

The word superiority effect in brief visual displays: Elimination by vocalization*

JOSEPH J. MEZRICH

The University of Michigan, Ann Arbor, Michigan 48104

A variation on the experiments of Reicher (1968) and Wheeler (1970) was performed to expose the difference between the perceptual processing of words and that of single letters that leads to the word superiority effect in brief visual displays. The innovation in the present experiment is the use of a 1.5-sec interval between stimulus offset and choice onset during which Ss were required to vocalize the stimulus. A visual noise mask was presented during this interval. A control group was run with the conditions of choice delay and vocalization requirement absent, i.e., they were tested under the conditions used by Reicher and Wheeler. For the control group, recognition accuracy for words was 6.1% greater than for letters, while for the experimental group, recognition accuracy for letters was 5.6% greater than for words.

Before the experiments of Reicher (1968) and Wheeler (1970), it would have been unreasonable to expect that the recognizability of four-letter words should be better than that of single letters when both are tachistoscopically displayed for the same duration. It may still be unreasonable to expect this result, but it is a well-established fact.

The recognition measure used in their experiments was response accuracy in a two-alternative forced-choice situation. Single-letter displays were followed immediately by a noise mask and two single-letter choices. Word displays were followed by the same type of scene. For example, the stimulus WORD would be followed by the choices D and K, the S being required to determine which letter was in the displayed word. For all choices following word stimuli, either letter could be used in the critical letter position to make a word. All letter positions were used equally often. The word-letter accuracy difference was found to be about 8% by Reicher and about 10% by Wheeler.

The intuitions that are violated by this result seem to be based on the premise that similar perceptual strategies are used to recognize both types of stimuli. Under the assumption of similar strategies, at least as much cognitive work is required to recognize a word as a single letter, or equivalently, at least as much time is required to recognize a word as a letter. If this is so, then it is difficult to understand why recognition accuracy should be better for words exposed for the same duration as single letters.

When pilot Ss, in a replication of the Reicher and Wheeler experiments, were questioned about their

response strategies, striking differences were reported between the way a decision was made in choosing between the alternatives following a word display and the alternatives following a single-letter display. For choices following word stimuli, there was never an attempt at a physical, or feature, match to determine which test letter was the correct one. The representation in memory of the word perceived, which was not necessarily the word displayed, was inspected for presence of one of the alternative choice letters. Decisions were difficult only when the word remembered was not the word displayed, or when no word was remembered at all. In the latter case, pure guessing occurred, and in the former case, the choice was based on which letter would most likely fit into a word similar in spelling to the word recalled. That the alternative choices were ambiguous, i.e., either letter could have been used in the critical letter location to make a word, was rarely a source of confusion. In fact, when Ss reported being aware of this ambiguity, it was for only a few of the words presented. According to these reports, word stimuli were either perceived as complete words or not at all. For example, the briefly displayed stimulus READ might be remembered as READ, ROAD, or, if sufficiently unclear when viewed, perhaps as BEAD or BOND. It was rarely, if ever, the case that RE-D would be perceived and remembered; hence, a decision between the choices E and A following the stimulus READ might be difficult but not ambiguous in the sense described above. This is a classical effect first noted by Pillsbury (1897).

In contrast, these Ss claimed that, in responding to alternatives following a single-letter stimulus, attempts were made at matching critical physical features in the choice letters with those that were recognized as being present in the stimulus. For example, the presentation of the stimulus Z might evoke a search for a diagonal line in one of the alternatives. Thus, a choice between O and Z following the stimulus Z would be relatively easy.

*This study was supported by Public Health Service Grant GM-01231-08 to the University of Michigan. The encouragement and important suggestions of James R. Moeller and the valuable advice and assistance of Professors Robert G. Pachella and J. E. Keith Smith are gratefully acknowledged. In particular, the decision to require Ss to vocalize stimuli evolved from discussions with J. R. Moeller.

whereas a choice between E and F following the stimulus F would be relatively difficult. This experience suggests that the representation in memory of a briefly presented single letter is an aggregation of critical features not quite coalesced into the unitary image that we ordinarily associate with letters.

Sternberg (1969) has indicated that the memorial representation of a briefly presented digit may be a spatial pattern rather than a verbal code. The fact that a verbal code is available for a digit is not sufficient reason to suppose that in fact a verbal code is used. The experience of our pilot Ss suggests that, for briefly presented letters as well, verbal coding is not achieved. In contrast, the representation of words is primarily verbal.

If the word superiority effect is due to differences in processing modes, letters being processed in a primarily spatial manner and words being processed in a primarily verbal manner, then it follows that if Ss are forced to encode letters verbally, there should be a considerable relative enhancement in the recognition accuracy of these items. If both word and single-letter stimuli are encoded verbally, the recognition accuracy for words should not be higher than for letters.

It was hoped that by having to verbalize the single-letter stimuli, Ss would use the same type of perceptual processing used for words. It might be argued that verbalization occurs after perception, so the type of processing employed should not be affected. It was felt, however, that knowing that one has to immediately vocalize a stimulus primes the perceptual machinery for a type of processing that is not primarily spatial.

In order to insure that the manipulation affecting performance is verbalization, the experiment was also run under a condition of no forced verbalization, i.e., a replication of the Reicher and Wheeler experiments. We will refer to the condition with forced verbalization as the vocalized condition and to the control condition as the nonvocalized condition.

METHOD

Apparatus and Materials

The experiment was programmed on, and run by, a Digital Equipment Corporation LINC-8 computer. Ss viewed the displays, while seated in a room adjacent to the computer, on a Tektronix Type RM561A CRT equipped with a P7 double layer screen. Each painting of this screen yields a blue short time constant fluorescence and a yellow-green long time constant phosphorescence. According to manufacturer's specifications, the blue screen luminosity decays to 10% of luminosity at painting offset within 600 msec. A blue filter was used to block the yellow-green display, but no careful check was made of the spectral characteristics of the filter to see if it had good enough bandpass properties to completely block yellow-green while transmitting blue. As well, no precautions were taken to insure the absence of small fluctuations in screen luminosity not readily observable by the E. The bandpass properties of the blue filter appeared adequate, and the display intensity appeared stable. The word and single-letter stimuli were randomly interspersed. Since only differential, rather than absolute, recognition

accuracy for these two types of stimuli was of concern, the effects of small luminosity fluctuations and the effects of the blue filter having a small degree of transparency in the yellow-green portion of the spectrum were unimportant.

The painting cycle for all stimuli was 10 msec. Each letter took approximately 1.5 msec to be painted; consequently, there was an interval of 8.5 msec between paintings for single-letter stimuli and an interval of 4 msec between paintings for word stimuli. Each letter was constructed from a grid 15 dots wide and 21 dots high; the grid dimensions were .41 x .23 cm². All letters were capitalized. The width of each word was no larger than 1.5 cm. The S's forehead rested against a hood which kept his eyes at least 40 cm from the screen.

The mask used at stimulus offset was dynamic, as opposed to the static masks used by Reicher and Wheeler. Lines of random length, orientation, and location were painted continuously within a 2.5 x 4.0 cm² area that covered the stimulus location. The effect was somewhat like looking through a window at small sticks flying about. In the vocalized condition, there was a 1.5-sec interval between the offset of the stimulus and the onset of the two letter choices, during which only the mask appeared on the screen. At the end of this interval, the two letter choices appeared. The two choices were centered at 3 cm above and below and 2.4 cm to the right of the center of the stimulus location. The mask and the choices were displayed until a response was made. The poststimulus display for the nonvocalized condition was identical to that for the vocalized condition except for the absence of the 1.5-sec delay of choice onset. Further discussion of the use of a choice delay for the vocalized condition appears below.

A possible point of concern is that the type of mask employed here might interfere differently with the cognitive transformations following a briefly displayed letter than with the processes following a briefly displayed word. It should be clear, however, that this possibility is controlled for by testing both the vocalized and nonvocalized conditions with the same mask. If the dynamic mask interferes more with the processing of words than with letters, then we should find it difficult to replicate the results of Reicher and Wheeler in our nonvocalized condition. On the other hand, if the mask affects letters more than words, we should find it difficult to demonstrate a relative enhancement in letter recognition accuracy in the vocalized condition. Thus, verification of our prediction about the effects of verbalization, together with replication of the word superiority effect, does not demonstrate that there is no interaction between mask and type of stimulus, but it does indicate that such a differential interference is unimportant compared with the effects of encoding differences.

When the response was made, the mask and choices were replaced on the screen by a rating scale; the word GUESS appeared at the left of the scale, and the word SURE appeared at the right. The scale was broken up by 11 hash marks, numbered 0 through 10, to aid the S in assigning a rating. By rotating a knob next to the scope screen, he caused the pointed end of a vertically oriented arrow to glide along the scale. When the S felt that the arrow pointed to a position on the scale, not necessarily on one of the hash marks, that represented how sure he was of the response he had made for the stimulus on that trial, he pressed a button that removed the scale from the screen, replacing it by the fixation scene.

The fixation scene was a plus sign flanked on both sides by three dots. This display made up a 2-cm-long line which, if kept on the screen when the stimulus appeared, would directly underline the stimulus. During pilot studies, Ss pointed out that a centered fixation scene interfered to a certain extent with the stimuli, so the fixation scene just described was adopted.

The experimental room was kept dark except for overhead light from a 25-W bulb. Diffuse light from a 40-W bulb illuminated the scope screen to provide the appropriate contrast. A neutral density filter obliquely placed in the hood described above eliminated glare from the screen.

An intercom speaker was placed on a table next to the oscilloscope so that the E, in the room with the computer, could be sure that the S was vocalizing the stimuli for the vocalized condition. The vocalization task is more fully described below.

There were, of course, two types of stimuli, single letters and four-letter words. Words were used that could have their meaning altered by changing a critical letter. For example, a typical stimulus was EDIT followed by choices D and X. Each critical letter position was used equally often. All single-letter stimuli, and their respective incorrect choices, were randomly selected from the alphabet.

A potential problem is introduced by this method of selecting the distractors for a particular stimulus. The distractors for letter stimuli are randomly selected, whereas the distractors for words are not. Consequently, the choice situations for the two types of stimuli are not identical. This is, however, not a major concern in this study. The same stimuli and distractors were used in both the vocalized and the nonvocalized conditions, with different Ss used in the different conditions. If the choice situation for one stimulus type presents an easier decision, then we should either find it difficult to verify our prediction about the effect of verbalization or find it difficult to replicate the word superiority effect. Our technique, therefore, controls for this potential difficulty.

Procedure

In the vocalized condition of the experiment, Ss were required to vocalize all stimuli before the onset of the two alternative choices. A 1.5-sec delay between the offset of the stimulus and the onset of the choices was used to insure sufficient time to vocalize before attention was diverted by the choice display. A noise mask was displayed immediately at the stimulus offset and stayed on until a response was made. Even if the S had no idea what the stimulus was, he was required to utter something before the choices were displayed. Ss were encouraged to use their verbalization in deciding upon their responses. For the nonvocalized condition, the experimental setup and procedure were identical to that of the vocalized condition, except that no delay occurred between stimulus offset and choice onset and no vocalization was requested.

Wheeler found that a choice delay of 1 or 2 sec did not eliminate the word superiority effect, though a small, nonlinear decrease in the magnitude of the effect as a function of choice delay was found. We would, therefore, not expect a 1.5-sec choice delay without vocalization to yield an elimination of the word superiority effect. Under a condition of choice delay with no vocalization requirement, it would be impossible to tell if Ss were covertly verbalizing some or all of the stimuli. Interpretation of results where it is not known whether or not Ss maintain a uniform strategy is difficult at best. Because of the difficulty of interpreting the results of a condition of 1.5-sec choice delay with no vocalization requirement, and because Wheeler's results indicate that this condition maintains the word superiority effect, we decided not to include this condition as part of our experiment.

An experimental session, which lasted about 90 min, was broken up into three phases: a warm-up phase, a timing phase, and a data collection phase. The first 14 stimuli of the warm-up phase, half words and half letters, were exposed for 500 msec, with the E in the room with the S to explain the task. After these "slow" stimuli were shown, the E left the S and monitored the experiment from the computer room. The S then proceeded to run through 96 stimuli, 48 words and 48 letters, randomly ordered, at 60 msec exposure duration for the remainder of the warm-up phase. No data were collected during this phase.

At the end of the warm-up phase, there was a 2-min break. The timing phase, which followed, was used to determine an appropriate stimulus exposure duration for the S being tested to insure an overall recognition accuracy level in the neighborhood of 75%. Only single-letter stimuli were used during this phase.

Each sequence of two responses was examined, on line, by the computer. If both responses were hits, the exposure duration for the next pair of stimuli was decreased by 10 msec. If the two responses contained a miss, the exposure duration for the next pair was increased by 10 msec. If the first response of a pair was a miss, the exposure duration was increased without waiting for the outcome of the second response of the pair, since two consecutive hits were impossible for that pair. This hunting procedure continued until the exposure duration hopped back and forth between two 10-msec separated exposure durations four times in a row. The larger of the two exposure durations was taken to be the exposure duration for all stimuli for that S in the final, data collection phase of the experiment.

This scheme is based on a general Markov design procedure described by Smith (1961). The rationale behind the particular procedure employed here is simple. If p_i is the hit probability for exposure duration t_i , then our hunting procedure finds two exposure durations, t_0 and $t_0 + 10$ msec, such that about midway between them the probability of getting two consecutive hits equals the probability of not getting two consecutive hits; i.e., $p_1^2 = 1 - p_1^2$, or $p_1 = .707$. So an exposure duration of $t_0 + 5$ msec should yield about 70% hit accuracy. Due to constraints imposed by our display, we had to work with 10-msec increments, so $t_0 + 10$ msec was taken as the appropriate stimulus exposure duration for the S being tested. The main reason for illuminating the scope screen to decrease contrast, as mentioned above, was to insure a slowly rising psychometric function so that the hit accuracy at $t_0 + 10$ msec should not be too much greater than 70%.

One of the main advantages of this technique is that it consumes very little time. The initial exposure duration for all Ss in this phase was set at 60 msec, and the criterion for stopping typically was reached within less than 5 min.

At the end of the timing phase, the S was required to take a 5-min break to prevent fatigue. During this break, he was informed of a graduated payoff for accuracy to be used in the last phase of the experiment. To counter fatigue or boredom, each S was offered, in addition to the base pay of \$1.50/h, a bonus of \$1.00 for achieving an overall accuracy level on the last phase of the experiment of between 80% and 90% and a bonus of \$2.00 for overall accuracy exceeding 90%. Feedback about performance was given to the S after this last phase of the experiment, which collected responses to 96 words and 96 letters randomly ordered, was completed.

The confidence rating scale appeared on the screen only during the data collection phase of the experiment. The pace of the experiment was entirely up to the S being tested. Only the two breaks mentioned above were mandatory, but additional breaks were encouraged to prevent fatigue.

An a priori criterion of 60% overall accuracy for the data collection phase of the experiment was set for the inclusion of a S's data in the experiment. Pilot studies indicated that Ss falling below this accuracy level either became very fatigued or very bored sometime during the last phase of the experiment and found it very difficult to perform the vocalization task as required. For the sake of uniformity, the same criterion was used for both the vocalized and nonvocalized conditions.

Subjects

Eighteen right-handed female Ss were recruited from the paid S pool of the Human Performance Center of the University of Michigan. The pay rate was \$1.50/h plus the graduated bonus described above. All Ss were run individually. Inability to attain the 60% overall accuracy criterion described in the preceding paragraph forced the exclusion of 2 Ss from the vocalized condition and 2 Ss from the nonvocalized condition. Six of the remaining 13 Ss performed in the vocalized condition and 7 Ss performed in the nonvocalized condition.

Table 1
Results for Nonvocalized and Vocalized Condition

	Percent Correct		Confidence		Average Exposure Duration (Msec)
	Letters	Words	Letters	Words	
Nonvocalized	77.1	83.2	0.54	0.65	47
Vocalized	86.8	81.2	0.66	0.64	40

RESULTS AND DISCUSSION

The averaged results of the replication of the Reicher and Wheeler experiments, i.e., our nonvocalized condition, are given in the top row of Table 1. The accuracy level for words was found to be 6.1% higher than for letters ($p < .005$).¹ The confidence ratings in Table 1 have been normalized to vary between 0 and 1, a higher number indicating a higher level of confidence.

The effect found by Reicher and Wheeler is confirmed. The word-letter accuracy difference reported here is smaller than the values of 8% and 10% reported by Reicher and Wheeler, respectively. However, as Wheeler has pointed out, the magnitude of this effect varies with the overall accuracy level of the S, the maximum occurring at about 75% accuracy. Our use of a noncentered fixation scene could also have diminished the magnitude of the effect somewhat. Also, the imposed criterion of 60% overall accuracy for a S's data to be included in the experiment forced us to exclude Ss that would have put the magnitude of the word superiority effect we obtained in the range of the results of Reicher and Wheeler. What is interesting about this effect is not the degree to which it occurs, but the fact that it occurs at all. Since the effect was replicated in the nonvocalized condition, our procedures were deemed suitable for the vocalized condition.

The average results for the Ss in the vocalized condition of the experiment are given in the bottom row of Table 1. We find that the recognition accuracy for single-letter stimuli is 5.6% greater than for word stimuli

($p < .05$). There is no question that vocalization enhances the accuracy level of single-letter stimuli so that words are no longer more rapidly recognized than single letters. In fact, one is justified in claiming that vocalization completely reverses the word superiority effect, yielding a letter superiority effect.

REFERENCES

- Pillsbury, W. B. A study in apperception. *American Journal of Psychology*, 1897, 8, 315-393.
- Reicher, G. M. Perceptual recognition as a function of meaningfulness of stimulus material. Technical Report No. 7, The University of Michigan, Human Performance Center, 1968.
- Snedecor, G. W., & Cochran, W. G. *Statistical methods*. Ames, Iowa: Iowa State University Press, 1967. P. 255.
- Smith, J. E. K. Stimulus programming in psychophysics. *Psychometrika*, 1961, 26, 27-33.
- Sternberg, S. Memory-scanning: Mental processes revealed by reaction-time experiments. *American Scientist*, 1969, 57, 421-457.
- Wheeler, D. D. Processes in word recognition. *Cognitive Psychology*, 1970, 1, 59-85.

NOTE

1. All attained significance levels reported in this paper were computed by the method of combining sets of 2 by 2 contingency tables, as described in Snedecor and Cochran (1967).

(Received for publication June 19, 1972;
 revision received August 16, 1972.)