Notes and Comment

Letters are visual stimuli: A comment on the use of confusion matrices

GROVER C. GILMORE

*Case Western Reserve University
Cleveland, Ohio*

Letters, as a class of visual stimuli, have been used extensively in perception research. The question may focus specifically on how letters are recognized (e.g., Gibson, 1969; Townsend, 1971). Or the letters may be a vehicle for examining a higher process such as memory (e.g., Sperling, 1960). In spite of their broad usage, these stimuli remain as a source of uncontrolled variance in most studies. The purpose of this note is to call attention to the value of different letter confusion matrices and to defend one against an unwarranted criticism.

The suggestion by Gilmore, Hersh, Caramazza, and Griffin (1979) that greater specificity, and hence control, of letter stimuli was possible, virtually has gone unheeded. There are two consequences of this failure of control. The first is that the selection of letters for use as stimuli is arbitrary and hence is a potential source of error; letter confusions may be greater in one condition and thus yield unreliable data. The second, and most important, consequence is that a potentially important source of information to guide experimental design and interpretation has been overlooked. Gilmore (1980), for example, did use a multidimensional letter similarity metric and was able to demonstrate a facilitation in letter identification specific to a critical feature dimension. Without the specificity of the metric, he would have reached a different conclusion.

The root of the problem lies with the implicit assumption by investigators that letters are neutral stimuli that may be treated equivalently. However, each letter has a history and a physical structure that will influence its processing. The division of letters into vowel and consonant classes and the differences in frequency of occurrence in the language are but two factors that may influence response selection and, hence, accuracy in an identification task. Also, each letter is defined by a structure that varies in degree of uniqueness within a particular type font. Recent investigators have demonstrated that the psychological similarity among letters is highly related to the spatial frequency and phase content of the letters (Gervais, Harvey, & Roberts, 1984; Royer, Gruhn, Gilmore, & Thomas, 1985). This evidence is of great im-

portance, because it links the literature on letter and pattern processing to the neurophysiological and psychophysical evidence that the visual system performs a form of Fourier analysis.

The relevance of the spectral analysis of letters serves to illustrate the necessity of emphasizing the physical characteristics of letters over their nominal classification when comparing confusion matrices. It is reasonable that if two letters, or for that matter two people, have the same name and gross structural similarity they may still be judged differently. The identification of the letter or the face will depend on its physical characteristics and their degree of overlap with other members of the relevant set. Yet, the validity of confusion matrices has been questioned because of their low intercorrelations. Van der Heijden, Malhas, and van den Roovaart (1984) compared an uppercase confusion matrix they generated with ones produced by Townsend (1971) and Gilmore et al. (1979) and reported a common variance of 25% or less between the diagonals of the matrices.

It may be argued that comparison of the diagonals alone is inadequate for a valid comparison of the matrices. In Table 1, the correlations for the off-diagonal cells of the matrices, as well as correlations for the similarity and response bias values estimated with the Luce (1963) choice model, are presented.¹ It is clear that the off-diagonal cells are more related than the diagonals, but the relationships are still moderate. Conversion of the data to similarity estimates with the choice model strengthens the correlations but still accounts for only 25% to 50% of the variance. The most telling evidence for the hypothesis that the letter sets elicit different judgment behavior can be found in the low-to-moderate correlations between the bias values generated by the choice model. If subjects were predisposed in their choice of letters when guessing, then logically the response bias vectors should be highly correlated. The obtained correlations indicate that the bias vector was established by the characteristics of the individual letter sets.

Van der Heijden et al. (1984) considered several explanations for the low interrelationships between the diagonals of the confusion matrices and concluded, through the process of elimination, that the source might lie with the different retinal positions used in each study and, by inference, since the data are not provided in any of the studies, with differences in luminance levels. Although retinal projection and luminance level may affect letter confusions, the most likely source of the differences between the matrices is overlooked by van der Heijden et al. (1984). The physical characteristics of the letter sets are very different. Townsend's (1971) stimuli were typed with an IBM Executive Directrix typewriter; Gilmore et al. (1979) created letters on a CRT by illuminating dots in

Preparation of this paper was facilitated by a grant from the National Institute on Aging (AG03209). I thank Fred Royer for his comments on an earlier draft. Requests for reprints should be sent to Grover C. Gilmore, Department of Psychology, Case Western Reserve University, Cleveland, Ohio 44106.

Table 1
Correlations Among Three Confusion Matrices for Values from the Original Raw Matrices of Proportions and Similarity and Bias Values Estimated with the Luce Choice Model

	Proportions		Choice Model	
	Diagonal	Off-Diagonal	Similarity	Bias
Gilmore et al. $\times$ Townsend	.212	.485	.501	.496
Gilmore et al. $\times$ van der Heijden et al.	.503	.581	.656	.420
Townsend $\times$ van der Heijden et al.	.477	.585	.712	.247

a 7×5 matrix; and van der Heijden et al. (1984) presented continuous-line letters on a display screen.² The stimuli also differed in size. Townsend (1971) reported that the letters subtended a visual angle of $.5^\circ$; van der Heijden et al. (1984) used letters that subtended $.48^\circ \times .32^\circ$; and Gilmore et al.'s (1979) letters were only $.33^\circ \times .23^\circ$. Differences in font, stroke width, and construction by dots or lines should have marked effects on the structure of the confusion matrices because these are variables that do affect the spatial frequency and phase spectra of the letters. For instance, an increase in size or stroke is accompanied by an increase in the amplitude of low frequencies. Since it is the lower frequencies that account for the largest proportion of letter confusions (Royer et al., 1985), it is reasonable that a positive weighting of these frequencies will affect the letter confusions.

The lesson is clear. The validity of a confusion matrix cannot be established by comparing it to a matrix generated with a physically different letter set. The judgment of validity must await the testing of the information obtained from the matrices in independent investigations. As noted, Gilmore (1980) did use the Gilmore et al. (1979) data successfully. More recently, Gilmore, Tobias, and Royer (in press) have demonstrated that clusters of similar nontarget letters, as defined by the Gilmore et al. (1979) metric, speed target detection in a manner analogous to a homogeneous array of nontargets. The validity of this metric is established within the contexts of the identification and detection studies.

The three confusion matrices discussed can each improve the design and potential value of a study if they are applied properly. The primary rule is that the investigation use the original stimuli on which the matrix is based. Since the three matrices represent common methods of stimulus construction, an investigator should be able to adopt one of the letter sets. The standardization of letter

stimuli will permit the valid comparison of results across studies.

REFERENCES

- GERVAIS, M. J., HARVEY, JR., L. O., & ROBERTS, J. O. (1984). Identification confusions among letters of the alphabet. *Journal of Experimental Psychology: Human Perception and Performance*, 10, 655-666.
- GIBSON, E. (1969). *Principles of perceptual learning and development*. New York: Appleton-Century-Crofts.
- GILMORE, G. C. (1980). Letter interaction in brief visual displays. *Quarterly Journal of Experimental Psychology*, 32, 649-668.
- GILMORE, G. C., HERSH, H., CARAMAZZA, A., & GRIFFIN, J. (1979). Multidimensional letter similarity derived from recognition errors. *Perception & Psychophysics*, 25, 425-431.
- GILMORE, G. C., TOBIAS, T., & ROYER, F. L. (in press). Aging and similarity grouping in visual search. *Journal of Gerontology*.
- ROYER, F. L., GRUHN, J. J., GILMORE, G. C., & THOMAS, C. (1985). *Prediction of alphabetic letter confusions from similarity of Fourier spectra* (Tech. Rep. No. 1985-2). Cleveland: Case Western Reserve University, Perception Laboratory.
- SPELTING, G. (1960). The information available in brief visual presentations. *Psychological Monographs*, 74(11, Whole No. 498).
- TOWNSEND, J. T. (1971). Theoretical analysis of an alphabetic confusion matrix. *Perception & Psychophysics*, 9, 40-50.
- VAN DER HEIJDEN, A. H. C., MALHAS, M. S. M., & VAN DEN ROOVAART, B. P. (1984). An empirical interletter confusion matrix for continuous-line capitals. *Perception & Psychophysics*, 35, 85-88.

NOTES

1. To satisfy an assumption of the choice model, it was necessary to estimate the values of empty off-diagonal cells. An iterative procedure was used to choose the value that optimized the fit of the model. For the Townsend (1971), Gilmore et al. (1979), and van der Heijden et al. (1984) matrices, the values of .0059, .0013, and .0006, respectively, were used.

2. Representations of the original stimuli are given in the respective articles. The Townsend (1971) stimuli are an exception. However, IBM has stated that the font used in the Executive Directrix is closest to the commonly available Prestige Elite.