The Common Vector Approach and its Relation to Principal Component Analysis

M. Bilginer Gülmezoğlu, Vakıf Dzhafarov, and Atalay Barkana, *Member, IEEE*

***Abstract*—The main point of the paper is to show the close relation between the nonzero principal components and the difference subspace together with the complementary close relation between the zero principal components and the common vector. A common vector representing each word-class is obtained from the eigenvectors of the covariance matrix of its own word-class; that is, the common vector is in the direction of a linear combination of the eigenvectors corresponding to the zero eigenvalues of the covariance matrix. The methods that use the nonzero principal components for recognition purposes suggest the elimination of all the features that are in the direction of the eigenvectors corresponding to the smallest eigenvalues (including the zero eigenvalues) of the covariance matrix whereas the common vector approach suggests the elimination of all the features that are in the direction of the eigenvectors corresponding to the largest, all nonzero eigenvalues of the covariance matrix.**

***Index Terms*—Common vector approach, speech recognition, subspace methods.**

I. Introduction

A VOICE signal contains inter- and intra-speaker differences as well as the acoustical environment effects and the phase or temporal differences. A method has been presented to subtract all the differences among the various utterances of isolated words pronounced by different speakers in our previous work [1]. The difference vectors for each word belonging to a word-class in the training set constitute a unique difference subspace [1]. After all the differences for each word belonging to a certain word-class in the training set are subtracted and eliminated, only one vector containing the invariant features for each word-class remains, and we called this vector the "common vector" [1], [2]. Each common vector, which is unique [1] for each of the words in the training set, represents the common properties of the corresponding word-class. However, in the test set, when all the differences in the difference subspace are taken away, the leftover vectors are not unique, and we call those the "remaining vectors." The experimental studies have shown that the remaining vector is usually closer to the common vector of its own word-class than to the common vectors of other word-classes; therefore, the common vector approach (CVA) can be used in isolated word recognition.

In this section, short reviews of the previous work which include the computation of the difference subspace and the common vector [1] and, also a short review of the principal component analysis (PCA) and its application are given. In the second section, it is shown that the common vector satisfies the eigenvalue-eigenvector equation of the covariance matrix corresponding to the zero eigenvalues. It is also shown that the common vector can be obtained from the feature vector of a certain word-class by subtracting all of its components that are in the directions of the eigenvectors corresponding to the nonzero eigenvalues of the covariance matrix. The comparison of the CVA and the SELFIC (SELf-Featuring Information Comparison) which is the subspace method that uses the PCA is provided with an example in two-dimensional (2-D) feature space in the third section. In the fourth section, the experimental studies for the TI-digit database are given. The conclusion part gives the relation between the nonzero principal components and the difference subspace together with the relation between the zero principal components and the common vector. This part also emphasizes the differences between the results of the SELFIC method and of the CVA.

A. Difference Subspace

The feature vectors[1] for one of the word-classes in the training set are given as linearly independent vectors $\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_m$, where each $\mathbf{a}_i \in \mathbf{R}^n$ $(i = 1, 2, \ldots, m)$ is $n \times 1$ column vector and $m \leq n$. This n-dimensional feature space can be divided into an $(m-1)$ dimensional difference subspace $\mathbf{B}$ and an $(n-m+1)$ dimensional orthogonal indifference subspace $\mathbf{B}^{\perp}$ so that the direct sum of these two subspaces would cover the whole feature space.

One way to define the $(m-1)$ dimensional difference subspace $\mathbf{B}$ is to take the differences between the feature vectors, i.e.,

$$\mathbf{b}_i = \mathbf{a}_{i+1} - \mathbf{a}_1 \qquad (i = 1, 2, \ldots, m-1).$$

$\mathbf{B}$ is spanned by these difference vectors. Since $\mathbf{b}_1, \mathbf{b}_2, \ldots, \mathbf{b}_{m-1}$ are not expected to be orthonormal, an orthonormal basis vector set can be obtained by using Gram–Schmidt orthogonalization method [4], [5]. The basis vector set for $\mathbf{B}$ will be $\{\mathbf{z}_1, \mathbf{z}_2, \ldots, \mathbf{z}_{m-1}\}$ in this case. $\mathbf{B}$ is independent of the choice of subtrahand vector $\mathbf{a}_1$ [1].

[1]The vector that has features of a word as its elements is called the feature vector throughout this paper. For the sake of clarity in the notation, vectors will be referred to in boldface in the following discussion.

Manuscript received January 26, 2000; revised April 27, 2001. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Hynek Hermansky.

M. B. Gülmezoğlu and A. Barkana are with the Electrical and Electronics Engineering Department, Osmangazi University, Eskişehir, Turkey (e-mail: bgulmez@ogu.edu.tr).

V. Dzhafarov is with the Department of Mathematics, Anadolu University, Eskişehir, Turkey (e-mail: vcaferov@anadolu.edu.tr).

Publisher Item Identifier S 1063-6676(01)07434-X.

B. Computation of the Common Vector

If the common vector is called as $\mathbf{a}_{com}$, then each of the feature vectors can be written as

$$\mathbf{a}_i = \mathbf{a}_{i,\,dif} + \mathbf{a}_{com} + \epsilon_i \qquad (i = 1, 2, \ldots, m)$$

where ϵ_i is the error vector term. Even if all of the error vectors ϵ_i $(i = 1, 2, \ldots, m)$ are assumed to be zero, obviously there are m vector equations with $(m+1)$ unknown vectors. Therefore there are infinitely many solutions for the common and difference vectors. To obtain a unique solution for the common vector, one may make an assumption that the vectors $\mathbf{a}_{i,\,dif}$ are the projections of the feature vectors onto the difference subspace $\mathbf{B}$; that is

$$\mathbf{a}_{i,\,dif} = \langle \mathbf{a}_i, \mathbf{z}_1 \rangle \mathbf{z}_1 + \langle \mathbf{a}_i, \mathbf{z}_2 \rangle \mathbf{z}_2 + \cdots + \langle \mathbf{a}_i, \mathbf{z}_{m-1} \rangle \mathbf{z}_{m-1}$$

where $\langle \mathbf{a}, \mathbf{z} \rangle = a_1 z_1 + a_2 z_2 + \cdots + a_n z_n$ denote the scalar product of $\mathbf{a} = (a_1, a_2, \ldots, a_n) \in \mathbf{R}^n$ and $\mathbf{z} = (z_1, z_2, \ldots, z_n) \in \mathbf{R}^n$. Then a metric can be established to minimize the squares of the norms of the error vectors in order to obtain a unique solution for the common vector, that is

$$\mathbf{F} = \sum_{i=1}^{m} \|\epsilon_i\|^2 = \sum_{i=1}^{m} \|\mathbf{a}_i - \mathbf{a}_{i,\,dif} - \mathbf{a}_{com}\|^2$$

where $\|\epsilon_i\| = \langle \epsilon_i, \epsilon_i \rangle^{1/2}$ denote the Euclidean norm of the vector ϵ_i. It was shown that if the common vector $\mathbf{a}_{com}$ is chosen as

$$\mathbf{a}_{com} = \mathbf{a}_i - \mathbf{a}_{i,\,dif} \qquad \forall i = 1, 2, \ldots, m$$

then $\mathbf{F}$ is minimized and $\mathbf{F}_{\min} = 0$ [1].

The common vector does not depend on the choice of $\mathbf{a}_i$ or on the choice of the orthonormal basis vector set of $\mathbf{B}$ [1]. The common vector represents the common properties or invariant features of a spoken word. Since all of the feature vectors within one class yields the same common vector, the recognition rate for this class will always be 100%; that is, the method guarantees 100% recognition rate for each class in the training set under the condition $m \le n$. Obviously, if $m \ge n+1$, then the difference subspace would span the whole feature space and the common vector would become a zero vector.

C. Principal Component Analysis

Obtaining the eigenvalues and eigenvectors of the covariance matrices for normal distributions is also known as the PCA or the Karhunen–Loeve transforms (KLT). The PCA or KLT is also known as the optimal linear dimensionality reduction. The goal is to map vectors $\mathbf{a}_i$ $(i = 1, 2, \ldots, m)$ in an n-dimensional space onto another set of vectors in a k-dimensional subspace where $k < n$. But this usually causes some loss of the information which discriminates the different classes [6].

The covariance matrix $\mathbf{\Phi}$ of the feature vectors belonging to a word-class is defined as

$$\mathbf{\Phi} = \sum_{i=1}^{m} (\mathbf{a}_i - \mathbf{a}_{ave})(\mathbf{a}_i - \mathbf{a}_{ave})^T$$

where superscript T denotes the transpose and $\mathbf{a}_{ave}$ is the average vector of all feature vectors in the word-class

$$\mathbf{a}_{ave} = (\mathbf{a}_1 + \cdots + \mathbf{a}_m)/\,m.$$

Since the covariance matrix is real, symmetric and nonnegative, all of its eigenvalues are nonnegative, and its eigenvectors can be chosen to be orthonormal; that is

$$\mathbf{\Phi}\mathbf{u}_j = \lambda_j \mathbf{u}_j \quad \text{and} \quad \lambda_j \ge 0,$$
$$\langle \mathbf{u}_i, \mathbf{u}_j \rangle = \{1 \text{ if } i = j;\;\; 0 \text{ if } i \ne j\} \quad \text{for} \quad i, j = 1, 2, \ldots, n.$$

An error caused by dimensionality reduction can be defined and minimized by making use of the orthonormality relation between the eigenvectors [7]

$$\text{Error} = \tfrac{1}{2} \sum_{j=k+1}^{n} \mathbf{u}_j^T \mathbf{\Phi} \mathbf{u}_j = \tfrac{1}{2} \sum_{j=k+1}^{n} \lambda_j$$

which is just the minimization of the metric

$$\mathbf{F} = \tfrac{1}{2} \sum_{j=1}^{n} \mathbf{u}_j^T \mathbf{\Phi} \mathbf{u}_j.$$

The maximization of the metric $\mathbf{F}$ will yield

$$\mathbf{F}_{\max} = \tfrac{1}{2} \sum_{j=1}^{k} \mathbf{u}_j^T \mathbf{\Phi} \mathbf{u}_j = \tfrac{1}{2} \sum_{j=1}^{k} \lambda_j.$$

The minimum error is obtained by choosing the $(n-k)$ smallest (zero in our case) eigenvalues and their corresponding eigenvectors as the ones to discard [8]. Since the number of largest (nonzero) eigenvalues is limited by the number $(m-1)$ when $m \le n$, the dimension k of the subspace spanned by the eigenvectors corresponding to the largest eigenvalues can be extended up to $(m-1)$.

The PCA is used for classification purposes that are known as the subspace methods.

D. Subspace Method: SELFIC

The subspace methods for pattern recognition are introduced by Watanabe [9] and widely used by Kohonen [10]–[12], and all the work is summarized by Oja [13]. One of these subspace methods is called SELFIC which uses the covariance matrix. In this method, the feature vectors are normalized first. Then, the average feature vector of the class is subtracted from each feature vector in the training set. In this method, the following metric is maximized [9]:

$$\mathbf{F} = \tfrac{1}{2} \sum_{j=1}^{n} \mathbf{u}_j^T \mathbf{Q} \mathbf{u}_j$$

where $\mathbf{Q}$ is the class-correlation matrix and $\mathbf{u}_j$s are its eigenvectors. Since the average feature vector of the class is subtracted from each feature vector first, the correlation and covariance matrices will become the same for the SELFIC method, that is,

$\mathbf{Q} = \mathbf{\Phi}$. Therefore the SELFIC method turns out to have the same metric with the PCA. If $\mathbf{F}$ is maximized, then

$$\mathbf{F}_{\max} = \tfrac{1}{2} \sum_{j=1}^{m-1} \mathbf{u}_j^T \mathbf{\Phi} \mathbf{u}_j = \tfrac{1}{2} \sum_{i=1}^{m} \|\mathbf{a}_i - \mathbf{a}_{ave}\|^2$$
$$= \tfrac{1}{2} (\lambda_1 + \lambda_2 + \cdots + \lambda_{m-1} + \lambda_m).$$

The classification is based on the maximization of the above metric in the literature [9], [13] whereas it could also be done by the minimization of the same metric, then it would be

$$\mathbf{F}_{\min} = \tfrac{1}{2} \sum_{j=m}^{n} \mathbf{u}_j^T \mathbf{\Phi} \mathbf{u}_j = \tfrac{1}{2} (\lambda_m + \lambda_{m+1} + \cdots + \lambda_n).$$

In this case the minimum of the metric would be zero since $\lambda_m = \lambda_{m+1} = \cdots = \lambda_n = 0$ which is missed in the earlier work [9], [13].

Another critical point is that the normalization of the feature vectors is not necessary. In fact, this should not be done because the radially aligned classes will overlap on the hypersphere with the normalization procedure. The experimental results of the SELFIC method are given in Section IV.

II. Relation Between the Common Vector and the Eigenvalues of the Covariance Matrix $\mathbf{\Phi}$

In this section, an approach which is different from the approach given in [1] to find the common vector will be given. Since the covariances characterize the variations of the vectors about their mean, as long as $m \leq n$, the nonzero eigenvalues of the covariance matrix $\mathbf{\Phi}$ should correspond to the eigenvectors which form an orthonormal basis for the difference subspace $\mathbf{B}$. The orthogonal complement $\mathbf{B}^\perp$ in this case is spanned by all the eigenvectors corresponding to the zero eigenvalues. Because the common vector is orthogonal to any vector in the difference subspace, the common vector must be some linear combination of the eigenvectors corresponding to the zero eigenvalues of $\mathbf{\Phi}$, or else, the common vector can also be obtained by subtracting all of the projections of a feature vector onto the eigenvectors corresponding to the nonzero eigenvalues from the feature vector itself.

The following definitions will be necessary to prove the subsequent statements.

The eigenvalues of the covariance matrix $\mathbf{\Phi}$ are nonnegative and they can be written in decreasing order: $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$. Let $\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n$ be the orthonormal eigenvectors corresponding to these eigenvalues.

Let Ker $\mathbf{\Phi}$ be the kernel of the covariance matrix $\mathbf{\Phi}$, and $\mathbf{B}^\perp$ be the orthogonal complement of the difference subspace $\mathbf{B}$. That is

$$\text{Ker}\,\mathbf{\Phi} = \left\{\mathbf{x} \in \mathbf{R}^{n\times 1}: \mathbf{\Phi}\mathbf{x} = \mathbf{0}\right\},$$
$$\mathbf{B}^\perp = \left\{\mathbf{x} \in \mathbf{R}^{n\times 1}: \langle \mathbf{x}, \mathbf{b}\rangle = 0 \qquad \forall\, \mathbf{b} \in \mathbf{B}\right\}.$$

The kernel space of $\mathbf{\Phi}$ is the space of all eigenvectors corresponding to the zero eigenvalues of $\mathbf{\Phi}$. Further, since the space $\mathbf{B}$ is $(m-1)$ dimensional, the space $\mathbf{B}^\perp$ is $(n-m+1)$ dimensional [14].

Theorem 1: The following is true:

$$\mathbf{B}^\perp \subset \text{Ker}\mathbf{\Phi}. \tag{1}$$

Proof: See the Appendix. ■

Theorem 2: The following is true:

$$\text{Ker}\mathbf{\Phi} \subset \mathbf{B}^\perp.$$

Proof: See the Appendix. ■

From the Theorems 1 and 2, one concludes that

$$\text{Ker}\mathbf{\Phi} = \mathbf{B}^\perp.$$

From this last equality and remembering that $\mathbf{B}^\perp$ is $(n-m+1)$ dimensional, it follows that the last $(n-m+1)$ eigenvectors of the set $\{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n\}$ correspond to the zero eigenvalues. Therefore

$$\mathbf{B}^\perp = \text{span}\{\mathbf{u}_m, \mathbf{u}_{m+1}, \ldots, \mathbf{u}_n\}.$$

A. Relation Between the Common Vector and the Zero Eigenvalues

First of all, the following theorem should be given.

Theorem 3: Under the previous definitions for all $i, j = 1, 2, \ldots, m$, the following is true:

$$\langle \mathbf{a}_i - \mathbf{a}_j, \mathbf{a}_{com}\rangle = 0. \tag{2}$$

Proof: Since $(\mathbf{a}_i - \mathbf{a}_j) \in \mathbf{B}$, then the following can be written:

$$\mathbf{a}_i - \mathbf{a}_j = \langle \mathbf{a}_i - \mathbf{a}_j, \mathbf{z}_1\rangle \mathbf{z}_1 + \cdots + \langle \mathbf{a}_i - \mathbf{a}_j, \mathbf{z}_{m-1}\rangle \mathbf{z}_{m-1}.$$

The definition of the common vector states that

$$\mathbf{a}_{com} = \mathbf{a}_j - \langle \mathbf{a}_j, \mathbf{z}_1\rangle \mathbf{z}_1 - \cdots - \langle \mathbf{a}_j, \mathbf{z}_{m-1}\rangle \mathbf{z}_{m-1}.$$

The scalar multiplication of the last two equality gives (2). ■

In the following theorem, it will be shown that the common vector satisfies the eigenvalue–eigenvector equation for the covariance matrix $\mathbf{\Phi}$ for the zero eigenvalues.

Theorem 4: The following is true:

$$\mathbf{\Phi}\mathbf{a}_{com} = \sum_{i=1}^{m} \left[(\mathbf{a}_i - \mathbf{a}_{ave})(\mathbf{a}_i - \mathbf{a}_{ave})^T \mathbf{a}_{com}\right]$$
$$= \sum_{i=1}^{m} \left[(\mathbf{a}_i - \mathbf{a}_{ave})(0)\right] = \mathbf{0}. \tag{3}$$

Proof: From Theorem 3, the following can be written:

$$\left\langle \left(\frac{\mathbf{a}_1 + \mathbf{a}_2 + \cdots + \mathbf{a}_m}{m} - \mathbf{a}_1\right), \mathbf{a}_{com}\right\rangle = 0.$$

Since the first term in the parenthesis is the average of all $\mathbf{a}_i$ $(i = 1, 2, \ldots, m)$, the following can be written:

$$\langle (\mathbf{a}_1 - \mathbf{a}_{ave}), \mathbf{a}_{com}\rangle = 0. \tag{4}$$

Similarly

$$\langle (\mathbf{a}_i - \mathbf{a}_{ave}), \mathbf{a}_{com}\rangle = 0. \qquad (i = 2, 3, \ldots, m). \tag{5}$$

From (4) and (5), it follows that (3) holds. ■

B. Relation Between the Common Vector and the Nonzero Eigenvalues

The first $(m-1)$ eigenvectors of the covariance matrix correspond to the nonzero eigenvalues. Now it will be shown that $\mathbf{B}$ is the span of the eigenvectors corresponding to the nonzero eigenvalues.

Theorem 5: The following is true:

$$\mathbf{B} = \text{span}\{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_{m-1}\}.$$

Here $\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_{m-1}$ are the eigenvectors corresponding to the nonzero eigenvalues of $\mathbf{\Phi}$.

Proof: For any $\mathbf{x} \in \text{span}\{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_{m-1}\}$

$$\mathbf{x} \perp \mathbf{B}^\perp = \text{span}\{\mathbf{u}_m, \mathbf{u}_{m+1}, \ldots, \mathbf{u}_n\}.$$

Therefore, $\mathbf{x} \in (\mathbf{B}^\perp)^\perp$ or $\mathbf{x} \in \mathbf{B}$, since $\mathbf{B}$ is finite dimensional and therefore $(\mathbf{B}^\perp)^\perp = \mathbf{B}$ [4]. Thus, $\text{span}\{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_{m-1}\} \subset \mathbf{B}$.

Since the $\text{span}\{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_{m-1}\}$ and $\mathbf{B}$ have the same $(m-1)$ dimensions; from this, we get

$$\text{span}\{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_{m-1}\} = \mathbf{B}.$$

■

By combining the Theorems 3–5, any feature vector $\mathbf{a}_i$ can be written as

$$\mathbf{a}_i = \langle \mathbf{a}_i, \mathbf{u}_1 \rangle \mathbf{u}_1 + \cdots + \langle \mathbf{a}_i, \mathbf{u}_{m-1} \rangle \mathbf{u}_{m-1} + \langle \mathbf{a}_i, \mathbf{u}_m \rangle \mathbf{u}_m + \cdots + \langle \mathbf{a}_i, \mathbf{u}_n \rangle \mathbf{u}_n \tag{6}$$

or

$$\mathbf{a}_i = \mathbf{a}_i^* + \mathbf{a}_i^\perp$$

where

$$\mathbf{a}_i^* = \langle \mathbf{a}_i, \mathbf{u}_1 \rangle \mathbf{u}_1 + \cdots + \langle \mathbf{a}_i, \mathbf{u}_{m-1} \rangle \mathbf{u}_{m-1} \qquad \mathbf{a}_i^* \in \mathbf{B}$$

and

$$\mathbf{a}_i^\perp = \mathbf{a}_{com} = \langle \mathbf{a}_i, \mathbf{u}_m \rangle \mathbf{u}_m + \cdots + \langle \mathbf{a}_i, \mathbf{u}_n \rangle \mathbf{u}_n \qquad \mathbf{a}_i^\perp \in \mathbf{B}^\perp. \tag{7}$$

From (6), for any feature vector $\mathbf{a}_i$, the common vector $\mathbf{a}_{com}$ can also be written as

$$\mathbf{a}_{com} = \mathbf{a}_i - \mathbf{a}_i^*.$$

As seen from the above derivations, the common vector can be determined by using the eigenvectors corresponding to the zero or nonzero eigenvalues of the covariance matrix $\mathbf{\Phi}$. But it must be noted that the common vector does not include information in the directions corresponding to the nonzero eigenvalues. The CVA also saves computation time by selecting zero or nonzero eigenvalues whichever is the smaller set.

Remark: Similar derivations can be carried out when the set of feature vectors $\{\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_m\}$ are linearly dependent. If r out of m feature vectors in one class are linearly independent, the difference subspace $\mathbf{B}$ will have $(r-1)$ dimensions. The common vector $\mathbf{a}_{com}$ will correspond to the eigenvectors belonging to the $(n-r+1)$ zero eigenvalues of the covariance matrix $\mathbf{\Phi}$.

C. Numerical Examples

Example 1: Let the following feature vectors represent the training set of a class:

$$\mathbf{a}_1 = [2/3 \quad 1 \quad 1/3]^T \qquad \mathbf{a}_2 = [1 \quad 2 \quad 0]^T.$$

Then, the difference vector $\mathbf{b}_1$ will be

$$\mathbf{b}_1 = \mathbf{a}_2 - \mathbf{a}_1 = [1/3 \quad 1 \quad -1/3]^T.$$

The unit vector in the direction of $\mathbf{b}_1$ is

$$\mathbf{z}_1 = \frac{\mathbf{b}_1}{\|\mathbf{b}_1\|} = \left[1/\sqrt{11} \quad 3/\sqrt{11} \quad -1/\sqrt{11}\right]^T.$$

The common vector $\mathbf{a}_{com}$ is then

$$\begin{aligned} \mathbf{a}_{com} &= \mathbf{a}_1 - \langle \mathbf{a}_1, \mathbf{z}_1 \rangle \mathbf{z}_1 = \mathbf{a}_2 - \langle \mathbf{a}_2, \mathbf{z}_1 \rangle \mathbf{z}_1 \\ &= [4/11 \quad 1/11 \quad 7/11]^T = [0.3636 \quad 0.0909 \quad 0.6364]^T. \end{aligned}$$

The average vector is

$$\mathbf{a}_{ave} = (\mathbf{a}_1 + \mathbf{a}_2)/2 = [5/6 \quad 3/2 \quad 1/6]^T.$$

The covariance matrix is

$$\mathbf{\Phi} = \begin{bmatrix} 1/18 & 1/6 & -1/18 \\ 1/6 & 1/2 & -1/6 \\ -1/18 & -1/6 & 1/18 \end{bmatrix}.$$

The covariance matrix has the following eigenvectors and eigenvalues

$$\begin{aligned} \mathbf{u}_1 &= [0.3015 \quad 0.9045 \quad -0.3015]^T & \lambda_1 &= 0.6111 \\ \mathbf{u}_2 &= [0.9487 \quad -0.3162 \quad 0]^T & \lambda_2 &= 0 \\ \mathbf{u}_3 &= [-0.0953 \quad -0.2860 \quad -0.9535]^T & \lambda_3 &= 0. \end{aligned}$$

The common vector $\mathbf{a}_{com}$ in this case is the following:

$$\mathbf{a}_{com} = \mathbf{a}_1 - (\mathbf{a}_1^T \mathbf{u}_1)\mathbf{u}_1 = [0.3636 \quad 0.0909 \quad 0.6364]^T$$

or

$$\mathbf{a}_{com} = \mathbf{a}_2 - (\mathbf{a}_2^T \mathbf{u}_1)\mathbf{u}_1 = [0.3636 \quad 0.0909 \quad 0.6364]^T.$$

The common vector $\mathbf{a}_{com}$ is also the summation of the projections of any feature vector $\mathbf{a}_1$ or $\mathbf{a}_2$ onto the $\mathbf{u}_2$ and $\mathbf{u}_3$

$$\begin{aligned} \mathbf{a}_{com} &= (\mathbf{a}_1^T \mathbf{u}_2)\mathbf{u}_2 + (\mathbf{a}_1^T \mathbf{u}_3)\mathbf{u}_3 = (\mathbf{a}_2^T \mathbf{u}_2)\mathbf{u}_2 + (\mathbf{a}_2^T \mathbf{u}_3)\mathbf{u}_3 \\ &= [0.3636 \quad 0.0909 \quad 0.6364]^T. \end{aligned}$$

Example 2: Let the feature vectors be

$$\mathbf{a}_1 = [1 \quad 2 \quad 0]^T \quad \mathbf{a}_2 = [1 \quad 1 \quad 0]^T \quad \mathbf{a}_3 = [0 \quad 0 \quad 1]^T$$

belonging to the same class.

The difference vectors between these feature vectors will be

$$\begin{aligned} \mathbf{b}_1 &= \mathbf{a}_1 - \mathbf{a}_2 = [0 \quad 1 \quad 0]^T \\ \mathbf{b}_2 &= \mathbf{a}_3 - \mathbf{a}_2 = [-1 \quad -1 \quad 1]^T. \end{aligned}$$

The orthonormal basis vector set that will define the subspace $\mathbf{B} = \text{span}\{\mathbf{b}_1, \mathbf{b}_2\}$

$$\mathbf{d}_1 = \mathbf{b}_1 \qquad \mathbf{z}_1 = \frac{\mathbf{d}_1}{||\mathbf{d}_1||} = [0 \quad 1 \quad 0]^T$$

$$\mathbf{d}_2 = \mathbf{b}_2 - \langle \mathbf{b}_2, \mathbf{z}_1 \rangle \mathbf{z}_1 \qquad \mathbf{z}_2 = \frac{\mathbf{d}_2}{||\mathbf{d}_2||} = \left[-1/\sqrt{2} \quad 0 \quad 1/\sqrt{2}\right]^T.$$

The common vector $\mathbf{a}_{com}$ can be obtained from any of the feature vectors $\mathbf{a}_i$ $i = 1, 2, 3$

$$\mathbf{a}_{com} = \mathbf{a}_1 - \langle \mathbf{a}_1, \mathbf{z}_1 \rangle \mathbf{z}_1 - \langle \mathbf{a}_1, \mathbf{z}_2 \rangle \mathbf{z}_2 = [0.5 \quad 0 \quad 0.5]^T$$

$$\mathbf{a}_{com} = \mathbf{a}_2 - \langle \mathbf{a}_2, \mathbf{z}_1 \rangle \mathbf{z}_1 - \langle \mathbf{a}_2, \mathbf{z}_2 \rangle \mathbf{z}_2 = [0.5 \quad 0 \quad 0.5]^T$$

$$\mathbf{a}_{com} = \mathbf{a}_3 - \langle \mathbf{a}_3, \mathbf{z}_1 \rangle \mathbf{z}_1 - \langle \mathbf{a}_3, \mathbf{z}_2 \rangle \mathbf{z}_2 = [0.5 \quad 0 \quad 0.5]^T.$$

The average vector $\mathbf{a}_{ave}$ is

$$\mathbf{a}_{ave} = \tfrac{1}{3}(\mathbf{a}_1 + \mathbf{a}_2 + \mathbf{a}_3) = [2/3 \quad 1 \quad 1/3]^T.$$

The covariance matrix is

$$\mathbf{\Phi} = \begin{bmatrix} 2/3 & 1 & -2/3 \\ 1 & 2 & -1 \\ -2/3 & -1 & 2/3 \end{bmatrix}.$$

The eigenvalues of $\mathbf{\Phi}$ are $\lambda_i = 3.1196, 0.2137, 0.0000$ and the corresponding eigenvectors are

$$\mathbf{u}_1 = [-0.4389 \quad -0.7840 \quad 0.4389]^T \qquad \lambda_1 = 3.1196$$

$$\mathbf{u}_2 = [0.5544 \quad -0.6207 \quad -0.5544]^T \qquad \lambda_2 = 0.2137$$

$$\mathbf{u}_3 = [-0.7071 \quad 0 \quad -0.7071]^T \qquad \lambda_3 = 0.$$

The projections of $\mathbf{a}_1$ onto $\mathbf{u}_1$ and $\mathbf{u}_2$ are the following:

$$(\mathbf{a}_1^T \mathbf{u}_1)\mathbf{u}_1 = [0.8809 \quad 1.5735 \quad -0.8809]^T$$

$$(\mathbf{a}_1^T \mathbf{u}_2)\mathbf{u}_2 = [-0.3809 \quad 0.4265 \quad 0.3809]^T.$$

Their summation is

$$(\mathbf{a}_1^T \mathbf{u}_1)\mathbf{u}_1 + (\mathbf{a}_1^T \mathbf{u}_2)\mathbf{u}_2 = [0.5 \quad 2 \quad -0.5]^T.$$

The common vector in this case is

$$\mathbf{a}_{com} = \mathbf{a}_1 - (\mathbf{a}_1^T \mathbf{u}_1)\mathbf{u}_1 - (\mathbf{a}_1^T \mathbf{u}_2)\mathbf{u}_2 = [0.5 \quad 0 \quad 0.5]^T$$

which is in accord with the theorem.

The alignment of $\mathbf{a}_{com}$ and $\mathbf{u}_3$ is easily seen to be the same and the following is true:

$$\mathbf{a}_{com} = (\mathbf{a}_1^T \mathbf{u}_3)\mathbf{u}_3 = (\mathbf{a}_2^T \mathbf{u}_3)\mathbf{u}_3 = (\mathbf{a}_3^T \mathbf{u}_3)\mathbf{u}_3 = [0.5 \quad 0 \quad 0.5]^T.$$

The calculation of the covariance matrix and its eigenvalues and eigenvectors is not necessary for both the common vector and the SELFIC approach. Since the common vector is the projection of any of the feature vectors onto the indifference subspace, the common vector can be obtained by just a subtraction of the difference projection from the feature vector itself.

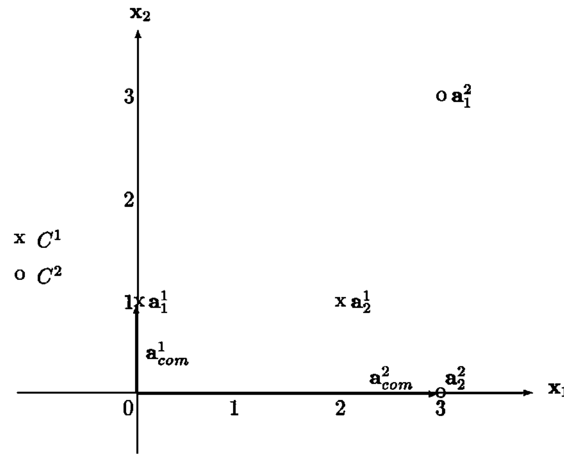

Fig. 1. Application of the SELFIC method and the CVA in 2-D feature space.

III. Comparison of the Common Vector and the SELFIC Method for Classification

In this section, the common vector and the SELFIC method which uses the nonzero principal components are compared for the classification of two classes which are shown in Figs. 1 and 2 in 2-D feature space. In Fig. 1, one of the classes (C^1) has the feature vectors of $\mathbf{a}_1^1$ and $\mathbf{a}_2^1$, and the other class (C^2) has the feature vectors of $\mathbf{a}_1^2$ and $\mathbf{a}_2^2$. The common vectors of two classes are also shown as $\mathbf{a}_{com}^1$ and $\mathbf{a}_{com}^2$ in Fig. 1.

First of all, let us apply the CVA to classify each of the feature vectors. The direction of the eigenvector corresponding to the zero eigenvalue for each class is the same as the direction of the common vector; that is, the direction of the eigenvector corresponding to the zero eigenvalue for the class C^1 is the vertical direction, and so is the direction of the common vector $\mathbf{a}_{com}^1$ in Fig. 1. The directions for the class C^2 are the horizontal directions in Fig. 1. The Euclidean distance between the projection of all feature vectors ($\mathbf{a}_i^j$) of a given class (C^j) onto the eigenvector corresponding to a zero eigenvalue of the same class and the common vector ($\mathbf{a}_{com}^j$) of the same class gives a zero value. But the Euclidean distance between the projection of all feature vectors ($\mathbf{a}_i^j$) of a given class (C^j) onto the eigenvector corresponding to a zero eigenvalue of the other class (C^k, $k \neq j$) and the common vector ($\mathbf{a}_{com}^k$) of the other class gives the values of 3, 1, 2 and 1 (all of which are greater than zero) for the feature vectors $\mathbf{a}_1^1$, $\mathbf{a}_2^1$, $\mathbf{a}_1^2$ and $\mathbf{a}_2^2$ respectively. Therefore, since we consider the minimum values of the distances obtained for each feature vector for classification purposes in the CVA, all of the feature vectors are correctly classified. A 100% recognition rate is always guaranteed for the training set of the classes unless they yield the same common vector. Also, a feature vector may be unrecognized if one of the feature vectors of a given class is aligned with the line combining two feature vectors of the other class.

If the SELFIC method, an application of the PCA, is applied without normalization of the feature vectors, the projection of all feature vectors of a given class onto the eigenvector corresponding to nonzero eigenvalue of the same class must be determined. For the example considered here, the values of 0, 2, 3, and 0 are obtained in magnitudes for the feature vectors $\mathbf{a}_1^1$, $\mathbf{a}_2^1$,

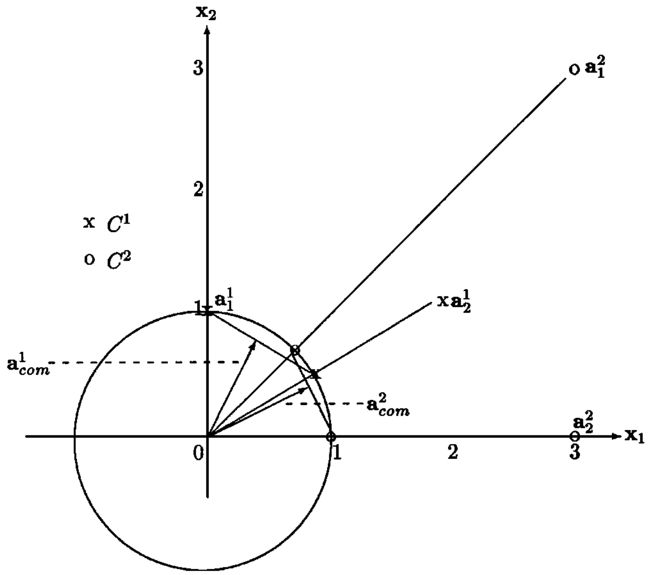

Fig. 2. Direct application of the SELFIC method and the CVA after normalization.

$\mathbf{a}_1^2$ and $\mathbf{a}_2^2$ respectively. Also the projection of all feature vectors of a given class onto the eigenvector corresponding to nonzero eigenvalue of the other class must be determined to classify the feature vectors. Then the values of 1, 1, 3 and 3 are obtained in magnitudes for the feature vectors $\mathbf{a}_1^1$, $\mathbf{a}_2^1$, $\mathbf{a}_1^2$ and $\mathbf{a}_2^2$, respectively. Therefore, $\mathbf{a}_1^1$ and $\mathbf{a}_2^2$ are misclassified; $\mathbf{a}_1^2$, not classified; and $\mathbf{a}_2^1$, correctly classified since we consider the maxima of these values obtained for each feature vector.

In the SELFIC method with normalization, the feature vectors are mapped onto the unit circle as shown in Fig. 2. Then the nonzero principal components are considered to classify each feature vector on the unit circle. In this case, the feature vectors $\mathbf{a}_2^1$ and $\mathbf{a}_1^2$ are correctly classified, but the feature vectors $\mathbf{a}_1^1$ and $\mathbf{a}_2^2$ are misclassified.

If the CVA is applied to classify each of the feature vectors on the unit circle (after normalization), still the Euclidean distance between the projection of all feature vectors of a given class onto the eigenvector corresponding to a zero eigenvalue of the same class and the common vector of the same class gives a zero value. But the Euclidean distance between the projection of all feature vectors of a given class onto the eigenvector corresponding to a zero eigenvalue of the other class and the common vector of the other class gives the values greater than zero. Therefore, all of the feature vectors are correctly classified since the minimum distances are used for classification in the CVA.

IV. Experimental Study for Isolated Work Recognition

When all the differences in the difference subspaces are taken away from the words in the test set, the leftover vectors are called the remaining vectors [1], [3]. In fact for a feature vector $\mathbf{a}_x$ of an unknown word-class, the remaining vector ($\mathbf{a}_{x,\,rem}^l$) can be written as

$$\mathbf{a}_{x,\,rem}^l = \mathbf{a}_x - \sum_{j=1}^{m-1} \langle \mathbf{a}_x^T,\, \mathbf{u}_j^l \rangle \mathbf{u}_j^l = \sum_{j=m}^{n} \langle \mathbf{a}_x^T,\, \mathbf{u}_j^l \rangle \mathbf{u}_j^l$$

where l denotes the class. The Euclidean distance between the common vector and the remaining vector is used as the decision criterion, and it is given as

$$\mathbf{C}^* = \arg\min_l \left\| \mathbf{a}_{x,\,rem}^l - \mathbf{a}_{com}^l \right\| \tag{8}$$

that is, if the feature vector $\mathbf{a}_x$ belongs to the class C^l, the distance between $\mathbf{a}_{x,\,rem}^l$ and $\mathbf{a}_{com}^l$ is expected to be a minimum just as it is the case for the training set.

In the recognition process of the SELFIC method, the digits in the database are classified according to maximum value of the following criterion:

$$\mathbf{C}^* = \arg\max_l \left\| \sum_{j=1}^{k} (\mathbf{a}_x^T \mathbf{u}_j^l)^2 \right\|.$$

Experimental studies on the TI digit database have shown that the remaining vector is usually closer to the common vector of its own word-class than it is to the common vectors of the other word-classes, and therefore, it can be used in isolated word recognition [1]. In the analysis of the TI database, at first, silence regions at the beginning and at the end of each digit are removed by using energy and zero-crossing thresholds. Then the speech frames consisting of 256 samples are pre-emphasized and analyzed to calculate the 11th-order root-melcep parameters. Thus the feature vector for each repetition of each digit is constructed with these parameters. After this process, the longest feature vector in the TI database has 407 parameters or features ($n = 407$), and the shortest feature vector has 110 parameters. Therefore, the dimensions of the feature vectors which have fewer than 407 parameters are extended to 407 by inserting random values obtained with the command "randn" in Matlab. The covariance matrix $\mathbf{\Phi}$ for each digit is then a 407×407 matrix. The covariance matrix $\mathbf{\Phi}$ and its eigenvalues and eigenvectors are found by using 224 feature vectors obtained from 224 repetitions of each digit (112 speakers say each digit twice) in the training set ($m = 224$). 223 eigenvalues, that is, $(m-1)$ eigenvalues, out of 407 eigenvalues are found to be nonzero and the rest $(407 - 223 = 184)$ of the eigenvalues are zero. Since the eigenvalues change abruptly, e.g., from 1.22×10^{-12} to 0.219 for the digit eight, the values under 1×10^{-10} are assumed to be zero. In fact, the number of zero eigenvalues is always equal to $407 - 223 = 184$ for all classes. The common vector for each digit is obtained from (7) by using the eigenvectors corresponding to 184 zero eigenvalues. If the decision criterion given in (8) is applied to the training set, a recognition rate of 100% is obtained for each digit. If the test set formed from 226 repetitions (113 speakers say each digit twice) is used in the recognition process, recognition rates of 97% and 96%, which are the best performances on the test set, are obtained on the average for the normalized and nonnormalized feature vectors respectively. The results are given in Table I for each digit.

If the SELFIC method is applied to the TI database, the maximum recognition rate of 91% in the average is obtained by using the eigenvectors corresponding to the 115 largest of the 223 nonzero eigenvalues for each digit in the training set. The maximum recognition rate of 75% on the average is obtained by using the eigenvectors corresponding to the 74 largest eigen-

TABLE I
CORRECT RECOGNITION RATES IN AVERAGE OBTAINED BY USING COMMON VECTOR APPROACH AND THE SELFIC METHOD EXPRESSED AS PERCENTAGES

	TRAINING SET			TEST SET		
	COM. VECT.		SELF	COM. VECT.		SELF
	Nor.	NNor.	Nor.	Nor.	NNor.	Nor.
One	100	100	93	97	97	55
Two	100	100	95	97	96	78
Three	100	100	97	93	93	82
Four	100	100	43	97	90	25
Five	100	100	100	95	96	80
Six	100	100	94	99	96	83
Seven	100	100	100	98	98	91
Eight	100	100	96	97	97	64
Nine	100	100	91	97	97	90
Ow	100	100	100	96	96	93
Zero	100	100	95	100	99	88
Average	100	100	91	97	96	75

values for each digit in the test set. If the eigenvectors corresponding to all 223 nonzero eigenvalues are used in the test set, the recognition rate decreases to 70%. In the SELFIC method, the dimensions of the feature vectors which have fewer than 407 parameters are also extended to 407 by inserting random values. The results of the SELFIC method are given in Table I for each digit.

V. CONCLUSIONS

Throughout this paper, the CVA is discussed, and the relation between the CVA and the principal components is given. The relation between the common vector $\mathbf{a}_{com}$ and the eigenvectors of the covariance matrix of any feature vector set is established by the theorems in Section II. Two numerical examples are provided in the same section. The most important conclusion is that the common vector satisfies the eigenvalue-eigenvector equation of the covariance matrix corresponding to the zero eigenvalues, i.e., the common vector is orthogonal to all the eigenvectors that correspond to the nonzero eigenvalues of the covariance matrix.

The following results should be highlighted.

1) The PCA is usually carried out to reduce the dimensions of a given feature vector set. The features that are eliminated correspond to the smallest (including zero) eigenvalues of the covariance matrix. This may cause some loss of information which discriminates the different classes. The SELFIC method which is used for discriminating classes and which is just a direct application of the PCA verifies this situation.
2) The CVA method is used for discriminating classes. According to the CVA, the process in the SELFIC method must be reversed to increase the power of discrimination between the different classes; that is, the common vector is obtained by eliminating all the features in the difference subspace $\mathbf{B}$. This also eliminates all the components of a feature vector that are along the eigenvectors corresponding to the nonzero eigenvalues of the covariance matrix.

Seemingly contradictory, these two results stem from the fact that the goals in using the PCA and the CVA are completely different; that is, the PCA is used for reducing the dimensions whereas the CVA is used for recognition purposes. These two results are true in case the dimension n of the feature vectors is larger than or equal to m which is the number of data available for any class in the training set or when some of the eigenvalues of the covariance matrix are zero. This case is true for many speech and pattern recognition applications.

These conclusions may be pushed a little further. The common vector that also corresponds to the zero eigenvalues of the covariance matrix contains the invariant features of any class which is highly important from the recognition or discrimination point of view. This result is important for information scientists since classes with a huge number of features can be represented with unique common vectors, each of which is unique for its class and contains all the common invariant features of its own class.

Obviously, the method fails when the number of feature vectors m is at least one larger than their dimensions n, that is, the method fails when $m \geq n+1$. This is a subject under investigation, but in our future work, it will be shown that the common vector $\mathbf{a}_{com}$ will be just the projection of the average vector $\mathbf{a}_{ave}$ of all the intra-class feature vectors onto the indifference subspace under certain conditions.

APPENDIX

Proof of Theorem 1: If any $\mathbf{x} \in \mathbf{B}^{\perp}$, then $\mathbf{x} \perp \mathbf{B}$. From here the following scalar multiplication must be zero:

$$\langle \mathbf{a}_i - \mathbf{a}_j, \mathbf{x} \rangle = 0 \quad \text{for} \qquad i, j = 1, 2, \ldots, m$$

or

$$\langle \mathbf{a}_i - \mathbf{a}_{ave}, \mathbf{x} \rangle = 0 \quad \text{for} \qquad i = 1, 2, \ldots, m.$$

Therefore

$$\begin{aligned} \mathbf{\Phi x} &= \sum_{i=1}^{m} \left[(\mathbf{a}_i - \mathbf{a}_{ave})(\mathbf{a}_i - \mathbf{a}_{ave})^T \mathbf{x} \right] \\ &= \sum_{i=1}^{m} \left[(\mathbf{a}_i - \mathbf{a}_{ave}) \langle \mathbf{a}_i - \mathbf{a}_{ave}, \mathbf{x} \rangle \right] = \mathbf{0} \end{aligned}$$

or $\mathbf{x} \in \mathrm{Ker}\mathbf{\Phi}$. Since the element $\mathbf{x} \in \mathbf{B}^{\perp}$ was arbitrary, then (1) is true.

Proof of Theorem 2: If $\hat{\mathbf{a}}_i = \mathbf{a}_i - \mathbf{a}_{ave}$ for $i = 1, 2, \ldots, m$, and if the matrix $\mathbf{A}$ is defined as $\mathbf{A} = (\hat{\mathbf{a}}_1, \ldots, \hat{\mathbf{a}}_m)$, then the covariance matrix $\mathbf{\Phi}$ can be written as

$$\mathbf{\Phi} = \mathbf{A}\mathbf{A}^T.$$

Let $\mathbf{u}_i$ be one of the eigenvectors corresponding to a zero eigenvalue. Thus

$$\mathbf{\Phi u}_i = \mathbf{0}$$

or

$$\mathbf{A}\mathbf{A}^T\mathbf{u}_i = \mathbf{0}.$$

By premultiplying the above equality with $\mathbf{u}_i^T$, we get

$$\mathbf{u}_i^T\mathbf{A}\mathbf{A}^T\mathbf{u}_i = 0$$

or

$$\|\mathbf{A}^T\mathbf{u}_i\| = 0.$$

From here

$$\hat{\mathbf{a}}_1^T\mathbf{u}_i = \hat{\mathbf{a}}_2^T\mathbf{u}_i = \cdots = \hat{\mathbf{a}}_m^T\mathbf{u}_i = 0. \tag{A1}$$

It can be easily shown that the space spanned by $\{\hat{\mathbf{a}}_1, \hat{\mathbf{a}}_2, \ldots, \hat{\mathbf{a}}_m\}$ has $(m - 1)$ dimension. On the other hand, the dimension of $\mathbf{B}$ is also $(m-1)$ and since $\hat{\mathbf{a}}_i \in \mathbf{B}$ then according to (A1) we obtain $\mathbf{u}_i \perp \mathbf{B}$ or equivalently $\mathbf{u}_i \in \mathbf{B}^\perp$.

Thus all the eigenvectors $\mathbf{u}_i$ corresponding to zero eigenvalues must belong to $\mathbf{B}^\perp$. Since the set of the eigenvectors $\mathbf{u}_i$ corresponding to zero eigenvalues is the basis of Ker $\mathbf{\Phi}$, then

$$\text{Ker}\mathbf{\Phi} \subset \mathbf{B}^\perp.$$

ACKNOWLEDGMENT

The authors would like to thank H. H. Erkaya for extensive English editing of the manuscript, and the three anonymous reviewers for their critical comments, which greatly helped improve the presentation of the paper.

REFERENCES

[1] M. B. Gülmezoğlu, V. Dzhafarov, M. Keskin, and A. Barkana, "A novel approach to isolated word recognition," *IEEE Trans. Speech Audio Processing*, vol. 7, pp. 620–628, Nov. 1999.

[2] M. Keskin, M. B. Gülmezoğlu, O. Parlaktuna, and A. Barkana, "Isolated word recognition by extracting personal differences," in *Proc. 6th Int. Conf. Signal Processing Applications Technology*, Boston, MA, Oct. 1995, pp. 1989–1992.

[3] A. Barkana, M. B. Gülmezoğlu, R. Edizkan, Ü. Künkçü, and S. E. Künkçü, "Recognition of limited number of words by computer," Sci. Tech. Res. Council, Turkey, Project Rep. EEEAG-82, 1995.

[4] C. H. Edwards and D. E. Penney, *Elementary Linear Algebra*. Englewood Cliffs, NJ: Prentice-Hall, 1988.

[5] M. Keskin, "Orthogonalization process of vector space in Work Station and Matlab medium," Elect. Electron. Eng. Dept., Osmangazi Univ., Eskişehir, Turkey, July 1994.

[6] D. F. Morrison, *Multivariate Statistical Methods*. New York: McGraw-Hill, 1967, pp. 156–195.

[7] C. M. Bishop, *Neural Networks for Pattern Recognition*. Oxford, U.K.: Clarendon, 1995, pp. 312–314, 454–456.

[8] E. L. Bocchieri and G. R. Doddington, "Frame-specific statistical features for speaker independent speech recognition," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-34, pp. 755–764, Aug. 1986.

[9] S. Watanabe, P. F. Lambert, C. A. Kulikowski, J. L. Buxton, and R. Walker, "Evaluation and selection of variables in pattern recognition," in *Computer and Information Sciences II*, J. T. Tou, Ed. New York: Academic, 1967, pp. 91–122.

[10] T. Kohonen, G. Nemeth, K.-J. Bry, M. Jalanko, and H. Riittinen, "Spectral classification of phonemes by learning subspaces," in *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing*, 1979, pp. 97–100.

[11] T. Kohonen, H. Riittinen, M. Jalanko, E. Reuhkala, and S. Haltsonen, "A thousand-word recognition system based on the learning subspace method and redundant hash addressing," in *Proc. 5th Int. Conf. Pattern Recognition*, 1980, pp. 158–165.

[12] T. Kohonen, H. Riittinen, E. Reuhkala, and S. Haltsonen, "On-line recognition of spoken words from a large vocabulary," *Inf. Sci.*, vol. 33, pp. 3–30, 1984.

[13] E. Oja, *Subspace Methods of Pattern Recognition*. New York: Wiley, 1983, pp. 73–108.

[14] K. Hoffman and R. Kunze, *Linear Algebra*. Englewood Cliffs, NJ: Prentice-Hall, 1961, pp. 229–233.

M. Bilginer Gülmezoğlu received the B.S. and M.S. degrees in electrical engineering from the Middle East Technical University, Ankara, Turkey, in 1983 and 1985, respectively, and the Ph.D. degree from Anadolu University, Eskişehir, Turkey, in 1992.

From 1983 to 1986, he was a Research Assistant at Anadolu University, where he worked on electric machinery and power electronics. Currently, he is an Assistant Professor with the Electrical and Electronics Engineering Department, Osmangazi University, Eskişehir. His current interests include speech production and analysis, speech recognition, and signal processing.

Vakıf Dzhafarov received the Dipl.Math. degree from Azerbaijan State University, Baku, Azerbaijan, in 1975 and the Ph.D. degree in applied mathematics from USSR Academy of Science, Sverdlovsk, Russia, in 1982.

From 1975 to 1982, he was a Research Assistant at the USSR Academy of Science. From 1985 to 1993, he served as Assistant Professor at the Academy of Science, Azerbaijan State University. In 1993, he joined Osmangazi University, Eskişehir, Turkey. Presently, he is a Professor with the Department of Mathematics, Anadolu University, Eskişehir, Turkey. His research interests include theory of differential equations, differential games and optimization, robust stability, and pattern recognition.

Atalay Barkana (S'71–M'74) received the B.S. degree in electrical engineering from Robert College, Istanbul, Turkey, in 1969, and the M.S. and Ph.D. degrees in electrical engineering from the University of Virginia, Charlottesville, in 1971 and 1974, respectively.

From 1974 to 1986, he worked on linear and nonlinear control theory. His current interests include speech recognition, pattern analysis, neural networks, and statistical signal processing. Since 1993, he has been with the Electrical and Electronics Engineering Department, Osmangazi University, Eskişehir, Turkey, where he is Professor.

Dr. Barkana is a member of the IEEE Signal Processing Society.