AD-A045 066 MELPAR FALLS CHURCH VA F/G 17/1

THE FAST FOURIER-HADAMARD TRANSFORM AND ITS USE IN SIGNAL REPRE--ETC(U)

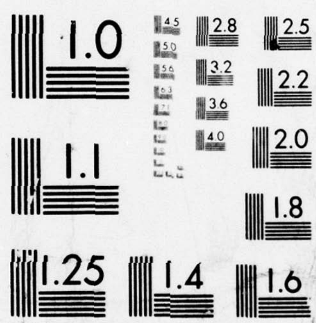
1968 J E WHELCHEL, D F GUINN

UNCLASSIFIED PRC-68-11 NL

| OF |

AD
A045 066

END
DATE
FILMED
11-77
DDC



MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

PRC-68-11

THE FAST FOURIER-HADAMARD TRANSFORM AND ITS USE IN SIGNAL REPRESENTATION AND CLASSIFICATION

1968

14p.

Mr. J. E. Whelchel, Jr. [redacted] D. F. Guinn

Melpar, Inc.
7700 Arlington Boulevard
Falls Church, Virginia 22046

THE FAST FOURIER-HADAMARD TRANSFORM AND ITS USE IN SIGNAL REPRESENTATION AND CLASSIFICATION

By: Mr. J. E. Whelchel, Jr. and Mr. D. F. Guinn, Melpar, Inc.

ABSTRACT

A discrete time transform has been *studied and applied to* the representation and discrimination of digitized signals. The transform consists of an orthogonal (Hadamard) matrix whose elements are all ones and minus ones. To facilitate implementation, a fast Hadamard transform (FHT) has been developed requiring only NlogN rather than N^2 algebraic additions. Several properties of the FHT are revealed, including the nature of its presence in the fast Fourier transform, in which it performs the additive operations as shown by further decomposing the product of matrices representing the FFT. Among other properties discussed are the Hadamard-sampled Walsh relationship, the Hadamard analog to the "shift" property of the Fourier transform, and the structure of covariance matrices diagonalized by the Hadamard transform. The efficiency of the Hadamard transform is compared with that of optimum transforms for signal representation and discrimination. For signal representation, the optimum transform is the set of eigenvectors which diagonalize the sample correlation matrix of the signals. For signal discrimination, the optimum transform for Nilsson's likelihood decision technique is that which simultaneously diagonalizes two different covariance matrices. The inductive capability (training set vs. unknown data) of discriminants estimated from the transform coefficients by parametric techniques is described. Discriminants based on sample means generalize better than those based on probability densities.

INTRODUCTION

The impact of the fast Fourier transform on signal processing is just beginning to be felt, particularly the "fast" convolution property of the transform. In this paper, a discrete time transform, termed the Hadamard transform, is analyzed and some relationships between it and the discrete time Fourier transform are developed. In addition, the effectiveness of the transform for two application areas of interest, signal representation and discrimination, are considered in detail. Potential application areas best suited for the transform are suggested and contrasted with those being exploited by the fast Fourier transform.

DISCRETE TIME FOURIER TRANSFORM

The primary transform in signal analysis has been historically the Fourier transform pair,

$$F(w) = \int_{-\infty}^{\infty} f(t)\, e^{-jwt}\, dt$$

and its inverse.

For signals that are of an echo or burst nature, extending over some finite time epoch 0 to T, it is customary to consider the signals as periodically extended functions of period T for which the above reduces to the Fourier series,

$$C_n = \frac{1}{T}\int_0^T f(t)\, e^{-jnw_0 t}\, dt \qquad w_0 = \frac{2\pi}{T}$$

with inverse

$$f(t) = \sum_{n=-\infty}^{\infty} C_n\, e^{jnw_0 t}\ .$$

We know, in addition, from the sampling theorem[1] that a signal band limited to W cycles over the interval 0 to T can be represented by a 2TW dimensional signal vector $\vec{s}$ where $\vec{s} = (s_1, s_2, s_3, \ldots, s_{2TW})$. The increased use of sampled and digitized signals for spectrum analysis has led to the further *modification known as the discrete time Fourier transform*[2] defined as:

$$f_n = \sum_{k=0}^{N-1} s_k\, e^{-j2\pi nk/N} \qquad N = 2TW, \quad n = 0,1,\ldots,N-1$$

with inverse,

$$s_k = \frac{1}{N}\sum_{n=0}^{N-1} f_n\, e^{j2\pi nk/N}\ .$$

The Fourier transform becomes an N by N unitary matrix F performing as a linear operator on the sampled time series to obtain the "spectrum" vector f,

$$\vec{f} = F\vec{s},$$

where

$$F F^* = F^* F = NI.$$

The scope of this paper will include the relationships between the Fourier matrix F and an N by N orthogonal matrix, H, known as a Hadamard matrix to be described next.

DISCRETE TIME HADAMARD TRANSFORM

The Hadamard matrix[3] is generated recursively from successive Kronecker (direct) products, $H_N = H_2 \times H_{N/2}$ for

$N = 4, 8, 16, \ldots$ and where $H_2 = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}$. For example,

$$H_4 = H_2 \times H_2 = \begin{pmatrix} H_2 & H_2 \\ H_2 & -H_2 \end{pmatrix} = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & -1 & 1 & -1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 \end{pmatrix}.$$

and in general $H_{2N} = \begin{pmatrix} H_N & H_N \\ H_N & -H_N \end{pmatrix}$.

It is clear that both the orthogonal and symmetric properties of $\frac{1}{\sqrt{2}} H_2$ are preserved in each step of the recursion so that $\left(\frac{1}{\sqrt{N}}\right) H_N$ is both orthonormal and symmetric $\left(\frac{1}{N}\right) HH = I$ and $H = H^T$.

In the analysis of the rest of this paper, we will be concerned only with $N = 2^P$, $P = 1,2,3,\ldots$ as the dimensionality of both the F and H matrices.

A matrix T which relates F and H can be computed from the product F H since from $F = TH$, $FH^{-1} = T = \left(\frac{1}{N}\right) FH$. The structure of the T matrix is rather obscure, however, and a much simpler and appealing relationship will be shown to exist between the fast Fourier transform and a "fast" Hadamard transform to be developed next.

FAST HADAMARD TRANSFORM (FHT)

The recent innovation of the fast Fourier transform in digital signal processing has reduced the requirement for computing the Fourier transform by orders of magnitude from N^2 real products and adds to 2NlogN. A similar algorithm exists for the Hadamard transform, the nature of which we discuss next.

To develop the algorithm, we note[4],

$$\text{Diag } (H_N, H_N) = I_2 \times H_N = \begin{pmatrix} H_N & 0 \\ 0 & H_N \end{pmatrix}$$

and note that $H_2 \times I_N = \begin{pmatrix} I_N & I_N \\ I_N & -I_N \end{pmatrix}$,

$I_N \triangleq$ N by N diagonal matrix. Then a recursion formula can be developed from the following property of direct products:

$$H_N = H_2 \times H_{N/2} = (I_2 \times H_{N/2})(H_2 \times I_{N/2})$$

thus

$$H_4 = (I_2 \times H_2)(H_2 \times I_2) = \begin{pmatrix} H_2 & 0 \\ 0 & H_2 \end{pmatrix} \begin{pmatrix} I_2 & I_2 \\ I_2 & -I_2 \end{pmatrix}$$

$$H_8 = (I_2 \times H_4)(H_2 \times I_4)$$

$$= (I_4 \times H_2)(I_2 \times H_2 \times I_2)(H_2 \times I_4)$$

$$= \begin{pmatrix} H_2 & & & 0 \\ & H_2 & & \\ & & H_2 & \\ 0 & & & H_2 \end{pmatrix} \begin{pmatrix} H_2 \times I_2 & 0 \\ 0 & H_2 \times I_2 \end{pmatrix} \begin{pmatrix} I_4 & I_4 \\ I_4 & -I_4 \end{pmatrix}$$

continuing we have

$$H_{16} = (I_2 \times H_8)(H_2 \times I_8)$$

$$= (I_8 \times H_2)(I_4 \times H_2 \times I_2)(I_2 \times H_2 \times I_4)(H_2 \times I_8)$$

and in general

$$H_N = \prod_{i=1,2,3,}^{\log N} (I_{2^{\log N - i}} \times H_2 \times I_{2^{i-1}}) = \prod_{i=1,2,\ldots}^{\log N} E_{2^i}.$$

Since H_N is symmetric as well as each matrix E_{2^i} in its product decomposition, it follows that

$$H_N = \prod_{i=1,2,3,}^{\log N} E_{2^i}$$

$$= (H_N)^T$$

$$= E_N^T \ldots E_4^T E_2^T$$

$$= E_N \ldots E_4 E_2 .$$

Note that there are N add opeartions in each E matrix with logN such matrices in each product giving NlogN algebraic operations.

We next show the marked similarity of the above recursion to one which generates the fast Fourier transform.

FAST FOURIER TRANSFORM

To do this we return to the matrix F in more detail, devoting our attention primarily to the complex form of the Fourier transform. Hamming[5] and others have analyzed the finite Fourier transform as a real orthogonal matrix whose rows consist of the sampled sines and cosines, but the Fourier-Hadamard relationship is more cumbersome to develop in this form.

If we let $W = e^{-j2\pi/N}$, then W^n becomes the n^{th} harmonic (n^{th} row) of the discrete Fourier transform:[6]

so that $\quad f_n = \sum_{k=0}^{N-1} s_k W^{nk} \qquad n = 0, 1, \ldots, N-1$

and for N = 4, $F = \begin{pmatrix} W^0 & W^0 & W^0 & W^0 \\ W^0 & W^1 & W^2 & W^3 \\ W^0 & W^2 & W^0 & W^2 \\ W^0 & W^3 & W^2 & W^1 \end{pmatrix}$ where $W^{nk} = W^{nk \bmod N}$.

Note that F is symmetric since reversing the exponents of W^{nk} in the summation does not alter the values of the terms in F. To illustrate the FFT using direct products, we alter for convenience the designation of W to lower case $w = e^{-j2\pi}$ so that $w^{1/n}$ becomes the n^{th} root of unity.

Given $F_2 = \begin{pmatrix} w^0 & w^0 \\ w^0 & w^{1/2} \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} = H_2$

by rearranging the row ordering for F_4 it can be shown that (designating $F_N = P F_N'$ where P is a symmetric row permutation matrix to be discussed later),

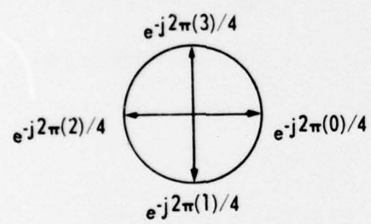
$$F_4' = \begin{pmatrix} F_2 & F_2 \\ G_2 & -G_2 \end{pmatrix}$$

where $G_2 = F_2 D_2 = F_2 \begin{pmatrix} w^0 & 0 \\ 0 & w^{1/4} \end{pmatrix}$

and

$$F_4' = \begin{pmatrix} w^0 & w^0 & w^0 & w^0 \\ w^0 & w^{1/2} & w^0 & w^{1/2} \\ w^0 & w^{1/4} & -w^0 & -w^{1/4} \\ w^0 & w^{3/4} & -w^0 & -w^{3/4} \end{pmatrix} = \begin{pmatrix} w^0 & w^0 & w^0 & w^0 \\ w^0 & w^{1/2} & w^0 & w^{1/2} \\ w^0 & w^{1/4} & w^{1/2} & w^{3/4} \\ w^0 & w^{3/4} & w^{1/2} & w^{1/4} \end{pmatrix}$$

using the symmetry properties of the rotating vector W^{nk}.



$$\therefore F_4' = \begin{pmatrix} F_2 & 0 \\ 0 & F_2 \end{pmatrix} \begin{pmatrix} I_2 & 0 \\ 0 & D_2 \end{pmatrix} \begin{pmatrix} I_2 & I_2 \\ I_2 & -I_2 \end{pmatrix} = (I_2 \times H_2)\, \hat{D}_2\, (H_2 \times I_2)$$

where $\hat{D}_2 \triangleq \mathrm{Diag}\,(I_2, D_2)$

in general one can show the recursion for the Fourier transform:*

$$F_N' = (I_2 \times F_{N/2}')\, \hat{D}_{N/2}\, (H_2 \times I_{N/2})$$

* Since completion of this paper, it has been found that a similar recursion is also developed by M. C. Pease in the Jour. A.C.M. April 1968, pp. 252-264.

where

$$\hat{D}_{N/2} = \left(\begin{array}{c|cccc} I_{N/2} & & 0 & & \\ \hline & w^0 & & & \\ & & w^{1/N} & & \\ 0 & & & w^{2/N} & 0 \\ & & 0 & \ddots & \\ & & & & w^{1/2-1/N} \end{array}\right)$$

Using a property of direct products,[7]

$$I_N \times (A_1 A_2 A_3) = (I_N \times A_1)(I_N \times A_2)(I_N \times A_3)$$

one has for example:

$$F_8' = I_2 \times \left[(I_2 \times H_2)\, \hat{D}_2\, (H_2 \times I_2)\right] \hat{D}_4\, (H_2 \times I_4)$$

$$= \underline{\underline{(I_4 \times H_2)}}\, (I_2 \times \hat{D}_2)\, \underline{\underline{(I_2 \times H_2 \times I_2)}}\, \hat{D}_4\, \underline{\underline{(H_2 \times I_4)}}$$

where $I_2 \times \hat{D}_2 = \mathrm{Diag}\left[1,1,w^0,w^{1/4},1,1,w^0,w^{1/4}\right]$

$\hat{D}_4 = \mathrm{Diag}\left[1,1,1,1,w^0,w^{1/8},w^{2/8},w^{3/8}\right]$.

The matrices present in the FHT can be seen to be embedded in the product form of the FFT above for N = 8. Each doubling of N contributes two new terms and modifies those already present by the direct product with I_2. The general product term becomes:

$$F_N = P F_N'$$

$$= P \prod_{i=1,2,3}^{\log N} \left(I_{2^{\log N - i}} \times \hat{D}_{2^{i-1}}\right)\left(I_{2^{\log N - i}} \times H_2 \times I_{2^{i-1}}\right)$$

$$= P \prod_{i=1,2,3}^{\log N} U_{2^i} E_{2^i}$$

where E_{2^i} make up the FHT. This expansion is equivalent to the "decimation in frequency" form of the FFT.[8] Since F_N is a symmetric matrix one can again reverse the ordering (noting also that $E_{2^i} = E_{2^i}^T$ and $U_{2^i} = U_{2^i}^T$ since U_{2^i} is diagonal),

$$(F_N)^T = \left(\prod_{i=1,2,\ldots}^{\log N} U_{2^i} E_{2^i}\right)^T P^T \qquad P^T = P$$

$$= E_N U_N \ldots E_4 U_4 E_2 U_2 P.$$

This form of the FFT has been programmed in FORTRAN II with provision made for optional computation of the FHT.

Each matrix in the product form of the FFT derived by Brigham and Morrow[6] and by Anderson[9] can also be split into two factors where each product pair is $E_{2^i} P U_{2^{\log N - i}} P$.

Before proceeding, consider the permutation matrix P which related F′ to F after each iteration. We can designate the harmonic number of each row of F_N by subscripting that row vector with the equivalent binary number thus the 2^{nd} harmonic of F_4, the row vector $(w^0\ w^{1/2}\ w^0\ w^{1/2})$ is designated $\vec{f}_{10}$. Then F_2 consists of $\begin{pmatrix}\vec{f}_0\\ \vec{f}_1\end{pmatrix} = \begin{pmatrix}w^0 & w^0\\ w^0 & w^{1/2}\end{pmatrix}$. Each recursion in the FFT "doubles" the harmonic number to obtain even harmonics in the top half of F_N and adds one to each to obtain the respective odd harmonics in the lower half; thus F_4 and F_8 obtained by the recursion has the following ordering:

$$\begin{array}{ccccc} & & & F_8 & \\ & & & \vec{f}_{000} & \\ & F_4 & & \vec{f}_{100} & \\ F_2 & \vec{f}_{00} = \vec{f}_0 & & \vec{f}_{010} & \\ \vec{f}_0 \longrightarrow & \vec{f}_{10} = \vec{f}_2 & \longrightarrow & \vec{f}_{110} & \text{etc.} \\ \vec{f}_1 & \vec{f}_{01} = \vec{f}_1 & & \vec{f}_{001} & \\ & \vec{f}_{11} = \vec{f}_3 & & \vec{f}_{101} & \\ & & & \vec{f}_{011} & \\ & & & \vec{f}_{111} & \end{array}$$

In each case, it is evident that the natural ordering is the reverse ordering of each binary subscript in the above arrays (see also references 6 and 9). The permutation matrix P which reorders the Fourier coefficients to natural order has in each row a 1 in the column defined by the above relationship. Thus for N = 4

$$P = \begin{array}{c} 00\\ 01\\ 10\\ 11 \end{array}\begin{pmatrix} 1&0&0&0\\ 0&0&1&0\\ 0&1&0&0\\ 0&0&0&1 \end{pmatrix}.$$

A further discussion of the P matrix with its significance in the Hadamard-sampled Walsh relationship is presented in appendix A.

COMPARISON WITH OPTIMUM TRANSFORMS

SIGNAL REPRESENTATION AND DISCRIMINATION

The Fourier transform provides a useful and familiar reference with which to compare and relate the properties of the Hadamard transform. There are several application areas, however, such as signal representation and discrimination for which neither transform is "optimum" in the sense of efficiency of representation or classification accuracy.

For this reason, the Hadamard transform was compared also with the following two transforms: the orthogonal matrix consisting of the eigenvectors of the sample correlation matrix of a set of digitized echoes, and a nonsingular matrix which performs the simultaneous diagonalization of two different covariance matrices. The first transform is optimum in the mean square sense for signal representation[10,11] while the second is optimum within the context of a specified decision criterion known as Nilsson's likelihood scheme to be described later in this paper. The optimality property follows directly from the attainment of statistically independent variables through the diagonalization procedure.

A. Signal Representation:

1. Representation of signal sets

It is known that the most efficient set of basis functions for representing a class of signals is determined by the Karhunen-Loeve expansion.[12] For sampled and digitized signals, $\vec{x} = (x_1 \dots x_N)$ where N = 2TW, the optimum coordinant system is readily shown to be the eigenvectors of the sample correlation matrix B,

$$B = ((b_{ij})) \qquad i = 1,2,\dots,N \text{ and } j = 1,2,\dots,N$$

where the correlation matrix is estimated from an ensemble of n column signal vectors $\vec{x}_k$ using the relations

$$B = \frac{1}{n}\sum_{k=1}^{n} \vec{x}_k \vec{x}_k^T \quad \text{and} \quad b_{ij} = E\{x_i x_j\}.$$

For a choice of dimensionality, $r < 2TW$, the coefficients of the signals expanded over the eigenvectors corresponding to the r largest eigenvalues of B will represent the signals with least loss of energy on the average. To show this variational property one finds the stationary values of the quadratic form

$$Q(\vec{x}) = \vec{x}^T B \vec{x} \ ,$$

where $\vec{x} \triangleq$ column vector subject to the constraint

$$0 = N(\vec{x}) = \vec{x}^T\vec{x} - 1 \ ,$$

by extremizing the functional

$$H(\vec{x}) = Q(\vec{x}) - \lambda N(\vec{x}) \ .$$

From calculus of variations, the set of Euler-Lagrange equations are:[13]

$$\frac{\partial H}{\partial x_i} = \frac{\partial}{\partial x_i} Q(\vec{x}) - \lambda \frac{\partial}{\partial x_i}(\vec{x}^T\vec{x} - 1) = 0 \qquad i = 1,2,\dots,N$$

from which

$$\frac{\partial}{\partial x_i}(\vec{x}^T\vec{x}) = 2x_i, \text{ and } \frac{\partial}{\partial x_i}(\vec{x}^T B\vec{x}) = 2\sum_{j=1}^{N} b_{ij} x_i$$

$$\therefore \lambda x_i - \sum_{j=1}^{N} b_{ij} x_j = 0 \qquad i = 1,2,\dots,N$$

or $A\vec{x} = \lambda\vec{x}$.

The actual maximum is of course λ_{max} since for any eigenvector, $\vec{\xi}$, $\vec{\xi}^T B \vec{\xi} = \vec{\xi}^T \lambda \vec{\xi} = \lambda$. It follows from this that the squares of the projections of the signal vectors on the eigenvector, $\vec{y}$, corresponding to λ_{max} are maximized due to the nature of the correlation matrix. That is, let the signals be column vectors $\vec{x}_1, \ldots, \vec{x}_n$ then,

$$\lambda_{max} = \vec{y}^T B \vec{y} = \frac{1}{n}\sum_{k=1}^{n} \vec{y}^T (\vec{x}_k \vec{x}_k^T)\, \vec{y} = \frac{1}{n}\sum_{k=1}^{n} (\vec{y}^T \vec{x}_k)^2$$

the expression on the right being the sum of the squares of the signal vectors projected on $\vec{y}$.

For a comparative measure of the representation efficiency obtained with the Hadamard transform, the ratio "r_e" between the average energy retained by a subset of Hadamard coefficients and the total energy (full set of coefficients) was computed for several signal sets with dimensionality 2TW = 32.

The same information for the eigenvectors of the correlation matrix exists in the corresponding eigenvalue spectrum.

A comparison of the number of coefficients required by each of the two transforms to retain the same amount of signal energy on the average (r_e) is a measure of the relative efficiency of the Hadamard transform.

Spectral plots of the eigenvectors and the Hadamard codes are shown in figure 1 for a representative data set consisting of 52 signals. The relative amplitudes for the Hadamard codes are the mean of the squares of the values of the 11 largest coefficients. For the eigenvectors the function illustrated in this figure is the relative magnitude of each eigenvalue plotted against its position in a list of the 11 largest eigenvalues when placed in order of decreasing values.

The efficiency is such that 82% of the sum of the expected squares of the coefficients of all 52 signal vectors making up the data set are contained in a three-dimensional space represented by the projections of the signals across three eigenvectors. In comparison, the binary coordinate systems required seven components to achieve equal representation for this data set.

Figure 2 shows graphically the reconstruction efficiency for one particular echo in the signal set.

The reconstruction of a signal based on the most representative coefficients of a set of signals tended to overemphasize the lower frequency Hadamard rows possibly due to a "washout" effect of the signal ensemble on the relatively large number of rows with considerable high frequency spectral content. One finds from the Hadamard-Walsh relationship (appendix A) that the spectra of each of fully half of the functions in the Hadamard transform is in the form of the lower sidebands of one of the other half modulating the sampled values of $\Psi_{100\ldots0}$ (equivalent to a sampled $\cos N\pi t$ carrier, $0<t<1$).

The eigenvector corresponding to λ_{max} tends to resemble (but not equal) the mean of a group of positive-going signals, thereby retaining a modest amount of high frequency information present in the mean.

In conclusion, the representation efficiency of the Hadamard transform (based on mean square criteria) was found to be surprisingly good relative to the eigenvectors of the sample correlation matrix.

2. Representation Efficiency of a Single Signal

Several interesting observations can be made regarding the Fourier vs. Hadamard representation of a single signal.

By thinking in general terms again about the Fourier spectra of sampled Walsh functions, one envisions each of those which modulate the sampled $\Psi_{100\ldots0}$ "carrier" as being linear combinations of the high frequency Fourier components of the signal with bandwidth 2W. Therefore, while a signal having considerable energy spread over the higher frequency range is by definition not efficiently represented by its Fourier coefficients, it often is by its Hadamard coefficients.

In the extreme, signals which undergo spectral widening transformations, such as clipping or phase modulation, are often more efficiently represented by their Hadamard coefficients.

In figure 3-A, a signal with comparative Fourier and Hadamard reconstruction efficiency (using eight coefficients) is shown, with the Fourier line spectra of one of the best Hadamard coefficients shown in figure 3-B. In figure 4 the reconstruction of a signal with 2TW = 256, using the FHT, is shown, using the best 32 coefficients.

SIGNAL DISCRIMINATION, THE PATTERN RECOGNITION PROBLEM

The problem of discriminating between two or more classes of signals can be subdivided into two distinct phases, the estimation problem and the classification problem. In the estimation phase, the parametric technique for arriving at a discriminant configuration was chosen.

The parametric approach consists of the techniques of statistical estimation, estimation of probability densities, computation of sample means, etc. The statistical parameters are computed using a set of data known as the training set, and the results are considered as estimates of the true parameters describing a larger class of data generated by mechanisms assumed to be similar to those which generated the training data. This larger class represents the real world or working phase in the pattern recognition problem.

The degree of sophistication employed (i.e., sample means vs. probability densities) is usually determined by how accurately these parameters, as estimated from the training set, are actually expected to describe the unknown data. This generalization problem is of central importance in any pattern recognition problem.

Three estimation and discrimination schemes were studied in regard to their capabilities with the training set only, and with a small degree of generalization (50% training set, 50% unknown), using approximately 250 signals, each having a dimensionality of 32 time samples.

These methods were:

a. *The matched filter, based on sample means.*

b. Discriminants, derived from independent probability density estimates (simultaneous diagonalization of two different covariance matrices).

c. Discriminants derived from probability densities at the outputs of the Hadamard transform.

ESTIMATION BASED ON SAMPLE MEANS

Matched Filter: The matched filter (linear discriminant) configuration consists of the difference of the mean vectors of the two classes. The term "matched filter," is a specialization in the sense that only the correlation for zero delay or dot product between the signal vector $\vec{s}$ and a stored vector is implied rather than a full convolution operation.

The estimation problem involves merely the computation of the sample mean vector for each class in the training set. The classification problem for which this procedure is optimum is that in which the covariance matrices, K_i, are alike and are assumed to be scalar matrices. The i^{th} class consists of a mean vector $\vec{m}_i$ in a 2TW dimensional signal space surrounded by a spherical cluster of vectors $\vec{n}_j$, $j=1,2,\ldots,r$ each representing $\vec{m}_i$ perturbed by independent additive Gaussian noise with zero mean (figure 5-A). For this paper, we will consider only discrimination between two classes, $i=1,2$. Since the covariance matrix for the multivariate Gaussian distribution is the same for each class, the conditional a priori probability density is[1]

$$P(\vec{x}|i) = \left(\frac{1}{2\pi N_0 W}\right)^{TW} \exp -1/2 \left\{(\vec{x}-\vec{m}_i)^T K^{-1} (\vec{x}-\vec{m}_i)\right\}$$

where $K^{-1} = \frac{1}{\sigma^2} I_{2TW}$ $\quad I_{2TW} \triangleq$ identity matrix

$\sigma^2 = N_0 W$ $\quad N_0 \triangleq$ noise power spectral density.

The log likelihood decision is $g(\vec{x}) = \ln \frac{P(\vec{x}|1)}{P(\vec{x}|2)} \gtrless 0$

The surface $g(\vec{x}) = 0$, is a hyperplane separating the clusters in figure 5-A between the means $\vec{m}_1$ and $\vec{m}_2$. The likelihood function is simply, let $\sigma = 1$, for convenience

$$g(\vec{x}) = \left[2\vec{x}^T \cdot (\vec{m}_1 - \vec{m}_2) - \vec{m}_1 \cdot \vec{m}_1 + \vec{m}_2 \cdot \vec{m}_2\right] \gtrless 0.$$

An inspection of figure 5-B and the above expression for two classes shows that the decisions are amplitude sensitive due to the scalar terms which are functions of the energy content of the mean vector $\vec{m}_i$. Often amplitude invariance is desired so a modified (not necessarily optimum) discriminant is employed, resulting from the normalization of $\vec{m}_i$ (division by $||\vec{m}_i||$ $i=1,2$). The resulting correlation discriminant $\vec{x}^T \cdot (\vec{m}_1 - \vec{m}_2) \gtrless 0$ yields decisions affected only by signal shape information (direction and not length of $\vec{x}$ as in figure 5-C).

If $\vec{y} = H\vec{x}$, and $\vec{u}_i = H\,\vec{m}_i$, a discriminant with similar properties operating on the Hadamard mean vectors $\vec{u}_1$ and $\vec{u}_2$ can be expressed by $\vec{y}^T \cdot (\vec{u}_1 - \vec{u}_2) \gtrless 0$, where $||\vec{u}_i|| = \sqrt{N}\,||\vec{m}_i||$. In addition, since the Hadamard (or any nonsingular) transform on the data effects a similarity transform on the covariance matrix, K, it is left unchanged,

$$(H^{-1})^T (\tfrac{1}{N} I) H^{-1} = \frac{1}{N} I H H^T = \frac{1}{N} I .$$

The actual discriminant can often be simplified by truncating the vector representing the difference of the normalized mean vectors $\vec{u}_1$ and $\vec{u}_2$ of the two classes making up the training set. In figure 6-A a discriminant is shown (in the time domain) resulting from the linear combination of those Hadamard rows corresponding to, and weighted by, the seven largest elements of $\vec{d}$, where $\vec{d} = \vec{u}_1 - \vec{u}_2$. A comparison of this vector with the vector $\vec{d}$, figure 6-B, shows the small effect of truncation for this data set (data set 1). Note the relative lack of high-frequency detail in the discriminant, a property which will be shown later to be in marked contrast to the spectral content of discriminants based on probability density estimates (simultaneous diagonalization).

ESTIMATION AND DISCRIMINATION USING PROBABILITY DENSITIES

The more general problem of discrimination between two multivariate Gaussian processes[14] with estimated parameters $N(\vec{m}_i, K_i)$, $i = 1,2$, involves the determination of the conditional probabilities $p(\vec{x}|i)$

$$p(\vec{x}|i) = \frac{1}{(2\pi)^{TW} |K_i|^{1/2}} \exp\left\{-\frac{1}{2}(\vec{x}-\vec{m}_i)^T K_i^{-1} (\vec{x}-\vec{m}_i)\right\}$$

with $K_1 \neq K_2$ in general. The log likelihood decision for equal p(i) becomes

$$g(\vec{x}) = \ln \frac{p(\vec{x}|1)}{p(\vec{x}|2)} \gtrless 0 \quad \text{or} \quad g(\vec{x}) = \frac{1}{2}\left[\ln \frac{|K_2|}{|K_1|} + Q_2 - Q_1\right] \gtrless 0$$

where Q_i is the term in the exponent above.

A linear nonsingular transformation A on the data vector $\vec{x}$ results in the vector $\vec{y}$ with Gaussian parameters $N(\vec{u}_i, \Sigma_i)$ where Σ_i results from a similarity transformation on K_i by A. Of particular theoretical interest is the transformation R which simultaneously diagonalizes K_1 and K_2 resulting in:[15]

$$g(\vec{x}) = \frac{1}{2}\left[\ln|\Lambda| + \vec{r}_2^T \Lambda^{-1} \vec{r}_2 - \vec{r}_1^T \cdot \vec{r}_1\right] \gtrless 0$$

where Λ is K_2 diagonalized, $\vec{r}_i = R(\vec{x} \cdot \vec{m}_i)$ and K_1 is transformed to the identity matrix. The coefficients of this transform can then be processed as independent univariate Gaussian variables.

An application of this procedure in an asymptotic sense to the class of covariance matrices describing noncyclic stationary processes was described by Capon[15] who employed sampled power spectra obtained from the Fourier transform.

One observes the presence of a bias term and energy measurement terms in the likelihood function above, resulting in amplitude sensitive decisions, an undesirable feature in many pattern recognition applications. Furthermore, one of the exploitable features of the Hadamard transform, from an implementation standpoint, is its binary nature. Because of this it is often desirable to process the Hadamard coefficients as random variables, by clipping them, followed by an estimation scheme requiring only the area under the positive portion of the probability density function of each coefficient.

Such a scheme has been reported by Nilsson[16] which in turn develops a set of weights by which the clipped coefficients of unknown signals are linearly combined as a discriminator. To develop the weighted discriminant, the assumption of statistical independence between the variables is made which, for the usual case of nondiagonal covariance matrices, is unjustified. Because of the implementation simplicity of the method, however, it was desirable to compare the results of using Nilsson's approach on the coefficients of both the Hadamard and "R" transform, the latter satisfying the requirement of statistical independence.

Nilsson's method is summarized here.

Let p(1) = p(2), and consider from Bayes theorem the ratio,

$$\frac{p(1|\vec{x})}{p(2|\vec{x})} = \frac{p(\vec{x}|1)}{p(\vec{x}|2)} \lessgtr 1 \quad \text{or} \quad g(\vec{x}) = \ln \frac{p(\vec{x}|1)}{p(\vec{x}|2)} \lessgtr 0$$

where $\vec{x} \triangleq$ transform coefficient vector. With statistical independence we have:

$$p(\vec{x}|1) = \prod_{i=1}^{n} p(x_i|1) \quad \text{etc.} \qquad \therefore g(\vec{x}) = \sum_{i=1}^{n} \ln \left\{ \frac{p(x_i|1)}{p(x_i|2)} \right\} \lessgtr 0$$

By clipping the transformation coefficients, the notation complexity reduces from the analog case to the following:

$$p(x_i = 1|1) \triangleq p_i\,, \quad p(x_i = 0|1) \triangleq 1 - p_i$$

$$p(x_i = 1|2) \triangleq g_i\,, \quad p(x_i = 0|2) \triangleq 1 - g_i$$

$$\therefore g(\vec{x}) = \sum_{i=1}^{n} x_i \ln\left[\frac{p_i(1-g_i)}{g_i(1-p_i)}\right] + \sum_{i=1}^{n} \ln \frac{(1-p_i)}{(1-g_i)} \gtrless 0$$

which is a linear discriminant estimated from the cumulative probability for x_i.

Note that the presence of the bias term does not introduce amplitude sensitivity in the decision, due to clipping. For a predetermined transform, (Hadamard, Fourier) the clipping also renders the estimation scheme amplitude insensitive on a signal-to-signal basis.

SIMULTANEOUS DIAGONALIZATION

While the "R" matrix is known to be the Fourier transform for stationary processes, most data encountered in pattern recognition problems belong to the more general class of non-stationary processes. For such processes *a matrix is derived*[17] which performs a linear transformation of the signal data in such a way that the covariance matrix of class 1 is reduced to the identity matrix while the covariance matrix describing class 2 is diagonalized. For class 1, the random variables obtained by correlating each row of the filter matrix with the signals will be independent with unit variance about the transformed mean. For class 2 the outputs are again statistically independent, but with variances proportional to the respective eigenvalues. The technique is accomplished in the following manner.

The parameters of a set of multivariate Gaussian signals, $\vec{X}$, are described by a mean vector $\vec{m}_1$ and a covariance matrix, R, where

$$R = \left(\left(\sigma_{ij}\right)\right), \quad i = 1,2,\ldots,N \text{ and } j = 1,2,\ldots,N\,.$$

The covariance matrix can be estimated from a set of signals using the relation:

$$\sigma_{ij} = E\left[(x_i - m_i)(x_j - m_j)\right],$$

In the procedure, one computes the product $(R_1^{-1/2})^T R_2 R_1^{-1/2}$ where the matrices R_1 and R_2 arise in the following manner.

We let R_2 be the covariance matrix of class 2 and then determine $R_1^{-1/2}$ by first diagonalizing the covariance matrix (R_1) of class 1. The i^{th} column of the column eigenvector matrix Q, which performs the diagonalization, is multiplied by $\frac{1}{\sqrt{\lambda_i}}$ ($\lambda_i \triangleq i^{th}$ eigenvalue of R_1) yielding the matrix $R_1^{-1/2}$, known as the whitening filter for class 1, since $(R_1^{-1/2})^T R_1 R_1^{-1/2} = I$.

The symmetric matrix $(R_1^{-1/2})^T R_2 R_1^{-1/2}$ results when the transformation which whitens the covariance statistic of class 1 is applied to the covariance matrix of class 2. When a second transformation (M) is found that diagonalizes $(R_1^{-1/2})^T R_2 R_1^{-1/2}$, it can be seen that it will not affect the statistics of class 1, since $M^T I M = I$ where $M^T M = I$, the matrix M consisting of the column eigenvectors of $(R_1^{-1/2})^T R_2 R_1^{-1/2}$. We then obtain, finally, a linear transformation R which is equal to:

$$\vec{X} = QHM\vec{y} \qquad \text{where } QH = R^{-\frac{1}{2}}$$

or

$$\vec{y} = (QHM)^{-1}\vec{X}$$

$$\vec{y} = R\vec{X}$$

$$R = M^T H^{-1} Q^T.$$

The R matrix then yields a set of 32 time-domain filters which yield outputs that are statistically independent for both classes and, thus, could be thought of as the optimum set of linear observables for discriminating between two stationary or nonstationary Gaussian processes.[18]

Figure 7 shows one row of the "R" transformation obtained from the above procedure for a typical pair of classes. Note the predominant high frequency content in the spectra of the filter which is in direct contrast to the matched filter discriminants.

This phenomenon was noted in all rows of the "R" transformation matrix and occurred consistently for all data sets studied. It appears to be related to the decrease in statistical dependence with increasing frequency between the coefficients of the Fourier expansion of a stationary Gaussian time series.[19]

A comparison of the effectiveness of the Hadamard and "R" transforms for signal discrimination is illustrated in figure 8 for a representative signal set (2TW=32), consisting of 35 signals from class 1 and 35 from class 2 (data set 1). The Nilsson scheme of estimation and classification was employed on the transform coefficients in both cases. The relative classification accuracy is plotted against the number of coefficients with largest weights in Nilsson's method used to form the discriminant. The largest subset of coefficients shown is eight out of a possible 32. Although the absolute scale is not shown, all accuracies exceeded 50%.

While the difference between the best coefficient of each transform is not representative of all data sets studied the rate of increase and resulting divergence of the accuracies, as extra coefficients are added to the discriminant, is typical of the data and is due to statistical dependence between the Hadamard coefficients.

Although studies, performed with the training set only, provide useful clues on the relative effectiveness of probability density measures on the various transform coefficients, the generalization properties of the estimation procedures is of considerable importance and is described next.

GENERALIZATION PROPERTIES

Defining a generalization ratio g_r by:

$$g_r = \frac{\text{No. of signals in training set}}{\text{No. of unknown signals (working phase)}}$$

the Hadamard transform, "R" transform, and matched filter discriminant were evaluated against the training set only (g_r = 1.0) and for g_r = 0.5.

For the "R" transform, a technique sufficient for the purposes of testing relative generalization capability, was employed by summing selected "R" transform coefficients. This scheme (suggested by the statistical independence of the coefficients) allows for prediction of the classification performance on the training set since, for n independent Gaussian variables, $Var_T = \sum_{i=1}^{n} Var_i$ and $u_T = \sum_{i=1}^{n} u_i$. Classification accuracies predicted from the standard deviation tables for u_T and $(Var)_T$ agreed closely with the accuracies actually measured.

Table 1 shows a typical set of relative discrimination accuracies for studies involving 220 signals (2TW = 32), with g_r = 0.5. The eight best coefficients estimated from the training set for each transform were used. As before, while absolute accuracies are not indicated, all exceeded 50%.

TABLE 1

Comparison of Discriminant Techniques

	Training Set	Unknown
	Δ%	Δ%
Simultaneous Diagonalization	25	0
Matched Filter	12	8
Hadamard (Nilsson Technique)	11	0

From the table it can be seen that significant improvement over the matched filter in classification of the training set appears possible by obtaining independent probability measures of the two classes. The generalization capability of the matched filter, however, exceeded that obtained by the "R" transformation for the data studied, possibly due to the differences in spectral content of the two discriminants. Although the clipping operation in Nilsson's method prevents a time domain realization as a linear discriminant, those Hadamard codes included in the Nilsson discriminant exhibit a more balanced spectra (high and low frequency content).

The FHT was applied to a set of 3500 digitized signals each with 256 time samples under more severe generalization conditions (10% training set, 90% unknown). The Nilsson and the matched filter discriminants were estimated from the full transform and employed as classifiers on the best 32 Hadamard coefficients. It was found that the sample-means discriminant was again more effective from a generalization standpoint than those estimated from probability densities.

CONCLUSIONS AND APPLICATION AREAS

The Hadamard transform appears to have useful application in signal representation especially in the areas of data compression systems. Several investigators have noted the applicability of the transform for further processing the power spectral coefficients of the channel vocoder[20-22] before transmission over a channel. Such studies have emphasized reduced bit rate due to coarse quantization of selected coefficients rather than truncation of the number of coefficients.

Several investigators[23] have studied the relative merits of the "transformation compression" approach with other methods of data compression, studying in particular the Fourier and Karhunen-Loeve techniques using the coefficient truncation technique. In reference 23 it is concluded that the two techniques are efficient in performance but impractical in implementation. For those systems considered in reference 23 which include digital sensors or A/D conversion of the data, it would seem that perhaps the FHT would alleviate the implementation problem.

In signal discrimination it would appear that the use of probability densities on selected transform coefficients should be employed with generalization in mind for signals with appreciable variability in shape within a class. The means of the transform coefficients appear to be useful for many practical problems. Often in pattern recognition, one searches for transform properties which render it less sensitive to a particular type of variation of a signal such as translation, change in amplitude, etc. Such attacks against the generalization problem tend to overshadow in importance the excellent methods and insights given to us by theory for discriminating between training sets or even generalizing on the basis of partially known parameters.

One of the principle applications of the FFT in signal processing is in the performance of rapid convolution using frequency domain techniques[27] including matched filter processing of pulse compression radar echoes, etc. Because the Hadamard transform does not possess these translation-related properties it would not be expected to play a major role in this area (see appendix B).

Orthogonal multiplex data transmission systems have been discussed in references 24 and 25, with some comments made in the latter concerning the burst error immunity properties of such systems. It would appear that the FHT would have application in the formation of the transmission signal and in the receiver.

APPENDIX A

THE HADAMARD-WALSH RELATION AND THE SIGNIFICANCE OF P

The permutation matrix P can be shown to be identical to that which carries the ordering of the rows of the Hadamard matrix to the matrix whose rows are the sampled Walsh functions, termed the Walsh matrix[26]. The Walsh functions in turn have been defined[26] over the interval (0,1) as the successive modulation in various combinations of a set of square waves $\Psi_p(t)$, $p = 2^n$, $n = 0, 1, 2, \ldots r$, where p designates the number of periods in the interval. For $p = 0$, $\Psi_0(t) \triangleq 1$, $0<t<1$. Replacing p by its equivalent r bit binary number, the waveform of a particular Walsh function is specified as the product of those Walsh functions represented by each "1" in its binary subscript. Letting $r = 3$ for example, $\Psi_{011} = (\Psi_{001})(\Psi_{010})$ is seen to be the modulation of Ψ_{001} by Ψ_{010} (see figure 9).

The row vectors of the Walsh matrix of order $N = 2^r$ are designated as $\vec{\Psi}_{b_r \ldots b_0}$, $b_i = 0,1$, and can be generated from a recursion relation involving $\vec{\Psi}_0$ and $\vec{\Psi}_1$. For $N = 2$,

$$H_2 = \begin{pmatrix} \vec{\Psi}_0 \\ \vec{\Psi}_1 \end{pmatrix} \text{ where } \vec{\Psi}_0 = (1,1),\ \vec{\Psi}_1 = (1,-1).$$

Then it follows that:

$$\vec{\Psi}_{1\, b_r \ldots b_0} = \vec{\Psi}_{b_r \ldots b_0} X (1,-1) = \vec{\Psi}_{b_r \ldots b_0} X \vec{\Psi}_1$$

$$\vec{\Psi}_{0\, b_r \ldots b_0} = \vec{\Psi}_{b_r \ldots b_0} X (1,1) = \vec{\Psi}_{b_r \ldots b_0} X \vec{\Psi}_0$$

where $\vec{a} X \vec{b} = (a_1 b_1, a_1 b_2, a_2 b_1, a_2 b_2)$ $\quad \vec{a} = (a_1, a_2)$, $\vec{b} = (b_1, b_2)$.

From this it follows directly that

$$\vec{\Psi}_{b_n \ldots b_1 b_0} = \vec{\Psi}_{b_0} X \vec{\Psi}_{b_1} X \ldots X \vec{\Psi}_{b_n} \qquad b_i = 0,1$$

for example $\vec{\Psi}_{110}$ results from $\vec{\Psi}_{10}$ by the operation

$$\vec{\Psi}_{110} = \vec{\Psi}_0 X \vec{\Psi}_1 X \vec{\Psi}_1 = (1\ 1) X (1\ {-1}) X (1\ {-1}) = (1\ {-1}\ {-1}\ 1\ 1\ {-1}\ {-1}\ 1).$$

At the same time it can also be shown that

$$\vec{h}_{b_n \ldots b_1 b_0} = \vec{\Psi}_{b_n} X \ldots X \vec{\Psi}_{b_1} X \vec{\Psi}_{b_0},$$

where $\vec{h}_{b_n \ldots b_1 b_0}$ is the $b_n \ldots b_1 b_0$ row of H.

For example, $h_{110} = \vec{\Psi}_1 X \vec{\Psi}_1 X \vec{\Psi}_0$ is the only nonzero row of $E_1 H_2 X E_1 H_2 X E_0 H_2 = (E_1 X E_1 X E_0) H_8$ where $E_1 = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}$, $E_0 = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$

Thus the same permutation matrix specifies the Hadamard-sampled Walsh reordering and the required reordering necessary in the FFT. Because of this equivalence, other interpretations exist for the generation of each Fourier harmonic, (row of F). For example if N = 4 it is clear that $\vec{f}_3$ is the complex modulation of $\vec{f}_2$ with $\vec{f}_1$,

$$\vec{f}_{01} = \vec{f}_1 = (w^0, w^1, w^2, w^3) = (w^0, w^2) X (w^0, w^1)$$

$$\vec{f}_{10} = \vec{f}_2 = (w^0, w^2, w^0, w^2) = (w^0, w^0) X (w^0, w^2)$$

$$\vec{f}_{11} = \vec{f}_3 = \left[w^0 w^0, w^1 w^2, w^2 w^0, w^3 w^2 \right] = (w^0, w^3, w^2, w^1).$$

APPENDIX B

PROPERTIES OF THE HADAMARD TRANSFORM AND FOURIER-HADAMARD ANALOGIES

There are three properties of the Fourier transform which have contributed greatly to its usefulness in signal processing applications, the "shift" property and the resulting convolution theorem[27,28] being perhaps the most useful. The others include the invariance of the power spectrum to translation and finally the simultaneous diagonalization of all covariance matrices describing stationary processes.

The "shift" property of the Fourier trrnsform is stated mathematically be the statement,

$$\text{if } F(s) = \int_{-\infty}^{\infty} f(t)\, e^{-jst}\, dt$$

then the Fourier transform of $f(t - \tau)$ is $e^{-j\tau s} F(s)$.

In discrete time an equivalent expression for the shift property is the following, if $\vec{f} = F \vec{s}$, then $\vec{f}_1 = F\, T^1\, \vec{s}$

$$= \begin{pmatrix} W^0 & & & 0 \\ & W^1 & & \\ & & \cdot & \\ 0 & & & W^{N-1} \end{pmatrix} F \vec{s} \qquad \text{where } T^1 = \begin{pmatrix} 0 & 0 & . & . & 1 \\ 1 & 0 & . & . & 0 \\ 0 & 1 & . & . & 0 \\ . & . & . & . & . \\ 0 & . & . & 1 & 0 \end{pmatrix}$$

and $W = e^{j2\pi/N}$ as before.

Then $\vec{f_1} = \text{Diag}\left[W^0, W^1, \ldots, W^{N-1}\right] \vec{f} = D\ \vec{f}$.

The "shift" property becomes equivalent to the specification $F\,T^1 = D\ F$ or $F\,T^1\,F^* = D$

thus the shift property is exhibited by the set of eigenvectors of the translation operator T. By expanding $|\lambda I - T| = 0$ one obtains W^i, $i = 0, 1, \ldots, N-1$ as the set of eigenvalues, thus each row of the Fourier matrix is an eigenvector since, for example,

$$T^{N-1} \begin{pmatrix} W^0 \\ W^1 \\ \cdot \\ W^{N-1} \end{pmatrix} = W^1 \begin{pmatrix} W^0 \\ W^1 \\ \cdot \\ W^{N-1} \end{pmatrix} \text{ obviously holds.}$$

Since any translations less than the period is expressible as a power of T, the Fourier matrix is the set of eigenvectors diagonalizing T and thus exhibits the shift property in discrete time.

The invariance of the power spectrum to translation follows directly from thms property. The class of covariance matrices diagonalized by the Fourier matrix is also easily specified as linear combinatinns of the powers of T,

$$K = a_0 I + \sum_{i=1}^{N-1} a_i \left(T^i + T^{N-i}\right)$$

which is identified as the set of cyclic stationary processes.[30]

It is evident that the Hadamard transform does not possess these properties with respect to the translation operation. It does possess these properties with respect to an operator defined by the following relation

$$S_k = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}^{[k]} X\ I_{2^{\log N - k}} \qquad k = 1, 2, \ldots, \log N$$

where $A^{[k]}$ designates the Kronecker power of A, $\underbrace{A\ X\ A \ldots X\ A}_{k \text{ terms}}$, or products of the above terms. In particular, the linear combination of such products defines a class of covariance matrices diagonalized by the Hadamard matrix.

Letting $\hat{I}_2 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$, this follows from

$$H_N S_k H_N = \left(H_2^{[k]}\ X\ H_2^{[\log N - k]}\right) S_k \left(H_2^{[k]}\ X\ H_2^{[\log N - k]}\right),$$

$$= H_2^{[k]}\, \hat{I}_2^{[k]}\, H_2^{[k]}\ X \left(H_2^{[\log N - k]}\right)^2.$$

Using the property $(AB)^{[k]} = A^{[k]} B^{[k]}$, the above expression becomes $(H_2 \hat{I}_2\ H_2)^{[k]}\ X\ I_2^{[\log N - k]}$ which is a diagonal matrix for $k = 1, 2, \ldots, \log N$. For example, let $N = 4$, then

$$S_1 = \left(\begin{array}{cc|cc} & 0 & 1 & 0 \\ & & 0 & 1 \\ \hline 1 & 0 & & 0 \\ 0 & 1 & & \end{array}\right) \text{ etc.}$$

REFERENCES

1. H.W. Harmon, Principles of the Statistical Theory of Communication, McGraw Hill, 1963.
2. J. W. Cooley, P.A.W. Lewis, and P.D. Welch, "Application of the Fast Fourier Transform to Computation of Fourier Integrals, Fourier Series and Convolution Integrals," IEEE Transactions on Audio and Electro Acoustics, June 1967.
3. S.W. Golomb et al., Digital Communications, Prentice Hall, 1964.
4. R. Bellman, Introduction to Matrix Analysis, McGraw Hill, 1960.
5. R.W. Hamming, Numerical Methods for Scientists and Engineers, McGraw Hill, 1962.
6. E.O. Brigham and R.E. Morrow, "The Fast Fourier Transform," IEEE Spectrum, December 1967.
7. C.C. MacDuffee, The Theory of Matrices, Chelsea.
8. G-AE Subcommittee on Measurement Concepts, "What is the Fast Fourier Transform," IEEE Transactions on Audio and Electro-Acoustics, June 1967.
9. E.A. Robinson, Multichannel Time Series Analysis with Digital Computer Programs, Holden Day, 1967.
10. T.Y. Young and W.H. Huggins, "On the Representation of Electrocardiograms," IEEE Transactions on Bio-Medical Electronics, July 1963.
11. A.P. Kramer and M.V. Mathews, "A Linear Coding for Transmitting a Set of Correlated Signals," IRE Trans. on Information Theory, Vol. IT-2, pp. 41-46, Spetember 1956.
12. W.B. Davenport and W.L. Root, Introduction to Random Signals and Noise, McGraw Hill, 1963.
13. J.W. Dettman, Mathematical Methods in Physics and Engineering, McGraw Hill, 1962.
14. G.S. Sebestyen, Decision Making Processes in Pattern Recognition, Macmillan Company, New York 1962.
15. J. Capon, "An Asymptotic Simultaneous Diagonalization Procedure for Pattern Recognition," Information & Control, Vol. 8, pp 264-281, 1965.
16. Nils J. Nilsson, Learning Machines, McGraw Hill, 1965.
17. G. Hadley, Linear Algebra, University of Chicago, 1961.
18. T.T. Kadota and L.A. Shepp, "On the Best Finite Set of Linear Observables for Discriminating Two Gaussian Signals," IEEE Transactions on Information Theory, Vol. Vol. IT-13, No. 2, April 1967.
19. Athanasios Papoulies, Probability, Random Variables and Stochastic Processes, McGraw Hill, 1965.
20. C.M. Rader and W.R. Crowther, "Efficient Coding of Vocoder Channel Signals Using Linear Transformation," Proc. IEEE, pp. 1594-95, November 1966.

21. L.M. Goodman, "A Binary Linear Transformation for Redundancy Reduction," Proc. IEEE, pp. 467-67, March 1967.

22. B. Gold and C.M. Rader, "The Channel Vocoder," IEEE Transactions on Audio and Electro Acoustics, Vol. AU-15, No. 4, December 1967.

23. C.A. Andrews, J.M. Davies, and G.R. Schwartz, "Adaptive Data Compression," Proc. IEEE, Vol. 55, No. 3, March 1967.

24. H.F. Harmuth, "A Generalized Concept of Frequency and Some Applications," IEEE Transactions on Information Theory, Vol. IT-14, No. 3, May 1968.

25. R.F. Filipowsky, "Multiorthogonal Data Transmission Systems," IEEE International Conference on Communications, June 12-14, 1967. 10-5.

26. M. Corrington and R. Adams, "Applications of Walsh Functions to Nonlinear Analysis," R.C.A. sponsored by RADC, Contract No. F30(602)2484.

27. R. Bracewell, The Fourier Transform and Its Applications, McGraw Hill, 1965.

28. H.F. Davis, Fourier Series and Orthogonal Functions, Allyn Bacon, 1963.

29. M. Rosenblatt, Proceedings of the Symposium on Time Series Analysis, Brown University, 1962.

Acknowledgement

This work was sponsored in part by the Naval Ship Systems Command, Washington, D. C.

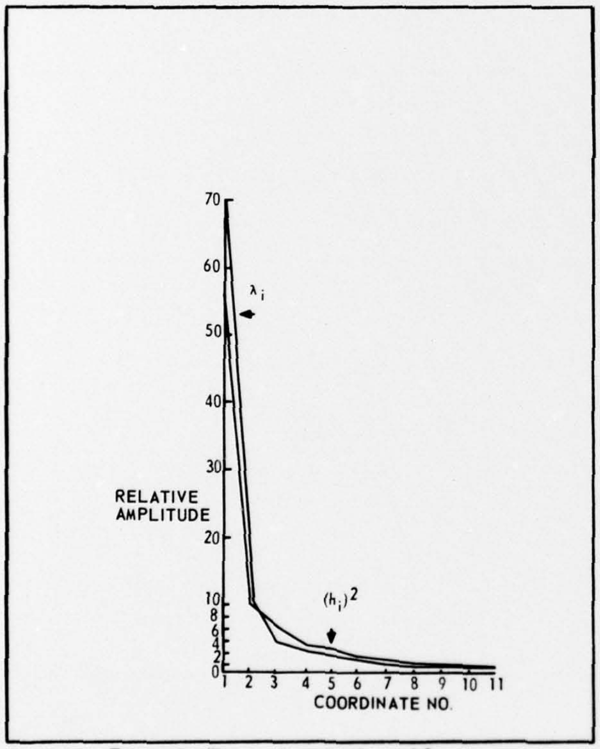


Figure 1. Eigenvalue - Hadamard Spectra

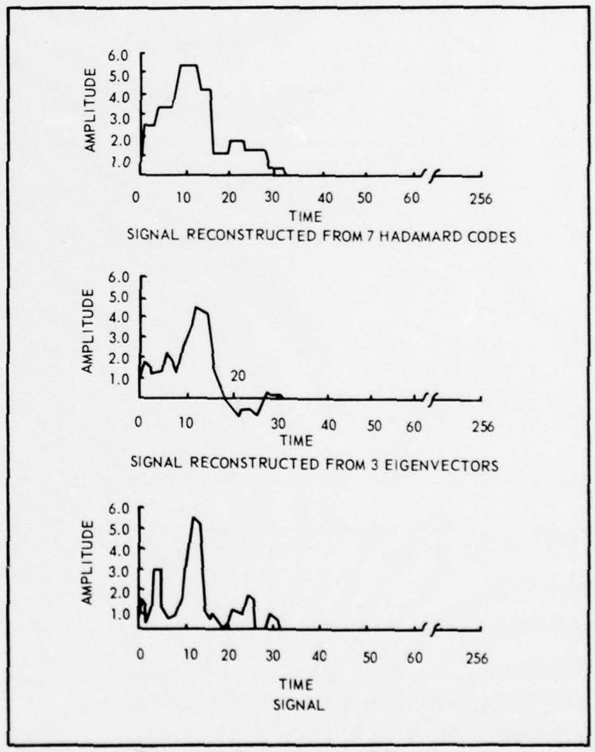


Figure 2. Reconstructed Efficiency

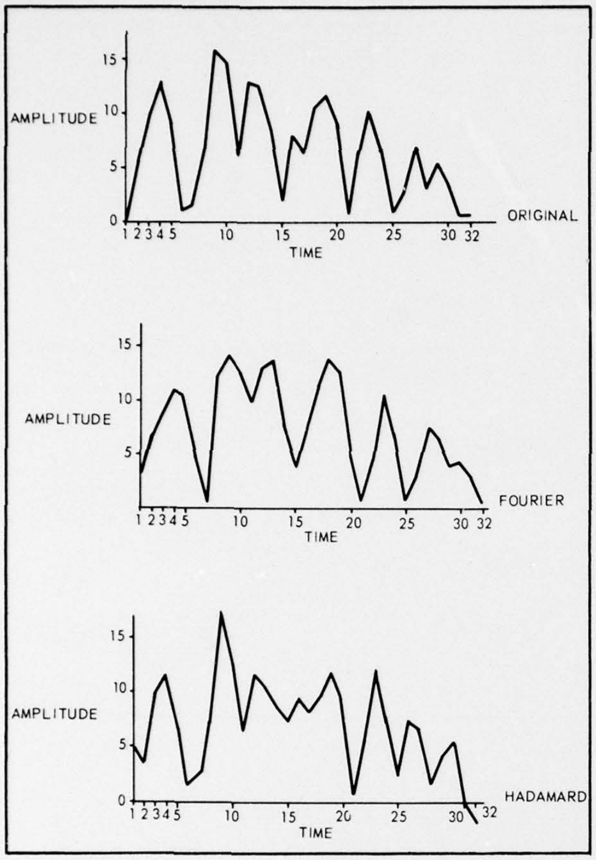


Figure 3a. Signal Reconstruction (2TW = 32)

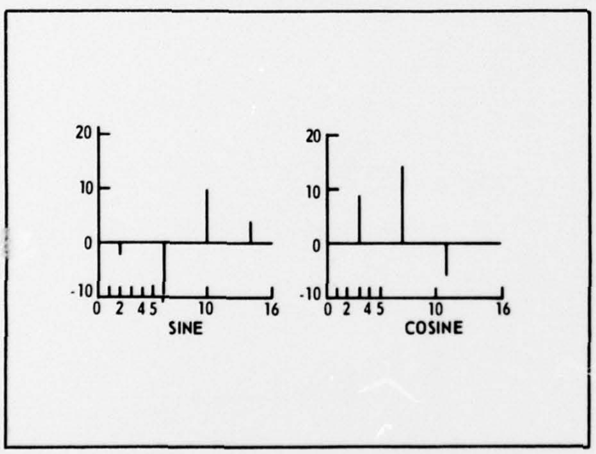


Figure 3b. Line Spectra of Hadamard Code No. 10

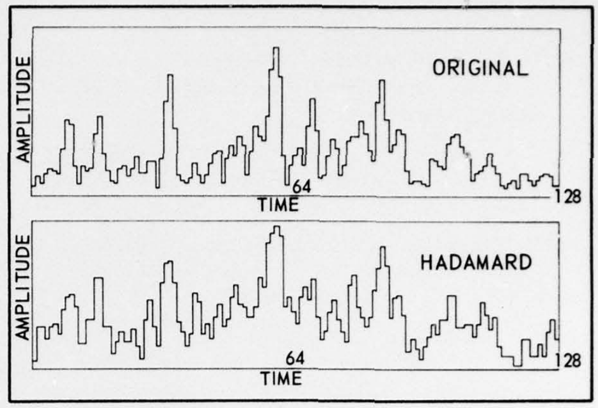


Figure 4. Signal Reconstruction Using the FHT (Showing 1/2 Original Time Window)

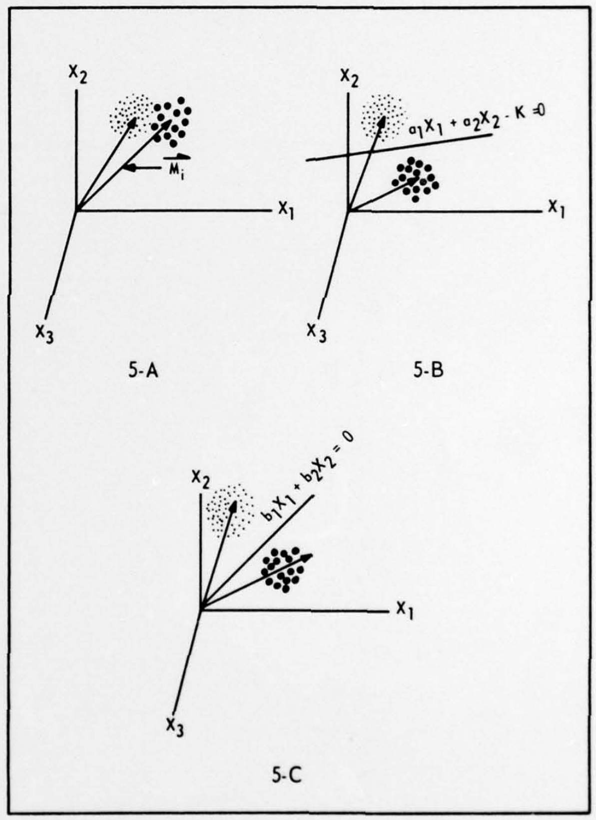


Figure 5. Linear Separable Classes

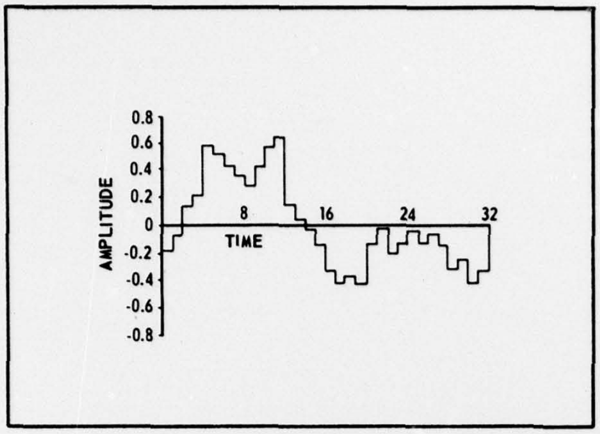


Figure 6-A. Discriminant d Truncated and Transformed to Time Domain

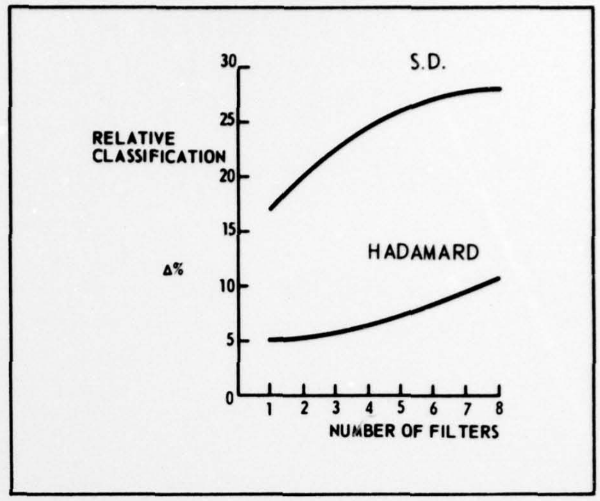


Figure 8. Nilsson Decision, Hadamard vs. S.D.

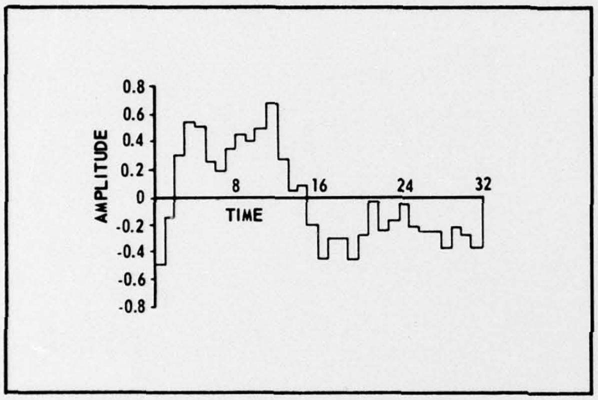


Figure 6-B. Difference of Means $\vec{M}_2 - \vec{M}_1$

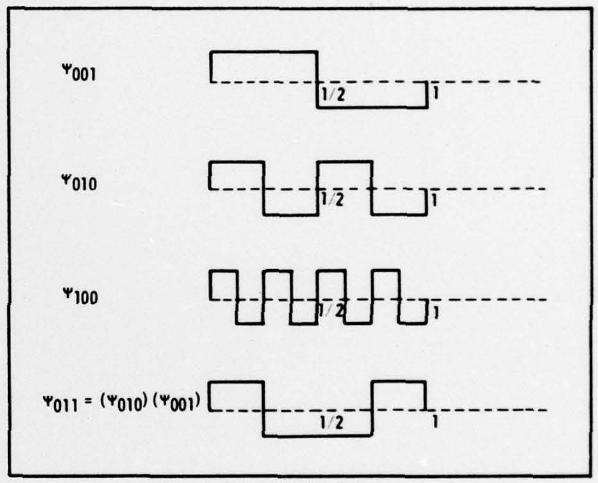


Figure 9-A. Walsh Functions

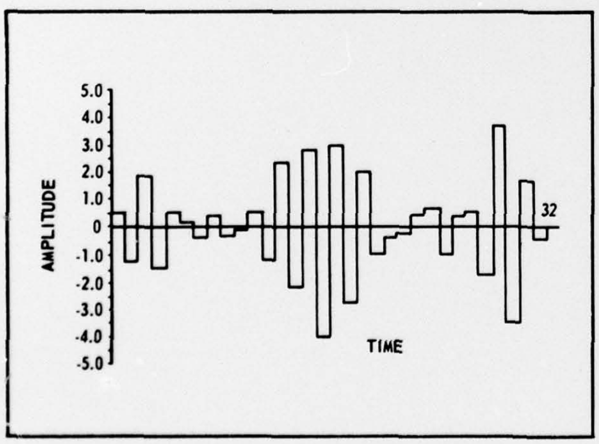


Figure 7. Simultaneous Diagonalization Discriminant

Ψ_{000}	1	1	1	1	1	1	1	1
Ψ_{001}	1	1	1	1	-1	-1	-1	-1
Ψ_{010}	1	1	-1	-1	1	1	-1	-1
Ψ_{011}	1	1	-1	-1	-1	-1	1	1
Ψ_{100}	1	-1	1	-1	1	-1	1	-1
Ψ_{101}	1	-1	1	-1	-1	1	-1	1
Ψ_{110}	1	-1	-1	1	1	-1	-1	1
Ψ_{111}	1	-1	-1	1	-1	1	1	-1

Figure 9-B. Walsh Matrix N = 8