INFORMATION AND CONTROL 3, 191–196 (1960)

A Note on Human Recognition of Hand-Printed Characters

ULRIC NEISSER AND PAUL WEENE

*Department of Psychology, Brandeis University, Waltham, Massachusetts**

Nine human observers were given the task of identifying isolated hand-printed characters. Their individual accuracies ranged from 94.9% to 96.5%, and even their pooled best guess was right only 96.8% of the time. These figures can serve as standards for the accuracy of mechanical devices for letter-recognition.

A good deal of research is currently being devoted to the development of machines that can read. The mechanical reading of *hand-printed* material presents a particularly interesting problem, in view of the variability exhibited by the form of any particular letter from time to time and from one person to the next.

What degree of accuracy can be expected from a recognizing device? Obviously less than 100%; not every hand-printed character is "legible." Some degree of illegibility or ambiguity can usually be overcome with the aid of context, but a device that can utilize context will necessarily be more complex than one working only with isolated characters. The appropriate performance standard for a simple character-recognizer would seem to be the accuracy with which human observers can perform the same task. The following experiment was carried out to obtain such a standard.

STIMULUS MATERIAL

Most of the hand-printed characters used in this study were selected from a pool originally obtained by Herbert Sherman of Lincoln Laboratory, who was kind enough to make the material available to us. He asked

* The present research was performed while the authors were staff members of the Lincoln Laboratory, a center for research operated by M.I.T. with the joint support of the U.S. Army, Navy, and Air Force under Air Force Contract AF 19(604)-5200.

passers-by at the main gate of the Laboratory to print their complete names and addresses on a specially prepared paper tape. A series of ¼ in. squares, spaced ⅛ in. apart, appeared on the tape; the participants were asked to put one letter in each square, using block capitals. About 200 persons cooperated, producing a pool of more than 5000 individual characters.

For the present study, every effort was made to select characters from this pool without bias either for, or against, legibility. Twenty exemplars of each of the letters from A to Z and each of the numerals from Ø to 9 were chosen. The following features of the sampling process are worth mentioning:

1. No sample (aggregate of letters written by a single participant) was used if it was in other than block capitals; if the characters were written sideways; if the printing was extremely faint; or, if the material written was other than a name or address (except in a few cases where not enough exemplars were available, especially as in item No. 5 below).

2. No more than four exemplars were taken from any one sample, and no two of those represented the same character.

3. No exemplar was used if any substantial proportion of it extended beyond the square; or if context did not permit a precise determination of the letter or numeral intended (except that two slightly ambiguous numerals were included).

4. The first three exemplars chosen from each sample were selected almost independently of the characters they represented, by a system based on their ordinal position in their sample (provision was made for rejecting characters of which 20 exemplars had already been chosen). The entire pool was then scanned a second time to fill the quota of 20 exemplars per character.

5. The quota for certain characters could not be filled, the entire pool not containing 20 exemplars. These were Q, X, Z, 2, 3, 4, 5, 6, 7, 8, 9, and Ø (zero). To fill these deficiencies, members of the Laboratory were asked to block-print either "QZ9Ø," "X234," or "5678" on similar tape, until enough additional material had been collected.

After the selection process, each exemplar chosen was assigned a number between 1 and 720, by a process which was essentially independent of both the character it represented and the sample in which it originated. These identifying numbers were written directly on the tape, immediately below the exemplar in question.

Each of the 720 exemplars was photographed on 16 mm film. A

4 × 6 in. print of each was prepared, on which the letter-square was enlarged to 2¾ × 2¾ in. The identifying number also appeared in the photograph. Specimen exemplars are shown in Fig. 1.

Upon inspection, some of the photographs proved unusable because of insufficient contrast between a lightly printed character and the rather "grainy" background. The two faintest exemplars of each character were therefore eliminated, leaving 18 that showed good contrast (through an oversight, only 17 exemplars of U were included.)

PROCEDURE

The subjects were nine college students or graduates, aware of the purpose of the experiment. (There were actually ten, but one was myopic and his performance was significantly poorer than that of the others; his data were not used.) Answer sheets were provided on which many iden-

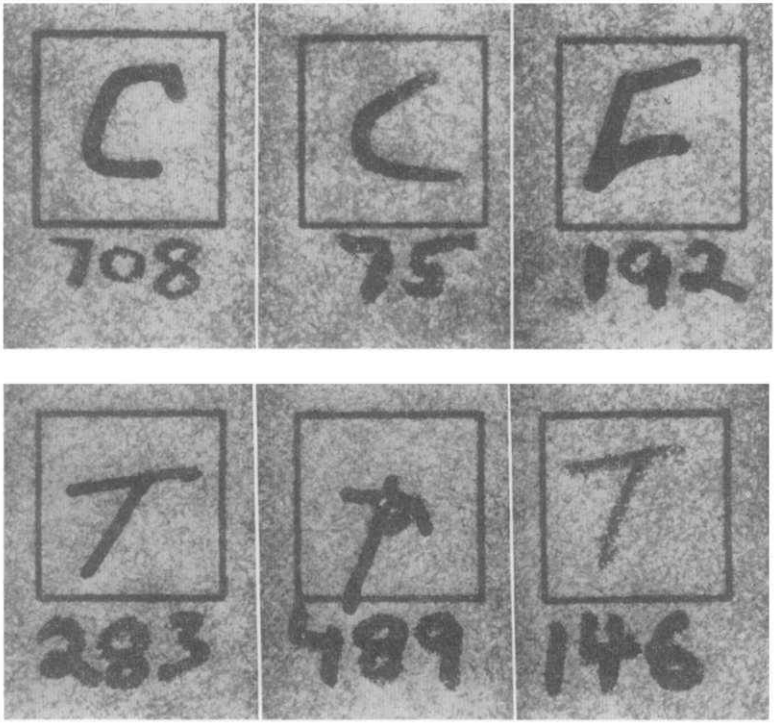

FIG. 1

tical rows of 36 characters each were typed (in alphanumerical order). The subjects sat two or three yards away from the experimenter, who showed the photographs one at a time, in the random order defined by the identifying numbers. As each was shown, the subjects were to cross out that character (in the appropriate row) which represented their best guess at the one in the photograph. The procedure took about 2½ hr, all the subjects being run simultaneously.

RESULTS

A preliminary analysis indicated that the subjects could not tell O from Ø at all, and could distinguish I from 1 only poorly. A majority responded "O" to 17 of the O-stimuli; but a majority also responded

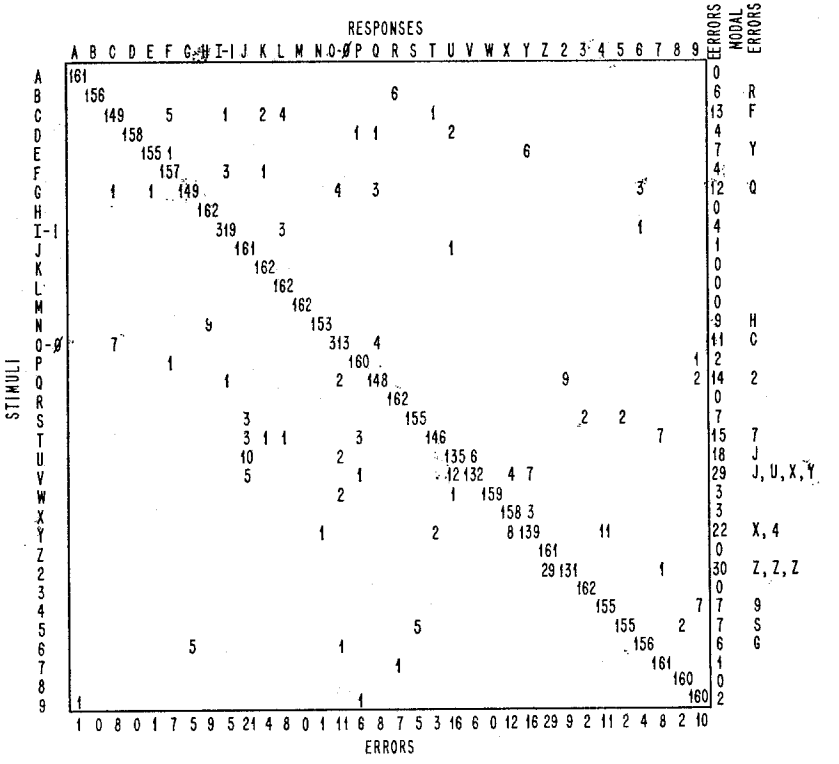

FIG. 2

"O" to 17 of the Ø-stimuli. On the other hand, a majority responded "I" to 12 of the I-stimuli and to only 3 of the 1-stimuli.

In what follows, confusions between I and 1, or between O and Ø, are not counted as errors. I-1 is treated as a single character with two permissible names appearing 36 times; O-Ø is treated analogously.

The results appear in Fig. 2 in the form of a confusion matrix. Entries represent the frequency with which each response was made to each stimulus; correct responses appear on the diagonal. The total number of errors made on each stimulus character appears in the right margin; the frequency with which each character appeared as an erroneous response appears in the lower margin. Since each character was presented 18 times to 9 subjects (except for U,I-1, and O-Ø, as previously noted), there are ordinarily 162 entries in each row. Slightly smaller totals sometimes appear because of individual failures to make any response at all to certain exemplars. (These failures represent inattentiveness rather than indecision; they did not occur on ambiguous exemplars.)

Note that entries in the matrix can represent several exemplars. One exemplar of S was called "J" by two subjects; a different exemplar of S was called "J" by one subject; these together make up the "3" in the matrix. There were 21 exemplars altogether on which more subjects made a particular erroneous response than the correct one, i.e., the best guess of the subjects taken as a group was wrong. The erroneous characters selected on these occasions are listed opposite the actual stimuli in the column labeled "modal errors."

TABLE I

ERROR RATES FOR INDIVIDUAL SUBJECTS

Subject	Errors	Responses	Percent-Error
1	23	647	3.6
2	22	645	3.5
3	28	643	4.4
4	33	644	5.1
5	29	646	4.5
6	27	647	4.2
7	25	647	3.9
8	27	647	4.2
9	23	645	3.6
		Mean of nine subjects	4.1%

It appears that the characters A, H, K, M, R, Z, 3, and 8 were recognized without error, while the most difficult characters were T, U, V, Y, and 2. Even for the latter, of course, correct recognitions predominated.

Specimens of easy and difficult stimulus exemplars are shown in Fig. 1. Of the three C's in the upper row, the first was correctly identified by all nine subjects, the second was called "C" by seven and "L" by two, and the third was a modal error, being called "F" by five subjects, "C" by two, "K" by 1, and "I" by one. The first of the three T's in the lower row elicited no errors, the second was called "T" by six subjects but "P" by three, and the third, a modal error, was called "7" by six subjects and "T" by three.

The best measures of over-all accuracy are the error rates for the individual subjects, presented in Table I. They average 4.1 %. A particularly meaningful minimum rate is the proportion of modal errors, 3.2 %. In a sense, this represents the degree of sheer illegibility in these exemplars of handprinted characters.