- ²⁰ Watt, B. C., and Merrill, A. L., *Composition of Foods: Agriculture Handbook No. 8* (Agricultural Research Service, United States Department of Agriculture, Washington, DC, 1963).
- ²¹ Food and Agriculture Organization, *Food Composition Tables for International Use* (FAO Nutritional Studies No. 3, Washington, 1949).
- ²² Platt, B. S., *Tables of Representative Values of Foods Commonly Used in Tropical Countries* (Med. Res. Council. Spec. Rep. Ser. No. 302, HMSO, London, 1962).
- ²³ Kent-Jones, D. W., and Amos, A. J., *Modern Cereal Chemistry*, 9 (1967).
- ²⁴ Fraser, J. R., and Holmes, D. C., *J. Sci. Food Agric.*, 7, 589 (1956).
- ²⁵ *Nature*, 149, 460 (1942).
- ²⁶ *Nature*, 150, 538 (1942).
- ²⁷ *Nature*, 151, 629 (1943).
- ²⁸ *Nature*, 154, 582 (1944).
- ²⁹ *Nature*, 154, 788 (1944).

GENERAL

The Use of Correlational Analysis for Pattern Recognition

In contrast to spectral analysis¹, by which the presence of rhythmic activity can be detected but not defined in the time domain of a set of data, correlational analysis can be used to identify the exact time when a particular pattern, rhythmic or otherwise, occurs in the data². This can be done by defining a set of data as a template and "scanning" the noisy data with this template, computing the cross-correlation coefficient ($|r|$) at each data point. A high value of the correlation coefficient implies that the noisy data are similar to the template. When searching for complex wave shapes this method can have serious limitations because there may be small but complex parts of the template that are important in the definition or recognition of the pattern but which may contribute little to the cross-correlation coefficient.

Thus, the correlation coefficient may be sensitive to one part of the pattern; this effectively limits the extent to which the whole pattern can be recognized. For example, this can occur when the template includes high frequency components in one part of the template superimposed on or added to a low frequency component, which incorporates most of the area of the template. In that case the template will correlate highly and "recognize" slow waves not having the fast components of the template.

To avoid this we have developed a technique which allows us to divide the pattern to be recognized into two templates and to compute separate correlations for each; the two correlation coefficients are then multiplied together. The chief effect of this is to ensure that both parts of the pattern are recognized equally. The Recognition Index (RI) is the product of the two correlation coefficients when they are both positive. All cases in which the coefficients are of opposite sign and those for which they are both negative are ignored. Consequently, the scale reflects the degree to which both parts of the pattern, that is, the two templates that define the pattern, are correlated with the data at the appropriate time. The RI is the product of two correlation coefficients each of which can only vary between 0.0 and 1.0; it is a second order exponentially accelerated function and is independent of the amplitude of the matched data points that enter into the correlation.

A program for a PDP 12 computer has been written with an automatic feature which designates the data that have been correlated with the template for each RI occurring above an arbitrarily set threshold. The plotter output is shown in Fig. 1; the second trace (C-C) is the simple cross-correlogram of the whole of template A against the data shown below. High positive values occur at several points. The third trace (RI) shows the recognition index when the same template is separated into two parts, B I and B II, comprising a high frequency complex

Fig. 1 Comparison of the cross-correlogram and the Recognition Index. Template A and template B encompass the same data; in template B the pattern to be recognized is separated into two features, B I and B II. Shaded areas encompass the parts of the data which enter into the correlations with template B. C-C, cross-correlation; RI, Recognition Index. Length of the base line of templates A and B = 2 s.

and a low frequency wave respectively, as shown below the data (template B). The data are the electrical activity of the brain recorded from the vertex and the pattern selected for recognition is one commonly known as the Contingent Negative Variation³; it includes a sharp evoked response (B I) and a slow wave (B II). The RI rises much more abruptly than the cross-correlogram because it is only when the evoked response and the slow wave are both coincident with the respective data that the correlations are high. This is clearly demonstrated by the patterns encompassed by radiating lines in the right half of the figure. The correlogram shows peaks at values of 0.65 and 0.60 whereas the RI shows peak values of 0.45 and 0.35, for the same patterns in the data. This occurs because the evoked potential encompassed by template B I is not present, although the slow waves are; consequently the RI is lower because its value, unlike the correlogram, is equally contingent on the presence of both features of the pattern. Further, measurement of the real time at which patterns similar to the template are present in the data is greatly enhanced by using the RI because of its greater discrimination³.

This work was supported by the National Research Council, Canada, and by the Clement and Jesse V. Stone Foundation.

H. WEINBERG*
R. COOPER

Burden Neurological Institute,
Bristol

Received January 17; revised May 4, 1972.

* Permanent address: Simon Fraser University, Vancouver, Canada.

¹ Blackman, R. B., and Tukey, J. W., *Measurement of Power Spectra* (Dover, New York, 1958).

² Adey, W. R., Walter, D. O., and Hendrix, C. E., *Exp. Neurol.*, 3, 501 (1961).

³ Walter, W. G., Cooper, R., Aldridge, V. J., McCallum, W. C., and Winter, A. L., *Nature*, 203, 380 (1964).