

Dividing attention between auditory and visual perception

DOMINIC W. MASSARO and DAVID S. WARNER

University of Wisconsin, Madison, Wisconsin 53706

To what extent is simultaneous visual and auditory perception subject to capacity limitations and attentional control? Two experiments addressed this question by asking observers to recognize test tones and test letters under selective and divided attention. In Experiment 1, both stimuli occurred on each trial, but subjects were cued in advance to process just one or both of the stimuli. In Experiment 2, subjects processed one stimulus and then the other or processed both stimuli simultaneously. Processing time was controlled using a backward recognition masking task. A significant, but small, attention effect was found in both experiments. The present positive results weaken the interpretation that previous attentional effects were due to the particular duration judgment task that was employed. The answer to the question addressed by the experiments appears to be that the degree of capacity limitations and attentional control during visual and auditory perception is small but significant.

One formulation of the concept of attention in an information processing framework proposes two defining criteria. First, processing capacity is limited, and second, some of this processing capacity can be allocated to enhance processing (Massaro, 1975, chap. 15). Either unlimited processing capacity or the inability to control allocation of a finite processing capacity would preclude attentional effects. It is also necessary, within the information processing model, to define the stage(s) of processing responsible for any observed attentional effects. The goal of this paper is to test the degree to which selective attention can enhance auditory (visual) perception relative to the situation in which attention must be divided between auditory and visual perception.

In the present paper, perception is defined as the primary recognition stage in an information processing model (Massaro, 1975, 1976). Primary recognition involves a resolution of the information held in preperceptual storage. Auditory recognition accomplishes the phenomenological outcome of perceiving a particular sound at a particular loudness and quality at some location in space. Visual recognition accomplishes an analogous visual experience. It is necessary to distinguish this recognition stage of processing from an earlier detection stage which is responsible for determining whether or not a stimulus is presented. Attention may play a different role in the detection of whether a stimulus is present than in the recognition of what stimulus is present (Massaro, 1975, chap. 15).

This research was supported, in part, by U. S. Public Health Service Grant MH-19399. Requests for reprints should be sent to Dominic W. Massaro, Department of Psychology, University of Wisconsin, Madison, Wisconsin 53706.

Massaro and Kahn (1973) asked observers to recognize just the timbre of a test tone or both the timbre and the duration of a visual character as short (100 msec) or long (160 msec). The processing time of the test tone was controlled by employing a backward recognition masking task. Subjects always identified the test tone before the duration of the visual character to eliminate any forgetting or response interference with the test tone recognition task in the divided attention condition. Performance improved with increases in the silent interval between the test and masking tones leveling off at 200 msec of processing time. Subjects also showed an overall 5.5% decrement in divided relative to selective attention. At a processing time of 100 msec, subjects produced a 10% decrement when overall performance averaged 77% correct. The smaller decrement averaged over the masking interval reflects the attenuation of the attention effect when overall performance of a given subject is near chance or close to perfect performance. It is possible to provide a rough measure of the size of the 10% attention effect by calculating the maximal attention effect that could be expected in a system in which only one of the two tasks could be performed on each trial. Assuming that the subjects attended to the auditory task on 50% of the divided attention trials, the predicted divided attention performance, $P(\text{DA})$, would be

$$P(\text{DA}) = .5[P(\text{SA})] + .5[P(\text{g})], \quad (1)$$

where $P(\text{SA})$ is performance on selective attention and $P(\text{g})$ is the probability of guessing the auditory stimulus correctly when attention was focused on the visual task. Performance was equal to 82% on selective attention trials and the guessing value of $P(\text{g})$ for

the two-alternative task is .5. Substituting these values in Equation 1 gives

$$P(\text{DA}) = .5(.82) + .5(.5) = .66. \quad (2)$$

Therefore, the maximal attention effect that could be expected is $.82 - .66$, or 16%. The 10% decrement shows that the attention effect is almost 2/3 of the maximal possible, arguing for a large attention effect at the primary auditory recognition stage of processing.

Shiffrin and his colleagues (Shiffrin, 1976; Shiffrin & Grantham, 1974) have argued that Massaro and Kahn's results are unique to a duration of judgment task. They claim that duration judgments require short-term memory and decision processes that cannot be divided without an attention decrement. Subjects deciding about the duration of the visual character would require dedicated processing time so that the relevant information about the tone percept would be lost before an appropriate decision could be made. In their view, the tone was resolved as well in the divided as in the selective attention situation, but there was not sufficient time to decide whether the percept was sharp or dull because of the duration task. Although duration judgments may be unique in processing capacity requirements, Shiffrin and Grantham (1974) did not have any direct evidence for this idea. The evidence presented by Massaro and Idson (1976), on the other hand, argues that duration judgments can be described in the same theoretical framework that has been developed for describing the processing of other stimulus attributes. Accordingly, we believe that the attention decrement in the Massaro and Kahn study reveals a true perceptual deficit and is not dependent upon the uniqueness of a duration judgment. To test between these two interpretations, we replicated the division of attention in auditory and visual perception in a task that does not require duration judgments. The Massaro and Kahn study was also extended to include judgments of visual perception in the selective attention situation. Processing time of both stimuli was controlled in a recognition masking task. The experiment therefore allows a direct assessment of attentional effects in both auditory and visual perception, with special concern for the stage of processing responsible for the result.

The present experiments control processing time by utilizing a backward recognition masking task. A short test stimulus is followed by a second masking stimulus after a variable blank interval. The second stimulus has been shown to interrupt perceptual processing of the first test stimulus (Massaro, 1975, chaps. 18 and 22). Accordingly, processing time can be held at intervals that are too short for switching of attention to be helpful. A second advantage of the task is that performance increases from near chance

performance to relatively good performance with increases in the processing time provided by the blank interval. This allows multiple tests of attentional effects across the different levels of performance. Floor and ceiling effects should attenuate any additional decrement at poor and good levels of performance, respectively. We predict that the attention decrement should interact with the overall level of performance with null or small attention effects at low and high levels and relatively large attention effects at the intermediate levels.

EXPERIMENT 1

Method

Subjects. Four students participated 1 h a day on 5 consecutive days. The subjects received extra credit in an introductory psychology course.

Stimuli. Observers were required to identify the test tone as high or low. The low tone was 800 Hz and the high tone was 880 Hz. The sine-wave test tone was presented for 20 msec. The masking tone was a 20-msec square-wave tone of 840 Hz. The test and masking tones were presented at a level of approximately 80 dB SPL. The test letter was to be identified as U or V. The test letter presentation lasted for a total of 11 msec and was turned on 9 msec after the onset of the test tone presentation, so that the offsets of the test stimuli were simultaneous. The physical differences between the alternative test stimuli had to be adjusted to maintain performance at 75% correct. In some sessions, the frequency of the high tone was changed to 890 Hz and the masking tone frequency was placed midway between the frequencies of the test tones. The test letter presentation usually involved a 4-msec test letter followed by 3 msec of the alternative letter followed by 4 msec of the test letter. In some sessions, the alternative letter filled 5 msec of the 11-msec test letter presentation. The visual mask consisted of alternating between 2-msec presentations of the letters U and V for a total of 20 msec. The test and masking letters subtended a visual angle of about 50' in height and 38' in width.

Procedure. The basic task was a recognition masking experiment. On each trial, a test tone and a test letter were simultaneously presented followed by a masking tone and masking letter, after a variable silent and blank interstimulus interval. Each trial began with a visual cue presentation informing the subject of the appropriate task on that trial. In the selective attention condition, the cue was either S or T, indicating that the subject should attend to and identify only the letter symbol or only the tone. The trial began with the cue B when the observer was required to identify *both* the letter stimulus and the tone test stimulus. The cue stimulus was presented for 500 msec followed by a 500-msec blank interval and then the test stimuli presentation. The masking stimuli followed the test stimuli presentation after an interstimulus interval of 0, 20, 40, 70, 140, 220, or 340 msec. No masking stimuli were presented on 1/8 of the trials. To compensate for the extra decision in the divided attention condition, the response interval was 2 sec in selective attention and 3 sec in divided attention. Visual feedback was presented on the stimulus or stimuli that had to be identified on each trial. The letters U or V indicated that the letters U or V had been presented, and the letters H or L indicated that the high or low tone had been presented. Two feedback letters were presented on divided attention trials and only one letter on selective attention trials. The feedback was presented for 500 msec followed by a 500-msec intertrial interval.

Observers made their responses by hitting one of four push-buttons on both selective and divided attention trials. The labels appropriate for selective attention trials were U, V, H, L, whereas

the labels appropriate for divided attention were UH, UL, VH, VL. Both sets of labels were always present. Observers had no trouble learning the stimulus-response mappings and responded quickly and appropriately in both the selective and divided attention conditions.

On the first day, the observers were familiarized with the procedure and apparatus. One hundred random presentations of the visual stimulus were given, followed by 100 presentations of the auditory stimulus. Feedback was given, but no responses were made. The observers then responded to 100 trials of the experiment with no masking stimuli present and then 100 trials of the experiment proper. On the 4 experimental days, two sessions of 250 trials each were presented each day. In the experiment proper, all 96 conditions (3 attention conditions by 2 test tones by 2 test letters by 8 masking conditions) were completely random and were programmed to occur equally often. The results were pooled over the 4 experimental days, eliminating the first 5 trials of each session. This gives about 31 observations for each subject at each condition of selective vs. divided attention by 2 test tones by 2 test letters by 8 masking conditions.

Apparatus. All experimental events were controlled by a PDP-8/L computer. Auditory stimuli were generated by a digitally controlled oscillator (Wavetek Model 155). The tones were presented binaurally over matched headphones (Grason-Stadler Model TDH-49). The visual stimuli, cues, and feedback were presented over light-emitting diodes (Monsanto Model MDA III). These displays are described in Nealis, Engelke, and Massaro (1973). The four observers were tested simultaneously in separate sound attenuated rooms.

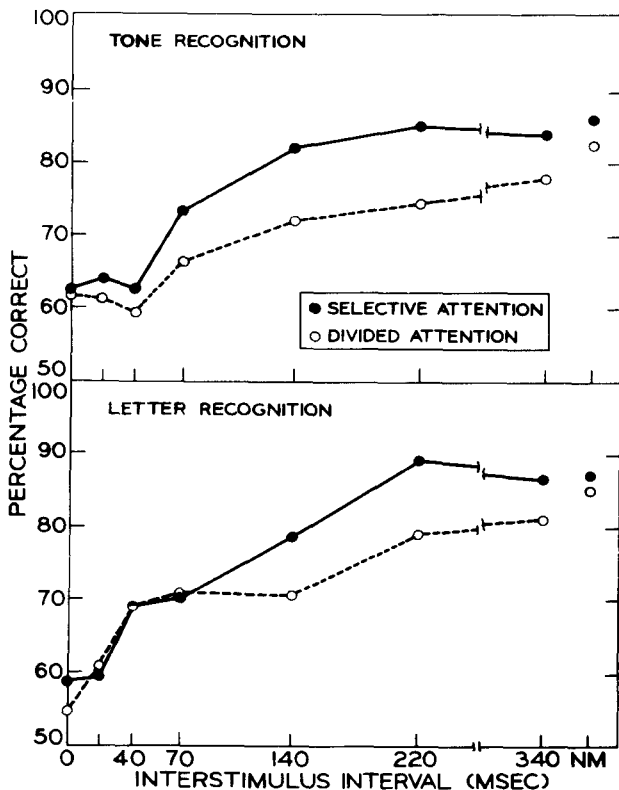


Figure 1. Percentages of correct recognitions of the test tone (top panel) and the test letter (bottom panel) as a function of the interstimulus interval between the test and masking stimuli under selective and divided attention (Experiment 1).

Results

An analysis of variance was carried out on the average percentage of correct responses with subjects, modality (visual or auditory), masking condition (7 ISIs and no mask), attention condition (selective vs. divided), and test stimulus (high tone/letter U, low tone/letter U, high tone/letter V, low tone/letter V).¹ Figure 1 shows that identification performance improved with increases in the interstimulus interval (ISI) for both the tone and letter test stimuli, $F(7,21) = 71.9$, $p < .001$. This result replicates earlier studies of backward recognition masking in both visual and auditory modalities (Massaro, 1970, 1973, 1975). Average performance improved about 23%, leveling off at an ISI of around 220 msec, roughly equal to performance under the no-mask condition.

Performance averaged about 4% better in the selective than in the divided-attention condition, $F(1,3) = 14.5$, $p < .05$. Although the interaction of ISI and attention condition was not significant, $F(1,11) = 1.19$, the difference between the attention conditions was related to the overall level of performance. The largest difference of 7.5% was observed at ISIs of 140 and 220 msec when performance averaged 79% correct. This result supports our interpretation of Massaro and Kahn's attention study and argues against the idea that the attentional effects found there were unique to processing stimulus duration. The attentional effect in duration judgments may have resulted from special demands on short-term memory and decision processes. This interpretation now seems unlikely given that similar attentional effects were observed in pitch and letter recognition.

EXPERIMENT 2

Experiment 2 replicates Experiment 1 while holding constant the number of decisions on each trial in the selective and divided attention conditions. Subjects in Experiment 1 had to indicate two choices in divided attention and just one in selective attention. One paradigm that eliminates response differences in selective and divided attention was developed by Shiffrin and Gardner (1972). Adopting this paradigm for the present task, we required the subjects to identify both visual and auditory stimuli on each trial, but now the stimuli are either presented simultaneously (divided attention) or successively (selective attention). In selective attention, the observer knows which stimulus will be presented first and has sufficient time (1,500 msec) between the two stimuli to switch attention to prepare for the second stimulus. The test letter and test tone are presented simultaneously in divided attention. The observer makes a

response identifying both stimuli on both selective and divided attention trials. Therefore, response requirements and short-term forgetting effects unique to identifying two stimuli instead of just one should not be responsible for any performance differences.

Method

Subjects. Twelve young adults from the University of Wisconsin community participated in the experiment 1 h a day for 5 consecutive days. The subjects were paid \$2 h.

Procedure. Stimuli were identical to those used in Experiment 1. On each trial, both an auditory and a visual stimulus were presented and the observers were instructed to identify both stimuli. Each trial began with a visual cue informing the observer of the experimental condition (SEQ or SIM). This cue had a duration of 500 msec, which was followed by a 500-msec blank interval prior to the test presentation. In the sequential condition, the test letter always preceded the test tone by 1500 msec. The simultaneous condition consisted of a phenomenologically simultaneous presentation of the test letter and test tone. The onset of the 11-msec test letter actually preceded the onset of the test tone by 16 msec. The masking stimuli were the same as in Experiment 1, and the ISIs define the blank interval between the offset of the visual (auditory) test stimulus and the onset of the visual (auditory) masking stimulus. The ISIs were 25, 40, 70, 120, 180, 250, and 340 msec. On 1/8 of the trials, no masking tone was presented. All 64 experimental conditions (2 test tones by 2 test letters by 2 attention conditions by 8 masking conditions) were presented randomly and were programmed to occur equally often. All other procedural details were the same as in Experiment 1. Pooling the results across the 4 experimental days gives about 122 observations for each subject at each of the 32 conditions of interest (2 modalities by 2 attention conditions by 8 masking conditions).

Results

There was a significant decrement in performance on the simultaneous relative to the sequential presentation of the signals, $F(1,11) = 13.0$, $p < .005$. Performance also improved with increases in the interstimulus interval for both the tone and letter recognition tasks, $F(7,77) = 34.2$, $p < .001$. Figure 2 shows that the attentional effect was relatively larger in the tone recognition task than in the letter recognition task, but this interaction was not significant, $F(1,11) = 2.7$, $p > .20$. The attention effect averaged 4.7% in tone recognition and 1.6% in letter recognition. Although the magnitude of the attention effect was largest at the intermediate levels of performance for tone recognition, the opposite was the case for letter recognition. These interactions were not statistically significant however.

The significant results support the hypothesis that the attention decrement is due to a limitation in the simultaneous perception of auditory and visual stimuli. The simultaneous-successive paradigm eliminates response differences in the two tasks, removing the possibility of a decrement at a response selection stage. The asymmetry in attention effects might be interpreted to mean that auditory resolution of a tone is more vulnerable to capacity limitations than is visual resolution of a letter character. One

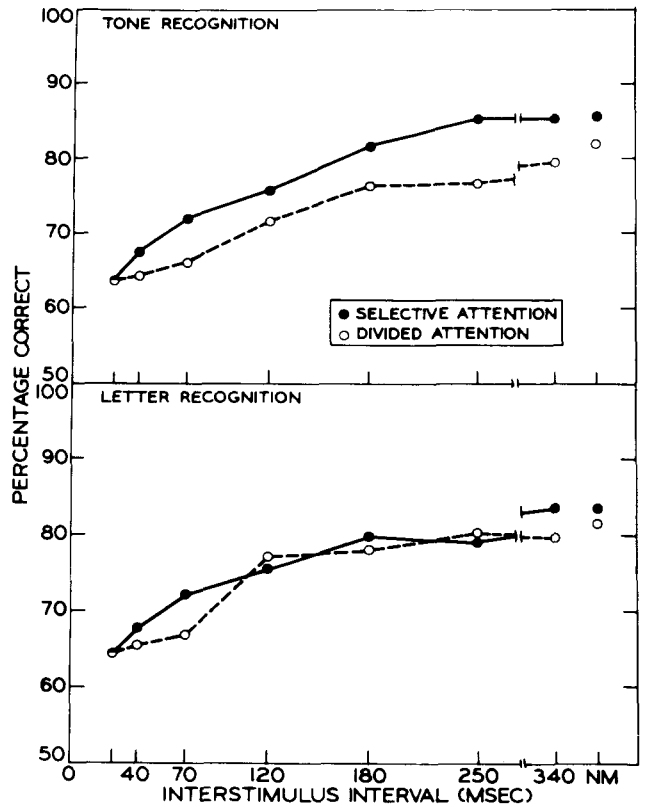


Figure 2. Percentage of correct recognitions of the test tone (top panel) and the test letter (bottom panel) as a function of the interstimulus interval between the test and masking stimuli under selective and divided attention (Experiment 2).

might have expected less of a decrement in audition than vision since the visual stimulus was always presented first in the successive presentation. Therefore, auditory perception was not truly selective in the successive presentation because the subjects had to remember the visual letter that was presented on that trial during the auditory processing task. In contrast, visual perception was truly selective in the successive presentation since no memory for the other dimension was necessary. It could be the case that subjects actually forgot the test letter during the tone recognition task in the successive presentation. This would have worked to offset any letter recognition advantage in the successive, relative to the simultaneous, presentation condition. Finally, the differences in the attention decrement in the two modalities may simply reflect how the limited capacity was allocated in divided attention. Subjects could have drawn on relatively more capacity for the visual than for the auditory tasks in divided attention, producing the observed asymmetry in attentional effects. Future research should be able to eliminate some of these alternative interpretations of this result.

GENERAL DISCUSSION

The results of Experiments 1 and 2 provide support for the role of limited-capacity and selective attention during visual and auditory perception. Tone recognition decreases when attention must be divided between tone and letter recognition. Similarly, selective recognition of a letter is better than it is when tone recognition is also required. Earlier positive attentional effects (Massaro & Kahn, 1973) in tone recognition as a function of simultaneous recognition of the duration of a visual stimulus do not appear to be unique to the duration estimation task.

It remains a possibility that the divided attention decrement is due to forgetting in short-term memory. Subjects had to make two choices in divided attention and may have forgotten some of the information that was encoded prior to the presentation of the mask. That is to say, although both signals were recognized and encoded equally in divided as in selective attention, the encoded information was more likely to be forgotten in divided than in selective attention. If short-term forgetting were responsible for the divided-attention decrement, however, it should always occur in this task. But Moore and Massaro (1973) did not find a similar divided-attention decrement in an equivalent task when subjects had to recognize both the loudness and timbre of a test tone relative to identifying only one of these dimensions. Measuring the attentional effect at an average performance level of 75% correct shows that the deficit in divided attention was about four times larger in the auditory-visual task than in the loudness-timbre task. This difference argues against a short-term forgetting explanation. (Massaro, 1976, presents a theoretical justification of why attentional effects are absent in multidimensional auditory perception although they are found in simultaneous auditory and visual perception.)

Shiffrin and Grantham (1974) found no selective attention effects in the detection of auditory, visual, and tactile stimuli. These results do not necessarily conflict with the present experiments, if one distinguishes between detection and recognition stages of processing and between capacity limitations and the ability to allocate a limited capacity (Massaro, 1975, chap. 15). With respect to detection and recognition, subjects may not be capable of improving detection of *whether or not* (detection) a tactile stimulus is present but may be able to enhance processing of *what* (recognition) stimulus is present. Massaro (1975, chap. 16) reviews evidence supporting the idea that the detection process may be limited in capacity but incapable of selectively allocating this limited capacity to a particular modality. Therefore, research has shown decrements in detection per-

formance when subjects must also perform a memory or imagination task (Broadbent & Gregory, 1963; Segal & Fusella, 1969, 1970), but no decrements in performance when detection must be shared within the process of detection itself. Subjects can detect an auditory signal at either of two frequencies and at either of two spatial locations as well as when attention is focused on just one frequency at one spatial location (Sorkin & Pohlmann, 1973). Shiffrin, Craig, and Cohen's (1973) observers were capable of monitoring three skin locations for detection of a tactile stimulus as well as monitoring just one. Shiffrin, Gardner, and Allmeyer (1973) and Shiffrin, McKay, and Shaffer (1976) found that observers could detect a dot in any of 4, 9, or 49 spatial locations as well as in the case of focused attention on just one spatial location. In summary, capacity limitations and attentional control may differ for detection and recognition, and either of these processes may have capacity limitations, even though its allocation is not always under attentional control.

REFERENCES

- BROADBENT, D., & GREGORY, M. Division of attention and the decision theory of signal detection. *Proceedings of the Royal Society*, 1963, **158**, 221-231.
- MASSARO, D. W. Preperceptual auditory images. *Journal of Experimental Psychology*, 1970, **85**, 411-417.
- MASSARO, D. W. Perception of letters, words, and nonwords. *Journal of Experimental Psychology*, 1973, **100**, 349-353.
- MASSARO, D. W. *Experimental psychology and information processing*. Chicago: Rand McNally, 1975.
- MASSARO, D. W. Auditory information processing. In W. K. Estes (Ed.), *Handbook of learning and cognitive processes* (Vol. 4) *Attention and memory*. Hillsdale, N.J.: Erlbaum, 1976.
- MASSARO, D. W., & INSON, W. L. Temporal course of perceived auditory duration. *Perception & Psychophysics*, 1976, **20**, 331-352.
- MASSARO, D. W., & KAHN, B. J. Effects of central processing on auditory recognition. *Journal of Experimental Psychology*, 1973, **97**, 51-58.
- MOORE, J. J., & MASSARO, D. W. Attention and processing capacity in auditory recognition. *Journal of Experimental Psychology*, 1973, **99**, 49-54.
- NEALIS, P. M., ENGELKE, R. M., & MASSARO, D. W. A description and evaluation of light-emitting diode displays for generation of visual stimuli. *Behavior Research Methods & Instrumentation*, 1973, **5**, 37-40.
- SEGAL, S. J., & FUSELLA, V. Effects of imaging and modes of stimulus onset on signal-to-noise ratio. *British Journal of Psychology*, 1969, **60**, 459-464.
- SEGAL, S. J., & FUSELLA, V. Influence of imaged pictures and sounds on detection of visual and auditory signals. *Journal of Experimental Psychology*, 1970, **83**, 458-464.
- SHIFFRIN, R. M. Capacity limitations in information processing, attention, and memory. In W. K. Estes (Ed.), *Handbook of learning and cognitive processes* (Vol. 4) *Attention and memory*. Hillsdale, N.J.: Erlbaum, 1976.
- SHIFFRIN, R. M., CRAIG, J. C., & COHEN, U. On the degree of attention and capacity limitations in tactile processing. *Perception & Psychophysics*, 1973, **13**, 328-336.

- SHIFFRIN, R. M., & GARDNER, G. T. Visual processing capacity and attentional control. *Journal of Experimental Psychology*, 1972, **93**, 72-82.
- SHIFFRIN, R. M., GARDNER, G. T., & ALLMEYER, D. H. On the degree of attention and capacity limitations in visual processing. *Perception & Psychophysics*, 1973, **14**, 231-236.
- SHIFFRIN, R. M., & GRANTHAM, D. W. Can attention be allocated to sensory modalities? *Perception & Psychophysics*, 1974, **15**, 460-474.
- SHIFFRIN, R. M., MCKAY, D. P., & SHAFFER, W. O. Attending to forty-nine spatial locations at once. *Journal of Experimental Psychology: Human Perception and Performance*, 1976, **2**, 14-22.
- SORKIN, R. D., & POHLMANN, L. D. Some models of observer be-

havior in two-channel auditory signal detection. *Perception & Psychophysics*, 1973, **14**, 101-109.

NOTE

1. Given that there were no significant response biases under any of the experimental conditions, equivalent conclusions would have been reached using d' as the dependent measure.

(Received for publication September 9, 1976;
revision accepted March 24, 1977.)