Performance Measure | Speaking Lips | | | | Chewing Lips | | | |
| | | Same Side | | Opposite Side | | Same Side | | Opposite Side | |
		M	SE	M	SE	M	SE	M	SE
Left	% Correct	58.4	4.6	79.1	3.2	45.6	3.0	60.4	3.5
	% Intrusion	5.5	1.4	1.1	0.4	3.6	0.6	2.7	1.0
Right	% Correct	59.6	4.1	78.6	3.2	53.1	3.1	60.6	3.4
	% Intrusion	3.3	1.0	1.1	0.4	3.6	0.7	4.1	1.6
Average	% Correct	59.0	2.6	78.9	2.2	49.3	2.0	60.5	2.3
	% Intrusion	4.4	1.1	1.1	0.3	3.6	0.5	3.4	1.1

when fixating meaningless chewing lips (M = 54.9%), resulting in a main effect of lips [$F(1,15)$ = 86.3, p < .0001]. This implies that the participants used lip-read information from the visual stream (when available) to facilitate their perception of the relevant auditory stream. The participants reported fewer words correctly in the same-side trials (M = 54.2%) than in the opposite-side trials (M = 69.7%) overall, resulting in a main effect of relative position [$F(1,15)$ = 69.7, p < .0001]. Thus, performance was worse when the participants fixated a visual stream at the same location as the irrelevant auditory stream.

The interaction between relative position and lips was also significant [$F(1,15)$ = 13.5, p = .002]. The same-side condition resulted in worse performance than the opposite-side condition (at p < .01, by t test) for both speaking lips and chewing lips. The interaction arose because relative position had a larger effect with speaking lips than with chewing lips (mean opposite-side minus same-side differences of 14.3% and 8.1%, respectively).[2]

None of the other effects or interactions in the analysis of the percent correct data were significant [for hemispace, $F(1,15)$ = 1.3, p = .27; for lips × hemispace, $F(1,15)$ = 1.5, p = .24; for hemispace × relative position, $F(1,15)$ = 1.5, p = .24; for lips × hemispace × relative position, $F(1,15)$ = 2.1, p = .17]. An analogous ANOVA on the intrusion data (i.e., the number of responses where the participants mistakenly reported a word from the irrelevant auditory stream rather than the relevant auditory stream) revealed a main effect of relative position [$F(1,15)$ = 6.0, p = .03], with more intrusion errors in the same-side condition (M = 4.0%) than in the opposite-side condition (M = 2.3%). The interaction between relative position and lips also approached significance [$F(1,15)$ = 4.5, p = .05]. This was again due to a larger influence of relative position with speaking lips (mean same-side − opposite-side difference of 3.3%; p < .01, by t test) than with chewing lips (mean difference of 0.2%; n.s.), as in the percent correct data.[3] None of the other effects or interactions were significant in the intrusion data [for lips, $F(1,15)$ = 1.4, p = .26; for hemispace × lips, $F(1,15)$ = 1.4, p = .26; for hemispace × relative position, $F(1,15)$ = 2.9, p = .11; for all other effects and interactions, F < 1].

We turn now to examine any effects of the sudden switch in location of the auditory distractors on the 7th triplet in each block. In order to examine whether this switch exogenously drew attention to the distracting auditory stream, we conducted a further analysis that considered the serial position of each triplet of words within a block. This included all 13 triplets (previously the 7th "switch" triplet had been excluded) but now treated each triplet position separately. The percentages of correct reports of relevant words for each condition are shown in Figure 2, as a function of the position of the triplet within the experimental block (note that same-side condition vs. opposite-side condition switched within each block, along
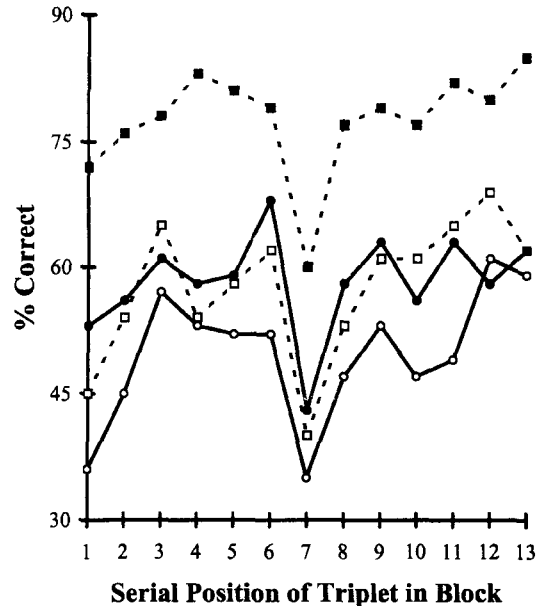

Figure 2. The mean percentages of correct responses, as a function of the serial position of the triplet in the block. Square symbols connected by dotted lines represent the speaking-lips condition, and circles connected by solid lines represent the chewing-lips condition. Open symbols represent same-side trials, and closed symbols represent opposite-side trials. Note that the data for the first 6 triplets in each condition and for the subsequent 7 triplets actually came from different blocks, since the condition always changed (from same side to opposite side, or vice versa) on the critical 7th "switch" triplet.

with the location of the 7th triplet; the graph pools across blocks). Two aspects of the graphed data are immediately striking. First, there is a drop in performance on the 7th switch triplet for all four conditions. Second, the differences between conditions that were described above seem apparent for every single one of the 13 successive triplets, even when considered individually.

This pattern was confirmed by a three-way within-participants ANOVA (lips × relative position × triplet number) on the total correct data, which showed the same main effects and lips × relative position interaction as reported in the preceding analyses, plus with a significant effect of triplet number [$F(12,180)$ = 14.5, p < .0001]. This was due to the poorer performance on the critical 7th triplet, where the position of the irrelevant auditory stream switched (all pairwise comparisons of this triplet with any of the others were significant at p < .01, by Newman–Keuls tests). The critical effect of relative position and its interaction with speaking lips versus chewing lips as described above apparently held across all the serial positions of triplets within each block, even on the 7th switch triplet (see Figure 2). An ANOVA on the accuracy data for just this 7th triplet found a main effect of lips [$F(1,15)$ = 9.6, p < .01], with better performance for speaking lips, and a main effect of relative position [$F(1,15)$ = 21.4, p < .001], with better performance in

the opposite-side condition in which the distractor sounds were now far from the attended visual stream. Finally, the interaction term was marginally significant [$F(1,15) = 3.5, p = .08$], due to the usual pattern of a larger effect from relative position with speaking lips than from chewing lips. Thus, our basic effects are still apparent even when one considers just $1/13$ of the data. The robustness of this pattern across the 13 positions suggests that effects of endogenous attention (i.e., those caused by where the participant deliberately attended in vision) were additive with the exogenous effect caused by the sudden change in the position of irrelevant sounds on the 7th triplet.

Discussion

Several important findings emerge from Experiment 1. First, selective shadowing was consistently better when relevant lip-read information was added, as observed previously (e.g., Driver & Spence, 1994; Reisberg, 1978, 1987; Sumby & Pollack, 1954; see also Grant & Seitz, 1998; Summerfield, 1992). This is perhaps unsurprising, since appropriate lip-read information adds phonological information to constrain identification of the relevant words (cf. Dodd & Campbell, 1987; McGurk & McDonald, 1976). Second, and more importantly, reports of the relevant words were more accurate when the distracting sounds were placed farther away in space from the fixated visual stream (while holding all other spatial factors constant). There were also fewer intrusion errors in this situation. These influences of relative position confirm that it is harder to select relevant sounds and ignore distracting sounds when the latter are placed at a fixated visual position. Third, relative position had a stronger effect when the visual stream was actively attended to process useful lip-read information, rather than merely being passively fixated as with the meaningless chewing lip movements.

These results imply that people find it difficult to ignore speech sounds that they are passively looking at and more difficult still to ignore speech sounds presented near lip movements that they are visually attending in order to extract lip-read information. The effect from the direction of passive fixation (with chewing lips) supports some previous suggestions that shadowing performance can be impaired by looking in the direction of the distracting sounds (e.g., Gopher, 1973; Morais & Bertelson, 1973; Morais et al., 1980; Reisberg et al., 1981). This outcome was so clear and consistent in our experiment that we can only summize that occasional failures to observe this pattern in the past (e.g., Wolters & Schiano, 1989) must have been caused by some insensitivity in the measures that were used, due perhaps to relatively small changes in the direction of fixation (by contrast, in the present experiment, there was a substantial 76° difference in the direction of gaze between conditions).[4]

The most novel finding in our experiment was that the spatial separation of auditory distractors from the visual stream had a significantly larger impact when the visual stream was actively attended for lipreading (speaking lips) than when it was merely passively fixated (chewing lips). Proof that the participants were indeed actively processing the visual stream in the speaking-lips conditions comes from the fact that overall performance was significantly better than in the chewing-lips condition, due to the extraction of useful lip-read information. The greater impact of spatial separation from the auditory distractors with an active rather than passive visual task suggests that people find it particularly difficult to ignore an irrelevant auditory stream when they must deliberately attend that same location in vision rather than merely fixating it.

The analysis of performance as a function of the serial position of a triplet in each block revealed that the participants found it particularly difficult to ignore the irrelevant stream when it suddenly changed position (on the 7th triplet of every block). The impaired performance on this switch triplet is presumably caused by exogenous orienting of attention toward sudden sounds at a novel location (Spence & Driver, 1994; see also Näätänen, 1992, and Näätänen et al., 1993, for evidence of attention capture by unexpected sounds in event-related potential studies using scalp recordings of neural activity, and Wood & Cowan, 1995, p. 253, for related findings). Recent behavioral studies have shown that the sudden presentation of a peripheral sound invariably leads to a short-lasting shift of exogenous attention in its direction (e.g., Spence & Driver, 1994, 1997a; Spence et al., 1998). In the present experiments, the participants' auditory attention should therefore have been drawn in this automatic manner away from the endogenously attended location behind their heads toward the changed position of the distractor sound for a brief time. Such a short-lasting shift of auditory attention toward the new source of the irrelevant message would be expected to briefly disrupt shadowing performance, consistent with the abrupt drop in accuracy we found on the switch triplet.[5]

Interestingly, the disruptive effect of suddenly shifting the distractor sounds was additive with the effect of spatial separation between auditory distractors and visual stream and likewise with the interaction between this factor and active versus passive visual task (see Figure 2). This additivity implies that exogenous attention capture by the sudden shift in distractor location did not change the influence of more endogenous factors, such as where the participant was deliberately attending or fixating in vision. From the perspective of Sternberg's (1969) additive-factors logic, this suggests that endogenous mechanisms of attention involve processes separate from exogenous mechanisms, adding further support to many recent claims that these two forms of spatial attention are qualitatively distinct and may be served by separate neural substrates (e.g., Briand, 1998; Butter, 1987; Klein, 1994; Klein, Kingstone, & Pontefract, 1992; Riggio & Kirsner, 1997; Robinson & Kertzman, 1995; Spence & Driver, 1994, 1996, 1997a).

Are These Cross-Modal Spatial Interactions Specific to Cross-Modal Speech Perception?

Thus far, we have interpreted our findings in terms of a general difficulty in ignoring a distracting auditory stream when presented at the current focus of visual attention. However, a somewhat different account could be put forward in terms of audiovisual integration during speech perception. The decline in performance when the irrelevant auditory stream is placed at the same location as the relevant visual lip-read information (as in the same-side speaking-lips condition) might be caused by some tendency to integrate the relevant visual lip-read information with the mismatching irrelevant auditory stream in this particular situation. This could arise if any tendency for inappropriate audiovisual integration were spatially constrained, being more likely to arise for lip movements and mismatching sounds that appear at the same location rather than at different positions.

If any such spatial constraint on inappropriate audiovisual integration were caused by cross-modal links in spatial attention, this would, of course, reduce to our own account. However, one might argue that inappropriate integration of speech sounds with mismatching lip movements could be spatially constrained for entirely nonattentional reasons (e.g., because lip movements and speech sounds emanating from the same external location usually do share a common source and, thus, typically should be integrated). In fact, recent evidence from studies of audiovisual integration in speech perception suggest that there is no such spatial constraint on the integration of mismatching lip movements and speech sounds, thus arguing against any nonattentional account for our results. This evidence comes from studies of the well-known "McGurk effect," which arises when visual lipread information for one speech sound (e.g., "aga") is paired concurrently with a somewhat different speech sound (e.g., "aba"). The classic finding is that the nature of the lip movement changes the perceived speech sound (e.g., a blend such as "ada" may be heard; see McGurk & McDonald, 1976). Jones and Munhall (1997) and Bertelson, Vroomen, and de Gelder (1995; see also Bertelson, 1998) have recently reported that this form of inappropriate integration is not constrained spatially. That is, McGurk effects are just as pronounced when the mismatching auditory and visual stimuli come from widely separated sources as when they share the same spatial location.

Given that inappropriate integration of mismatching auditory and visual streams appears to be unaffected by spatial separation, we thought it unlikely that any such integration between the lip movements and the mismatching irrelevant sounds could explain the greater effect of relative position for the speaking-lips condition than for the chewing-lips conditions in Experiment 1. However, the most straightforward way to assess this is to test whether we can obtain similar results when the active visual task is entirely unrelated to the shadowing task performed on the heard speech. If we could replicate the same pattern

with a nonlinguistic visual task, then clearly the result cannot be specific to cross-modal integration for speech perception.

EXPERIMENT 2

In Experiment 2, the lip movements used in Experiment 1 were replaced by a continuous stream of meaningless visual shapes (generated by a grid of LEDs), presented from the same two possible spatial locations as before. Thus, all of the visual events were now entirely unrelated to the auditory speech streams presented for the shadowing task. Participants either were required to actively monitor the visual stream of shapes for a specific target stimulus (active condition) or were simply required to passively monitor the alternation of two nontarget shapes in a similar stream (passive condition; see Driver & Spence, 1994, Experiments 2 and 3, for similar task requirements). As before, participants had to shadow the auditory stream of words presented from behind their heads, while simultaneously ignoring the irrelevant spoken words presented auditorily from in front of them on one side. In addition, they either fixated or actively monitored a visual stream on the same side of space as the irrelevant auditory words or the opposite side of space to the irrelevant auditory words.

If the pattern of results from Experiment 1 is specific to the particular linguistic task of lipreading while listening to speech, then it should not be replicated in Experiment 2. If instead the pattern reflects a more general rule that irrelevant sounds are hard to ignore when appearing at fixated locations and even harder to ignore when that fixated location is actively attended, then the same pattern should again be found (provided, of course, that the new active visual task is at least as demanding of visual attention as the lipreading had been in Experiment 1).

Method

Participants. Ten undergraduate students (5 men and 5 women) at Oxford University were recruited to take part in this experiment. Their mean age was 20 years (range, 19–20 years). All participants were right-handed by self-report. The participants were naive as to the purpose of the experiment, and all were native English speakers.

Apparatus and Materials. The participant was seated on a chair in a darkened room. The two auditory messages was transferred from the videotapes used in Experiment 1 onto the left and right channels of an audiotape, which was played back on a Yamaha K-340 stereo cassette deck. The nonlinguistic visual displays were generated by two grids of nine yellow LEDs, with each grid arranged in a 3 × 3 square configuration at a distance of 96 cm from the participant on either side of fixation (i.e., at the same position as the lips had been presented in Experiment 1). These two grids of yellow LEDs were located at ear level, and each grid subtended a visual angle of 3.9° vertically and horizontally. The luminance of each LED was 41.4 cd/m². A red LED was placed immediately under each grid of lights to provide feedback following an incorrect response on the visual task. Only one grid was used during each block of trials, and the central LED within each grid was never used in this experiment.

Targets for the active visual monitoring task consisted of a square shape, created by illuminating just the four corner LEDs together in

the fixated grid of lights. There were eight possible nontarget stimuli for the active visual monitoring task, consisting of the illumination of all but the central LED, the illumination of three of the four corner LEDs, the illumination of the four noncorner LEDs, or the illumination of two diagonally opposite corner LEDs. The sequence of visual stimuli at the fixated grid in the active monitoring task was random. The visual stimuli for the passive fixation task consisted of the illumination of both the upper row and the lower row of LEDs in the fixated grid, alternating with both its left column and its right column of LEDs. The visual stimuli in both tasks were presented for 250 msec each, with successive stimuli presented at stimulus onset asynchronies (SOAs) of 350 msec. Only one of the grids of LEDs was illuminated in each block (analogous to only one monitor being used to present lip-read information in Experiment 1). White noise was presented throughout the experiment from a loudspeaker cone placed underneath the participants' chair at 52 dB (A), to ensure that the auditory shadowing task was not at ceiling.

The participant responded to visual targets in the active visual task by pressing a response button placed in the right hand. RTs were measured in milliseconds from target onset using an 82C54 interval-timer chip on an input–output card (Blue Chip Technology, parts DCM-16 and DOP-24) that interfaced to the LEDs and response keys and could also switch which of two loudspeakers presented the irrelevant speech sounds from the audiotape. Timing of the stimuli and responses was controlled by a Viglen 3/33 microcomputer (IBM 386 compatible), using a custom program written in Turbo Pascal. The computer automatically shifted the irrelevant speech signal from one loudspeaker cone to the other in front of the participant, depending on the current condition.

Design. There were again three within-participants factors of interest. One factor was relative position of fixation, which could either be on the same side as the irrelevant auditory stream or on the opposite side. The second factor was analogous to the previous speaking lips versus chewing lips comparison: The visual task was either active monitoring (for the specified square target in a stream of nontargets) or passive fixation of an alternating stream of visual stimuli. The third factor was whether the irrelevant auditory stream was presented on the left or the right. This was now switched between blocks, rather than halfway through each block, due to the way in which the computer controlled the visual events and the location of the auditory stimuli. These three factors were crossed to yield eight conditions, which were each presented once in the eight practice blocks and twice in the 16 experimental blocks. The conditions were ordered such that the same-side versus opposite-side factor alternated every block, whereas active versus passive visual task alternated every other block, and the direction of fixation alternated every 4 blocks. The starting condition was randomized across participants.

Procedure. The visual stream was started shortly before the onset of the first triplet of auditorily presented words and was stopped after the participants had made their verbal responses to the last triplet of words in the block. The probability of a particular visual stimulus being a target within the random stream of shapes for the active condition was $\frac{1}{75}$. Thus, on average, a visual target was presented every 26 sec during the active task. The participants were allowed a maximum of 3 sec from target onset in the active task to initiate a buttonpress response before a "missed-target" response was scored. Correct responses, missed targets, and false alarms in the active visual task were all recorded automatically by the computer controlling the experiment. Feedback following a false alarm or following a missed target was provided by illumination of the red LED under the grid for 2,000 msec.

The participants were instructed to repeat as many of the relevant auditory words as possible in the shadowing task, while also responding to all the visual targets they detected. It was stressed that they should fixate the illuminated grid on one side throughout each block, while keeping their heads pointing straight ahead. Ad-

herence to these instructions was monitored by the experimenter throughout. There was a short break between each condition.

In sum, there were only three major changes in procedure to Experiment 1. First, the visual stimuli were now always meaningless shapes created by the LEDs; no linguistic lipreading task was ever required. Second, the source of the irrelevant auditory stream now remained constant across each block of trials (i.e., it no longer switched on the 7th triplet), due to the computerized method for shifting the irrelevant auditory location, which could not be synchronized to a single auditory triplet but rather took place in the longer pause between blocks. Finally, intrusion errors were no longer recorded, with spoken responses simply being scored as correct or incorrect (because the majority of intrusion errors had occurred on the switch trial in Experiment 1, which was no longer present in this experiment).

Results

The interparticipant mean percentages of correct responses are shown for each condition in Table 2. It can be seen that performance was again worse overall in the same-side condition, in which the visual stream was at the same location as the irrelevant auditory stream that had to be ignored, than in the opposite-side condition. A three-way ANOVA on the total correct responses had the factors of task (active vs. passive), relative position (same side vs. opposite side), and hemispace of irrelevant auditory stream (left vs. right).

The participants reported fewer words correctly on same-side trials ($M = 60.9\%$) than on opposite-side trials ($M = 72.8\%$), resulting in a main effect of relative position [$F(1,9) = 25.2, p < .001$]. Thus, fixating visual events on the side of the irrelevant auditory stream again resulted in a significant decrement in shadowing performance for the attended auditory stream placed behind the participants' head. The participants also reported fewer words correctly overall in the active task ($M = 64.3\%$) than in the passive task ($M = 69.4\%$), resulting in a marginal effect of task [$F(1,9) = 1.9, p = .06$]. This is consistent with the participants' finding the active task more difficult and thus more "attention-demanding" than the passive task, as anticipated. However, in stark contrast to Experiment 1, there was no interaction between relative position and task [$F(1,9) < 0.5$]. That is, the participants found it no more difficult to ignore auditory events at fixation when actively attending there than when merely passively fixating there. This result contrasts with the sig-

Table 2
Mean Percentages of Correct Responses, With Their Standard Errors, as a Function of the Visual Task and the Relative Position of the Irrelevant Auditory Stream and the Fixated Visual Stream (Same vs. Opposite Side) for Irrelevant Auditory Streams Presented to the Left and the Right and Averaged Over Side in Experiment 2

| Irrelevant Auditory Stream | Active Task | | | | Passive Task | | | |
| | Same Side | | Opposite Side | | Same Side | | Opposite Side | |
	M	SE	M	SE	M	SE	M	SE
Left	56.2	6.4	70.3	6.1	61.0	5.8	74.9	5.2
Right	61.3	6.4	69.5	7.7	65.1	7.8	76.5	6.0
Average	58.7	6.4	69.9	6.5	63.1	6.5	75.7	5.5

nificant interaction reported in Experiment 1 and with our predictions. None of the other terms in the analysis of the shadowing data reached statistical significance [for hemispace, $F(1,9) = 2.2, p = .18$; for all other terms, $F < 1$].

For completeness, we also conducted statistical analyses of the target detection data from the active visual task conditions (the means from these data are presented in the Appendix). A two-way within-participants ANOVA on the RT data from the active visual conditions, with the factors of relative position (same side vs. opposite side) and hemispace (left vs. right; now with reference to the side of the visual stimulus) revealed that the participants responded more rapidly to targets presented on the right ($M = 548$ msec) than to targets on the left ($M = 578$ msec), resulting in a main effect of hemispace [$F(1,9) = 6.0$, $p = .04$]. A similar trend toward preferential responding to right over left visual targets has been reported in many previous studies (e.g., Spence & Driver, 1997c), although we note that it may be related here to our use of a right-hand buttonpress response. None of the other terms in the analysis of the RT data reached significance. Similar analyses of the percentages of targets missed and on the numbers of false alarms in each condition also revealed no significant main effects or interactions.

Discussion

The results for the shadowing task in Experiment 2 clearly demonstrate again that fixating visual events at the location of an irrelevant auditory stream makes it more difficult to ignore that stream than when fixating elsewhere. However, in contrast to the results of Experiment 1, we found no difference in the size of this spatial effect as a function of whether the participants merely fixated passively toward the distractor sounds or actively monitored the visual events there. This may suggest that the additional spatial effect that we had attributed to active visual attention in Experiment 1 may in fact be specific to lipreading while listening.

However, there are at least two alternative accounts for the present failure to find a stronger spatial effect in the active versus passive nonlinguistic visual task. First, it may simply be that the present active task of monitoring for an occasional square target (appearing only once every 26 sec, on average) was not as demanding on visual attention as the continuous lipreading had been in Experiment 1. This difference in attentional load might conceivably explain the different outcome for the two experiments rather than the contrast between linguistic and nonlinguistic tasks. Second, it may be that the particular visual stimuli used in Experiment 2, generated by grids of high-intensity LEDs and comprising many abrupt onsets and offsets in an otherwise darkened room, might be so attention capturing as to conceal any difference between the active and passive conditions. It seems unlikely that the continuously chewing lips used in Experiment 1 would have exerted the same exogenous pull on visual attention as the abrupt flashes in Experiment 2. Note that possible attention capture by these flashes may also complicate

interpretation of the same-side versus opposite-side effect for the passive task in Experiment 2 as a pure effect of fixation direction alone. This effect might conceivably be due instead to exogenous attention capture by the flashing grid of LEDs, which appeared only on the fixated side.

Our final experiment was designed to address these weaknesses and potential criticisms of the design used in Experiment 2. In order to address the possibility that the active task may not have been sufficiently demanding, we simply increased the rate of stimulus presentation in the visual stream from one stimulus every 350 msec to one stimulus every 200 msec. We also increased the level of the background white noise to make the shadowing task more difficult. Finally, we now presented simultaneous but independent streams of visual stimuli from both sides of space in all conditions, with the aim of equating any exogenous effects from the flashing grids across the two sides. With these changes, a stronger test should be provided for whether the pattern of Experiment 1 (i.e., a greater same-side vs. opposite-side difference in shadowing performance with an active rather than passive visual task) can be found even when the active visual task is nonlinguistic and is thus unrelated to the auditory shadowing task.

EXPERIMENT 3

Method

Participants. Twelve undergraduate students (8 men and 4 women) at Oxford University were recruited to take part in this experiment. Their mean age was 20 years (range, 18–21 years). All participants but 1 were right-handed by self-report. The participants were naive as to the purpose of the experiment, and all were native English speakers. None had taken part in Experiment 2.

Apparatus, Materials, Design, and Procedure. The only changes from Experiment 2 were as follows. First, the two grids of LEDs were now both used throughout, so that visual stimuli were always presented concurrently from both sides of space. The participants fixated toward one side or the other for the passive or active task, just as before. In the active task, the nontarget stimuli presented on the fixated side were chosen independently of those presented on the other side, and target stimuli (i.e., the square arrangement of illuminated corner LEDs) appeared only on the fixated side. Each visual stimulus was now presented for just 75 msec, with the SOA between successive visual stimuli being reduced to 200 msec.

Table 3
Mean Percentages of Correct Responses, With Their Standard Errors, as a Function of the Visual Task and the Relative Position of the Irrelevant Auditory Stream and the Fixated Visual Stream (Same vs. Opposite Side) for Irrelevant Auditory Streams Presented to the Left and the Right and Averaged Over Side in Experiment 3

| Irrelevant Auditory Stream | Active Task | | | | Passive Task | | | |
| | Same Side | | Opposite Side | | Same Side | | Opposite Side | |
	M	SE	M	SE	M	SE	M	SE
Left	35.4	4.3	45.2	3.3	43.0	4.4	50.4	3.8
Right	32.5	4.8	47.6	4.4	46.7	4.5	53.1	3.4
Average	33.9	4.2	46.4	3.7	44.8	4.0	51.8	3.3

Given the faster presentation rate in the present experiment, we added the constraint that successive targets could now appear only after at least 25 distractor stimuli had been presented since the onset of the preceding target. Aside from this constraint, the probability of a particular visual stimulus being a target was again 1/75. Finally, the level of white noise underneath the participant's chair was now set to 58 dB (A).

Results

The interparticipant mean percentages of correct responses are shown for each condition in Table 3. Note that the overall level of performance on the shadowing task was worse than in Experiment 2, presumably due to the higher level of white noise. More important, it can be seen that performance was again worse overall in the same-side condition, in which the fixated visual stream was at the same location as the irrelevant auditory stream, than in the opposite-side condition. A three-way ANOVA on the total correct responses had the factors of task (active vs. passive), relative position (same side vs. opposite side), and hemispace of the irrelevant auditory stream (left vs. right).

The participants reported fewer words correctly on same-side trials ($M = 39.4\%$) than on opposite-side trials ($M = 49.1\%$) overall, resulting in a main effect of relative position [$F(1,11) = 55.9, p < .0001$]. This shows that fixating the side of the irrelevant auditory stream again produced a significant decrement in shadowing performance for the attended auditory stream. The participants reported fewer words correctly in the active task ($M = 40.1\%$) than in the passive task ($M = 48.3\%$) overall, resulting in a highly significant main effect of task [$F(1,11) = 21.2, p < .001$]. This is consistent with the active condition being more difficult and thus more attention demanding than the passive condition. The most important result is the significant interaction between relative position and task [$F(1,11) = 7.7, p = .02$]. That is, the participants now found it significantly more difficult to ignore auditory events at fixation when actively attending the fixated location (mean difference between same- and opposite-side conditions of 12.5%; $p < .01$, by t-test pairwise comparison) than when merely passively fixating there (mean difference of just 7%; $p < .01$). None of the other terms in the ANOVA on shadowing accuracy reached statistical significance [for side by relative position × task, $F(1,11) = 1.7, p = .22$; for all other terms, $F < 1$].

We also conducted a two-way within-participants ANOVA on the RT data from the conditions in which the participants performed the active visual task, with the factors of relative position and hemispace (see the Appendix for a fuller presentation of these data). The participants again showed a trend toward responding more rapidly to visual targets presented on the right ($M = 826$ msec) than on the left ($M = 985$ msec); however, this difference just failed to reach statistical significance [$F(1,11) = 3.2, p = .10$]. None of the other terms in the analysis of the RT data reached significance, nor did the terms in similar analyses of the percentages of targets

missed or the numbers of false alarms in the active visual task. Note that performance in the active visual task was much worse (slower RTs, lower percent correct, and more false alarms) in Experiment 3 than in Experiment 2, as expected given the changes we had introduced to make the visual task harder (see Table A1 in Appendix).

Discussion

The results of Experiment 3 again demonstrate that people find it difficult to ignore irrelevant auditory stimuli when fixating visual events at the same location than when fixating away. Since visual stimuli were now presented from both sides throughout, this effect within the passive visual task (i.e., mere fixation) was presumably due to the direction of gaze alone (whereas, in Experiments 1 and 2, it might conceivably have been attributed to the unilateral visual stimulation that was used).

More important, the participants found it significantly more difficult to ignore the irrelevant auditory stream when they were actively attending there (for the monitoring task) than when simply passively fixating there, just as was found in Experiment 1. Taken together, our results thus show that while people find it difficult to ignore sounds at fixation, this is even more difficult when they are actively attending there to perform a visual task. Since a similar pattern was found in Experiments 1 and 3, this pattern of results is evidently not specific to the particular case of cross-modal integration between lipreading and auditory speech perception. Instead, the pattern reflects cross-modal links in spatial attention that operate even when the auditory and visual tasks are entirely unrelated.

The absence of this pattern in Experiment 2 was presumably due to the active visual task being insufficiently demanding and/or to the flashing LED grid on only one side capturing attention regardless of the specified task. Several previous authors (Greenwald, 1972; Pashler, 1998) have suggested that cross-modal links in attention may become more apparent in more demanding tasks. This accords with the contrast between Experiments 2 and 3; Experiment 3 was more demanding in terms of both the faster visual presentation rate and the use of bilateral rather than unilateral stimulation.

GENERAL DISCUSSION

Our results show that increasing the spatial separation between relevant and irrelevant streams allows more efficient selective processing, even when the streams in question are in different sensory modalities. This extends the spatial-separation principle for efficient selective attention from the strictly unimodal settings where it has previously been studied (e.g., Broadbent, 1958; Driver & Grossenbacher, 1996; Eriksen & Eriksen, 1974; Johnston & Dark, 1986) to a multimodal case. Our findings complement those of Driver and Spence (1994), who found that increasing the spatial separation between

a relevant auditory stream and a relevant visual stream makes it more difficult to attend to both simultaneously (see also Spence & Driver, 1996). It is a well-known generalization (e.g., Kahneman & Treisman, 1984; Spence & Driver, 1997b) that factors that make *divided* attention more difficult (such as the separate positions in hearing and vision that gave relatively poor performance in Driver and Spence's, 1994, dual-task study) tend to make *selective* attention easier, as found here when irrelevant sounds were placed in a different location from the relevant sights.

Our results do not accord naturally with several recent claims that people have entirely independent modality-specific attention systems (e.g., Duncan, Martens, & Ward, 1997; Robin & Rizzo, 1989; Wickens, 1980). According to such accounts, which posit modality-specific resources, processing of the irrelevant auditory stream should have been quite unaffected by its locus relative to the current focus of visual attention in our study, because entirely separate resources should putatively have been involved for the two modalities. Our results fit better with hybrid models for cross-modal attention, which posit that there are some spatial links between auditory and visual attentional systems (see Posner, 1990; Spence & Driver, 1996) in addition to a degree of independence at earlier levels of representation.

It would be premature to speculate extensively on the possible neural basis for our findings. We note, however, that there are now several intriguing neurophysiological demonstrations of cells in cat auditory cortex, which only respond to sounds when the cat is looking in the direction of their source (e.g., Hubel, Henson, Rupert, & Galambos, 1959; see Stein & Meredith, 1993). It is unclear whether some of these neurophysiological findings are due merely to the cat passively fixating in the direction of the sound or instead to it actively attending in the direction of the sounds. When found in anesthetized animals, such results must presumably be due solely to passive fixation. However, it may be important to distinguish this from active visual attention in future neurophysiological work on cross-modal issues with behaving animals, given our present findings.

It would be interesting to know whether the other cross-modal interference effects highlighted in our introduction, such as cross-modal negative priming (e.g., Driver & Baylis, 1993) or cross-modal Stroop (e.g., Cowan, 1989; Cowan & Barron, 1987; Elliott et al., 1998; Greenwald, 1970, 1972; Shimada, 1990), will also show similar spatial effects, as suggested in a preliminary (albeit confounded) manner by Jones and Hapeshi (1991). It should be noted that evidence from our own work on the irrelevant-speech effect in serial recall (Spence & Driver, 1999) suggests that certain cross-modal phenomena may not show the spatial modulation; presumably, in the case of the irrelevant-speech effect, this may arise because the effect is not critically restricted by the ease of on-line perceptual selection, as discussed earlier.

Finally, it should be noted that, in addition to providing information about cross-modal spatial links during selective listening, which seem to be of considerable theoretical interest, the present findings may also help to provide some preliminary constraints for the effective design of multimodal interfaces. In particular, our findings highlight the potential tradeoff between arrangements that make it easier to attend jointly to information presented in different modalities yet, at the same time, can make it more difficult to ignore information in one of the modalities when it becomes irrelevant. Driver and Spence (1994) previously showed that it is easier to perform auditory and visual dual tasks when the relevant stimuli are presented from the same location across the two modalities. Here, we have demonstrated the unfortunate flip side of this for the first time: It can be more difficult to ignore distracting sounds when these are presented at a visually relevant location, particularly when that visual location is actively processed rather than merely passively fixated.

REFERENCES

BERTELSON, P. (1998). Starting from the ventriloquist: The perception of multimodal events. In M. Sabourin, F. I. M. Craik, & M. Robert (Eds.), *Advances in psychological science, Vol. 2: Biological and cognitive aspects* (pp. 419-639). Hove, U.K.: Psychology Press.

BERTELSON, P., VROOMEN, J., & DE GELDER, B. (1995, July). *Interaction of auditory and visual data in speech recognition and voice localization: McGurk interference versus ventriloquism.* Paper presented at the meeting of the Experimental Psychology Society, Birmingham, England.

BLAUERT, J. (1983). *Spatial hearing: The psychophysics of human sound localization.* Cambridge, MA: MIT Press.

BRIAND, K. A. (1998). Feature integration and spatial attention: More evidence of a dissociation between endogenous and exogenous orienting. *Journal of Experimental Psychology: Human Perception & Performance,* 24, 1243-1256.

BROADBENT, D. E. (1956). Listening between and during practised auditory distractions. *British Journal of Psychology,* 47, 51-60.

BROADBENT, D. E. (1958). *Perception and communication.* Elmsford, NJ: Pergamon.

BROWN, G. D. A. (1984). A frequency count of 190,000 words in the *London-Lund Corpus of English Conversation. Behavior Research Methods, Instruments, & Computers,* 16, 502-532.

BUTTER, C. M. (1987). Varieties of attention and disturbances of attention: A neuropsychological analysis. In M. Jeannerod (Ed.), *Neurophysiological and neuropsychological aspects of spatial neglect* (pp. 1-23). Amsterdam: Elsevier.

CHERRY, E. C. (1953). Some experiments upon the recognition of speech with one and two ears. *Journal of the Acoustical Society of America,* 25, 975-979.

COLLE, H. A. (1980). Auditory encoding in visual short-term recall: Effects of noise intensity and spatial location. *Journal of Verbal Learning & Verbal Behavior,* 19, 722-735.

COLLE, H. A., & WELSH, A. (1976) Acoustic masking in primary memory. *Journal of Verbal Learning & Verbal Behavior,* 15, 17-31.

COWAN, N. (1989). The reality of cross-modal Stroop effects. *Perception & Psychophysics,* 45, 87-88.

COWAN, N., & BARRON, A. (1987). Cross-modal, auditory-visual Stroop interference and possible implications for speech memory. *Perception & Psychophysics,* 41, 393-401.

DARWIN, C. J., & HUKIN, R. W. (1999). Auditory objects of attention: The role of interaural time-differences. *Journal of Experimental Psychology: Human Perception & Performance,* 25, 617-629.

DODD, B., & CAMPBELL, R. (1987). *Hearing by eye: The psychology of lipreading*. Hillsdale, NJ: Erlbaum.

DRIVER, J., & BAYLIS, G. C. (1993). Cross-modal negative priming and interference in selective attention. *Bulletin of the Psychonomic Society*, **31**, 45-48.

DRIVER, J., & GROSSENBACHER, P. G. (1996). Multimodal spatial constraints on tactile selective attention. In T. Inui & J. L. McClelland (Eds.), *Attention and performance XVI: Information integration in perception and communication* (pp. 209-235). Cambridge, MA: MIT Press.

DRIVER, J. S., & SPENCE, C. J. (1994). Spatial synergies between auditory and visual attention. In C. Umiltà and M. Moscovitch (Eds.), *Attention and performance XV: Conscious and nonconscious information processing* (pp. 311-331). Cambridge, MA: MIT Press.

DUNCAN, J., MARTENS, S., & WARD, R. (1997). Restricted attentional capacity within but not between sensory modalities. *Nature*, **387**, 808-810.

ELLIOT, E. M., COWAN, N., & VALLE-INCLAN, F. (1998). The nature of cross-modal, color–word interference effects. *Perception & Psychophysics*, **60**, 761-767.

ERIKSEN, B. A., & ERIKSEN, C. W. (1974). Effects of noise letters upon identification of a target letter in a nonsearch task. *Perception & Psychophysics*, **16**, 143-149.

GEISSLER, L. R. (1915). Sound localization under determined expectation. *American Journal of Psychology*, **26**, 268-285.

GOPHER, D. (1973). Eye-movement patterns in selective listening tasks of focused attention. *Perception & Psychophysics*, **14**, 259-264.

GRANT, K. W., & SEITZ, P. F. (1998). The use of visible speech cues (speechreading) for directing auditory attention: Reducing temporal and spectral uncertainty in auditory detection of spoken sentences. In P. K. Kuhl & L. A. Crum (Eds.), *Proceedings of the 16th International Congress on Acoustics and the 135th Meeting of the Acoustical Society of America* (Vol. 3, pp. 2335-2336). New York: Acoustical Society of America.

GREENWALD, A. G. (1970). Selective attention as a function of signal rate. *Journal of Experimental Psychology*, **86**, 48-52.

GREENWALD, A. G. (1972). Evidence of both perceptual filtering and response suppression for rejected messages in selective attention. *Journal of Experimental Psychology*, **94**, 58-67.

HISCOCK, M., HAMPSON, E., WONG, S. C. P., & KINSBOURNE, M. (1985). Effects of eye movements on the recognition and localization of dichotic stimuli. *Brain & Cognition*, **4**, 140-155.

HUBEL, D. H., HENSON, C. O., RUPERT, A., & GALAMBOS, R. (1959). "Attention" units in auditory cortex. *Science*, **129**, 1279-1280.

HUBLET, C., MORAIS, J., & BERTELSON, P. (1976). Spatial constraints on focussed attention: Beyond the right side advantage. *Perception*, **5**, 3-8.

HUBLET, C., MORAIS, J., & BERTELSON, P. (1977). Spatial effects in speech perception in the absence of spatial competition. *Perception*, **6**, 461-466.

JOHNSTON, W. A., & DARK, V. J. (1986). Selective attention. *Annual Review of Psychology*, **37**, 43-75.

JONES, D. (1993). Objects, streams, and threads of auditory attention. In A. Baddeley & L. Weiskrantz (Eds.), *Attention: Selection, awareness and control* (pp. 87-104). Oxford: Oxford University Press.

JONES, D. M. (1999). The cognitive psychology of auditory distraction. *British Journal of Psychology*.

JONES, D. M., & HAPESHI, K. (1991). *Information-processing under high workload* (final report of the contract). Farnborough, U.K.: Army Personnel Research Establishment.

JONES, D. M., MILES, C., & PAGE, J. (1990). Disruption of proof-reading by irrelevant speech: Effects of attention, arousal or memory? *Applied Cognitive Psychology*, **4**, 89-108.

JONES, J. A., & MUNHALL, K. G. (1997). The effects of separating auditory and visual sources on audiovisual integration of speech. *Canadian Acoustics*.

KAHNEMAN, D., & TREISMAN, A. (1984). Changing views of attention and automaticity. In R. Parasuraman & D. R. Davies (Eds.), *Varieties of attention* (pp. 26-61). San Diego, CA: Academic Press.

KLEIN, R. M. (1977). Attention and visual dominance: A chronometric analysis. *Journal of Experimental Psychology: Human Perception & Performance*, **3**, 365-378.

KLEIN, R. M. (1994). Perceptual-motor expectancies interact with covert visual orienting under conditions of endogenous but not exogenous control. *Canadian Journal of Experimental Psychology*, **48**, 167-181.

KLEIN, R. M., KINGSTONE, A., & PONTEFRACT, A. (1992). Orienting of visual attention. In K. Rayner (Ed.), *Eye movements and visual cognition: Scene perception and reading* (pp. 46-65). New York: Springer-Verlag.

LARMANDE, P., ELGHOZI, D., SINTES, J., BIGOT, T., & AUTRET, A. (1983). Test d'ecoute dichotique verbal et non verbal: Influence de l'etat d'activation hémisphérique chez le suject normal [Testing verbal and nonverbal dichotic listening: The influence of the state of hemispheric activation in the normal subject]. *Revue Neurologique*, **139**, 65-69.

LECOMPTE, D. C., NEELY, C. B., & WILSON, J. R. (in press). Irrelevant speech and irrelevant tones: The relative importance of speech to the irrelevant speech effect. *Journal of Experimental Psychology: Learning, Memory, & Cognition*.

MCGURK, H., & MCDONALD, J. (1976). Hearing lips and seeing voices. *Nature*, **264**, 746-748.

MILES, C., MADDEN, C., & JONES, D. M. (1989). Cross-modal, auditory-visual Stroop interference: A reply to Cowan and Barron (1987). *Perception & Psychophysics*, **45**, 77-81.

MORAIS, J., & BERTELSON, P. (1973). Laterality effects in diotic listening. *Perception*, **2**, 107-111.

MORAIS, J., CARY, L., VANHAELEN, H., & BERTELSON, P. (1980). Postural determinants of frontal-position advantage in listening to speech. *Perception & Psychophysics*, **27**, 141-148.

MORTON, J. (1969). Categories of interference: Verbal mediation and conflict in card sorting. *British Journal of Psychology*, **60**, 329-346.

NÄÄTÄNEN, R. (1992). *Attention and brain function*. Hillsdale, NJ: Erlbaum.

NÄÄTÄNEN, R., PAAVILAINEN, P., TIITINEN, H., JIANG, D., & ALHO, K. (1993). Attention and mismatch negativity. *Psychophysiology*, **30**, 436-450.

NELSON, W. T., BOLIA, R. S., ERICSON, M. A., & MCKINLEY, R. L. (1998). Monitoring the simultaneous presentation of multiple spatialized speech signals in the free field. In P. K. Kuhl & L. A. Crum (Eds.), *Proceedings of the 16th International Congress on Acoustics and the 135th Meeting of the Acoustical Society of America* (Vol. 4, pp. 2341-2342). New York: Acoustical Society of America.

PASHLER, H. E. (1998). *The psychology of attention*. Cambridge, MA: MIT Press.

PLENGE, G. (1974). On the differences between localization and lateralization. *Journal of the Acoustical Society of America*, **56**, 944-951.

POSNER, M. I. (1978). *Chronometric explorations of mind*. Hillsdale, NJ: Erlbaum.

POSNER, M. I. (1990). Hierarchical distributed networks in the neuropsychology of selective attention. In A. Caramazza (Ed.), *Cognitive neuropsychology and neurolinguistics: Advances in models of cognitive function and impairment* (pp. 187-210). Hillsdale, NJ: Erlbaum.

REISBERG, D. (1978). Looking where you listen: Visual cues and auditory attention. *Acta Psychologica*, **42**, 331-341.

REISBERG, D. (1987). Easy to hear but hard to understand: A lip-reading advantage with intact auditory stimuli. In B. Dodd & R. Campbell (Eds.), *Hearing by eye: The psychology of lip-reading* (pp. 97-113). Hillsdale, NJ: Erlbaum.

REISBERG, D., SCHEIBER, R., & POTEMKEN, L. (1981). Eye position and the control of auditory attention. *Journal of Experimental Psychology: Human Perception & Performance*, **7**, 318-323.

RIGGIO, L., & KIRSNER, K. (1997). The relationship between central cues and peripheral cues in covert visual orientation. *Perception & Psychophysics*, **59**, 885-899.

ROBIN, D. A., & RIZZO, M. (1989).The effect of focal cerebral lesions on intramodal and cross-modal orienting of attention. In T. E. Prescott (Ed.), *Clinical aphasiology* (Vol. 18, pp. 61-74). Boston: College-Hill.

ROBINSON, D. L., & KERTZMAN, C. (1995). Covert orienting of attention in macaques. III. Contributions of the superior colliculus. *Journal of Neurophysiology*, **74**, 713-721.

SALAMÈ, P., & BADDELEY, A. D. (1982). Disruption of short-term memory by unattended speech: Implications for the structure of working memory. *Journal of Verbal Learning & Verbal Behavior*, **21**, 150-164.

SHIMADA, H. (1990). Effect of auditory presentation of words on color

naming: The intermodal Stroop effect. *Perceptual & Motor Skills*, **70**, 1155-1161.

SIMON, J. R., & CRAFT, J. L. (1970). Effects of an irrelevant auditory stimulus on visual choice reaction time. *Journal of Experimental Psychology*, **86**, 272-274.

SPEITH, W., CURTIS, J. F., & WEBSTER, J. C. (1954). Responding to one of two simultaneous messages. *Journal of the Acoustical Society of America*, **26**, 391-396.

SPENCE, C. J., & DRIVER, J. (1994). Covert spatial orienting in audition: Exogenous and endogenous mechanisms. *Journal of Experimental Psychology: Human Perception & Performance*, **20**, 555-574.

SPENCE, C., & DRIVER, J. (1996). Audiovisual links in endogenous covert spatial attention. *Journal of Experimental Psychology: Human Perception & Performance*, **22**, 1005-1030.

SPENCE, C., & DRIVER, J. (1997a). Audiovisual links in exogenous covert spatial orienting. *Perception & Psychophysics*, **59**, 1-22.

SPENCE, C., & DRIVER, J. (1997b). Cross-modal links in attention between audition, vision, and touch: Implications for interface design. *International Journal of Cognitive Ergonomics*, **1**, 351-373.

SPENCE, C., & DRIVER, J. (1997c). On measuring selective attention to an expected sensory modality. *Perception & Psychophysics*, **59**, 389-403.

SPENCE, C., & DRIVER, J. (1999). *Spatial location of irrelevant speech does not modulate the unattended speech effect in immediate serial recall*. Unpublished manuscript.

SPENCE, C., NICHOLLS, M. E. R., GILLESPIE, N., & DRIVER, J. (1998). Cross-modal links in exogenous covert spatial orienting between touch, audition, and vision. *Perception & Psychophysics*, **60**, 544-557.

STEIN, B. E., & MEREDITH, M. A. (1993). *The merging of the senses*. Cambridge, MA: MIT Press.

STERNBERG, S. (1969). The discovery of processing stages: Extensions of Donders' method. *Acta Psychologica*, **30**, 276-315.

STEVENS, S. S., & NEWMAN, E. B. (1936). The localization of actual sources of sound. *American Journal of Psychology*, **48**, 297-306.

SUMBY, W. H., & POLLACK, I. (1954). Visual contribution to speech intelligibility in noise. *Journal of the Acoustical Society of America*, **26**, 212-215.

SUMMERFIELD, Q. (1992). Lipreading and audio-visual speech perception. *Philosophical Transactions of the Royal Society of London: Series B*, **335**, 71-78.

THACKRAY, R. I., & JONES, K. N. (1971). Level of arousal during Stroop performance: Effects of speed stress and "distraction." *Psychonomic Science*, **23**, 133-135.

THEEUWES, J. (1994). Endogenous and exogenous control of visual selection. *Perception*, **23**, 429-440.

TREISMAN, A. (1969). Strategies and models of selective attention. *Psychological Review*, **76**, 282-299.

WARD, L. M. (1994). Supramodal and modality-specific mechanisms for stimulus-driven shifts of auditory and visual attention. *Canadian Journal of Experimental Psychology*, **48**, 242-259.

WICKENS, C. D. (1980). The structure of attentional resources. In R. S. Nickerson (Ed.), *Attention and performance VIII* (pp. 239-257). Hillsdale, NJ: Erlbaum.

WOLTERS, N. C. W., & SCHIANO, D. J. (1989). On listening where we look: The fragility of a phenomenon. *Perception & Psychophysics*, **45**, 184-186.

WOOD, N. L., & COWAN, N. (1995). The cocktail party phenomenon revisited: Attention and memory in the classic selective listening procedure of Cherry (1953). *Journal of Experimental Psychology: General*, **124**, 243-262.

NOTES

1. Sounds presented from directly behind a person's head can sometimes be mislocalized as having been presented in front (e.g., Geissler, 1915; Stevens & Newman, 1936), especially under conditions in which head movements are prevented, as in the present study. If participants in this study mislocalized the relevant auditory stream behind them as coming from directly in front, then the effective spatial separation between relevant and irrelevant auditory streams would be only 38° rather than the 142° that physically separated the sources. However, none of the participants in the present study reported any such front–back mislocalizations on debriefing. Indeed, it would be unusual for speech stimuli to be mislocalized in the median plane, given Blauert's (1983, pp. 95–107) point (see also Plenge, 1974) that front–back confusions occur only when people must localize unfamiliar sounds that have very narrow frequency spectra. Moreover, Geissler (1915) reported that, even for impoverished auditory stimuli, front–back localization errors can be reduced when participants direct their attention to the appropriate source location (as in the present study, in which the true source of the relevant sounds was known). We are therefore quite confident that front–back confusions did not arise, so the relevant and irrelevant sounds did appear to be widely separated.

2. This interaction might conceivably still be due merely to differences in the overall level of performance between speaking lips and chewing lips (i.e., essentially, it would reduce to the main effect), arguing that the better shadowing in the speaking-lips condition might simply lie on a more sensitive section of the performance curve, thus rendering the relative position effect more detectable. We examined this possibility by comparing the relative position effect for the speaking-lips condition versus the chewing-lips condition in a subset of the data that gave more similar overall performance for speaking lips and chewing lips. To this end, we compared speaking-lips data for the 8 of the 16 participants whose overall performance in the speaking-lips condition was lowest with the chewing-lips data from the 8 participants whose overall performance in the chewing-lips condition was highest. A mixed ANOVA was conducted on this subset of the data, with the between-participants factor of lips condition and the within-participants factors of hemispace and relative position. Unsurprisingly, given the data selection, overall performance in the speaking-lips condition ($M = 60\%$) was no longer higher than that in the chewing-lips condition ($M = 63\%$); the trend was slightly the other way, but nonsignificant [$F(1,14) < 0.8$, $p = .40$]. Nevertheless, the interaction between lips and relative position remained close to significance [$F(1,14) = 3.5, p = .08$] because bringing the irrelevant auditory stream closer to the relevant visual stream still had a larger effect on performance in the speaking-lips condition (reducing performance from 71% correct in the opposite-side condition to just 49% correct in the same-side condition) than in the chewing-lips condition (in which performance dropped from 71% to 56% correct). The reduced significance of the interaction in this subanalysis is presumably attributable to only half as many data points being analyzed, relative to the main analysis, and to the lips factor now being between rather than within participants, further reducing power. Nevertheless, the important point to note for present purposes is that relative position still had a greater effect in the speaking-lips condition than in the chewing-lips condition, even when overall performance levels were approximately matched, arguing against sensitivity account for the interaction.

3. It might be suggested that the intrusion error effects could account for the overall effects on the percentage of words correctly reported. That is, there might be no significant difference between the overall number of presented words that were reported for each condition, with the differences between conditions merely reflecting the relative percentages of words reported from the relevant stream (coded as correct) versus the irrelevant stream (coded as intrusion errors). Visual inspection of Table 1 suggests that this is not the case (intrusion errors are too rare to account for all of the effects in the percent correct data). Further statistical analysis, when reports of either relevant or irrelevant words were now both scored as correct, confirm the same pattern of results as reported in just the percent correct analysis. Thus, although intrusion errors did vary systematically with condition and in an inverse manner to the number of relevant words that were correctly reported, these two outcomes cannot be reduced to each other. Instead, the two measures corroborate which conditions were the most difficult.

4. It might be suggested that the participants looked directly at the active video monitor only when it presented relevant lip-read information, fixating a less eccentric location in the chewing-lips condition because the peripheral visual stream was uninformative. This might then account for the reduced performance difference between same- and opposite-side conditions for chewing lips versus speaking lips. However, it should be noted that the participants were regularly instructed to look directly at the active video monitor for all conditions, and moni-

toring by the experimenter throughout the experimental session confirmed that all participants followed these instructions. Moreover, Driver and Spence (1994, Experiment 2) monitored gaze direction with video in their study, which also required participants to fixate eccentric chewing or speaking lips. They found indistinguishable patterns of fixation for the two types of visual stream. Taken together, these observations suggest that it is unlikely that very different patterns of fixation could account for the differences between speaking- and chewing-lips conditions in the present study.

5. Further support for this conclusion comes from an additional analysis on intrusion errors, which showed that the participants made significantly more intrusion errors on the 7th switch triplet ($M = 4.6\%$) than on the immediately preceding 6th triplet ($M = 2.3\%$). A one-way ANOVA conducted on these data revealed a significant difference [$F(1,15) = 6.6, p = .02$], consistent with the participants' attention being exogenously drawn toward the irrelevant auditory stream when its location moved.

APPENDIX

The visual target detection data from the active visual conditions in Experiments 2 and 3 are shown in Table A1 for completeness. The most striking observation is that performance was worse overall in Experiment 3 than in Experiment 2: The participants responded more slowly, detected fewer targets correctly, and made more false alarms in Experiment 3. This was expected given the faster presentation rate and the use of bilateral rather than unilateral presentation, which we had introduced with the deliberate aim of making the active visual task more difficult in Experiment 3 than in Experiment 2. There were some trends for reduced accuracy in the same-side condition relative to the

APPENDIX (Continued)

opposite-side condition (see Table A1), but none reached significance (see fuller statistical accounts in main text); in any case, RTs tended to trend slightly against this, again nonsignificantly.

Table A1
Mean Visual Reaction Times (RTs), Percentages of Correct Responses, and Numbers of False Alarms, as a Function of the Relative Position of the Irrelevant Auditory Stream in Experiments 2 and 3

| Visual Task Performance Measure | Fixated Side | | | |
| | Left | | Right | |
	Same Side	Opposite Side	Same Side	Opposite Side
Experiment 2				
RT	570	586	539	557
% Correct	74.8	80.1	75.0	84.0
False Alarms	2.8	3.1	1.7	1.7
Experiment 3				
RT	937	1,034	745	906
% Correct	54.0	51.3	47.8	54.4
False Alarms	15.8	16.0	16.8	16.4