short (or the reliable range of a linear detector may be short), and it may be found that the range of channel intensities is greater than the range of the bench. In such a case, flatten the channel with a color-balancing filter, which may be left in until all the profiles are measured. For example, a lamp distribution high in the red combined with a photocell response high in the blue may produce a resultant green peak. A magenta filter will flatten this peak and reduce the range of intensities to be accommodated by the bench (or linear detector).

The spacing constants for the profiles are measured by holding wavelength constant ($\lambda = 550$) and measuring densities between the θs at which the profiles were measured, again by finding positions of the bench lamp that produce the same deflections as successive wedge positions. Measurement of these densities yields the necessary data for computation of the $\theta D\lambda$ solid in the manner described previously.

A geometric method for indexing stimulus relations in association

RICHARD P. ALLEN, *BALTIMORE CITY HOSPITALS, 4940 Eastern Avenue, Baltimore, Maryland 21124*

Methods for indexing stimulus relations in association involve measuring the overlap of the response set and naturally lend themselves to a set-theory method. A geometric method using associative frequencies as distance measures might be suited better to associative data. Two geometric measures are advanced, and it is argued that their accuracy for extreme cases exceeds that for measures from a set-theory approach.

One of the current methodological problems in word association involves indexing relations among stimulus words. A recent article by Johnson and Collier (1969) reviewed four indices that have been advanced and suggested two new measures based upon set theoretic considerations. All six indices assume that the relations between stimulus words are best determined by the similarity of responses as indicators of some sort of stimulus equivalence involving semantics or usage within linguistic frameworks. Thus, for a sentence with a word omitted, the likely guesses at the omitted word will be those stimuli with fairly similar responses in a free-association test.

The notion of stimulus equivalence based upon similarity of responses in association seemed a natural condition for the application of set theory to measure the overlap of responses. The article by Johnson and Collier (1969), indeed, comes as the logical conclusion to the approach started by Deese (1962). The use of set theory, however, disregards much of the metric information available from word-association probability that has proved useful in research on perception (O'Neil, 1953), latencies (Thumb & Marbe, 1901), and aphasia (Howes, 1965; Allen, 1967). It is proposed in the following paper that a geometric approach provides a measure of stimulus relations better suited to the associative data because it takes account of the metric data and, coincidentally, minimizes the problem of the representational response.

For geometric considerations, each response word can be used as a dimension in a space with the frequency of response indicating the location along the dimension. It is assumed as a first-order approximation that the metric is linear with frequency, the response dimensions are orthogonal, and the dimensions involving each of the stimulus words can be ignored as irrelevant. The distance "D" between two stimulus words is then defined as:

$$D^2 = \sum_i (f_{1,i} - f_{2,i})^2,$$

$f_{j,i}$ = the frequency for Response i to Stimulus j. Defining the metric in probability terms adjusts for sample sizes, thus: $p_{j,i} = f_{j,i}/n_j$, n_j = the sample size defined as the number of Ss who each gave one association response to the stimulus "j." The linear distance measure, "ν" now becomes $\nu^2 = \sum(p_{1,i} - p_{2,i})^2$; if $n = n_1 = n_2$, then $\nu^2 = 1/n^2 \sum(f_{1,i} - f_{2,i})^2$. As an example, consider that the stimulus "CASE" produced 15 associates of "BOX," 4 of "CARTON," and 1 of "TRIAL," and that the stimulus "CARTON" produced 10 associates of "BOX," 8 of "CASE," and 3 of "PACKAGE," then: $\nu^2 = (15/20 - 10/20)^2 + (1/20 - 0/20)^2 + (0/20 - 3/20)^2 = 0.09$.

Obviously, the assumptions made above are only approximately correct. Since the sum of the response probabilities must be 1.0, the dimensions are not independent and accordingly not orthogonal. But for large sets of responses, orthogonality seems justified considering the mathematical simplification it introduces. A more important assumption to consider would be the linear metric. The literature reviewed by Allen (1967) demonstrates that the associative frequency metric is approximately logarithmic: the difference between frequencies of 3 and 2 is more like that between 30 and 20 than that between 30 and 29.

Accepting a logarithmic metric poses the serious problem of handling zero frequencies. It is assumed here that a zero frequency may be arbitrarily set at ½; this assumption requires, of course, caution since it essentially asserts that, if the sample size were doubled, about half of these zero-frequency responses would occur at least once. The test of this assumption here will be merely the appropriateness of the measure of stimulus relations. The logarithmic distance, "T," is: $T^2 = \sum(\log p_{1,i} - \log p_{2,i})^2 = \sum \log^2(p_{1,i}/p_{2,i})$; if $n = n_1 = n_2$, then $T^2 = \sum \log^2(f_{1,i}/f_{2,i})$. The maximum value of "T" unfortunately depends upon the sample sizes for the stimuli. At least to partially reduce the sample-size complication, the logarithmic geometric measure "τ" is defined as the ratio of the value of "T" to its maximum value. Thus, $\tau^2 = T^2/M^2$, $M^2 = N_1 \log^2(2 \times n_1/n_2) + n_2 \log^2(2 \times n_2/n_1)$, for $n = n_1 = n_2$ $\tau^2 = [1/(2n \log^2 2)] \times \sum \log^2(f_{1,i}/f_{2,i})$, and using common logs, $\tau^2 = (1/0.18 \times n) \times \sum \log^2(f_{1,i}/f_{2,i})$.

To compare these measures with the six previously advanced, consider the two extreme cases of identical associative responses: Case 1, the two stimuli never elicit each other but have identical

48

Behav. Res. Meth. & Instru., 1970, Vol. 2 (2)

Table 1
Measures of Relation for Two Extreme Cases

	$S_1 = S_0$	D	MRx2	ITA	$\nu \times (1/\sqrt{2})$	τ
Case 1	0.33	0.50	0.50	0.50	0.00	0.00
Case 2	1.00	1.00	1.00	0.00	0.00	0.00

response distributions; and Case 2, the two stimuli elicit only each other, respectively. The indices of relation for each case are indicated in Table 1. In reading Table 1, recall that, for the six measures previously defined, the larger the number, the more related the stimuli. The data in the following tables have been adjusted where required to give each index a maximum range of 0.0 to 1.0. In the tables below, S_0 and S_1 are the measures advanced by Johnson and Collier (1969) and for only two stimuli are identically defined as the ratio of the overlap to the total set of elements; D (Deese, 1962) is the ratio of the overlap to the geometric mean of the sample sizes; MR (Jenkins & Cofer, 1957) is the ratio of the overlap to the sum of the sample sizes; ITA (Marshall & Cofer, 1963) is defined as the sum of the frequencies for each response word given for both of the two stimuli divided by the sum of all response frequencies. The representational responses are assumed for all of the indices above except ITA and, of course, ν and τ.

According to S, D, and MR, the Case 2 stimuli are considerably more related than the Case 1 stimuli, while for ITA, the reverse is true. ν and τ provide the minimum distance measure for each case. Since both of these cases were selected to represent stimulus pairs with identical response characteristics, it would seem indeed desirable that both be considered as examples of maximum stimulus relations; the ν and τ measures accordingly provide the most reasonable assessment of the stimulus relations for these extreme or limiting cases.

But are these extreme cases realistic? Perhaps linguistic constraints prohibit approximations to one case or the other. Table 2 provides two examples of each case. The first three word pairs in the table are ordered according to a subjective assessment of decreasing stimulus relation agreed upon unanimously by three independent judges; conceptual similarity of opposites such as the fourth pair is, of course, harder to judge. The three pairs of synonyms, however, provide examples of both Case 1 and Case 2 stimuli. Each of the four pairs, of course, represent two words closely related conceptually. S_1, S_0, D, and MR indices indicate a high degree of stimulus relation for only the fourth pair, setting the first three pairs aside as considerably less related. ITA yields a high relation index only for the two Case 1 stimulus pairs. Again, only the ν and τ measures indicate that all of these pairs are closely related. It is worth noting that for these pairs, τ proves more satisfactory than ν, giving an order for the three synonym pairs in agreement with subjective ratings in contrast with the results for ν.

While any of the indices of stimulus relations have some problems, these limited data and the theoretical consideration above suggest that the geometric-distance measures "ν" and "τ" provide a better assessment of stimulus relations than do any of the other measures. Extending the geometric measures to include more than two stimuli follows the procedures for the similar problems for points in a space.

Table 2
Stimulus Relation Indices (Data from Allen, 1967)

	Stimulus Pair	$S_1 = S_0$	D	MRx2	ITA	$\nu \times (1/\sqrt{2})$	τ
Case 1	can't:cannot	0.14	0.29	0.30	0.77	0.18	0.33
	abode:residence	0.13	0.26	0.26	0.71	0.16	0.34
Case 2	house:home	0.18	0.37	0.30	0.08	0.15	0.73
	long:short	0.49	0.65	0.64	0.11	0.13	0.49

REFERENCES

ALLEN, R. P. Some quantitative properties of word association. Unpublished doctoral dissertation, Cambridge University, 1967.

DEESE, J. On the structure of associative meaning. Psychological Review, 1962, 69, 169-175.

HOWES, D. H. Application of word-frequency concept to aphasia. In Disorders of language, Ciba Foundation symposium. London: Churchill, 1965. Pp.

JENKINS, P., & COFER, C. N. An exploratory study of discrete free association to compound verbal stimuli. Psychological Reports, 1957, 3, 599-602.

JOHNSON, P. E., & COLLIER, R. O., JR. A note on methods of indexing associative relatedness. Behavior Research Methods & Instrumentation, 1969, 7, 140-142.

MARSHALL, G. R., & COFER, C. N. Associative indices as measures of word relatedness: A summary and comparison of ten methods. Journal of Verbal Learning & Verbal Behavior, 1963, 1, 408-421.

O'NEIL, W. M. The effect of verbal association on tachistoscopic recognition. American Journal of Psychology, 1953, 5, 42-44.

THUMB, A., & MARBE, K. Experimentelle Untersuchungen über die psychologischen Grundlagen der sprachlichen Analogiebildung. Leipzig: Engelmann, 1901.

Behav. Res. Meth. & Instru., 1970, Vol. 2 (2)

49