

R. M. FRUMKINA - P. F. ANDRUKOVICH - A. YU. TEREKHINA

COMPUTATIONAL METHODS IN THE ANALYSIS  
OF VERBAL BEHAVIOUR

The paper attempts to contribute to the development of the mathematical models of verbal behaviour by demonstrating the use of multi-dimensional individual scaling methods for differential representation of verbal perceptual structures.

The method is illustrated with data on computational analysis of perception of letters of Russian alphabet.

The main concern of computational linguistics (at least, of its theoretically oriented branch) has been the automatic analysis or generation of written texts. "Computational linguistic analysis" has thus become a tool for the validation of linguistic methods and theories.

Another branch of computational linguistics – a more empirical and more practically-oriented one – has been largely restricted to data processing and information – retrieval problems.

It is safe to say that the full extent of the potential influence of the computational approach upon the study of language functioning, that is of speech perception and verbal behaviour has not been generally recognized. Unlike psychologists, linguists have been rather slow to adopt computers for the development and evaluation of mathematical models of verbal behaviour. The present paper attempts to contribute to the development of such models by demonstrating the use of some computational approaches for differential representation of verbal perceptual structures.

Speech perception may in a broad sense be defined as that part of the communication process taking place within the receiver. We should try to reach a more detailed view of the speech perception mechanism, that is to suggest some "white box" instead of the "black box".

The common methodological premise for a model which accounts for various aspects of the speech perception problem is a representation of any speech unit (a sound, a syllable, a word etc.) as a point in a multidimensional perceptual space with the perceived difference between

such stimuli represented by the distance between the stimulus points.

To develop a model for this psychological space we have: 1) to determine the dimensionality of the space; 2) to describe the position of the speech stimuli in question along the dimensions obtained; 3) to determine the position of the Ss perceiving the given set of speech stimuli with respect to these dimensions; 4) to suggest some linguistic and/or psychological interpretation of the discriminative dimensions and of groupings (if any) of individuals having different "viewpoints" about stimulus interrelations.

From the manner in which Ss make perceptual similarity judgements about verbal stimuli it is possible to infer the dimensions of the given set of stimuli which account for the responses obtained.

Methods of multidimensional scaling provide us with highly sophisticated tools for making this inference. The problem of multidimensional scaling broadly stated is to find  $n$  points whose interpoint distances match in some sense the experimental similarities of  $n$  objects (instead of similarities the experimental measurements may be dissimilarities, confusion probabilities or other measures). Hence, we view multidimensional scaling as a problem of statistical fitting - the similarities are given, and we wish to find the point configuration whose distances fit them best.

The well-known method of principal components can be used to find out the most discriminative response dimensions (W. S. TORGERSON, 1952). Given data which represent nonmetric information concerning perceived similarity of stimuli, this method aims at constructing a configuration of those stimuli in a best-fitting Euclidean subspace. The dimensions obtained are oriented along the latent vectors corresponding to the most valued latent roots of similarity matrix. Method of principal components gives a linear orthogonal projection in a best fitting subspace.

There exists, however, another class of methods which provide for the mapping of raw data by some kind of non-linear transformation. For the analysis of perceptual structures in verbal behaviour the present authors have used a modification of the "individual multidimensional scaling" methods suggested by B. BLOXOM (1968), C. HORAN (1969) and J. D. CARROL and J. J. CHANG (1970). This approach to the analysis of experimental data permits us to describe the individual differences between Ss in stimuli perception. In our previous research we have applied some methods of multidimensional scaling for studying the subjective probability estimates of Russian words. In the present

paper for purposes of illustration we perform an analysis of similarity judgements among block letters of the Russian alphabet according to the above mentioned model.

Measures of perceived similarity among stimuli may be obtained by several experimental procedures. This study uses the data of the experiment, in which the stimuli – 32 block letters of Russian alphabet – were presented pair-wise (E. N. GERGANOV, P. F. ANDRUKOVICH, A. P. VASILEVICH, 1972; the same procedure was used by T. KÜNNAPAS, 1966). Ss have been asked to judge each pair of letters as “alike” or “not alike”. Each comparison was made by at least 50 Ss, the “point of view” of each S being represented by similarity matrix with 0 for “alike” and 1 for “not alike” as entries.

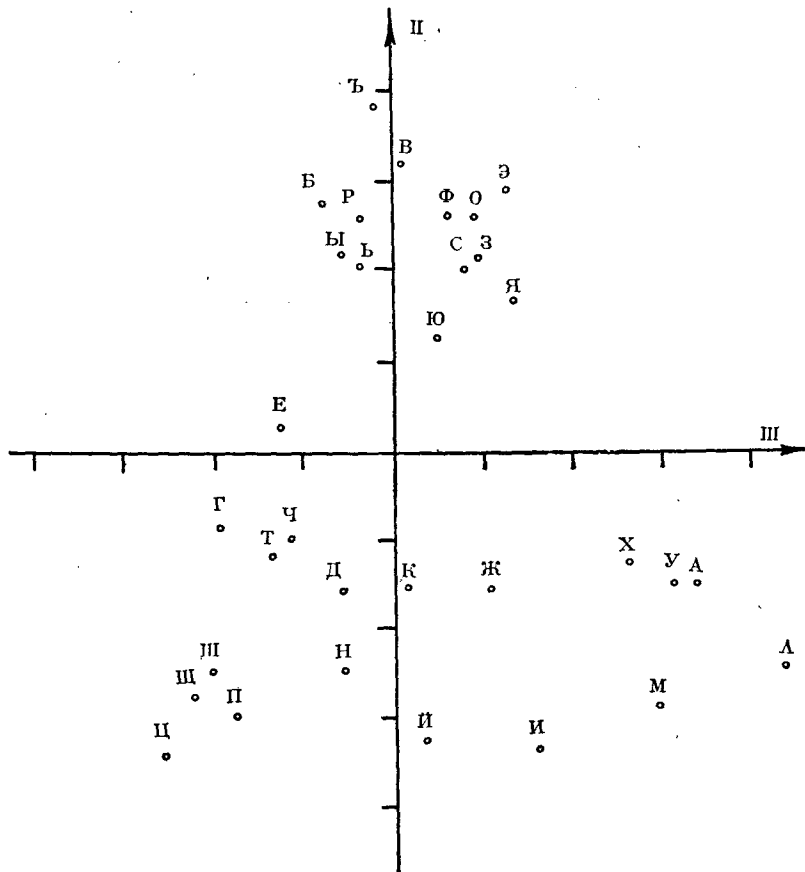


Fig. 1.

After these data had been processed by the method of principal components we have obtained the plotting of stimuli on the planes corresponding to various pairs of latent vectors. We have excluded from further analysis the first latent vector because it has more or less constant elements and may be regarded as allowing for the mean value of distance between all the elements of pairs (J. C. GOWER, 1966).

The projection of letters obtained is presented in Fig. 1 and 2.

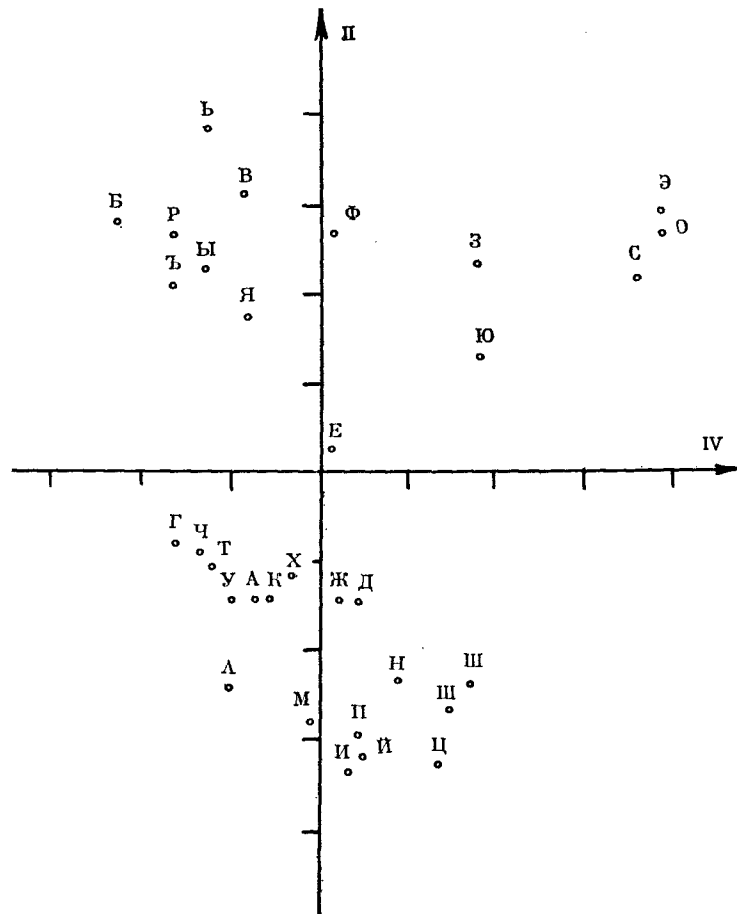
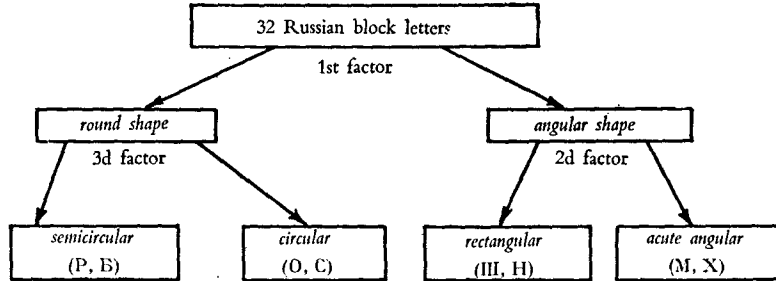


Fig. 2.

An analysis of these configurations resulted in discovering the three main factors underlying the perceptual behaviour of the Ss:



A more detailed analysis can be made from the data based upon the projections on successive coordinate axes allowing for progressively less important differences among letters.

Now we turn to individual scaling method which, as we have stated earlier, permits us to uncover the individual perceptual structures. According to this approach individuals are assumed differentially to weight the several dimensions of the “common psychological space”.

Let us take an assumption about the existence of a “true” point configuration in  $k$ -dimensional Euclidean space, where a set of  $k$  dimensions or “factors” are common to all individuals, but the weights they assign to these factors are different. Then, for any individual the “weighted” Euclidean distance is given by

$$d_{hij}^2 = \sum_{l=1}^k w_{hl}^2 (x_{il} - x_{jl})^2$$

where  $x_{il}, x_{jl}$  – values of the  $j$ -th and  $i$ -th stimulus on the  $l$ -th dimension,  $w_{hl}$  – the weight which represents the salience of the  $l$ -th dimension for the  $h$ -th individual. If for a given  $S$  the  $l$ -th dimension have no importance at all,  $w_{hl}$  will be equal to zero.

Let us denote by  $D_h$  the similarity matrix for  $h$ -th individual. Now our goal will be to find out such point configuration and such set of weights to minimize some function  $f(D_h, d_h)$  used as a criterion of goodness of fit. One of the present authors has suggested the following criterion:

$$S_h = \sum_{i>j} \begin{cases} \frac{(D_{ij} - d_{ij})^2}{D_{ij}}, & \text{for } d_{ij} \geq D_{ij}, \\ (D_{ij} - d_{ij})^2 \cdot D_{ij}, & \text{for } d_{ij} < D_{ij}. \end{cases}$$

This criterion applied, the distortion of the small distances makes them smaller, and the distortion of the greater ones makes them greater, thus permitting better discrimination between groups of objects.

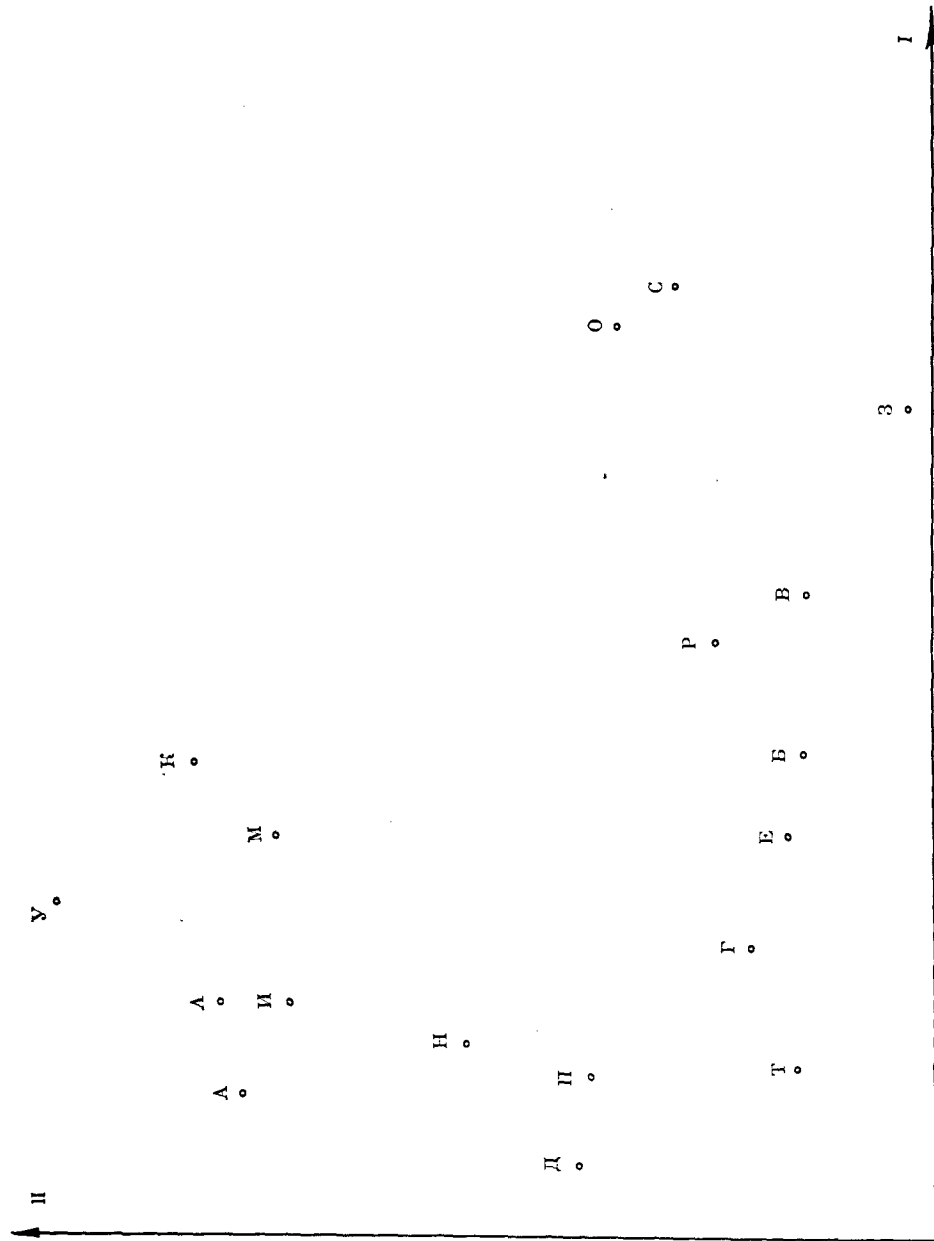


Fig. 3.

To evaluate the amount of discrepancies for all the judges, we can take the average of  $S_h$  thus defining the following divergency criterion:

$$S = \frac{1}{m} \sum_{h=1}^m S_h$$

The gradient method has been used for obtaining the numerical values of  $x_{ii}$  and  $w_{hi}$  which minimize the value of  $S$ . Limiting ourselves to the two-dimensional representation of the  $S_s$ ' perceptual spaces we may plot our data on the plane, taking values of  $x_{ii}$  as coordinates of the stimuli and  $w_{hi}$  as those of the  $S_s$ . Fig. 3 and 4 represent two interrelated point configurations.

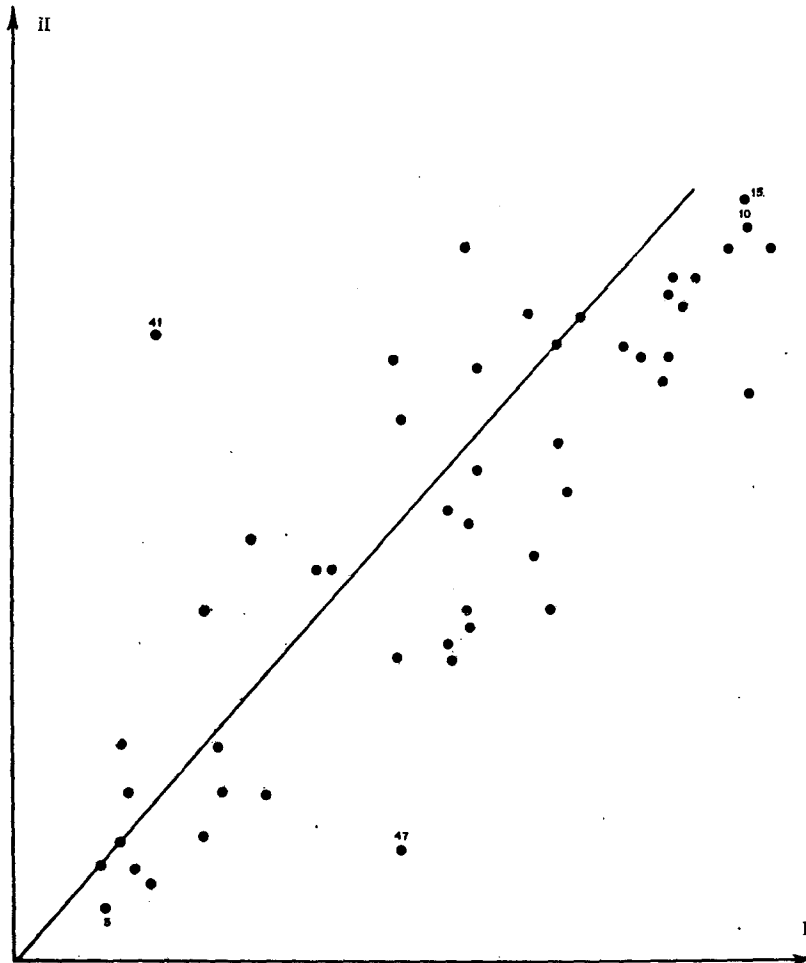


Fig. 4.

It is evident, from Fig. 3, that there are groupings of letters according to their subjective intersimilarity, thus providing us with two psycholinguistically interpretable dimensions. The analysis reveals that dimension 1 corresponds to the opposition "letters with straight elements - letters without straight elements". The dimension 2 corresponds to the opposition "letters with acute-angular elements - letters without acute-angular elements".

Method of individual scaling also provides us with a mapping of Ss into two-dimensional subject space. Coordinates of the point for a given S in this space correspond to the weights of the various dimensions in the stimulus space. Fig. 4 gives a visual impression of the one-two plane of the subjects space. We see that the Ss can be first of all contrasted with respect to *magnitude of the weights* assigned by them to the 1-st and 2-nd dimensions. The analysis of individual similarity matrixes revealed that the Ss who tended to choose answers "alike", weighted equally low both the 1-st and the 2-nd dimensions, while the Ss with tendency to choose answers "not alike" weighted heavily both dimensions. For instance, subject N. 5, who attached very low weight to both dimensions, gave 117 answers "alike" and 36 answers "not alike". A good contrast to the subject N. 5 is provided by the subjects N. 10 and N. 15, who answered "alike" only 3 times from 153 comparison judgements.

However, the most important outcome of our analysis is the fact that the same dimensions being present they have *different relative importance for different Ss*. One group of Ss attaches maximal weights to the 1-st dimension (in Fig. 4 the corresponding points are under the diagonal), while another group attaches maximal weights to the 2-nd dimension. S N. 47 and S N. 41 provide a good contrast in this respect. S N. 47 weights dimension I considerably more than dimension II while S N. 41 shows the opposite tendency. Fig. 5 and 6 contrast the "perceptual spaces" for these two subjects (coordinate axes are transformed by multiplication to the corresponding weights).

Still, for the greater part of the Ss there seems to be almost no difference in the weights for the dimensions (the corresponding points are plotted in Fig. 4 along the diagonal). The place of the study of individual perceptual structures underlying similarity judgements of Russian letters is, of course, restricted to the value of illustrative example. However, trivial could seem the resulting letter classification, even in this very simple case the method of individual scaling has allowed



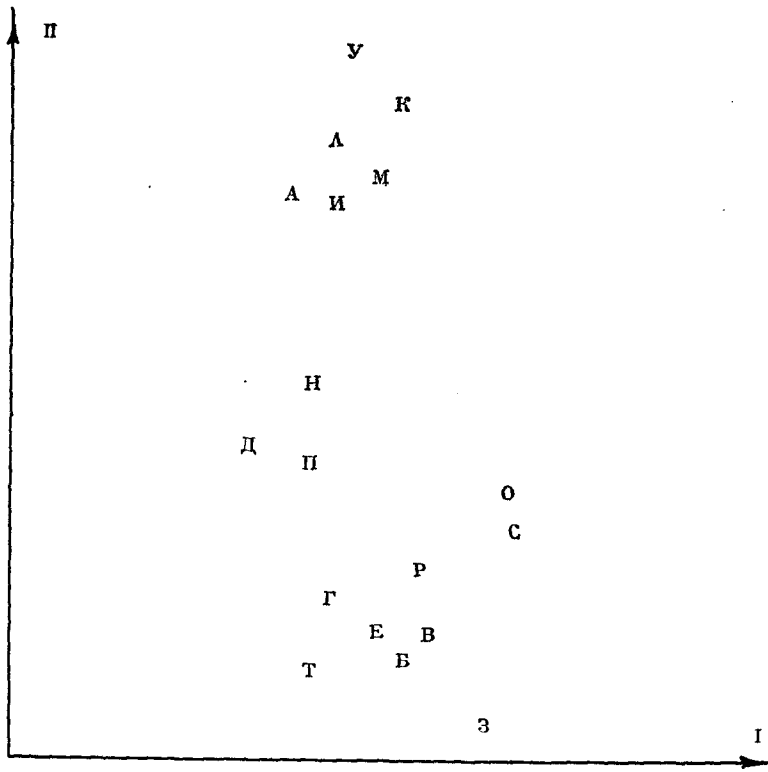


Fig. 5 (S 41)

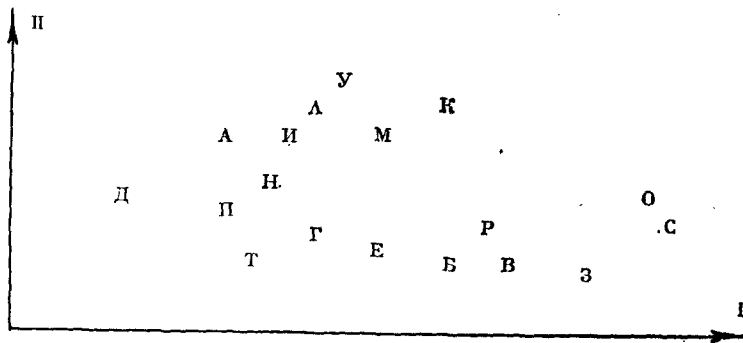


Fig. 6. (S 47)

\*

us to obtain highly non-trivial results concerning the communality and differences of Ss' perceptual subspaces.

We would like to stress that in the overall context of verbal behaviour research the possibilities given to a linguist by the model described could not be overestimated. Perhaps, one of the strongest points of this method being applied to the analysis of verbal behaviour phenomena is its potential generalization to discovering socially determined cognitive superstructures underlying the individual behaviour.

To give no more than one example the individual scaling method makes it possible to analyze confusions data for children at different stages of native language acquisition, thus providing some experimental data which, we hope, could throw some light on the problem of the internalized and unconscious speaker-hearer's knowledge of his language (N. CHOMSKY, 1965), which serves as a base for distinction between "competence" and "performance".

## REFERENCES

- B. BLOXOM, *Individual differences in multidimensional scaling*, Educational Testing Service, Princeton (N.J.), Research Bulletin 68-45, (1968).
- J. D. CARROLL, J. J. CHANG, *Analysis of individual differences in multidimensional scaling via an  $n$ -way generalization of « Eckart-Young » decomposition*, in « Psychometrika », XXXV (1970) 3.
- N. CHOMSKY, *Aspects of the theory of syntax*, Cambridge (Mass.), 1965.
- E. N. GERGANOV, P. F. ANDRUKOVICH, A. VASILEVICH, *On graphical resemblance of Russian block letters*, in *Sinkhronicheski-tipologicheskie i istorikotipologicheskie issledovania*, Institute of Linguistics, the Academy of the USSR, 1972.
- J. C. GOWER, *Some distance properties of latent root and vector methods used in multivariate analysis*, in « Biometrika », LIII (1966), p. 3.
- C. HORAN, *Multidimensional scaling: combining observation when individuals have different perceptual structures*, in « Psychometrika », XXXIV (1969) 2.
- T. KÜNNAPAS, *Visual perception of capital letters*, in « Scand. J. Psychol. », VII (1966).
- W. S. TORGERSON, *Multidimensional Scaling, Theory and Method*, in « Psychometrika », XVII (1952) 4.

