Seeing or hearing?
Perceptual independence, modality confusions, and crossmodal congruity effects with focused and divided attention

AXEL LARSEN
University of Copenhagen, Copenhagen, Denmark

WILLIAM McILHAGGA
University of Bradford, Bradford, England

JEROEN BAERT
University of Leuven, Leuven, Belgium

and

CLAUS BUNDESEN
University of Copenhagen, Copenhagen, Denmark

Observers were given brief presentations of pairs of simultaneous stimuli consisting of a visual and a spoken letter. In the visual focused-attention condition, only the visual letter should be reported; in the auditory focused-attention condition, only the spoken letter should be reported; in the divided-attention condition, both letters, as well as their respective modalities, should be reported (forced choice). The proportions of correct reports were nearly the same in the three conditions (no significant divided-attention decrement), and in the divided-attention condition, the probability that the visual letter was correctly reported was independent of whether the auditory letter was correctly reported. However, with a probability much higher than chance, the observers reported hearing the visual stimulus letter or seeing the spoken stimulus letter (modality confusions). The strength of the effect was nearly the same with focused as with divided attention. We also discovered a crossmodal congruity effect: Performance was better when the two letters in a stimulus pair were the same than when they differed in type.

Many human activities require concurrent intake of information from ear and eye, but bimodal perception has not been investigated very much (see Pashler, 1998). Both in vision and in hearing, systematic interference has been demonstrated when two inputs in the same modality must be identified at the same time (see, e.g., Bundesen, 1990; Duncan, 1980; Duncan et al., 1999; Moray, 1969). As a simple example, in the group of normal control participants tested by Duncan et al., the probability that a briefly presented visual letter was correctly reported decreased from a value of .80 when the letter was presented alone to a value of .62 when the letter was accompanied by an additional visual letter to be reported. Comparatively few studies have investigated identification of simultaneous inputs to vision and hearing, and the results of these studies have been conflicting (see Duncan, Martens, & Ward, 1997;

Jolicœur, 1999; and related studies analyzed in the Discussion section).

In this article we report an experiment in which participants were presented with pairs of concurrent stimuli consisting of a visual and a spoken letter. In the divided-attention condition, both letters, as well as their respective modalities, should be reported (forced choice). In the visual focused-attention condition, only the visual letter should be reported, and in the auditory focused-attention condition, only the auditory letter. The stimuli were degraded in such a way that the proportions of correct reports were nearly the same in the two focused-attention conditions. With this calibration, the proportions of correct reports of visual and spoken letters were almost the same in the divided-attention condition as in the focused-attention conditions (*no divided-attention decrement*). Moreover, in the divided-attention condition, the probability that the visual letter was correctly reported was independent of whether the auditory letter was correctly reported (*stochastic independence*). However, we discovered that with a probability much higher than chance, the participants reported hearing the visual stimulus letter or see-

ing the spoken stimulus letter (*modality confusions*). Furthermore, performance was substantially more accurate when the two letters in a stimulus pair were identical than when they differed in type (*crossmodal congruity effect*).

METHOD

Participants

Eight students (7 females, 1 male) with a mean age of 23.5 years ($SD = 2.6$ years) participated in the study. They were paid by the hour.

Stimuli and Apparatus

The stimuli were visual and spoken letters of types A, B, C, I, O, P, S, and T. The visual and the spoken member of a stimulus pair were presented at the same time. The spoken letter was presented through the loudspeakers of a Power Macintosh 7500 computer, and the visual letter was shown at the center of the screen of an Apple Multiple Scan 17 Display unit connected to the computer.

The spoken letters were pronounced in Danish by a male voice. The original vocal recordings were made using the Apple microphone in the Power Macintosh computer. The recordings were done at a sampling rate of 22 kHz using 8 bits of resolution for amplitude. By use of Sound Edit 16 software, the recorded letters were compressed or extended in time until each letter had a duration of 150 msec. The resulting letters were clearly audible and highly similar to the original spoken letters. Each letter was then normalized to an initial maximum sound pressure level of 64 dBA and mixed with white noise of 20% of the maximum sound pressure level.

The visual stimulus letters (Chicago font) were exposed in black on a white background and were followed by a 500-msec mask of random letter fragments created by XOR-ing the letters in the stimulus set in a Macintosh Pict-file. Mean height and width of a letter were 0.9 and 1.2 cm, respectively, and the viewing distance was 60 cm. Each letter was degraded by randomly removing 70% of its pixels. Initially the exposure duration was set at 200 msec.

Design

Each participant served in 1,024 trials in the visual focused-attention condition, 1,024 trials in the auditory focused-attention condition, and 2,048 trials in the divided-attention condition. On each trial, the participant was presented with a visual letter and a spoken letter. The letters were drawn at random, with replacement, from the eight-letter set {A, B, C, I, O, P, S, T}, so that the a priori probability that any particular combination of letters (e.g., a visual A and a spoken S) would appear on a given trial was kept constant at 1/64. The total set of 4,096 trials was divided into 16 blocks. Each of the 16 blocks consisted of three subblocks: a subblock with visual focused attention (64 trials), a subblock with auditory focused attention (64 trials), and a subblock with divided attention (128 trials). The ordering of the three subblocks was counterbalanced across the 16 blocks. All randomizations were done anew for each participant.

Throughout the experiment, the exposure duration of the visual letters and the sound pressure level of the spoken letters were adjusted between blocks of trials so that for each participant, the proportion of correct reports was approximately .75 in both the visual and the auditory focused-attention condition. The experimental sessions were preceded by a practice session of about 1 h, which served to familiarize the participant with the apparatus and procedure and make initial adjustments of the exposure duration of the visual letters and the sound pressure level of the spoken letters.

Procedure

The participants were tested individually. Each participant was seated in front of the computer at the viewing distance of 60 cm from the screen. In all conditions, the participants were instructed to fixate a small fixation cross at the center of the display screen before initiating the next trial by pressing the space bar. When the bar was pressed, the stimulus pair was presented with a latency of 200 msec. In the visual focused-attention condition, the participants were instructed to report what they saw on the display screen, and only what they saw. In the auditory focused-attention condition, they were asked to report what they heard, and only what they heard. In the divided-attention condition, they were asked to report both what they saw and what they heard, in free order. The instructions emphasized that a response was considered erroneous unless both the modality (visual or auditory) and the letter identity were correct, but response order was immaterial.

The participants responded by a forced choice (pointing and clicking with the mouse) of one of the eight possible stimuli, which were displayed as printed letters on the buttons of two horizontal response panels. The panels were displayed 100 msec after mask offset, one near the top and the other near the bottom of the display screen. The horizontal locations of the labeled response buttons were the same in the two panels. The visual response panel was marked by a black picture (icon) of an eye, which was dimmed to gray to signal irrelevance in the auditory condition. Similarly, the auditory response panel was marked by a black picture of an ear, which was dimmed to gray in the visual condition. For five participants the visual response panel was near the top and the auditory panel near the bottom of the display screen; for the remaining participants the locations were reversed.

In the visual focused-attention condition, a correct response elicited a high-pitched tone, and an error was followed by silence. In the auditory focused-attention condition, a correct response elicited a low-pitched tone, and again an error was followed by silence. In the divided-attention condition, a correct response to the visual letter elicited a high-pitched tone and a correct response to the spoken letter elicited a low-pitched tone; the feedback was not provided until both responses had been made.

Participants were free to take breaks between trials. Each participant served for about 5 h during each of 2 experimental days separated by at most 1 week. Participants usually completed 6 to 8 blocks of trials during their 1st experimental day and 8 to 10 blocks during their 2nd day.

RESULTS

Tables 1 and 2 show stimulus–response confusion matrices for focused- and divided-attention conditions averaged across the 8 participants. Our calibrations of the exposure duration of the visual letters and the sound pressure level of the spoken letters induced almost equal proportions of correct reports in the two focused-attention conditions (.742, $SE = 0.03$, in the visual condition; .747, $SE = 0.03$, in the auditory condition). When participants divided attention and reported the stimulus characters in both modalities, proportions of correct responses to the visual and spoken stimuli were also nearly the same (.738, $SE = 0.02$, for the visual letters; .736, $SE = 0.03$, for the spoken letters). More importantly, dividing attention between the two modalities and reporting both the visual and the spoken letter rather than focusing attention on just one of the letters caused no reliable cost (no significant divided-attention decrement). By a two-way (attentional focusing × stimulus modality) repeated measures analysis of variance (ANOVA), neither attentional focusing [focused vs. divided attention, $F(1,7) = 4.33$, $p = .08$] nor stimulus modality (visual vs. auditory, $F < 1$) had reliable effects

Table 1
Visual Stimulus–Response Confusion Matrices

Response	Visual Stimulus Letter							
	A	B	C	I	O	P	S	T
Focused Attention								
A	.372	.085	.004	.007	.070	.007	.004	.007
B	.274	.501	.020	.026	.102	.029	.023	.009
C	.025	.043	.891	.012	.057	.012	.016	.009
I	.008	.010	.005	.845	.008	.004	.006	.087
O	.136	.141	.044	.012	.685	.008	.007	.002
P	.066	.060	.019	.018	.029	.886	.013	.028
S	.104	.154	.010	.026	.043	.039	.915	.018
T	.014	.007	.006	.053	.006	.014	.016	.841
Divided Attention								
A	.406	.103	.003	.009	.072	.009	.004	.004
B	.249	.471	.017	.016	.119	.033	.025	.011
C	.028	.045	.881	.013	.051	.014	.011	.009
I	.007	.007	.005	.844	.013	.005	.012	.106
O	.135	.161	.066	.013	.671	.010	.003	.004
P	.062	.057	.016	.032	.025	.884	.012	.022
S	.102	.145	.011	.031	.043	.032	.922	.019
T	.011	.010	.002	.042	.007	.013	.011	.824

Note—Confusion matrices are averaged across subjects. Diagonal entries show proportions of correct responses. Off-diagonal entries show proportions of errors.

on report accuracy.[1] Furthermore, attentional focusing showed no interaction with modality ($F < 1$).

Stochastic Independence

In the divided-attention condition, the probability that a letter in a given modality (visual or auditory) was correctly reported was independent of whether the letter in the other modality was correctly reported. For each participant, Table 3 shows the observed probability that the visual letter was correctly reported in the divided-attention condition, regardless of whether the spoken letter was correctly reported; the observed probability that the spoken letter was correctly reported, regardless of whether the visual letter was correctly reported; and observed and expected probabilities that both the visual and the spoken letter were correctly reported. The expected probability that both the visual and the spoken letter were correctly reported was based on the assumption that correct report of the visual letter and correct report of the spoken letter were stochastically independent events. The expected probability equaled the product of the observed probability that the visual letter was correctly reported and the ob-

Table 2
Auditory Stimulus–Response Confusion

Response	Spoken Stimulus Letter							
	A	B	C	I	O	P	S	T
Focused Attention								
A	.731	.044	.002	.192	.000	.191	.017	.001
B	.091	.804	.000	.094	.002	.268	.015	.003
C	.003	.011	.746	.016	.000	.012	.005	.030
I	.076	.059	.005	.542	.008	.183	.013	.001
O	.011	.004	.008	.011	.981	.014	.013	.006
P	.059	.073	.008	.124	.007	.296	.017	.000
S	.028	.005	.009	.012	.000	.030	.919	.005
T	.001	.001	.223	.009	.003	.007	.002	.954
Divided Attention								
A	.747	.040	.003	.238	.001	.191	.013	.001
B	.077	.778	.002	.096	.002	.287	.009	.001
C	.005	.002	.756	.015	.000	.011	.007	.046
I	.071	.073	.002	.488	.007	.175	.012	.001
O	.007	.005	.008	.019	.984	.016	.021	.004
P	.061	.096	.003	.117	.002	.277	.014	.001
S	.031	.004	.008	.021	.003	.035	.916	.004
T	.001	.001	.217	.006	.000	.008	.009	.942

Note—Confusion matrices are averaged across subjects. Diagonal entries show proportions of correct responses. Off-diagonal entries show proportions of errors.

Table 3
Probabilities of Correct Reports of Visual and Auditory Stimuli With Divided Attention

Participant	Visual	Auditory	Both[a]
He	.747	.736	.548 (.550)
Me	.758	.722	.540 (.547)
Th	.750	.770	.586 (.578)
Ch	.688	.761	.530 (.524)
Ma	.769	.755	.579 (.581)
Gu	.732	.728	.537 (.533)
Le	.707	.674	.479 (.476)
Mu	.739	.745	.549 (.551)
Overall	.736	.736	.544 (.542)

[a]Probability that both the visual and the auditory stimulus were correctly reported. The expected probability given stochastic independence is indicated in parentheses.

served probability that the spoken letter was correctly reported. As can be seen from the table, the observed probability that both the visual and the spoken letter were correctly reported was remarkably close to the expected value for each individual participant. Despite the large number of observations (more than 16,000 trials), deviations between the observed and expected values were not statistically significant. The data for each participant were submitted to a chi-square test with 1 df and summing across the 8 participants [$\chi^2(8) = 10.06, p = .26$].

Modality Confusions

Participants tended to make modality errors—that is, report a letter presented in one modality (visual or auditory) as if it had been presented in the other modality. To measure the strength of this tendency, we considered the probability that an erroneous report of a letter in a given modality (visual or auditory) was identical in type to the letter presented in the other modality (i.e., the probability that a report that mismatched the letter in the target modality matched the letter presented in the other modality).

In the visual focused-attention condition, the observed probability that an erroneous report was identical in type to the spoken letter was .179 (mean of means for individual participants). The observed probability exceeded the chance level of .125 ($\frac{1}{8}$) for all of the 8 participants ($p < .005$ by a one-tailed sign test). In the auditory focused-attention condition, the probability that an erroneous report was identical in type to the visual letter was .170. In this condition the probability exceeded the chance level for 7 out of the 8 participants ($p < .05$). In the divided-attention condition, the probability that an erroneous report of a visual letter was identical in type to the spoken letter was .172; the probability exceeded the chance level for all of the 8 participants ($p < .005$). In the same condition, the probability that an erroneous report of a spoken letter was identical in type to the visual letter was .185; the probability exceeded the chance level for 7 out of the 8 participants ($p < .05$).

The overall probability that an erroneous report of a letter in a given modality (visual or auditory) was identical in type to the letter presented in the other modality was

nearly the same with focused attention (.175) as with divided attention (.178). The probability was higher in the focused-attention condition for 4 out of the 8 participants and higher in the divided-attention condition for the remaining 4 participants.

The participants showed no systematic tendency to make modality reversals in the divided-attention condition—that is, to let a modality error with respect to one of the letters (e.g., a report of a visual S on a trial with a visual A and a spoken S) be accompanied by a modality error with respect to the other letter (a report of a spoken A). There were 1,052 trials on which the visual and the spoken letter were different in type and both letters were erroneously reported. For any stimulus pair in which the visual and the spoken letter were different in type (e.g., a visual A and a spoken S), a modality reversal was one among $7 \times 7 = 49$ possible reports in which both letters were in error. Out of the 1,052 trials, a total of 24 trials (2.3%) showed modality reversals, so the proportion of trials with modality reversals was close to the chance level of $\frac{1}{49}$. For each individual participant, we made a binomial test of the null hypothesis that the probability of a modality reversal was $\frac{1}{49}$ on any trial in which the visual and the spoken letter were different in type and both letters were erroneously reported. An overall p value was computed by converting the p value for each participant to a value of chi square with 2 df and summing the chi squares across the 8 participants (see Winer, 1971, p. 49). The null hypothesis accounted well for the data [$\chi^2(16) = 18.62, p = .29$].

Crossmodal Congruity Effect

We found substantial effects of congruity between the visual and the spoken member of a stimulus pair. The proportion of correct reports averaged .789 when the two letters in a stimulus pair were the same in type (congruent) and .735 when the letters were different in type (incongruent). The effect of congruity was roughly the same for reports of visual as for reports of auditory stimuli and roughly the same whether attention was focused or divided. Table 4 shows the proportions of correct reports of visual and spoken letters in focused- and divided-attention conditions with congruent and incongruent stimuli in the other modality. By a three-way (congruency × attentional focusing × stimulus modality) repeated measures ANOVA,

Table 4
Probabilities (Ps) of Correct Reports of Visual and Auditory Stimuli, and Standard Errors, in Focused- and Divided-Attention Conditions With Congruent and Incongruent Stimuli in the Other Modality

Other Modality	Focused		Divided			
	P	SE	P	SE	M	SE
Visual						
Congruent	.797	.050	.776	.031	.786	.041
Incongruent	.737	.026	.731	.028	.734	.027
Auditory						
Congruent	.778	.040	.804	.044	.791	.043
Incongruent	.745	.023	.727	.032	.736	.028

the effect of congruity was highly significant [$F(1,7) = 110.8, p < .001$]. The main effects of attentional focusing [$F(1,7) = 1.15, p = .32$] and stimulus modality ($F < 1$) were not significant, and there were no significant interactions.

DISCUSSION

No Divided-Attention Decrement

In the present experiment, the proportion of correct reports of letters presented in a given modality (visual or auditory) was essentially the same when only the letters in the given modality should be reported (focused attention) as when letters in both modalities should be reported (divided attention). This finding supports the notion of separate pools of attentional processing resources for visual and auditory perception (see Duncan et al., 1997). Specifically, the finding suggests that there was little or no overlap between the processing resources that limited the accuracy of the report of visual letters and the resources that limited the accuracy of the report of spoken letters. For, suppose there were a certain resource such that both the accuracy of the report of visual letters and the accuracy of the report of spoken letters were limited by the scarcity of this resource (resource-limited in the sense of Norman & Bobrow, 1975). Then this resource should be allocated selectively to processing of visual letters in the visual focused-attention condition and to processing of spoken letters in the auditory focused-attention condition, which would imply that performance on the selected modalities in the focused-attention conditions would become superior to performance in the divided-attention condition. Thus, the fact that we found no divided-attention decrement suggests that accuracy of visual letter identification and accuracy of auditory letter identification were limited by separate types of processing resources.

Relation to previous findings. Previous experiments on perception of bimodal stimuli have yielded conflicting results. Some studies have supported the notion that visual and auditory attention relate to separate pools of processing resources (e.g., Allport, Antonis, & Reynolds, 1972; Duncan et al., 1997; Shiffrin & Grantham, 1974; Treisman & Davies, 1973). For example, extending and refining the study by Treisman and Davies, Duncan et al. (1997) presented concurrent visual and auditory streams of stimuli in which occasional target words should be identified. The identification responses were nonspeeded and typed in after the streams finished, using keys appropriately labeled for the targets ("cot," "cod," "nab," and "nap"). Identification of a target in a given modality (visual or auditory) produced a sustained reduction in the ability to identify a second target in the same modality, the period of interference lasting for more than 300 msec. In contrast, no such time-locked interference was found between targets in different modalities.

Other studies have shown that dividing attention between seeing and hearing may lead to decrements in performance (e.g., Arnell & Jolicœur, 1999; Jolicœur, 1999;

Lindsay, Taylor, & Forbes, 1968; Long, 1975; Massaro & Warner, 1977; also see Pashler, 1998, pp. 157–161). However, in all of these studies, the tasks performed by the participants seem to have required more attention-demanding control processes than did the tasks we used in the present experiment (see Atkinson & Shiffrin, 1968; Shiffrin & Grantham, 1974). In most of the studies, one or both of the component tasks (the visual task and the auditory task) seemed to require exacting comparisons of stimuli against standards held in short-term memory. For example, in the studies by Lindsay et al., Long, and Massaro and Warner, the auditory task was to make absolute judgments of tones that required fine discrimination (i.e., a difficult and unfamiliar type of stimulus identification). Thus, Lindsay et al. requested participants to assign different responses to tones of 496 and 500 Hz, respectively; Massaro and Warner used 20-msec tones of 800 and 880 Hz, and Long used tones of 1096 and 1116 Hz. In the studies by Lindsay et al. and Long, in which divided-attention decrements were relatively severe, the visual tasks were to make absolute judgments of spatial position (Lindsay et al., 1968) and luminance (Long, 1975) that also required fine discriminations. In the study by Massaro and Warner, in which divided-attention decrements were modest, the visual task was to discriminate between letters U and V.

Arnell and Jolicœur (1999) and Jolicœur (1999) used bimodal versions of the attentional blink paradigm (Raymond, Shapiro, & Arnell, 1992) and demonstrated crossmodal attentional blinks. They interpreted the crossmodal interference as evidence of central, amodal processing resources limiting performance in both the visual task and the auditory attention task. The participants' task in the attentional blink paradigm is much more complex than the task in the present experiment, so the difference in results is not surprising (see Arnell & Jolicœur, 1999, p. 647, for a discussion of the relationship of their experiment to that of Duncan et al., 1997).

Stochastic Independence

In the divided-attention condition, the probability that a letter in a given modality (visual or auditory) was correctly reported was independent of whether the letter in the other modality was correctly reported. Thus, the results provided strong evidence that correct report of the visual letter and correct report of the spoken letter were stochastically independent events. We are not aware of previous analyses for stochastic independence in bimodal perception, but the results fit in with demonstrations of stochastic independence in recent studies of extraction of features from pairs of simultaneously presented visual objects (see Bundesen, Kyllingsbæk, & Larsen, in press, for recent results and further references).

Modality Confusions

Participants tended to make modality confusions (i.e., report a letter presented in one modality as if it had been presented in the other modality). The strength of the tendency was measured by the probability that an erroneous

report of a letter in a given modality (visual or auditory) was identical in type to the letter presented in the other modality. By this measure, the tendency was virtually as strong when attention was focused on the given modality as when attention was divided between the two modalities. This finding suggests that the processing resources allocated to the irrelevant modality in the focused-attention condition were nearly identical to the processing resources allocated to the same modality (vision or hearing) in the divided-attention condition.

Whereas the correctness of a report of a letter in a given modality was independent of the correctness of the report of the letter in the other modality (stochastic independence), the type of errors made in response to a letter in a given modality depended on the letter presented in the other modality (modality confusions). This combination of results can be explained by assuming that encoding of the visual letter and encoding of the spoken letter were mutually independent, but information from the spoken letter was used in selecting (guessing) among visual letter types that were equally likely on the basis of the information extracted from the visual letter, and information from the visual letter was used in selecting (guessing) among spoken letter types that were equally likely on the basis of the information extracted from the spoken letter.

As a simple example, consider a so-called high-threshold model in which (1) the probability that, for instance, the visual letter is detected is independent of whether the spoken letter is detected; and (2) detected letters are reported; but (3) when, say, the spoken letter has been detected, but the visual letter fails to be detected, the most available (and, therefore, most likely) guess at the identity of the visual letter is the identity of the spoken letter. By the high-threshold property, the participant gains complete information about the identity of the visual letter on trials in which the visual letter is *detected* but no information about the identity of the visual letter when it fails to be detected. Hence, when detection fails, the probability of guessing the correct identity is $\frac{1}{8}$, regardless of the participant's strategy for guessing among the eight possible identities. By Assumption 3, participants should tend to report the identity of the spoken letter in response to the presentation of the visual letter (modality confusion). However, the tendency should be found only on trials in which detection of the visual letter failed, so the tendency to make modality confusions should not affect the probability of reporting the visual letter correctly.

The high-threshold model also explains why the proportion of trials with modality reversals was close to the chance level of 1/49 of those divided-attention trials in which the visual and the spoken letter were different in type and both letters were erroneously reported. The model implies no effects on responses of stimuli that fail to be detected. On trials in which both the visual and the spoken letter were erroneously reported, neither letter was detected, so neither letter had any effects on responses. Hence, on trials in which the visual and the spoken letter were different in type and both letters were erroneously

reported, the probability of a modality reversal was at chance level.

Crossmodal Congruity Effect

We discovered a crossmodal congruity effect: Performance was better when the visual and the spoken letter in a pair were the same than when they differed in type. The crossmodal congruity effect can be explained by the tendency to make modality confusions. When the letters in a stimulus pair are different in type (incongruous), modality confusions are scored as errors. But when the letters in a stimulus pair are identical in type (congruous), modality confusions (i.e., reports of letters presented in one modality as if they had been presented in the other modality) are scored as correct reports. In agreement with this explanation, the crossmodal congruity effect was roughly the same for reports of visual as for reports of auditory stimuli and roughly the same whether attention was focused or divided (like the tendency to make modality confusions).

Our crossmodal congruity effect is reminiscent of crossmodal auditory–visual priming effects (see, e.g., Driver & Baylis, 1993; Greenwald, 1970; Lewis, 1972) and crossmodal Stroop effects (Cowan, 1989; Cowan & Barron, 1987; Elliot, Cowan, & Valle-Inclan, 1998). However, Stroop (1935) effects and the Stroop-like interference effects seen in the flankers task of Eriksen and Eriksen (1974) are found in speeded responses but not in nonspeeded responses (see Egeth & Santee, 1981). Unlike these effects, our crossmodal congruity effect is found in nonspeeded responses.

CONCLUSION

In the reported experiment, we presented observers with pairs of concurrent stimuli consisting of a visual and a spoken letter. In the divided-attention condition, both letters, as well as their respective modalities, should be reported. In the visual focused-attention condition, only the visual letter should be reported, and in the auditory focused-attention condition, only the auditory letter. The stimuli were constructed so that the proportions of correct reports were nearly the same in the two focused-attention conditions (.74 and .75, respectively). With this calibration, the experiment yielded four main findings. First, the proportions of correct reports of visual and spoken letters were almost the same in the divided-attention condition (.74 for either modality) as in the focused-attention conditions (*no divided-attention decrement*). Second, in the divided-attention condition, the probability that a letter in a given modality (visual or auditory) was correctly reported was independent of whether the letter in the other modality was correctly reported (*stochastic independence*). Third, with a probability much higher than chance, the observers reported hearing the visual stimulus letter or seeing the spoken stimulus letter (*modality confusions*). Fourth, performance was substantially more accurate when the two letters in a stimulus pair were identical than when they differed in type (*crossmodal congruity effect*). The findings

are consistent with a high-threshold model in which visual and auditory letter encoding processing are entirely independent (separate processing resources) but guessing at the identity of a letter presented in one of the modalities is biased in favor of the identity of the letter presented in the other modality.

The absence of any reliable divided-attention decrement suggests that the amount of processing resources allocated to the relevant modality in the focused-attention condition was not noticeably greater than the amount of processing resources allocated to the same modality (vision or hearing) in the divided-attention condition. Our finding that the tendency to make modality confusions was virtually as strong with focused as with divided attention suggests that the amount of processing resources allocated to the irrelevant modality in the focused-attention condition was not noticeably smaller than the amount of processing resources allocated to the same modality (vision or hearing) in the divided-attention condition. The finding that the crossmodal congruity effect was roughly the same with focused as with divided attention supports the same interpretation. Taken together the findings of (1) no divided-attention decrement and (2) comparable modality confusions and comparable crossmodal congruity effects with focused and divided attention suggest that the processing resources allocated to a given modality were nearly the same in the three attention conditions (visual focused attention, auditory focused attention, and divided attention).

REFERENCES

ALLPORT, D. A., ANTONIS, B., & REYNOLDS, P. (1972). On the division of attention: A disproof of the single channel hypothesis. *Quarterly Journal of Experimental Psychology*, **24**, 225-235.

ARNELL, K. E., & JOLICŒUR, P. (1999). The attentional blink across stimulus modalities: Evidence for central processing limitations. *Journal of Experimental Psychology: Human Perception & Performance*, **25**, 630-648.

ATKINSON, R. C., & SHIFFRIN, R. M. (1968). Human memory: A proposed system and its control processes. In K. W. Spence & J. T. Spence (Eds.), *The psychology of learning and motivation* (Vol. 2, pp. 89-195). New York: Academic Press.

BUNDESEN, C. (1990). A theory of visual attention. *Psychological Review*, **97**, 523-547.

BUNDESEN, C., KYLLINGSBÆK, S., & LARSEN, A. (in press). Independent encoding of colors and shapes from two stimuli. *Psychonomic Bulletin & Review*.

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.

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

DUNCAN, J. (1980). The locus of interference in the perception of simultaneous stimuli. *Psychological Review*, **87**, 272-300.

DUNCAN, J., BUNDESEN, C., OLSON, A., HUMPHREYS, G., CHAVDA, S., & SHIBUYA, H. (1999). Systematic analysis of deficits in visual attention. *Journal of Experimental Psychology: General*, **128**, 450-478.

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

EGETH, H. E., & SANTEE, J. L. (1981). Conceptual and perceptual components of interletter inhibition. *Journal of Experimental Psychology: Human Perception & Performance*, **7**, 506-517.

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 the identification of a target letter in a nonsearch task. *Perception & Psychophysics*, **16**, 143-149.

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

JOLICŒUR, P. (1999). Restricted attentional capacity between sensory modalities. *Psychonomic Bulletin & Review*, **6**, 87-92.

LEWIS, J. L. (1972). Semantic processing with bisensory stimulation. *Journal of Experimental Psychology*, **96**, 455-457.

LINDSAY, P. H., TAYLOR, M. M, & FORBES, S. M. (1968). Attention and multidimensional discrimination. *Perception & Psychophysics*, **4**, 113-117.

LONG, J. (1975). Reduced efficiency and capacity limitations in multidimensional signal recognition. *Quarterly Journal of Experimental Psychology*, **27**, 599-614.

MASSARO, D. W., & WARNER, D. S. (1977). Dividing attention between auditory and visual perception. *Perception & Psychophysics*, **21**, 569-574.

MORAY, N. (1969). *Listening and attention*. Harmondsworth, U.K.: Penguin.

NORMAN, D. A., & BOBROW, D. G. (1975). On data-limited and resource-limited processes. *Cognitive Psychology*, **7**, 44-64.

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

RAYMOND, J. E., SHAPIRO, K. L., & ARNELL, K. M. (1992). Temporary suppression of visual processing in an RSVP task: An attentional blink? *Journal of Experimental Psychology: Human Perception & Performance*, **18**, 849-860.

SHIFFRIN, R. M., & GRANTHAM, D. W. (1974). Can attention be allocated to sensory modalities? *Perception & Psychophysics*, **15**, 460-474.

STROOP, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology*, **18**, 643-662.

TREISMAN, A. M., & DAVIES, A. (1973). Divided attention to ear and eye. In S. Kornblum (Ed.), *Attention and performance IV*. New York: Academic Press.

WINER, B. J. (1971) *Statistical principles in experimental design* (2nd ed.). New York: McGraw-Hill.

NOTE

1. All analyses of variance were done on arcsin transformed proportions of correct responses (x), according to the formula: $x' = 2 \sin^{-1}(\sqrt{x})$.