PRINCIPAL COMPONENT ANALYSIS OF THREE-MODE DATA BY MEANS OF ALTERNATING LEAST SQUARES ALGORITHMS

PIETER M. KROONENBERG AND JAN DE LEEUW

UNIVERSITY OF LEIDEN

A new method to estimate the parameters of Tucker's three-mode principal component model is discussed, and the convergence properties of the alternating least squares algorithm to solve the estimation problem are considered. A special case of the general Tucker model, in which the principal component analysis is only performed over two of the three modes is briefly outlined as well. The Miller & Nicely data on the confusion of English consonants are used to illustrate the programs TUCKALS3 and TUCKALS2 which incorporate the algorithms for the two models described.

Key words: three-mode principal component analysis, alternating least squares, factor analysis, multidimensional scaling, individual differences scaling, simultaneous iteration, confusion of consonants.

1. Three-Mode Models and Their Solutions

The three-mode model—here referred to as the Tucker3 model—was first formulated by Tucker [1963], and subsequently extended in articles by Tucker [1964, 1966], and Levin [1963, Note 5] especially with respect to the mathematical description and programming aspects of the model. In the context of multidimensional scaling, references to this model occur frequently [Harshman, 1970, Note 2; Jennrich, 1972, Note 3; Carroll & Chang, 1972, Note 1; Takane, Young & de Leeuw, 1977], since the Tucker3 model is the general model comprising various individual differences models. A discussion of the relationships between multidimensional scaling and three-mode principal component analysis can be found in Tucker [1972], Carroll & Wish [1974], and Takane, Young & de Leeuw [1977].

The algorithms developed by Tucker [1966] are used to solve the three-mode model in all cases. References to computer programs based on these algorithms are Wainer et al. [1971, 1974], Walsh [1964], Walsh & Walsh [1976], and one such program is embodied in the statistical package SOUPAC developed at the University of Illinois. Numerous similar programs have been written, and they are mostly referred to in passing in applied articles. In his 1966 article Tucker remarks that his procedures "do not produce a least squares approximation to the data. Investigations of the mathematics of a least squares fit for three-mode factor analysis indicate a need for an involved series of successive approximations." The procedures described in the sequel are designed to provide least squares estimates of the parameters in the three-mode model. The alternating least squares approach used can also be extended to accommodate other levels of measurement, as has been recently demonstrated by Sands & Young [1980] for a more restricted model.

2. The Description of the Tucker3 Model

The Tucker3 model deals with data which can be arranged in a three-dimensional block or a so-called three-mode matrix. Specifically, a $l \times m \times n$ three-mode matrix Z is

defined as the collection of elements

$$\{z_{ijk}|i = 1,\cdots,l; j = 1,\cdots,m; k = 1,\cdots,n\}.$$

The elements are placed in the three-dimensional block such that the index i runs along the vertical axis, the index j along the horizontal axis, and the index k along the "depth" axis. We will use the word "mode" to indicate a collection of indices by which the data can be classified. For instance, in semantic differential studies [Osgood, Tannenbaum & Suci, 1957] one collects scores of a number of persons on a set of bipolar scales for a collection of attributes. These data can be classified by persons, scales, and attributes; each of these therefore determine a mode of the data.

We will only use real matrices here, and in general the number of rows will be larger than the number of columns. We will use $R^{n \times m}$ for the class of real $n \times m$ matrices, and $K^{n \times m}$ for the class of columnwise orthonormal matrices, and $R^{l \times m \times n}$ for the class of all $l \times m \times n$ three-mode matrices.

Using the above definitions we can formulate the Tucker3 model as the factorization of the three-mode data matrix $Z = \{z_{ijk}\}$, $Z \in R^{l \times m \times n}$ such that

(1)
$$z_{ijk} = \sum_{p=1}^{s} \sum_{q=1}^{t} \sum_{r=1}^{u} g_{ip}h_{jq}e_{kr}c_{pqr}$$

for $i = 1,\cdots,l; j = 1,\cdots,m; k = 1,\cdots,n$, where the coefficients g_{ip}, h_{jq}, and e_{kr} are the elements of the component matrices $G \in K^{l \times s}$, $H \in K^{m \times t}$, and $E \in K^{n \times s}$ respectively, and s, t, and u are the number of components of the first, second, and third mode. The c_{pqr} are the elements of the so-called three-mode core matrix $C \in R^{s \times t \times u}$. In the matrix Z each element represents a specific combination of categories of the original variables. In the same way each element of the core matrix C represents a unique combination of categories of the components. One could conceive of the core matrix as describing the basic relations that exist between the various collections of variables.

A matrix formulation of the Tucker3 model is

(2)
$$Z = GC(H' \otimes E'),$$

where $Z \in R^{l \times nm}$ and $C \in R^{s \times tu}$ are now ordinary (two-mode) matrices by making use of so-called combination modes [Tucker, 1966, p. 281], and $\otimes$ denotes the Kronecker product. By symmetry there are two other matrix formulations (see also Section 4). We will not introduce special notation to distinguish between the two-mode and three-mode versions of Z and C, as the appropriate version is indicated by the real space of which it is an element.

It should be noted that our formulation of the three-mode model corresponds to the model Tucker treats in pages 294ff of his 1966 paper when he describes ways to estimate the parameters of his model. In the earlier theoretical part of his paper Tucker gives a more general formulation with G, H, and E as full column rank matrices, rather than orthonormal ones. Both computational expedience, and the desire to formulate conditions for a unique solution motivated us to describe the model entirely in terms of columnwise orthonormal matrices. Once a solution has been obtained we can transform G, H, and E by either orthonormal transformations and/or by non-singular transformations of the appropriate rank without affecting the loss function (3) defined below, provided we counterrotate the core matrix. In fact we have included in the TUCKALS2 program (see Section 9) a procedure for orthonormal transformation of the core matrix [for details see Kroonenberg & de Leeuw, 1977, Note 4, Appendix A], and are in the process of including a non-singular transformation routine as well. Similarly transformation routines will be eventually included in the TUCKALS3 program as well.

If we computed all the principal components, i.e. $s = l$, $t = m$, and $u = n$, we could decompose any data matrix exactly into its components. However, in practical applications one is just interested in the two, three, or four first principal components. In general this precludes finding an exact factorization of Z in G, H, E, and C. One therefore has to be satisfied with an approximation, $\tilde{Z} = GC(H' \otimes E')$, i.e. finding G, H, E, and C such that the difference between the model and the data is minimal according to some loss function. In slightly different terms, we have to look for the best approximate factorization $\tilde{Z}$ of the three-mode matrix Z into G, H, E, and C according to the Tucker3 model.

In our case, as in many similar situations, we define a mean-squared loss function. We then search for an approximate factorization $\tilde{Z}$ such that

$$(3) \qquad f(G,H,E,C) = \|Z - \tilde{Z}\|^2 = \|Z - GC(H' \otimes E')\|^2$$

is minimal, where $\| \cdot \|$ denotes the euclidean norm. The minimization has to be carried out under the restrictions of the model, i.e. G, H, and E must be columnwise orthonormal matrices. The Z for which f attains its minimum will be designated as $\hat{Z} = \hat{G}\hat{C}(\hat{H}' \otimes \hat{E}')$, and the variables with the carets are the least squares estimators of the model parameters.

3. The Existence of a Best Approximate Solution

In this section we will show that there always exist some G, H, E, and C such that f attains its (global) minimum. Essentially the proof comes down to first showing that C can be uniquely expressed in terms of G, H, E, and Z, and secondly, using the resulting $\hat{C}$, showing that f must have a minimum because it is a continuous bounded function on a finite-dimensional real space.

In fact there exists a unique best C, called $\hat{C}$, such that for fixed G, H, and E, f attains its minimum for this $\hat{C}$ which has as its elements

$$\hat{c}_{pqr} = \sum_{a=1}^{l} \sum_{b=1}^{m} \sum_{d=1}^{n} g_{ap} h_{bq} e_{dr} z_{abd} ,$$

or

$$(4) \qquad \hat{C} = G'Z(H' \otimes E').$$

To prove the above assertion we use a simplified version of a lemma by Penrose [1955], which is presented as Lemma 3.2 in Kroonenberg & de Leeuw [1977, Note 4]. This lemma states that there exists a unique $\hat{C}$, such that the function h,

$$(5) \qquad h(C) = \|Z - \tilde{Z}\|^2 = \|Z - GCF'\|^2,$$

is as small as possible. This $\hat{C}$ is equal to $G'ZF$, and the absolute minimum, i.e. 0, is reached if and only if $Z = GG'ZFF'$. If we write $H \otimes E$ for F in (5), we may conclude that $\hat{C}$ as in (4) minimizes f for fixed G, H, and E, and that

$$(6) \qquad h(\hat{C}) = 0 \text{ iff } Z = GG'Z(H \otimes E)(H' \otimes E').$$

The minimization of the loss function, therefore, is really only dependent upon G, H, and E. Once we have found the appropriate $\hat{G}$, $\hat{H}$, and $\hat{E}$, we can reconstruct $\hat{C}$ via (4).

To proceed with the minimization of f we substitute (4) into (3), call the rewritten function g, and thus

$$(7) \qquad g(G,H,E) = \|Z - \tilde{Z}\|^2 = \|Z - GG'Z(H \otimes E)(H' \otimes E')\|^2$$

$$= \|Z - GG'Z(HH' \otimes EE')\|^2.$$

As the domain S of the function g is

$$(8) \qquad S = \{s \mid s = (G,H,E), \ G \in K^{l \times s}, \ H \in K^{m \times t}, \ E \in K^{n \times u}\} ,$$

we can see that S is a compact subset in a finite-dimensional real space. Using the fact that g is a bounded continuous function on S $(0 \leq g \leq ||Z||^2)$, we can conclude that there exists a point $\hat{s} = (\hat{G}, \hat{H}, \hat{E})$ in S, such that g attains its minimum. In other words the minimization problem always has a solution.

4. A Solution of the Minimization Problem

In this section we will give some details of a solution to the minimization problem (3). In order to do this we will convert the minimization problem into a maximization problem. This done, we will show that the component matrices $\hat{G}$, $\hat{H}$, and $\hat{E}$ of a solution are each nothing but the eigenvectors corresponding to the largest eigenvalues of suitably constructed cross-products of the data matrix Z and the other two component matrices.

To convert the minimization problem into a maximization problem we rewrite (7) using traces instead of norms, and manipulate the various terms somewhat, i.e.

$$g(G,H,E) = \text{tr}(Z - \tilde{Z})(Z - \tilde{Z})' = \text{tr}(ZZ' - \tilde{Z}Z' - Z\tilde{Z}' + \tilde{Z}\tilde{Z}')$$

$$(9) \qquad = \text{tr}ZZ' - 2\text{tr}\tilde{Z}Z' + \text{tr}\tilde{Z}\tilde{Z}'.$$

Expanding each term in turn, and adding them as in (9), we get

$$g(G,H,E) = \text{tr}ZZ' - 2\text{tr}GG'Z(HH' \otimes EE')Z' + \text{tr}G'Z(HH' \otimes EE')Z'G$$

$$(10) \qquad = \text{tr}ZZ' - \text{tr}G'Z(HH' \otimes EE')Z'G.$$

We define p to be equal to the last term on the right-hand side of (10),

$$(11) \qquad\qquad p(G,H,E) = \text{tr}G'Z(HH' \otimes EE')Z'G.$$

Clearly the minimization of g comes down to the same as the maximization of p, as both are bounded. For the sequel it will be convenient to rewrite p a bit further,

$$p(G,H,E) = \text{tr}G'\{Z(HH' \otimes EE')Z'\}G$$

$$= \text{tr}G'PG,$$

with

$$(12) \qquad\qquad P = P(H,E) = Z(HH' \otimes EE')Z', \text{ and } Z \in R^{l \times mn}.$$

So far we have always placed H and E in the Kronecker-product term, but we could equally well have done so with G and E, or G and H. Such substitutions entail only a change in form, but not in the model itself. The model is indifferent to such notational changes as can be clearly seen from (1). In the following we will also need the other forms:

$$p(G,H,E) = \text{tr}H'QH$$

with

$$(13) \qquad\qquad Q = Q(E,G) = Z(EE' \otimes GG')Z', \text{ and } Z \in R^{m \times nl};$$

$$p(G,H,E) = \text{tr}E'RE$$

with

$$(14) \qquad\qquad R = R(G,H) = Z(GG' \otimes HH')Z', \text{ and } Z \in R^{n \times lm}.$$

The maximization of p is of course not unconstrained, but restricted to the set S. We can incorporate the constraints in the maximization problem by using Lagrange multiplier matrices L, M, and N, and obtain $\tilde{p}$,

$$\tilde{p}(G,H,E,L,M,N) = p(G,H,E) - \text{tr}L(G'G - I_t) - \text{tr}M(H'H - I_t) - \text{tr}N(E'E - I_u),$$

where I_a is the $a \times a$ identity matrix. The maximum of p follows from the requirement that the first order partial derivatives of $\bar{p}$ are simultaneously zero at the maximum of p and that the Hessian is negative. We will state here the exact nature of the solution as Theorem 1, and refer the reader for a proof to the Appendix.

Theorem 1

Let Z be a three-mode matrix, and let p, P, Q, R, and S be defined as in (11), (12), (13), (14), and (8) respectively, and finally let U, V, and W be defined as follows: U is an eigenvector matrix of P, V an eigenvector matrix of Q, W is an eigenvector matrix of R, and $(U, V, W) \in S$. Then (a.) $(\hat{G}, \hat{H}, \hat{E}) \in S$ is a stationary point of p if and only if $\hat{G} = U$, $\hat{H} = V$, and $\hat{E} = W$, or some orthonormal rotation thereof. (b.) $(\hat{G}, \hat{H}, \hat{E}) \in S$ maximizes p if and only if their columns are eigenvectors corresponding to the largest s, t, and u eigenvalues of $P(\hat{H}, \hat{E})$, $Q(\hat{E}, \hat{G})$, and $R(\hat{G}, \hat{H})$ respectively, or orthonormal rotations thereof.

The following theorem provides the necessary and sufficient conditions for the existence of an exact solution to the minimization problem (3), and indicates the form of such a solution. The proof of part (a.), and of part (b.)1 follow directly from the definitions, and the proof of part (b.)2 can be found in the Appendix.

Theorem 2

(a.) Let Z be a three-mode data matrix, and let f, g, and p be defined as above. Furthermore let $(\hat{G}, \hat{H}, \hat{E}, \hat{C})$ satisfy the constraints of (3), and let $\hat{C}$ be defined as $\hat{C} = G'Z(\hat{H} \otimes \hat{E})$. Then the following statements are equivalent:

    1. $f(\hat{G}, \hat{H}, \hat{E}, \hat{C}) = 0$,

    2. $g(\hat{G}, \hat{H}, \hat{E}) = 0$,

    3. $p(\hat{G}, \hat{H}, \hat{E}) = \mathrm{tr} ZZ'$ with $Z \in R^{l \times mn}$,

(15)   4. $Z = \hat{G}\hat{G}'Z(\hat{H}\hat{H}' \otimes \hat{E}\hat{E}')$ with $Z \in R^{l \times mn}$, and

    5. $(\hat{G}, \hat{H}, \hat{E}, \hat{C})$ is an exact solution of (3) .

(b.) 1 Let $(\hat{G}, \hat{H}, \hat{E}, \hat{C})$ be an exact solution of (3) . Then

    $\hat{G}$ is the eigenvector matrix (or an orthonormal rotation thereof) corresponding to the p non-zero eigenvalues of ZZ' with $Z \in R^{l \times mn}$,

    $\hat{H}$ is the eigenvector matrix (or an orthonormal rotation thereof) corresponding to the q non-zero eigenvalues of ZZ' with $Z \in R^{m \times nl}$,

    $\hat{E}$ is the eigenvector matrix (or an orthonormal rotation thereof) corresponding to the r non-zero eigenvalues of ZZ' with $Z \in R^{n \times lm}$, and $\hat{C} = \hat{G}'Z(\hat{H} \otimes \hat{E})$.

(b.) 2 On the other hand, if $\hat{G}$, $\hat{H}$, $\hat{E}$, and $\hat{C}$ are defined as in (b.)1, the eigenvalues associated with $\hat{G}$, $\hat{H}$, $\hat{E}$, are different for each matrix separately, and (15) is satisfied, then $(\hat{G}, \hat{H}, \hat{E}, \hat{C})$ is the exact unique solution.

It should be noted that statement (b.)2 is not as strong as one would like to have it, as any $(\hat{G}, \hat{H}, \hat{E}, \hat{C})$ which satisfies (15) determines an exact solution. A more satisfactory statement, however, has not been found yet.

5. Towards an Algorithm for the Solution of the Tucker3 Model

Obviously we would like to construct an algorithm for the maximization of p that converges to a global maximum of p. Unfortunately p is the cross-product term of a multivariate polynomial of the sixth degree, and in general it is not possible to prove that methods to solve such nonlinear problems attain a global maximum. In the present case this also seems to be true. We will have to be satisfied with proving that the algorithm outlined

below will converge to some stationary point which is not a minimum, rather than a global maximum.

The method to be described utilizes the so-called alternating least squares (ALS) technique. The essential feature of the ALS approach is that in solving optimization problems with more than one set of parameters, each set is estimated in turn by applying least squares procedures holding the other sets fixed. After all sets have been estimated once, the procedure is repeated again and again until convergence. Further details and references to applications of the ALS approach can, for instance, be found in Young, de Leeuw & Takane [1980].

In order to see how the ALS approach can be applied in the present context, let us return briefly to (7):

$$g(G,H,E) = \|Z - GG'Z(HH' \otimes EE')\|^2 .$$

Clearly the sets of parameters are here G,H, and E. Minimizing g over G holding H and E fixed is identical to solving one least squares problem, minimizing over H with E and G fixed, and minimizing over E with G and H fixed are the two others. That we are in practice maximizing p does not prevent the problem from being an ALS one.

From the above discussion a rough outline for an algorithm is readily deduced. First choose an arbitrary H_0 and E_0 yielding a new G_1, maximize subsequently over H with the just computed G_1 and E_0 fixed yielding a new H_1, and finally maximize p over E with G_1 and H_1 fixed yielding a new E_1, and iterate this procedure until—one hopes—convergence. According to Theorem 1 the maximizations are essentially identical to searches for eigenvectors and eigenvalues of matrices of the order l, m, and n respectively. As l, m, and n can be quite large, while s, t, and u are typically very small, say 2, 3, or 4, we want to use a technique for solving the eigenvector-eigenvalue problem (or eigenproblem for short) which is particularly efficient in finding just a few eigenvectors.

A very appropriate technique in this situation is the so-called simultaneous iteration method [or Treppen (= staircase) iteration] of Bauer-Rutishauser [Rutishauser, 1969]. For further details on this method see Section 6.

The maximization of p consists thus of an, in principle, infinite iteration process, in which at each step three eigenproblems have to be solved. Clearly, solving these eigenproblems by an infinite iteration process has its drawbacks. The whole procedure is likely to become computationally burdensome. In order to avoid this we perform only one single step towards the solution of the eigenproblems, instead of the complete iterations. A similar approach has been applied by de Leeuw and others in a number of cases when using an ALS technique. The experience has been that carrying out the complete iteration for solving the eigenproblem only serves to decrease the overall efficiency of the procedure, while it has no effect on the eventual convergence point if one uses only one step [Takane, Young & de Leeuw, 1977 p. 59]. They suggest that the reason for this behavior might be found in the same reasons that often cause relaxation procedures to be more efficient than non-relaxation procedures.

6. The Bauer-Rutishauser Method

As the algorithm is based on the method of Bauer-Rutishauser for computing eigenvectors and eigenvalues it seems in order to describe this method in some detail. In addition, some of the formulations developed here will be used in the rest of this paper.

Let $A \in R^{n \times n}$ be a symmetric positive definite matrix, and p the desired number of eigenvectors. Furthermore let $X \in R^{n \times p}$ be defined as the matrix which has as its columns

the iteration vectors. If we write X after i iterations as X_i then the method of Bauer-Rutishauser is defined as follows.

(i)   Choose an arbitrary orthonormal X_0,

(ii) $Y_i = AX_i$, and

(iii) $B_i = Y_i'Y_i$.

(iv) Solve the eigenproblem for B_i, i.e. determine an orthonormal T_i, and a diagonal L_i with $l_1 \geq l_2 \geq \cdots \geq l_p$, such that $T_i'B_iT_i = L_i$, and T_i is the eigenvector matrix of B_i, and then define

(v) $X_{i+1} = Y_iT_iL_i^{-1/2}T_i'$.

Schwartz et al. [1968] show that for $i \rightarrow \infty$, $L_i^{-1/2}$ converges to the matrix with the largest p eigenvalues of A on the diagonal, and the columns of X_i converge to the associated eigenvectors, provided A is positive definite, and the columns of X are not orthogonal to one or more of the eigenvectors, and in addition the pth and $(p + 1)^{st}$ eigenvalues are different. We may write (ii) through (v) somewhat more concisely as

$$X_{i+1} = Y_iT_iL_i^{-1/2}T_i' = AX_iB_i^{-1/2} = AX_i(X_i'A^2X_i)^{-1/2} .$$

With a view to what follows it will be convenient to define the function

(16) $$\phi(X_i) = AX_i(X_i'A^2X_i)^{-1/2} .$$

When we use in the sequel a recursive formula like (16) we mean to say that $X_{i+1} = \phi(X_i)$ can be computed by carrying out one step of the Bauer-Rutishauser method. It should be noted that the inverse square root of $X'A^2X$ exists, and is uniquely defined, if the expression is positive definite. This implies that in such a case ϕ is well-defined, and it can be proved that ϕ is continuous as well (see Appendix). As will be shown in Section 8 rather strong convergence theorems can be used for the algorithm to be described if ϕ is continuous. It seems therefore worthwhile to take measures in constructing the algorithm to ensure the positive definiteness of $X'A^2X$. An inspection of the method to arrive at (16) shows that in fact only the inverse square root is taken of the eigenvalues of B_i. One, therefore, only has to check in each iteration step if all eigenvalues are larger than zero, or in practice larger than some very small number. If one of the eigenvalues is too small, one can restart the iteration procedure with a smaller number of components. There is, however, no guarantee that this will solve the singularity problem. On the other hand if no singularities have occurred one knows that at each step ϕ must have been uniquely defined and continuous on $R^{n \times p}$. As we have taken the above precaution in the program we will from now on assume that expressions like $X'A^2X$ are positive definite.

7. The TUCKALS3 Algorithm

In this section we will describe the algorithm to solve the maximization of p, as well as give some consideration to the initialization of the algorithm. Here Z is again defined as the $l \times m \times n$ three-mode data matrix, and s, t, and u will be the desired number of components for the three-component matrices. Furthermore the orthonormal matrices G, H, and E will be the matrices with as their columns the iteration vectors. We will write G, H, and E as they are after i iteration steps as G_i, H_i, and E_i. One main iteration step of the TUCKALS3 algorithm is then defined by (17) through (22).

G substep

(17) $$P_i = Z(H_iH_i' \otimes E_iE_i')Z'   \text{ with } Z \in R^{m \times nl}$$

(18) $$G_{i+1} = \phi_1(G_i) = P_i G_i (G_i' P_i^2 G_i)^{-1/2}$$

H substep

(19) $$Q_i = Z(E_i E_i' \otimes G_{i+1} G_{i+1}' G_{i+1}')Z' \qquad \text{with } Z \in R^{l \times mn}$$

(20) $$H_{i+1} = \phi_2(H_i) = Q_i H_i (H_i' Q_i^2 H_i)^{-1/2}$$

E substep

(21) $$R_i = Z(G_{i+1} G_{i+1}' \otimes H_{i+1} H_{i+1}')Z' \qquad \text{with } Z \in R^{n \times lm}$$

(22) $$E_{i+1} = \phi_3(E_i) = R_i E_i (E_i' R_i^2 E_i)^{-1/2}$$

As mentioned before, each G, H, and E substep is one step of an inner iteration to find the eigenvectors of P, Q, and R respectively, and together they define one step of the main iteration.

Because we want to discuss the properties of the TUCKALS3 algorithm in the sequel, it is useful to introduce some notation first.

$F: S \rightarrow S$ is a function on S, and F defines a complete step of the main iteration, and S is defined as in (8).
$F = F_3 \cdot F_2 \cdot F_1$ with $F_i: S \rightarrow S$ for $i = 1,2,3$ such that
$\quad F_1(G_i, H_i, E_i) = (\phi_1(G_i), H_i, E_i) = (G_{i+1}, H_i, E_i),$
$\quad F_2(G_{i+1}, H_i, E_i) = (G_{i+1}, \phi_2(H_i), E_i) = (G_{i+1}, H_{i+1}, E_i),$ and
$\quad F_3(G_{i+1}, H_{i+1}, E_i) = (G_{i+1}, H_{i+1}, \phi_3(E_i)) = (G_{i+1}, H_{i+1}, E_{i+1}).$
Thus $F(s_i) = F(G_i, H_i, E_i) = (G_{i+1}, H_{i+1}, E_{i+1}) = s_{i+1}.$

In Section 6 we remarked that ϕ as defined in (16) was a continuous function, and thus ϕ_1, ϕ_2, and ϕ_3 are continuous functions. Because F is a composite of continuous functions, F is continuous as well.

It can be shown that both at each step of the main iteration and at each substep the value of p is increased (see Appendix). Thus

$$p(F(s_i)) = p(s_{i+1}) \geq p(s_i).$$

If p is not increased strictly, i.e. $p(F(s_i)) = p(s_i)$, the algorithm stops. In that case $(\hat{G}, \hat{H}, \hat{E})$ satisfies the necessary conditions of Lemma 3 (see Section 8). Consequently we can assume without loss of generality that the algorithm generates an infinite sequences with $p(F(s_i)) > p(s_i)$.

Obviously we need some G_0, H_0, and E_0 to initialize the procedure. It seems sensible to choose them in such a way that they are optimal in some sense. We chose such an initialization that it would solve the maximization problem exactly if the problem had such a solution. In other words the eigenvector matrices mentioned in Theorem 2 (b.) 1 were used as initializations. Comparing this with Method 1 of Tucker [1966, p. 297], we note that the initialization is nothing but Tucker's final solution. In practice we do not need to know the eigenvectors exactly as they are only used to initialize, and therefore we made only five iteration steps towards their solution, using the Bauer-Rutishauser method.

8. The Convergence of the TUCKALS3 Algorithm

It is, of course, of prime importance to show that the algorithm outlined in (17)–(22) converges, and moreover that it converges to a maximum of p, or at least not to a minimum.

The algorithm considered here is a type of algorithm that has been described in the nonlinear programming literature, and in that field various theorems about the con-

vergence of algorithms such as ours exist. The most appropriate one in our case is the following "fixed point" lemma described and proven by d'Esopo [1959].

Lemma 3.

Let F, p, S satisfy the following conditions.
1. a. S is a subset of a finite dimensional space,
   b. F is a continuous transformation of S to S,
   c. p is a real function defined and continuous for all $s \in S$,
2. $p(F(s)) \geq p(s)$,
3. if $p(F(s)) = p(s)$, then $F(s) = s$,
4. if the sequence $s_0, s_1, \cdots$ satisfies $p(s_{i+1}) \geq p(s_i)$ with $s_i \in S$, then for every limit point $\bar{s}$ of $s_0, s_1, \cdots$ $F(\bar{s}) = \bar{s}$.

In Section 3 we noted and discussed properties 1a, 1c, and in Section 7 we did the same for 1b, 2, 3, and 4. We may, therefore, conclude that Lemma 3 applies to the TUCKALS3 algorithm. As S is a bounded real subspace, any infinite sequence $s_0, s_1, \cdots$ is bounded, and thus the sequences generated by the algorithm are bounded as well. A theorem due to Weierstrass shows that such sequences have at least one limit point. It is shown in the Appendix that every point $\bar{s}$, such that $F(\bar{s}) = \bar{s}$ is a stationary point of p, and because we know that at every step p increases, we know that stationary points will not be minima.

As has been shown by Ostrowski [1966] the set of limit points of $\{s_i\}$ consists either of a single point or a continuum. The latter case, however, is a very unlikely one in practical applications, as is the occurrence of equal eigenvalues in real data. The above results imply, that from any arbitrary starting point s_0 the algorithm converges to a stationary point of p, but the algorithm "like all numerical methods based on local searches for solutions, can be best expected to yield local *(maxima)* ($\cdots$). Global *(maximality)* could be assured only by exhaustive searches over successively finer grids." [Meyer, 1970, p. 45].

9. Special Cases: The Tucker2 and the Tucker1 Model

The Tucker2 Model

An important special case of the Tucker3 model is obtained if the matrix E in (2) is taken to be the identity matrix. The resulting model, the Tucker2 model, can consequently be written as

$$（23） \qquad z_{ijk} = \sum_{p=1}^{s} \sum_{q=1}^{t} g_{ip} h_{jq} c_{pqk}$$

with $i = 1, \cdots, l, j = 1, \cdots, m$, and $k = 1, \cdots, n$, and with the same meaning of the restrictions on G and H as before. In Sands & Young [1980] this model is referred to as the "generalized subjective metrics model". The model can be written in matrix notation as

$$(24) \qquad Z = GCH',$$

where Z and C are three-mode matrices written with combination modes, or as

$$(25) \qquad Z_k = GC_kH',$$

where Z_k and C_k are two-mode matrices or so-called frontal planes for the kth individual of the data matrix and the core matrix respectively. Instead of specifying principal components of all three modes, the Tucker2 model only specifies them for two (say the first two) of the three modes. In other words, the third mode is not condensed and remains intact. This will enable one to study the interrelationships between the components of the

first two modes for each element (variable, individual, moment in time) of the third mode. The Tucker2 model has been independently formulated by Israelsson [1969], Carroll & Chang [1972, Note 1], and Jennrich [1972, Note 3].

The Tucker2 model has three important fields of application. In the first place it can be used in those analyses of data for which no natural condensation of the third mode can be defined. An obvious example would be the multivariate analysis of time series. In general no useful meaning can be attached to the components of a time mode. In certain other applications one is interested in persons as replications, and one does not want to investigate person components, but rather the interrelationships between the other two modes for each person. Secondly the model can be applied in individual differences scaling with asymmetric similarity matrices. A typical example, i.e. the Miller & Nicely confusion matrices [Miller & Nicely, 1955], will be treated in Section 10 as an example of both the Tucker3 and Tucker2 model. Finally, the Tucker2 model can be used to test the appropriateness of various individual differences models in multidimensional scaling, such as INDSCAL, IDIOSCAL, PARAFAC (also called the "weighted model" by Sands & Young, 1980). All these models can be seen as special cases of the Tucker2 model. For instance in the weighted model treated by Sands and Young [1980] it is assumed that the number of components in the first and second mode is equal (i.e. $s = t$), and that the core matrix is diagonal in each of its frontal planes C_k, i.e.

$$c_{pq}^{(k)} = 0, \text{ if } p \neq q \text{ for } p, q = 1, \cdots, s \text{ and } k = 1, \cdots, n.$$

The appropriateness of the weighted model could be investigated by searching for a $K \in K^{s \times s}$, and a $L \in K^{s \times s}$, such that

$$\|D_k - KC_kL'\|^2$$

is zero or appropriately small, where D_k is the diagonal matrix containing the diagonal elements of C_k for $k = 1, \cdots, n$. If such a K and L can be found, then the weighted model is appropriate for the data under investigation.

Technically the estimation of the parameters of the Tucker2 model poses no problems. In the algorithm outlined in Section 7 one simply leaves out the E substep, and inserts the identity matrix for E in the other substeps. Computationally it is, however, more efficient to solve the model directly by the analog of the TUCKALS3 algorithm, than solving the model through the TUCKALS3 algorithm itself. Because of the analogy the proofs of the properties of the TUCKALS2 algorithm are exactly the same as of the TUCKALS3 one. Details are given Kroonenberg & de Leeuw [1977, Note 4].

The Tucker1 Model

Instead of performing a principal component analysis over two or three modes, it is feasible to perform such an analysis over just one mode of the data. This would give the Tucker1 model

$$(25) \qquad\qquad z_{ijk} = \sum_{p=1}^{s} g_{ip}c_{pjk}$$

with $i = 1, \cdots, l, j = 1, \cdots, m$, and $k = 1, \cdots, n$, and with the standard meaning of and restrictions on G. The matrix formulation of this model becomes

$$(26) \qquad\qquad Z = GC,$$

where Z and C are three-mode matrices written with combination modes.

For the case that the horizontal planes of the data matrix Z are similarity matrices, the principal component analysis of the Tucker1 model is identical to the procedure de-

veloped by Tucker & Messick [1963, pp. 336ff]. The Tucker1 model has, of course, wider application as it does not restrict the horizontal planes to be similarity matrices.

There is no need to write a separate program to solve the Tucker1 model since the analysis can be carried out with any principal component program by properly organizing the data input.

10. An Example: The Miller & Nicely Data

The Data

The data from a classical study of confusions of English consonants will be used as an example to show a number of the features of the TUCKALS programs. The data consist of confusions among the 16 most used English consonants under 17 degrading conditions [Miller & Nicely, 1955]. Five North American female subjects served as talkers and as listening crew; when one talked, the other four listened. One syllable stimuli consisting of /a/ (as in f*a*ther) preceded by one of 16 consonants were spoken—/p/, /t/, /k/, /f/, /θ/ (as in *th*ought), /s/, /ʃ/ (as in *sh*ould), /b/, /d/, /g/, /v/, /ð/ (as in *th*at), /z/, /ʒ/ (as in vi*s*ion), /m/, and /n/. The consonants spoken were fed through a transmission circuit which was degraded each of the 17 times in a different way. Notably there were differences in signal-to-noise ratio (or masking), low-pass filtering, and high-pass filtering, some details of which are listed in Table 1. In each condition tested some 4000 observations were collected, but each consonant was not spoken equally often. In our analysis we first corrected for this by dividing each entry by its row total, as each row corresponds to the spoken consonant, while each column corresponds to the heard consonant. The entries in the matrix, therefore, indicate for that particular condition, the proportion of times each of the consonants was heard, when the consonant associated with that row was spoken.

In our analysis we added two more matrices to provide "zero-point" references, i.e. a matrix with perfect discrimination (only entries on the diagonal), and a matrix with total uniform confusion (equal entries in all cells). Strictly speaking the former matrix does not belong to any of the series degrading conditions, as perfect discrimination would probably require increasing both the signal-to-noise ratio above 12 db, and extending the frequency range on the high and the low side. With regard to the latter matrix, we could interpret it as referring to noise coming from just one frequency wave band for any signal-to-noise ratio, or as coming from any frequency band with very low signal-to-noise ratio. It, therefore, would fit any degrading series.

The Miller & Nicely data have been extensively used both in the field of phonetics as support or disproof of the distinctive feature theory, and as demonstration material for various scaling procedures. In the latter class fall most notably Shepard [1972, 1974], Wish [1970, Note 6], Carroll & Wish [1974], Smith [1973], and Smith & Jones [1975]. With respect to the structure of the consonant space we have not much to add to the very detailed analyses of Shepard [1972], and Soli & Arabie [1979]. We give, however, a new interpretation of the dimensions in the noise-condition space. We want to emphasize that our primary aim in presenting our analysis of these data is to demonstrate the developed computer programs, rather than provide a substantive contribution to (acoustic) phonetics. At the same time it should be realized that it is impossible in the present context to do full justice to all the various aspects of the two programs.

Stimulus Spaces

Inspection of the Tucker3 and Tucker2 models shows that the principal component matrices of the first and second mode, respectively spoken consonants and heard consonants, are treated separately. Therefore, it is possible to compare their configurations. It turns out that only small differences were present, indicating that the confusion matrices

TABLE 1

Degradation Conditions

Degradation condition	Signal-to-noise ratio (dB)	Bandwidth (Hz)	Amount of information per matrix
Noise masking conditions			
REF1	–	–	4.00
N1	12	200–6500	3.55
N2	6	200–6500	3.23
N3	0	200–6500	2.81
N4	– 6	200–6500	1.84
N5	–12	200–6500	0.96
N6	–18	200–6500	0.06
REF0	–	–	0.00
Low-pass filtering conditions			
REF1	–	–	4.00
L1 (=N1)	12	200–6500	3.55
L2 (=H1)	12	200–5000	3.20
L3	12	200–2500	2.83
L4	12	200–1200	2.38
L5	12	200– 600	2.18
L6	12	200– 400	1.67
L7	12	200– 300	1.15
REF0		–	0.00
High-pass filtering conditions			
H1 (=L2)	12	200–5000	3.20
H2	12	1000–5000	2.67
H3	12	2000–5000	1.59
H4	12	2500–5000	1.07
H5	12	3000–5000	0.62
H6	12	4500–5000	0.44
REF0	–	–	0.00

Based on Miller & Nicely [1955], and adapted from Carroll & Wish [1974]
REF1 = perfect intelligibility; REF0 = total uniform confusion

are rather symmetric. Therefore, we will discuss for the moment the stimulus spaces as if they were identical, and only come back to their differences later on.

In principal component analysis the number of components to retain is a primary problem, although not as much as in factor analysis. Essentially it amounts to deciding how much of the variation in the data is due to real structure, and how much is merely due to sampling. In three-mode principal component analysis the situation is, however,

TABLE 2

Rotated Stimulus Space for 'Spoken Consonants'

		1	2	3
1	t	22	42	6
2	k	18	42	6
3	p	17	38	5
4	f	4	18	− 2
5	θ	1	15	− 1
6	s	−20	14	− 1
7	ʃ	−85	14	5
8	ʒ	−26	−18	−19
9	z	− 1	−16	−19
10	ð	9	−16	−15
11	v	12	−15	−15
12	b	13	−14	−14
13	g	10	−22	−25
14	d	13	−27	−30
15	n	6	−28	55
16	m	6	−29	63
weights		13	12	10

The decimal points are omitted and the order of the consonants
is different from the one in Miller & Nicely [1955]. The sti-
mulus space was rotated to improve the diagonality of the core
matrix (see also section 2). The sum of all unrotated weights
is equal to one.

more problematic, because of the interwovenness of the three modes in the estimation
procedure. Changing the number of components in one mode implies immediately a dif-
ferent (whether substantially or not) solution of the other modes, as can be clearly seen
from (17) to (22). In the present example, however, it was surprising how stable the solu-
tion was to such changes. Possibly this could indicate a rather clear structure in the data.
The much used criterion based on the amount of variance explained by the components is
also problematic, as the components are eigenvectors not of the original inner-products of
the data, like the eigenvectors in Theorem 2, but they are eigenvectors of P, Q, and R (see
Section 4) which all are functions of the other modes as well. Because of this, care must be
taken when assigning importance of axes on the basis of their weights. The more so be-
cause also the core matrix contains information on the relative importance of the various
components. At present we have not worked out a satisfactory solution to this problem,
and in the case of the Miller & Nicely data we relied mainly on interpretability, fallible as
this may be.

In interpreting the output of analyses such as ours one can look for homogeneous
groups of variables (here: groups of consonants which are very often confused), or search
for meaningful directions (axes) in the stimulus space. We have used both approaches,
and it can be seen from Table 2 and Figure 3 that the consonant stimulus space has both
clearly interpretable axes and subgroups. Figure 3 shows the rotated solution of the
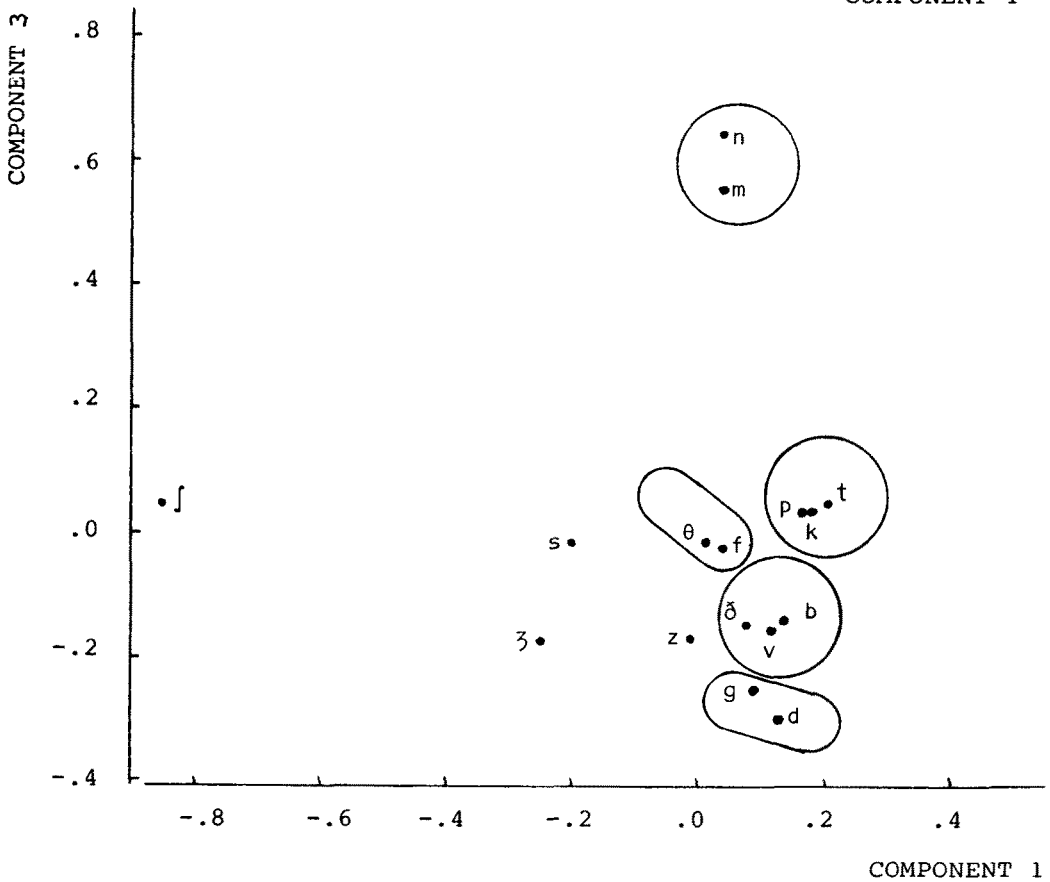TUCKALS2 analysis in three dimensions (the TUCKALS3 solution was virtually identi-
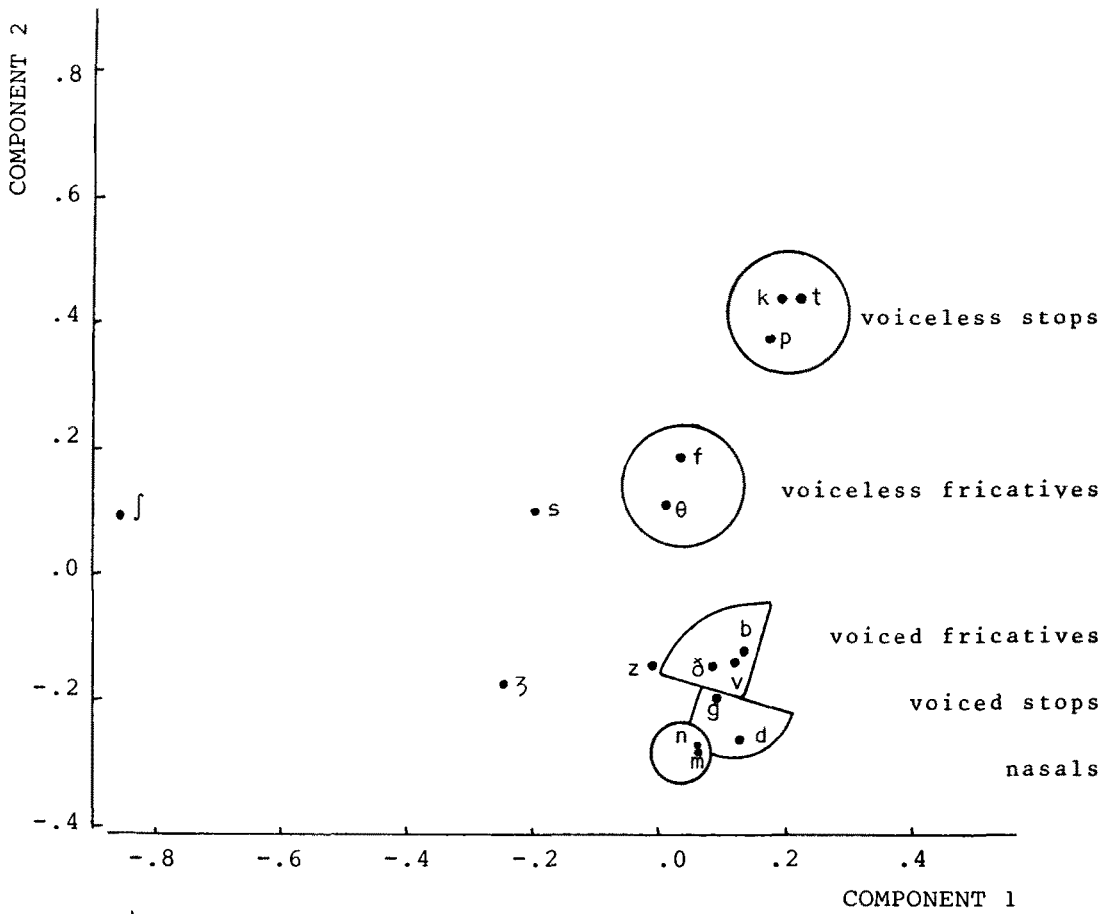
FIGURE 3
Stimulus space for mode 1 (spoken consonants)—rotated

cal). The first component roughly corresponds to the amount of spectrally dispersed acoustic energy located below 5kHz in the speech spectrum (see the discussion in Soli & Arabie, 1979, p. 53). We will refer to this component as "energy" for short. The second axis separates the voiceless stops, voiceless fricatives, voiced fricatives, and the voiced stops (except for /b/, the somewhat aberrant behavior of which was already noted by Shepard, 1972). This axis corresponds to, or is a slightly rotated version of, the "periodicity/burst order" axis from the analysis of Soli & Arabie [1979, p. 51, 52]. The third axis, finally, serves to set off the nasals from the rest, and in addition separates the voiceless and voiced consonants. This axis has no equivalent in the Soli & Arabie study, and we will refer to it as "nasality", realizing that this is an oversimplification. One of the reasons we did not find the same axis as Soli & Arabie in their paper referred to above, is that they used a different more restricted model (INDSCAL), and a log-transformation of the confusion matrices after symmetrization. In addition, they settled for a four-dimensional solution (which they found very interpretable), while our impression was that the fourth axis from a four-dimensional TUCKALS2 solution only served to set off the voiceless fricatives from the voiceless stops, a distinction already contained in the three-dimensional solution. Unfortunately, a detailed comparison could not be made as Soli & Arabie only give a figure of their stimulus space, and not the actual coordinates of the consonants.

Another problem—looking now at the more or less homogeneous subgroups—is to formally separate the groups of consonants. If some type of average similarity matrix would have been available an appropriate cluster analysis could be called to assistance (see, for instance, Shepard, 1972, for the use of such a procedure on the same data after symmetrization of the matrices). In our case we have used another feature of the TUCK-ALS2 program as a rough guide to the grouping. The program generates an "average" matrix on the basis of the components of the first two modes. In Table 4 this "average" confusion matrix—$G\bar{C}H'$—is given. A visual inspection indicates four major, partially overlapping clusters (voiced consonants, voiceless consonants, nasals, and sibilants), and some further distinction within the major clusters (/p/, /t/, /k/; /f/, /θ/; /g/, /d/; /ʃ/; and somewhat vaguely /v/, /b/, /ð/).

The above analysis has been carried out on per noise condition double-centered matrices, i.e.

$$\bar{z}_{ij}^k = z_{ij}^k - z_{.j}^k - z_{i.}^k + z_{..}^k .$$

Another way of looking at the same data using the same program, is to center the data for each heard consonant-spoken consonant combination over all noise conditions (a procedure, for instance, used by Tucker & Messick [1963] in their point-of-view analysis). This will show whether or not the various noise conditions treat the consonants differently. For instance, a consonant which is treated more or less the same in all conditions will now be located close to the center of the configuration. On the other hand, a consonant like /t/, which is treated differently by high-pass and low-pass filters, will have a high loading on one of the components. Figure 5 illustrates this for the Miller & Nicely data.

Figure 6 is a joint plot of the component matrices for the first and second mode of the main analysis showing how well the correspondence is between the understood and spoken consonants. The rather close similarity of the two stimulus spaces shows that symmetrization does not really violate the structure of the data. This is in accordance with the results of Hubert & Baker [1979], who investigated the symmetry of two of the Miller & Nicely confusion matrices, and found that they were not asymmetric. For details on the method to produce plots like Figure 6, see Kroonenberg & de Leeuw [1977, Note 4].

TABLE 4

"Average" Confusion Matrix

Heard consonants

		t	k	p	f	θ	s	ʃ	ʒ	z	ð	v	b	g	d	n	m
	t	11	11	10	4	3	1	−5	−7	−4	−3	−3	−2	−4	−6	−4	−4
	k	11	11	10	4	3	1	−4	−6	−4	−2	−3	−2	−4	−5	−4	−4
	p	10	10	9	4	3	1	−4	−6	−4	−2	−2	−2	−4	−5	−3	−4
	f	4	4	4	2	1	1	−0	−2	−3	−1	−1	−1	−1	−2	−3	−3
	θ	3	3	3	1	1	1	1	−1	−3	−1	−1	−1	−1	−2	−2	−3
spoken consonants	s	1	1	1	1	1	3	9	1	−1	−2	−2	−2	−2	−3	−3	−3
	ʃ	−5	−4	−4	−0	2	9	33	8	−1	−4	−6	−6	−5	−8	−4	−4
	ʒ	−7	−7	−6	−2	−1	2	8	6	3	1	1	1	3	4	−3	−4
	z	−4	−4	−4	−1	−1	−1	−1	3	3	2	3	2	4	5	−3	−3
	ð	−3	−3	−3	−1	−1	−2	−5	2	2	2	3	3	3	5	−1	−2
	v	−2	−3	−2	−1	−1	−2	−6	1	2	2	3	3	4	5	−1	−2
	b	−2	−2	−2	−1	−1	−2	−6	1	2	2	3	3	3	5	−1	−1
	g	−4	−5	−4	−2	−1	−2	−5	3	4	3	4	4	5	7	−3	−4
	d	−5	−5	−5	−2	−2	−3	−7	3	4	4	5	5	6	9	−4	−5
	n	−4	−4	−3	−3	−2	−3	−4	−3	−3	−1	−2	−1	−3	−4	18	21
	m	−4	−4	−3	−3	−2	−3	−4	−4	−3	−2	−2	−2	−3	−5	20	24

The "average" confusion matrix is constructed on the basis of the two com-
ponent matrices G and H, and the average frontal plane of the core matrix,
i.e. GCH'. Each entry indicates the weighted product of the row stimulus
and the column stimulus. High positive values indicate that the row and
column stimulus are often confused. High values on the main diagonal indi-
cate that the consonant is very distinct, and is seldom confused with other
consonants. The decimal points have been omitted from the body of the
table.

Noise Conditions

As is shown in Table 1 we can define three large groups of noise conditions—mask-
ing, filtering high frequencies, and filtering low frequencies. As mentioned by Miller &
Nicely [1955], and confirmed by Shepard [1972] low-pass filters and low signal-to-noise
conditions look somewhat alike, and are both different from high-pass filters. A two-di-
mensional rotated solution from the TUCKALS3 analysis is given in Table 7 and Figure
8. The space of the noise conditions has been rotated in such a way that one of the axes
passes through the two reference points, i.e. uniform total confusion and perfect in-
telligibility.

In their original publication Miller & Nicely used a measure of covariance between
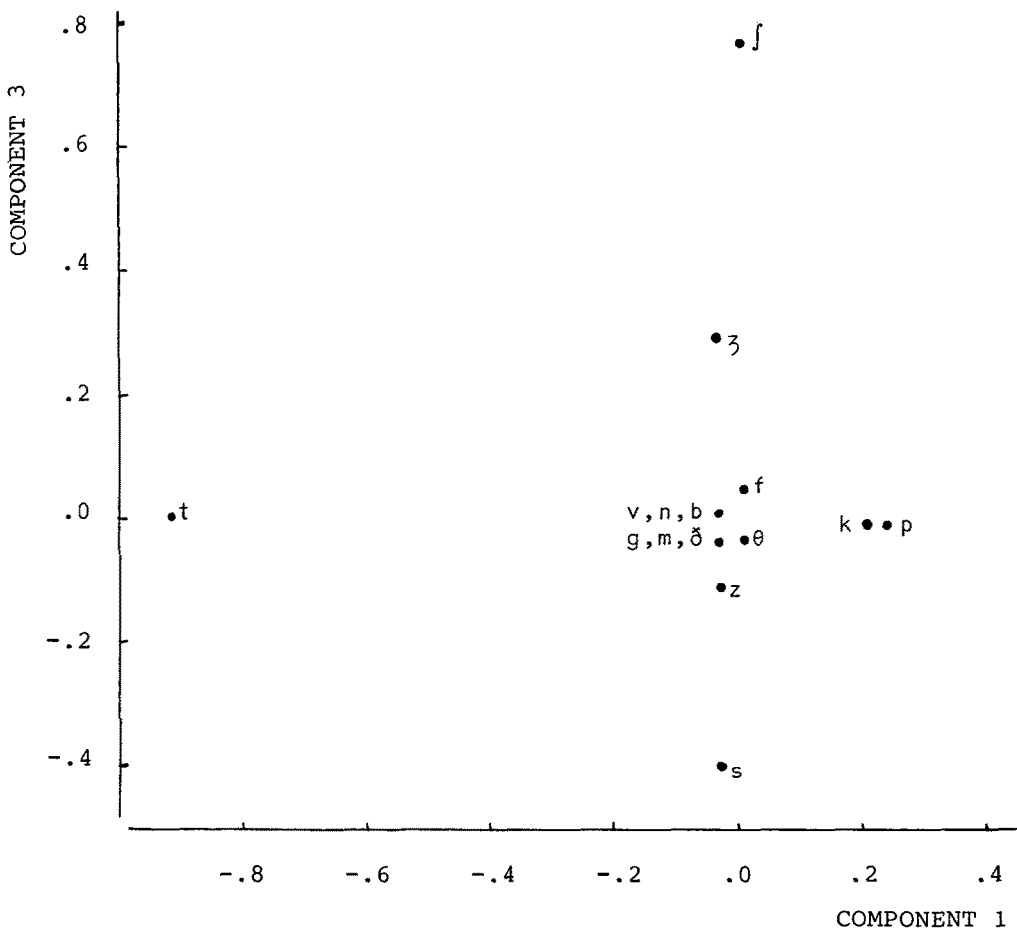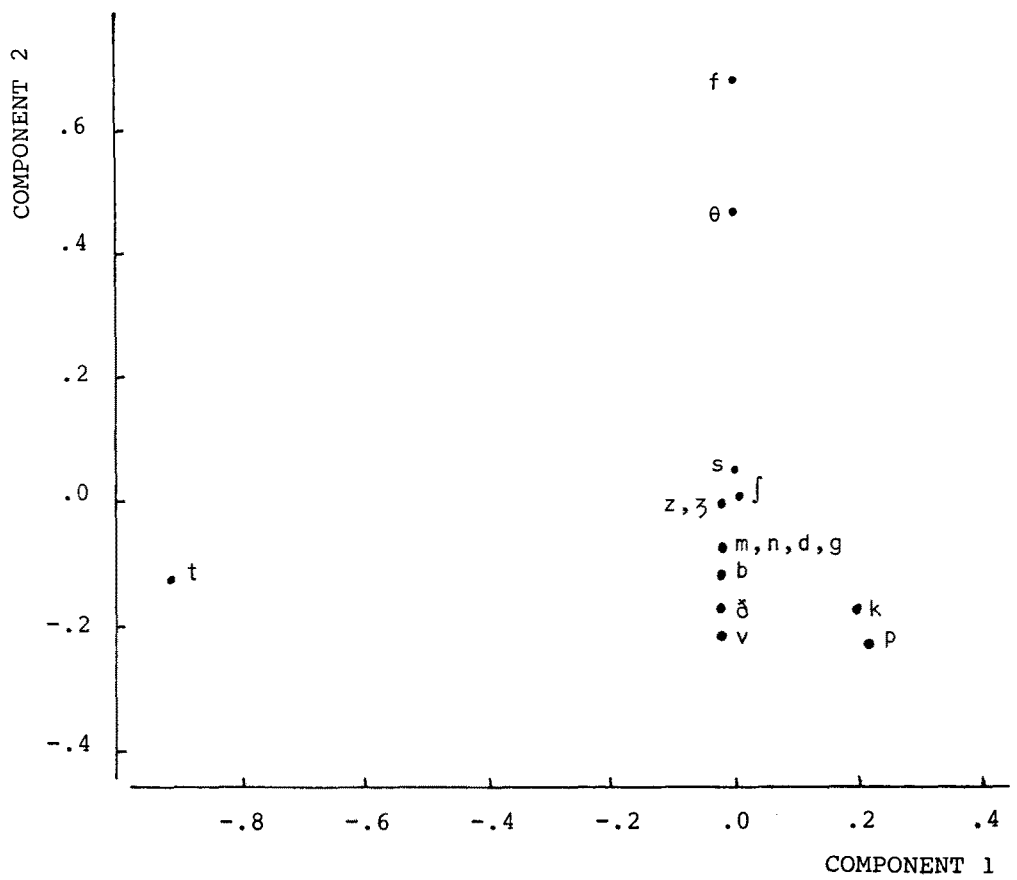
FIGURE 5
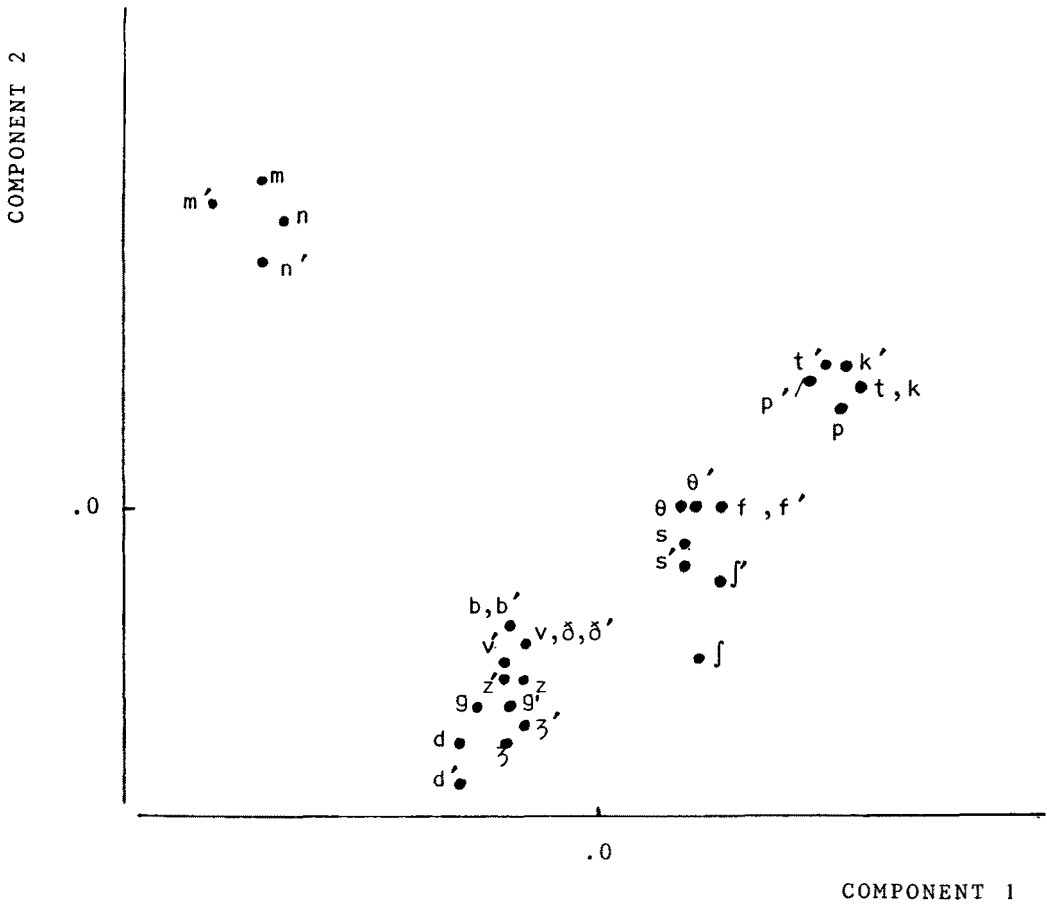Stimulus space for mode 1 (spoken consonants). Data centered per stimulus-stimulus combination.

FIGURE 6

Joint plot of the stimulus spaces of "spoken consonants" and "heard consonants" (the latter is indicated with an apostrophe). The third component is not shown.

input and output to classify the various noise conditions. In particular, this measure is

$$T(x, y) = -\sum_{i,j} p_{ij} \log \frac{p_i p_j}{p_{ij}},$$

where T is often referred to as the amount of information transmitted from input variable x to output variable y in bits per stimulus, and where it is assumed that x takes on discrete values (here: consonants) $i = 1, \cdots, k$ with probability p_i, and similarly y takes on the values $j = 1, \cdots, k$ with probability p_j, and p_{ij} is the probability of the joint occurrence of input i and output j. We have recalculated the values of T for the confusion matrices based on proportions, and these values are listed in Table 1. By trial-and-error a direction in the noise space can be found which corresponds to (a nonlinear transformation of) the amount of information contained in each matrix (see Figure 9B). This direction is indicated in Figure 8. One would have preferred this direction to be one of the axes of Figure 8, but this is unfortunately not the case as is confirmed by Figure 9A. As far as the other component of the rotated noise condition space is concerned, it seems to reflect something like the average frequency of the filtering or masking, but a proper measure to account for the numerical values is not known to us.

TABLE 7

Noise condition space

masking			low-pass filtering			high-pass filtering		
	1	2		1	2		1	2
REF1	30	0	REF1	30	0	REF1	30	0
N1 = L1	30	0	L1 = N1	30	0	H1 = L2	30	1
N2	30	1	L2 = H1	30	1	H2	30	− 7
N3	29	3	L3	28	6	H3	26	−20
N4	24	9	L4	18	32	H4	24	−28
N5	13	25	L5	18	33	H5	21	−40
N6	1	7	L6	18	27	H6	18	−42
REF0	0	0	L7	9	44	REF0	0	0
			REF0	0	0			

The noise condition space has been rotated, such that the first compo-
nent runs through the two reference points, REF0 and REF1. The compo-
nent weights of the unrotated solution equal .31 and .01 respectively,
and all weights sum to zero. The decimal points have been omitted
from the body of the table. The table has been split in three parts to
facilitate comparison within the type of degradation.
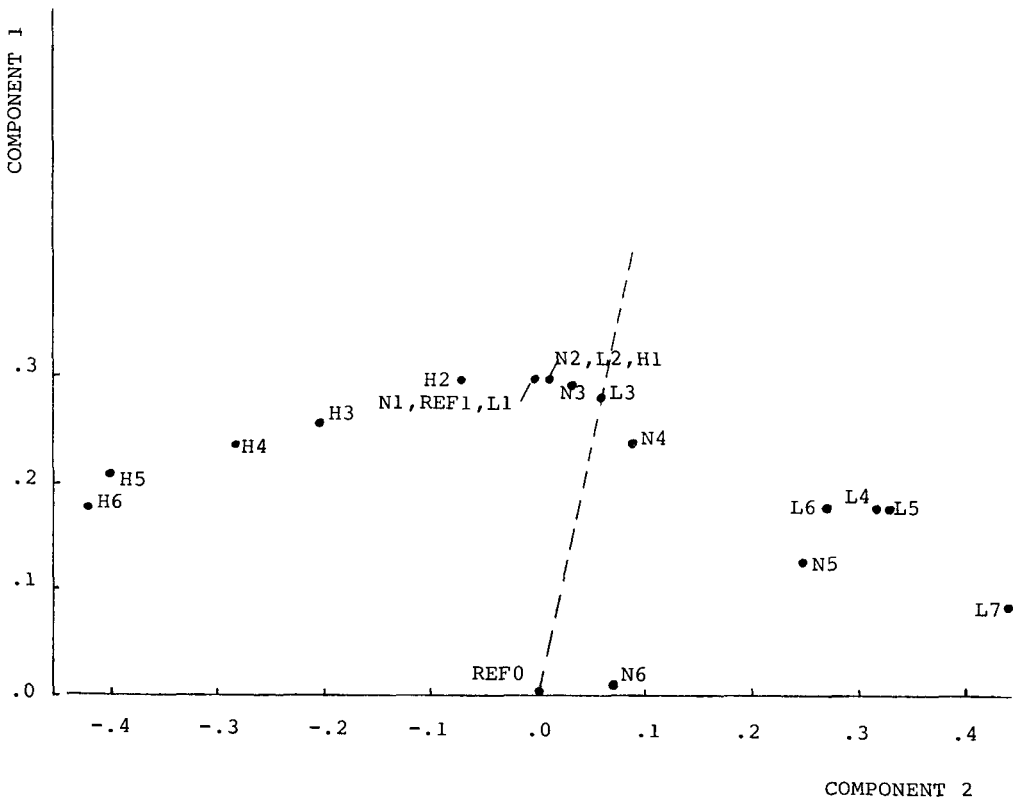

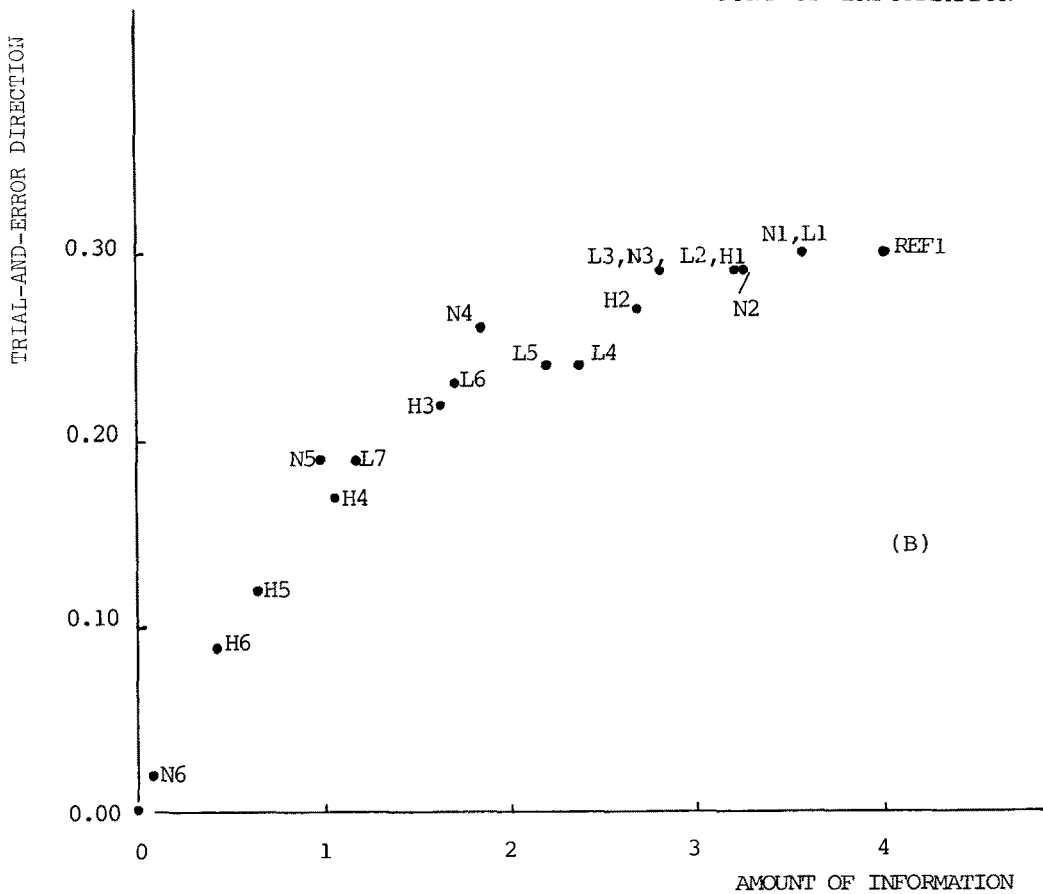FIGURE 8
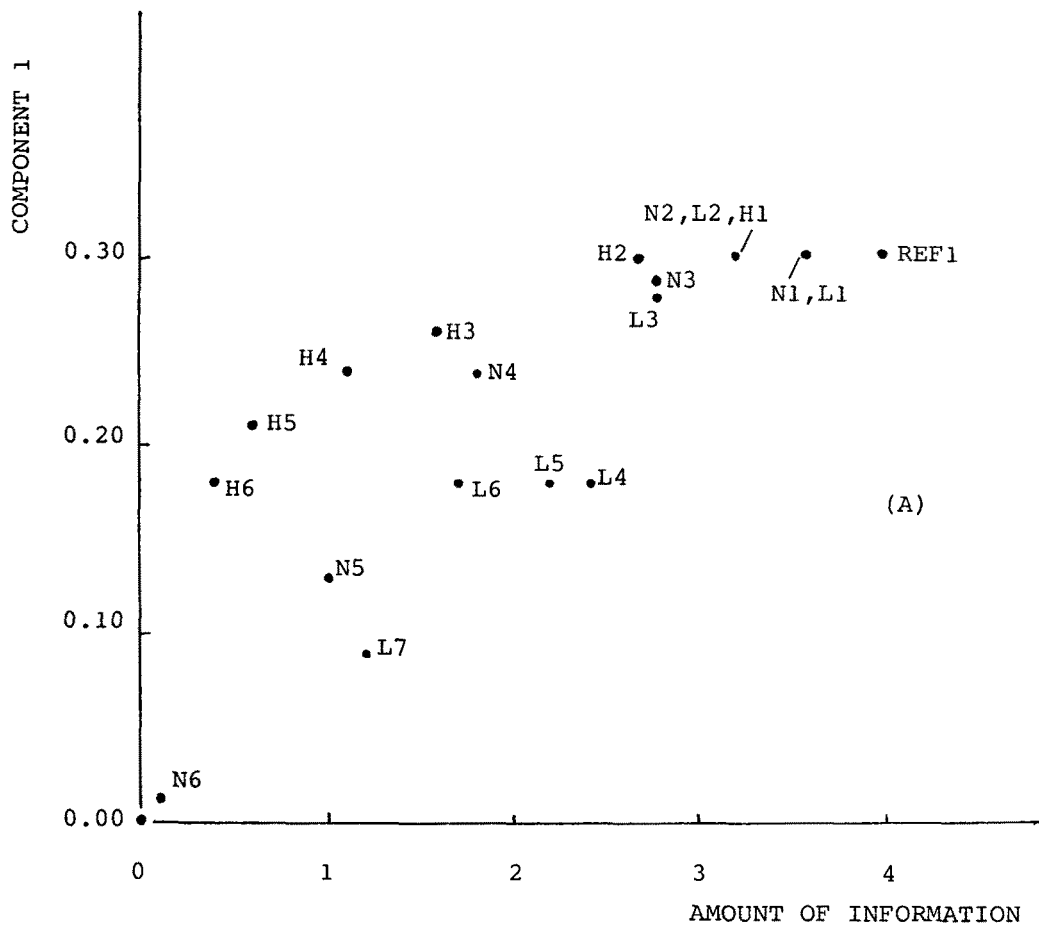Noise-condition space—rotated

FIGURE 9
Relation between the amount of information and directions in the noise-condition space

TABLE 10

Frontal Planes of the Core Matrix from the TUCKALS3 analysis

"amount of information"
(component 1 for noise conditions)

	energy 1	p/b order 2	nasality 3
energy	13	− 0	− 1
periodicity/burst order	1	5	0
nasality	1	− 0	4

"average frequency of unfiltered band"
(component 2 for noise conditions)

	energy 1	p/b order 2	nasality 3
energy	18	− 2	− 2
periodicity/burst order	2	23	1
nasality	2	− 2	23

Decimal points omitted,i.e. 13 is actually 1.3

Core Matrices

Finally we want to say something about the core matrices, both of the TUCKALS3 and the TUCKALS2 analyses. As mentioned above we have performed a number of rotations on the stimulus- and noise-condition spaces. In particular, the component matrices were rotated in such a way that the frontal planes of the core matrix were far more diagonal than before. At the same time this improved the interpretability of the axes of the stimulus spaces of the consonants. In the TUCKALS3 case the noise condition space was rotated in such a way that one of the axes went through the Point L3 (see Figure 8), with the appropriate counterrotations of the core matrix. The final effect of these rotations on the core matrix of the TUCKALS3 analysis is shown in Figure 10, where the frontal planes are shown, and where, as far as possible, the appropriate labels of the components have been added. The main pattern of the frontal planes is that each of the components of the first mode (spoken consonants) is predominantly related to the corresponding one of the second mode (heard consonants), thus although the consonants are often confused they are mainly confused on the basis of the characteristics specified by the components. Secondly all components have their largest loadings in the frontal plane we called for lack of anything better the "average frequency of the unfiltered band" plane, and thirdly that the "energy" component seems to be the only one substantially contributing to the amount-of-information distinctions, where as all three components play an equal role in the other noise-condition component. We are unfortunately not well enough versed in the substantive theory of acoustic phonetics to further interpret these findings. The frontal planes of the core matrix of the TUCKALS2 analysis provide us with the relations be-

tween the first and second mode components for each of the degrading conditions. In a sense these planes provide a summary of how the noise conditions affect the three major components of the stimulus space of the consonants. In Table 11 some typical examples are given.

TABLE 11

Rotated Frontal Planes of Core Matrix from the TUCKALS2 Analysis

Masking

	N6 -18db, 200-6500			N5 -12db, 200-6500		
	e 1	p 2	n 3	e 1	p 2	n 3
energy	1	- 1	1	26	- 4	3
p/b order	- 2	9	- 2	- 1	54	- 4
nasality	0	2	7	3	- 4	47

Low-pass filtering

	L7 12db, 200- 300			L4 12db, 200-1200		
	e 1	p 2	n 3	e 1	p 2	n 3
energy	15	- 5	2	36	2	3
p/b order	- 3	57	- 3	5	65	- 1
nasality	1	- 3	57	1	- 1	66

High-pass filtering

	H6 12db,4500-5000			H3 12db,2000-5000		
	e 1	p 2	n 3	e 1	p 2	n 3
energy	45	- 2	0	60	1	4
p/b order	- 2	10	1	1	47	- 4
nasality	2	1	2	2	0	37

Decimal points omitted, i.e. 26 is actually 2.6

11. Final Remarks

We have shown in the previous sections that it is possible to solve the estimation problem of the general Tucker principal component model with least squares techniques. In addition, the algorithm developed can be used to solve less general problems, as is here exemplified by the so-called Tucker2 model or "generalized subjective metrics" model. The programs based on the algorithm, TUCKALS2 and TUCKALS3, were written in FORTRAN for an IBM 370/158, and can be obtained from the first author.

A major shortcoming of the algorithm described is that it is only suitable for unconditional, metric data, but within the already existing ALS-framework of the algorithm it is possible to expand it with an optimal scaling phase to accommodate other types of data. For a discussion on conditionality and related measurement problems, and on the extension of metric procedures to data with other measurement characteristics the reader is referred to Young, de Leeuw & Takane [1980]. An example how the inclusion of optimal scaling works in the principal component analysis of three-mode data in a very specific case, is described by Sands & Young [1980].

Appendix

Stationary Points

For convenience we have used in this paper a more restricted definition of a stationary point of a function than is customary. If we let S be defined as in (8), and if we let h be a real continuous differentiable function on S, then $(\hat{G},\hat{H},\hat{E}) \in S$ is a stationary point if $(\hat{G},\hat{H},\hat{E})$ is a solution of the stationary equations

$$\frac{\delta}{\delta X}[h(G,H,E) - \mathrm{tr}L(G'G - I_s) - \mathrm{tr}M(H'H - I_t) - \mathrm{tr}N(E'E - I_u)] = 0 ,$$

with $X = G, H, E, L, M, N$ respectively, and L, M, and N are matrices of Lagrange multipliers.

Proof of Theorem 1 (Section 4)

(a.) 1. Let us first determine the stationary equations for

$$p(G,H,E) = \mathrm{tr}G'PG = \mathrm{tr}H'QH = \mathrm{tr}E'RE .$$

Incorporating the constraints on the parameter space into the function to be maximized, we get

$$\bar{p}(G,H,E,L,M,N) = \mathrm{tr}G'PG - \mathrm{tr}L(G'G - I_s) - \mathrm{tr}M(H'H - I_t) - \mathrm{tr}N(E'E - I_u).$$

Differentiating with respect to all parameter matrices, and setting all the derivatives equal to zero, we obtain the following set of equations which have to be solved simultaneously for all parameters.

(27) $\qquad P(\hat{H},\hat{E})\hat{G} = \hat{G}\hat{L}$ and $\hat{G}'\hat{G} = I_s$,

(28) $\qquad Q(\hat{E},\hat{G})\hat{H} = \hat{H}\hat{M}$ and $\hat{H}'\hat{H} = I_t$,

(29) $\qquad R(\hat{G},\hat{H})\hat{E} = \hat{E}\hat{N}$ and $\hat{E}'\hat{E} = I_u$.

To simplify the notation we will drop the carets from now on. Note that L, M, and N are necessarily symmetric, because, for instance, the restriction $g_i'g_j = \delta_{ij}$ is identical to the restriction $g_j'g_i = \delta_{ji}$, where g_i is the ith column of G, and δ_{ij} is the Kronecker delta.

(a.) 2. As G and L are solutions of (27), it follows from $PG = GL$, that $L = G'PG$. Furthermore L is positive definite (because P is positive definite), and in addition, because L is symmetric there exists a $F \in K^{s \times s}$ such that $L = F\Lambda F'$ with $\Lambda \in D^{s \times s}$, where $D^{s \times s}$ is the class of $s \times s$ diagonal matrices. Substituting this into (27), and postmultiplying with F we get $PGF = GF\Lambda$. By defining $U = GF$ (and thus $G = UF'$) it follows that Λ is an eigenvalue matrix of P, U is the associated eigenvector matrix, and G is an orthonormal transformation of U. Analogously there exists a $\bar{F} \in K^{t \times t}$ such that $H = V\bar{F}'$, and $M = \bar{F}\bar{\Lambda}\bar{F}'$, and there exists a $\tilde{F} \in K^{u \times u}$ such that $E = W\tilde{F}'$, and $N = \tilde{F}\tilde{\Lambda}\tilde{F}'$.

(a.) 3. Conversely, if we let U, V, W, and Λ, $\bar{\Lambda}$, $\tilde{\Lambda}$ be eigenvector matrices and eigenvalue matrices of P, Q, and R respectively, then (U,V,W) as well as their orthonormal transformations (G,H,E) with $G = UF'$, $H = V\bar{F}'$, and $E = W\tilde{F}'$ with the F's defined as above, satisfy (27) through (29), and thus are stationary points of p.

(b.) 1. Let T be defined as

$$T = \{t \mid t = (G,H,E); \; G, \; H, \text{ and } E \text{ are eigenvector matrices of } P, \; Q, \text{ and } R$$
respectively, or orthornormal transformations thereof$\}$.

We already know that there exists a $(G,H,E) \in S$ such that p attains its maximum (see Section 3). Now we can state that this maximum will, and can only be attained for some $(G,H,E) \in T$.

(b.) 2. Any $(G,H,E) \in T$ can be written as $(UF',V\bar{F}',W\tilde{F}')$ with the F's defined as above. Thus

$$p(G,H,E) = p(UF',V\bar{F}',W\tilde{F}') = \text{tr} F U'Z(V\bar{F}'\bar{F}V' \otimes W\tilde{F}'\tilde{F}W')Z'UF'$$

$$= \text{tr } U'Z(VV' \otimes WW')Z'U = p(U,V,W).$$

(b.) 3. Let $(G,H,E) \in T$ be the point at which p attains its maximum, then

$$p(\hat{G},\hat{H},\hat{E}) = p(\hat{U},\hat{V},\hat{W}) = \max_{U} p(U,\hat{V},\hat{W}) = \max_{U} \text{tr} U'P(\hat{V},\hat{W})U$$

$$= \max_{\lambda} \sum_{i=1}^{s} \lambda_i$$

(with the maximum taken over all possible ways to combine s of the total of l eigenvalues of $(P(\hat{V},\hat{W})$)

$$= \sum_{i=1}^{s} \hat{\lambda}_i$$

(where $\hat{\lambda}_i$ ($i = 1, \cdots , s$) are the s largest eigenvalues of $P(\hat{V},\hat{W})$).

Thus $\hat{U}$ must be the eigenvector matrix corresponding to these largest eigenvalues. Analogously $\hat{V}$ and $\hat{W}$ are the eigenvector matrices corresponding to the largest eigenvalues of $Q(\hat{W},\hat{U})$ and $R(\hat{U},\hat{V})$. The value of the maximum is

$$\sum_{i=1}^{s} \hat{\lambda}_i = \sum_{j=1}^{t} \hat{\mu}_j = \sum_{k=1}^{u} \hat{\nu}_k \, ,$$

where $\hat{\mu}$ and $\hat{\nu}$ are analogously defined as $\hat{\lambda}$.

(b.) 4. Conversely, let $\hat{U}$, $\hat{V}$, and $\hat{W}$ be the eigenvector matrices corresponding to the largest s, t, and u eigenvalues of $P(\hat{V},\hat{W})$, $Q(\hat{W},\hat{U})$, and $R(\hat{U},\hat{V})$. Furthermore let $\hat{G} = \hat{U}F'$, $\hat{H} = \hat{V}\bar{F}'$, and $\hat{E} = \hat{W}\tilde{F}'$ with the F's defined as above. Then

$$p(\hat{U},\hat{V},\hat{W}) = \sum_{i=1}^{s} \hat{\lambda}_i = \sum_{j=1}^{t} \hat{\mu}_j = \sum_{k=1}^{u} \hat{\nu}_k = \max_{(U,V,W)} p(U,V,W) .$$

As $p(\hat{G},\hat{H},\hat{E}) = p(\hat{U},\hat{V},\hat{W})$,

$$\max_{(U,V,W)} p(U,V,W) = \max_{(G,H,E)} p(G,H,E),$$

thus if $(\hat{U},\hat{V},\hat{W})$ maximizes p, then any orthornormal transformations of $\hat{U}$, $\hat{V}$, and $\hat{W}$ do so as well.

Proof of Theorem 2 (b.) 2 (Section 4)

Let $(\hat{G},\hat{H},\hat{E},\hat{C})$ be an exact solution of the minimization problem (3), then according to Theorem 2 part (a.):

$$(30) \qquad Z = \hat{G}\hat{G}'Z(\hat{H}\hat{H}' \otimes \hat{E}\hat{E}') \text{ with } Z \in R^{l\times mn}.$$

Theorem 1 part (b.) states that $\hat{G}$, $\hat{H}$, $\hat{E}$ are the eigenvector matrices associated with the s, t, and u largest eigenvalues of P, Q, and R respectively. If we define $\Lambda \in D^{s\times s}$ to be the eigenvalue matrix of P associated with $\hat{G}$, we have

$$(31) \qquad \hat{G}\Lambda\hat{G}' = P = Z(\hat{H}\hat{H}' \otimes \hat{E}\hat{E}')Z' \text{ with } Z \in R^{l\times mn}.$$

Pre- and postmultiplying (31) with $\hat{G}\hat{G}'$, and subsequently substituting this in (30) shows that

$$\hat{G}\Lambda\hat{G}' = ZZ'.$$

In other words $\hat{G}$ is the eigenvector matrix of ZZ', and Λ the associated eigenvalue matrix. Furthermore the rank of $\hat{G}$ ($= s$) is equal to that of ZZ', and thus the λ_i ($i = 1,\cdots,s$) are the s non-zero eigenvalues of ZZ'. The analogous result holds for $\hat{H}$ and $\hat{E}$.

The Monotonicity of F

   Theorem 4. Let ϕ be of the form

$$(32) \qquad \phi(X) = AX(X'A^2X)^{-1/2}$$

with $X'A^2X$ positive definite, and let h be defined as

$$(33) \qquad h(X,Y) = \text{tr} X'AY$$

with $X,Y \in K^{a\times b}$, and $A \in R^{a\times a}$ symmetric. If $Y = \phi(X)$, then

$$(34) \qquad h(Y,Y) \geq h(X,X),$$

with equality if, and only if $Y = X$.

   Proof. As $X'A^2X$ is positive definite (see also Section 6), its inverse exists, and thus ϕ is uniquely defined.

   a. We first show that for every $X \in K^{a\times b}$

$$(35) \qquad h(Y, X) = \max_{Z\in K^{a\times b}} h(Z, X).$$

To do this we incorporate the constraints on Z into the maximization

$$(36) \qquad h(Z, X) = h(Z, X) - \text{tr} M(Z'Z - I_b)$$

where M is a symmetric matrix of Lagrange multipliers. Differentiating with respect to Z and M, and setting all the partial derivatives equal to zero, we obtain the following set of equations which have to be solved simultaneously

$$(37) \qquad AX = 2ZM,$$

(38) $Z'Z = I_b.$

Say that some $(\hat{Z}, \hat{M})$ is a solution of the system. Then by premultiplying (37) by its transpose, reminding ourselves of the symmetry of A and $\hat{M}$, and substituting (38) into (37) we get

$$X'A^2X = 4\hat{M}^2,$$

and

$$\hat{Z} = AX(X'A^2X)^{-1/2} = \phi(X) = Y$$

according to (32). Thus for any $X \in K^{a \times b}$ Y maximizes h, or in other words

$$h(Y, X) = \max_{Z \in K^{a \times b}} h(Z, X) \geq h(X, X)$$

for all $X \in K^{a \times b}$.

  b. Next we show that

$$h(Y, X) \leq h(Y, Y)^{1/2} h(X, X)^{1/2}.$$

As A is symmetric it may be decomposed into $A = B'B$, where B is an upper-triangular matrix. Thus

$$h(Y, X) = \operatorname{tr} Y'AX = \operatorname{tr}(BY)'(BX).$$

The Cauchy-Schwarz inequality can now be applied

$$\operatorname{tr} Y'AX = \operatorname{tr}(BY)'(BX) \leq \{\operatorname{tr}(BY)'(BY)\}^{1/2} \{\operatorname{tr}(BX)'(BX)\}^{1/2}$$

$$= (\operatorname{tr} Y'AY)^{1/2}(\operatorname{tr} X'AX)^{1/2},$$

and thus

$$h(Y, X) \leq h(Y, Y)^{1/2} h(X, X)^{1/2}$$

  c. Now we can prove inequality (34):

$$h(X, X) \leq h(Y, X) \leq h(Y, Y)^{1/2} h(X, X)^{1/2}.$$

As h is always non-negative

$$h(X, X)^{1/2} \leq h(Y, Y)^{1/2},$$

and thus

$$h(X, X) \leq h(Y, Y).$$

  d. In the Cauchy-Schwartz inequality the equality sign holds if and only if X and Y are proportional. Inspection shows that the only possible proportionality constant is 1.

  The extension to the monotonicity of F is straightforward. The equality condition can be seen to hold if one applies Theorem 4 successively in each substep of the algorithm, arriving finally at the conclusion that the equality sign holds if and only if $F(s) = s$.

Continuity of F

  From the definitions of F, F_i, and ϕ_i it follows that if all ϕ_i are continuous, all F_i and F will be as well. It is thus sufficient to show that ϕ as defined in (32) is continuous for all $X \in K^{a \times b}$, as all ϕ_i are of the form (32).

Theorem 5. Let A be a given symmetric matrix, and let $X'A^2X$ be positive definite. Let ϕ be defined as in (32), then ϕ is continuous.

Proof. As $X'A^2X$ is positive definite, its inverse exists, and thus ϕ is uniquely defined. Let $\bar{X}$ be an arbitrary point in $K^{a\times b}$, and let X_0, X_1, X_2, $\cdots$ be a sequence in $K^{a\times b}$, which converges to $\bar{X}$, such that $X_i \neq \bar{X}(i = 0, 1, 2, \cdots)$. Define $Y_l = \phi(X_l)$, $l = 0, 1, 2, \cdots$. For each l part a of Theorem 4, shows that

$$\operatorname{tr} Y_l'AX_l \geq \operatorname{tr} Z'AX_l$$

for all $Z \in K^{a\times b}$. Because the sequence Y_0, Y_1, Y_2, $\cdots$ is defined on a compact set, there exists at least one limit point, say $\tilde{Y}$, in $K^{a\times b}$. In addition, there exists a subsequence Y_{l_0}, Y_{l_1}, Y_{l_2}, $\cdots$ which converges to $\tilde{Y}$. For such subsequences it is true that

$$\operatorname{tr} \tilde{Y}'A\bar{X} = \lim_{j\to\infty} \operatorname{tr} Y_{l_j}'AX_{l_j} \geq \lim_{j\to\infty} \operatorname{tr} Z'AX_{l_j} = \operatorname{tr} Z'A\bar{X}$$

for all $Z \in K^{a\times b}$, and thus

$$(39) \qquad \operatorname{tr} \tilde{Y}'A\bar{X} = \max_{Z\in K^{a\times b}} \operatorname{tr} Z'A\bar{X}.$$

In Theorem 4 part a it was shown that if $Y = \phi(X)$

$$\operatorname{tr} Y'AX = \max_{Z\in K^{a\times b}} \operatorname{tr} Z'AX$$

for each $X \in K^{a\times b}$. This also holds for $\bar{X}$ and $\bar{Y} = \phi(\bar{X})$,

$$(40) \qquad \operatorname{tr} \bar{Y}'A\bar{X} = \max_{Z\in K^{a\times b}} \operatorname{tr} Z'A\bar{X}.$$

Comparing (39) and (40) we may conclude that $\tilde{Y} = \bar{Y}$. Thus we know now that every convergent subsequence of Y_0, Y_1, Y_2, $\cdots$ has $\bar{Y}$ as its limit point, and therefore $\bar{Y}$ is the limit point of Y_0, Y_1, Y_2, $\cdots$ itself. Thus we may conclude that for each $\bar{X} \in K^{a\times b}$, and each X_0, X_1, X_2, $\cdots$ converging to $\bar{X}$, the sequence Y_0, Y_1, Y_2, $\cdots$ converges to the limit point $\bar{Y}$, which means that $\phi(X_0)$, $\phi(X_1)$, $\phi(X_2)$, $\cdots$ converges to $\phi(\bar{X})$. Recalling the definition of continuity of a function, we see that ϕ is continuous for each $\bar{X} \in K^{a\times b}$, and therefore ϕ is continuous on $K^{a\times b}$.

Limit Points of the Algorithm are Stationary Points of p

Theorem 5. Let Z, G, H, E, P, Q, R, and p be defined as in the previous sections. If $s = (G, H, E)$ is a limit point of the algorithm, then s is a stationary point of p.

Proof. Let $s = (G, H, E)$ be a limit point of the algorithm, then $F(s) = s$ (see Lemma 3), and thus

$$G = PG(G'P^2G)^{-1/2}$$

according to (18), and the parallel statements hold for H and E on the basis of (20) and (22). Define $L = (G'P^2G)^{1/2}$, then (G, L) is a solution of

$$(41) \qquad P\tilde{G} = \tilde{G}\tilde{L} \quad \text{and} \quad \tilde{G}'\tilde{G} = I_s.$$

As L is symmetric, there exists a $F \in K^{s\times s}$, such that $L = F\Lambda F'$ with $\Lambda \in D^{s\times s}$. Substituting this in (41) we get

$$PG = GF\Lambda F'$$

which leads to

$$P(GF) = (GF)\Lambda \quad \text{or} \quad P\hat{G} = \hat{G}\Lambda \quad \text{with} \quad \hat{G} = GF.$$

Thus $\hat{G}$ is a matrix with eigenvectors of P, and G is an orthonormal transformation of $\hat{G}$. The analogous result holds for H and E. Theorem 1 part (a) tells us that (G, H, E) is a stationary point of p.

REFERENCE NOTES

1. Carroll, J. D. & Chang, J. J. *IDIOSCAL: A generalization of INDSCAL allowing IDIOsyncratic reference systems as well as an analytic approximation to INDSCAL.* Paper presented at the Spring Meeting of the Psychometric Society, Princeton, N. J., March 1972.
2. Harshman, R. A. *Foundations of the PARAFAC procedure: Models and conditions for an "explanatory" multimode factor analysis* (Working Papers in Phonetics No. 16). Los Angeles: University of California, 1970.
3. Jennrich, R. *A generalization of the multidimensional scaling model of Carroll & Chang* (Working Papers in Phonetics No. 22). Los Angeles: University of California, 1972.
4. Kroonenberg, P. M. & de Leeuw, J. *TUCKALS2: A principal component analysis of three mode data* (Res. Bull. RB 001-77). Leiden: Department of Data Theory, University of Leiden, 1977.
5. Levin, J. *Three-mode factor analysis* (Unpublished doctoral thesis). Urbana, Ill.: University of Illinois, 1963.
6. Wish, M. *An INDSCAL analysis of the Miller & Nicely consonant confusion data.* Paper presented at meetings of the Acoustical Society of America. Houston, November, 1970.

REFERENCES

Carroll, J. D. & Chang, J. J. Analysis of individual differences in multidimensional scaling via an N-way generalization of "Eckart-Young" decomposition. *Psychometrika,* 1970, *35,* 283–320.

Carroll, J. D. & Wish, M. Models and methods for three-way multidimensional scaling. In D. H. Krantz, R. D. Luce, R. C. Atkinson, & P. Suppes (Eds.), *Contemporary developments in mathematical psychology* (Vol. II). San Francisco: W. H. Freeman, 1974.

d'Esopo, D. A. A convex programming procedure. *Naval Research Logistics Quarterly,* 1959, *11,* 33–42.

Hubert, L. J. & Baker, F. B. Evaluating the symmetry of a proximity matrix. *Quality & Quantity,* 1979, *13,* 77–84.

Israelsson, A. Three-way (or second order) component analysis. In H. Wold & E. Lyttkens (Eds.), Nonlinear iterative partial least-squares (NIPALS) estimation procedures. *Bulletin of the International Statistical Institute,* 1969, *43,* 29–51.

Meyer, R. R. The validity of a family of optimization methods. *SIAM Journal of Control and Optimization,* 1970, *15,* 699–715.

Miller, G. A. & Nicely, P. E. An analysis of perceptual confusion among some English consonants. *Journal of the Acoustical Society of America,* 1955, *27,* 338–352.

Osgood, C. E., Suci, G. J. & Tannenbaum, T. H. *The measurement of meaning.* Urbana, Ill.: University of Illinois Press, 1957.

Ostrowski, A. M. *Solution of equations and systems of equations.* New York: Academic Press, 1966.

Penrose, R. On the best approximate solutions of linear matrix equations. *Proceedings of the Cambridge Philosophical Society,* 1955, *51,* 406–413.

Rutishauser, H. Computational aspects of F. L. Bauer's simultaneous iteration method. *Numerische Mathematik,* 1969, *13,* 4–13.

Sands, R. & Young, F. W. Component models for three-way data: ALSCOMP3, an alternating least squares algorithm with optimal scaling features. *Psychometrika,* 1980, *45,* 39–67.

Schwartz, H. R., Rutishauser, H. & Stiefel, E. *Numerik. Symmetrischer matrizen.* Stuttgart: Teubner, 1968.

Shepard, R. N. Psychological representation of speech sounds. In E. E. David & P. B. Denes (Eds.), *Human communication. A unified view.* New York: McGraw Hill, 1972.

Shepard, R. N. Representation of structure in similarity data: Problems and prospects. *Psychometrika,* 1974, *39,* 373–421.

Smith, P. T. Feature-testing models and their application to perception and memory for speech. *Quarterly Journal of Experimental Psychology,* 1973, *25,* 511–534.

Smith, P. T. & Jones, K. F. Some hierarchical scaling methods for confusion matrix analysis II. Application to large matrices. *British Journal of Mathematical and Statistical Psychology,* 1975, *28,* 30–45.

Soli, S. D. & Arabie, P. Auditory versus phonetic accounts of observed confusions between consonant phonemes. *Journal of the Acoustical Society of America,* 1979, *66,* 46–59.

Takane, Y., Young, F. W. & de Leeuw, J. Non-metric individual differences multidimensional scaling: An alternating least squares method with optimal scaling features. *Psychometrika*, 1977, *42*, 7-67.

Tucker, L. R. Implications of factor analysis of three-way matrices for measurement of change. In C. W. Harris (Ed.), *Problems in measuring change*. Madison, Wis.: University of Wisconsin Press, 1963.

Tucker, L. R. The extension of factor analysis to three-dimensional matrices. In H. Gulliksen & N. Frederiksen (Eds.), *Contributions to mathematical psychology*. New York: Holt, Rinehardt & Winston, 1964.

Tucker, L. R. Some mathematical notes on three-mode factor analysis. *Psychometrika*, 1966, *31*, 279-311.

Tucker, L. R. Relations between multidimensional scaling and three-mode factor analysis. *Psychometrika*, 1972, *37*, 3-27.

Tucker, L. R. & Messick, S. An individual difference model for multidimensional scaling. *Psychometrika*, 1963, *28*, 333-367.

Wainer, H., Gruvaeus, G. & Blair, M. TREBIG: A 360/75 FORTRAN program for three-mode factor analysis for big data sets. *Behavioral Research Methods and Instrumentation*, 1974, *6*, 53-54.

Wainer, H., Gruvaeus, G. & Snijder, F. TREMOD: A 360/75 program for three-mode factor analysis. *Behavioral Science*, 1971, *16*, 421-422.

Walsh, J. A. An IBM 709 program for factor analyzing three-mode matrices. *Educational and Psychological Measurement*, 1964, *24*, 669-773.

Walsh, J. A. & Walsh, R. A revised Fortran program for three-mode factor analysis. *Educational and Psychological Measurement*, 1976, *36*, 169-170.

Young, F. W., de Leeuw, J. & Takane, Y. Quantifying qualitative data. In E. D. Lantermann & H. Feger (Eds.), *Similarity and choice*. Bern: Huber, 1980 (in press).