DASP99 PROCEEDINGS

Defense Applications of Signal Processing

Proceedings of the 1999 Workshop DASP99

Aug 22-27, 1999

LaSalle, Illinois USA

Edited by

Alan Lindsey
Bill Moran
Jim Schroeder
Mark Smith
Lang White

REPORT DOCUMENTATION PAGE

AFRL-SR-BL-TR-00-

0133

1. AGENCY USE ONLY *(Leave blank)*	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	1 September 1999	Final Technical Report 1 Jul 99 - 30 Nov 99

4. TITLE AND SUBTITLE
U. S./Australia Joint Workshop on Defense Applications of Signal Processing

5. FUNDING NUMBERS
F49620-99-1-0288

6. AUTHOR(S)
Mark J. T. Smith

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Georgia Tech
Continuing Education
613 Cherry Street, Swann Bldg
Atlanta, GA 30332-0385

8. PERFORMING ORGANIZATION REPORT NUMBER
X-20-5F3

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
AFOSR/NM
801 N. Randolph St, Rm 732
Arlington, VA 22203-1977

10. SPONSORING/MONITORING AGENCY REPORT NUMBER
F49620-99-1-0288

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION AVAILABILITY STATEMENT
Approved for public release; distribution unlimited.

12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 words)*

The 1999 Workshop on Defense Applications of Signal Processing (a.k.a DASP99) brought together approximately sixty leading researchers in defense signal processing from Australia and the United States on August 22-28 for technical presentations and informal discussions. Topics represented at the workshop included situation awareness, target recognition, sensor technology, support vector machines (SVM), time-frequency analysis, and GPS surveillance.

An important product of the workshop is the initiation of future collaborations and relationship building, both of which occurred. A formal Workshop Proceedings was produced and distributed to all participants and sponsors. Selected papers are currently being identified for expansion and journal publication in cooperation with Academic Press Inc.

14. SUBJECT TERMS
Workshop, conference, Defense related signal processing

15. NUMBER OF PAGES
223

16. PRICE CODE

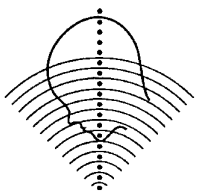
17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
UNCLASSIFIED	UNCLASSIFIED	UNCLASSIFIED	UL

DASP99

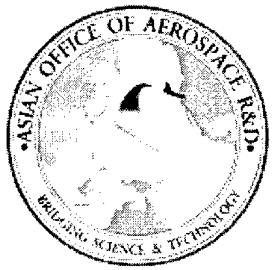
Defense Applications of Signal Processing
Proceedings of the 1999 Workshop

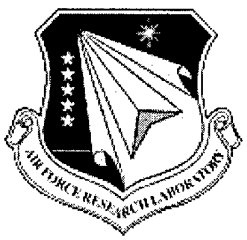
August 22-27, 1999

LaSalle, Illinois USA



Cooperative Research Centre
for Sensor Signal and
Information Processing





FOREWORD

Dear Workshop Participant

Welcome to the Second Workshop in Defense Applications of Signal Processing and to Starved Rock Resort. In the tradition of its predecessor, held at Victor Harbor, South Australia in June 1997, the objective is to gather world-class researchers for stimulating and thought provoking discussion on signal processing as it relates to defense applications. While DASP 97 succeeded in bringing together about 55 researchers from both defense laboratories and universities in the USA and Australia, on this occasion, there are about 60 participants, 25 from Australia and 35 from the USA, again with a mix of lab people and academics.

The motivation for this second workshop was developed at the first one. The venue, time of, and kind of conference has changed many times in the intervening two years, and, at times, it has seemed doomed not to happen. That it has is largely due to the efforts of a few people who have worked hard over the last few months to bring us all together there at this time. They are:

Mark Smith (Georgia Tech), Bill Moran (Flinders Uni and CSSIP), Alan Lindsey (Rome Lab), Jim Schroeder (CSSIP), Lang White (Adelaide Uni and CSSIP), Major Michele Gaudreault (AOARD), Marian Viola (DSTO), Jon Sjogren (AFOSR)

We also wish to pay tribute to the funding bodies for this Workshop. They are:

The Electronic and Surveillance Laboratories of DSTO,
US Defense Advanced Research Projects Agency (DARPA),
US Air Force Office of Scientific Research (AFOSR) and
Asian Office of Aerospace Research and Development (AOARD).

We wish to thank them all for their generosity in making this event possible.

The workshop will open with a welcome reception on Sunday evening. During the course of the workshop registrants may wish to engage in tennis or swimming, go horseback riding or hiking, or participate in any of the other recreational activities available at the Park. The formal program will conclude with lunch on Thursday. A field trip to Chicago for our Australian guests, complete with a boat tour of the city, shopping at the Navy Pier (ride the Ferris wheel!) and then Chicago-style pizza is planned for Thursday afternoon and evening.

Our aim, as last time, is to foster collaboration among the various communities represented here and, in particular, between the US and Australian researchers in defense signal processing. To this end, we have put together a formal program of lectures and poster sessions, together with several more informal events. We hope that you will work and play together over the next few days. Attend the talks and poster sessions, but above all interact with participants other than the ones you see regularly. Make sure that, when you leave here, you have found at least one person from the other side of the Pacific with whom you have agreed to ``Keep in touch" or ``Exchange ideas" or better still ``Collaborate". By doing so you help to make this workshop a success.

CONTENTS

Nonuniform Linear Antenna Array Design and Signal Processing for DOA Estimation of Gaussian Sources

Yuri I. Abramovich[1] and Nicholas K. Spencer

Cooperative Research Centre for Sensor Signal and Information Processing (CSSIP), SPRI Building, Technology Park, Mawson Lakes S.A. 5095, Australia

E-mail: yuri@cssip.edu.au and nspencer@cssip.edu.au

This paper discusses the problem of direction-of-arrival (DOA) estimation for Gaussian sources that are arbitrarily correlated — from independent to fully correlated. For independent sources, the antenna array design is governed by two competing considerations: *maximum aperture*, that inclines towards increasing sparsity for a given number of array sensors, and *identifiability*, that tends to exclude extreme sparsity. For fully correlated sources, these two competing criteria are augmented by a third which allows for the initialisation of DOA estimation by the generalised spatial smoothing (GSS) technique. The maximum number of fully correlated sources is in turn an important factor in the GSS algorithm and subsequent array geometry design. We present a geometry optimisation technique that permits accurate DOA estimation of arbitrarily correlated sources.

Key Words: sparse linear arrays, direction-of-arrival estimation, multimode, independent sources

1. INTRODUCTION

In many direction-finding applications, the number of antenna elements available for the construction of an array is limited, in which case the problem of optimum antenna geometry for a fixed number of elements M naturally arises. For linear arrays, solutions to this problem belong to the class of nonuniformly-spaced linear arrays (NLA's), also known as sparse or aperiodic arrays, and several different approaches currently exist which seek the "best" design.

Meanwhile, speculations regarding optimum sparse geometry have mostly been made for the independent (Gaussian) source model. In particular, the well-known suggestion that the minimum-redundancy criterion is appropriate for (integer) NLA geometry optimisation [1] is based on the simple fact that such geometries generate a contiguous ("gapless") set of spatial covariance lags. For independent Gaussian

[1]Supported in part by the INTAS SASPARC grant.

sources, this property immediately allows for the unambiguous estimation of up to

$$m < \frac{1}{2}M(M-1) \tag{1}$$

DOA's by the direct augmentation approach (DAA) of Pillai *et al.* [2].

On the other hand, it has been demonstrated in [3] that for arbitrary and in particular fully correlated Gaussian sources, manifold ambiguity leads to non-identifiability. Furthermore, it has been demonstrated by Proukakis and Manikas [4] that scenarios with an ambiguous manifold (i.e. with linearly dependent manifold "steering" vectors) always exist for sparse arrays.

This property obviously means that the number of identifiable arbitrary (fully) correlated sources in sparse antenna arrays is always significantly less than the number of antenna elements. Even for uniform M-element antenna arrays, the traditional spatial smoothing technique [5] allows for unambiguous DOA estimation of up to $\frac{M-1}{2}$ fully correlated sources.

For sparse antenna arrays, the spatial smoothing technique is not directly applicable. Consequently, some provisions are necessary to enable initialisation of the DOA estimation procedure at the very least. Obviously, the number of identifiable (and initialised) DOA's should be predefined and limited to

$$m \ll \frac{M}{2} \tag{2}$$

Thus, when non-uniform antenna geometry is optimised for DOA estimation of arbitrarily correlated sources, the following (competing) provisions are to be made:

1. A certain number of repetitive sub-arrays (partial arrays) are embedded into the original geometry, which allows for generalised spatial smoothing for a small predefined number of fully correlated sources. This allowance excludes non-redundant geometries and even those with the minimum redundancy.

2. Given some number of redundancies, we have to achieve:

(i) A fully augmentable geometry to provide identifiability for the maximal number of uncorrelated (Gaussian) sources.

(ii) Proper dimensionality of the (co-array) manifold of the co-array for the synthetic partial arrays so that the predefined number of uncorrelated (Gaussian) sources, m, is exceeded.

(iii) The maximal total aperture as well as the maximal aperture for the co-array of the synthetic partial array to enhance DOA estimation accuracy for the independent and fully correlated sources accordingly.

2. PARTIAL ARRAYS

Suppose that we have M antenna sensors and wish to construct the "optimal" NLA. Further suppose that the sensor positions $\boldsymbol{d} \equiv [0, d_2, d_3, \ldots, d_M]$ are restricted to integer values (usually measured in half-wavelength units). The meaning of "optimal" needs to be defined carefully from the perspective of the discussion above.

We consider the DOA estimation problem for some small pre-specified maximum number of coherent signals m (of arbitrary configuration) using the special class of *partial-array* NLA geometries and the corresponding *generalised spatial smoothing* (GSS) algorithm.

Let the *co-sequence* of an array $\boldsymbol{d}$ be its set of $M-1$ consecutive intersensor separations (*ie.* differences), while its *co-array* is the sorted set of $M(M-1)/2$ differences. We define a *partial array* to be a group of nonuniform linear non-contiguous sub-arrays of identical co-sequence structure [6]. Associated with each partial array are its *multiplicity* κ (number of occurrences or instances), *order* ℓ (number of co-sequence elements involved), and aperture a. A given NLA will have n embedded partial arrays, with a total of N instances. The GSS technique may be applied to a NLA providing it yields at least one partial array of multiplicity $\kappa \geq m$ and order $\ell \geq m$, where m is the number of fully correlated signals. Examples of partial arrays and their properties are more fully discussed in [7].

The GSS algorithm introduced in [6] consists of an initialisation step followed by local ML refinement. The initialisation step is based on the *PA-MUSIC* approach involving all appropriate partial arrays.

Suppose that an NLA yields a total of N partial arrays, each of multiplicity κ_i, order ℓ_i and aperture a_i $(i = 1, \ldots, N)$. Let $\boldsymbol{y}_{ij}$ be a (ℓ_i+1)-variate snapshot vector corresponding to the j^{th} instance $(j = 1, \ldots, \kappa_i)$ of the i^{th} partial array. If any instance of a partial array occurs as a mirror-image (*ie.* in reverse order), then the corresponding snapshot vector is observed by reversing the order of antenna samples and taking the complex conjugate of the vector. Thus for each partial array we may define the $(\ell_i+1) \times (\ell_i+1)$ partial array covariance matrix by spatial smoothing to be

$$\hat{R}_i = \sum_{j=1}^{\kappa_i} \boldsymbol{y}_{ij}\, \boldsymbol{y}_{ij}^H. \tag{3}$$

Let $\hat{G}_i$ be the noise eigen-subspace of $\hat{R}_i$, then $\hat{G}_i$ consists of at least one eigenvector (since $m \ll M$). The PA-MUSIC technique is:

$$\text{find} \quad \max_{\theta} f_{PA}(\theta) := \min_{\theta} \sum_{i=1}^{N} \boldsymbol{a}_i^H(\theta)\, \hat{G}_i\, \hat{G}_i^H\, \boldsymbol{a}_i(\theta) \tag{4}$$

where $\boldsymbol{a}_i(\theta)$ is the (ℓ_i+1)-variate manifold ("steering") vector which corresponds to the given partial array geometry. Evidently this approach eliminates non-coinciding ambiguities. More specifically, the co-array of the synthetic partial array that is constructed by all of the properly averaged covariance lags produced by all of the partial arrays, should have a manifold dimensionality that exceeds the pre-defined number of fully correlated sources. Thus the effectiveness of DOA estimation delivered by GSS is directly related to the number, variety and $\kappa\ell a$–properties of the available partial arrays. For this reason, the sum $\sum_{j=1}^{n} a_j^2 = A$ could be treated as a cost function for antenna geometry optimisation. Details of the three-stage optimisation approach appear in [7].

3. EXAMPLE RESULTS

The following example illustrates array geometry optimisation results for an $M = 16$ element array. The initial choice $M_1 = 10$ gives us the starting-point 10-element non-redundant Sverdlik array [8]

$$\boldsymbol{d}_{55}^{(0)} = [0, 1, 6, 10, 23, 26, 34, 41, 53, 55]\,. \tag{5}$$

The exhaustive tree search of stage two yields 37 candidate gap-free geometries, each with 14 elements and 36 redundancies. The integer programming maximisation of stage three finds that of these candidates, one in particular is the best (in the search range $\ell = 3$ and $c \in [1, 18]$), since with the addition of two sensors (8,19) it yields the 16-element NLA

$$\boldsymbol{d}_{55} = [0, 1, \underline{5}, 6, \underline{8}, 10, \underline{\underline{19}}, 23, 26, 34, \underline{37}, 41, \underline{44}, \underline{52}, 53, 55] \tag{6}$$

having the maximal cost function $A = 38467$ (and 65 redundancies). Thus we have partitioned our $M = 16$ elements in this example by $\{M_1 = 10,\ M_2 = 4,\ M_3 = 2\}$. Note that this three-stage optimisation search took a few days computing time on a modern workstation, even with its rather modest search range. At this point, we have no alternative but to assume that any NLA rich in partial arrays for a restricted search set $\{\ell, c\}$ will be similarly superior for a more expansive set.

Indeed, Table 1 shows the $\kappa\ell$-distribution and Fig. 1 illustrates the a-distribution of partial arrays for $\boldsymbol{d}_{55}$ for the expanded search range $\ell \in [3, 5]$ and $c \in [1, 30]$, whence we find $A = 99441$. This array performs better than the 16-element ULA because of the large numbers of embedded partial arrays, each of significant aperture. The minimum-redundancy array of comparable total aperture ($M_\alpha = 58$) has 13 elements , so we could consider $\boldsymbol{d}_{55}$ to be a type of "optimal" solution by the introduction of only three additional elements to the minimum-redundancy structure.

TABLE 1
Partial array distribution by multiplicity (κ) and order (ℓ) for $\boldsymbol{d}_{55}$ for $m = 3$ and the search range $\ell \in [3, 5]$ and $c \in [1, 30]$.

ℓ \ κ	3	4	5	6	7	8	9
3	40	9	0	0	0	0	0
4	33	3	0	0	0	0	0
5	12	0	0	0	0	0	0

4. PRACTICAL APPLICATION: FREE-CHANNEL ADVICE FOR HIGH-FREQUENCY SURFACE-WAVE RADARS

One of the areas for practical applications of non-uniform linear array methodology is in clear-channel advice for HF over-the-horizon radars, specifically for surface-wave radars. In most realistic situations, selection of the operational frequency for such radars should be performed with respect to the the directional spectrum of the external noise. Accordingly, clear-channel advice sub-systems should incorporate DOA estimation capabilities. Current HF surface-wave radars usually exploit

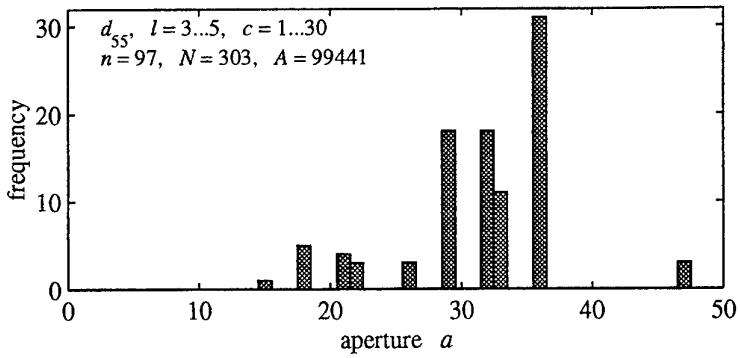


FIG. 1. Aperture histogram of partial arrays embedded in d_{55}.

uniform linear arrays with a "digital receiver per element" architecture in its main target detection module. It is, however, too expensive to duplicate this architecture for the clear-channel advice module. The number of digital receivers for such a module is limited and should be used with the maximal external noise DOA estimation efficiency.

Figure 2 illustrates the practical results on external noise DOA estimation, obtained from the experimental surface-wave over-the-horizon facility located in Northern Australia [9]. It depicts the MUSIC generated DOA's from a full 32-element uniform antenna array (top) compared to a similar set of DOA's produced form a 16-element non-uniform sub-array (middle) and from a 9-element sub-array (bottom). Both the 16- and 9-element sub-arrays have been selected from the original 32-element ULA within the original aperture in a fully augmentable manner.

The presented figure illustrate DOA's of multiple external noise sources as a function of the repetition period ("sweep") number. Direct comparison of this data makes it clear that all three sets of DOAs are operationally identical for this typical frequency.

5. SUMMARY

We have considered a problem involving nonuniformly spaced linear array geometry optimisation, in the context of enhancing the performance of modern super-resolution techniques in spatial spectrum (DOA) estimation. This optimisation problem has been reduced to a simplified form where effective techniques can be applied, based on dynamic programming principles.

Optimisation efficiency in terms of spectral (DOA) estimation accuracy has been analysed elsewhere [6], and in most cases is found to be very high and significantly superior to the conventional ULA geometry coupled with standard MUSIC-type routines.

Real-data processing involving external HF interference DOA estimation justifies the choice of NLA geometry for the frequency management subsystem of modern HF OTH radars.

REFERENCES

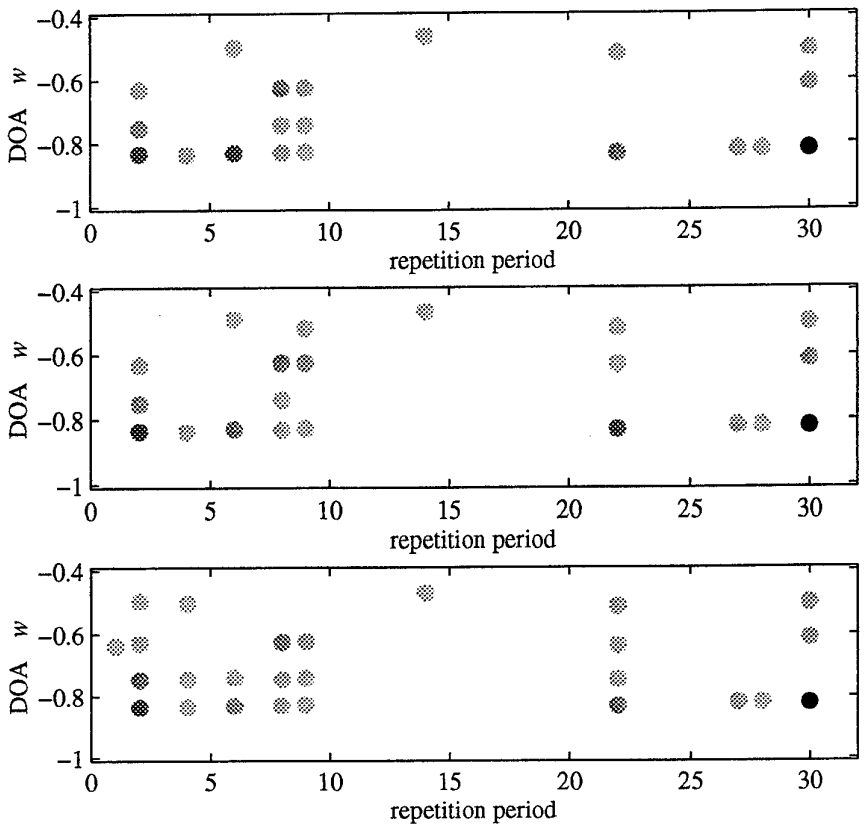


FIG. 2. Effect of interference DOA estimation by thinning a linear array: (upper) full array $M = 32$, (middle) partially thinned $M = 16$, (lower) highly thinned $M = 9$.

1. S.W. Lang, G.L. Duckworth, and J.H. McClellan, "Array design for MEM and MLM array processing", in *Proc. ICASSP-81*, 1981, pp. 145–148.

2. S. Pillai and F. Haber, "Statistical analysis of a high resolution spectrum estimator utilizing an augmented covariance matrix", *IEEE Trans. Acoust. Sp. Sig. Proc.*, vol. 35 (11), pp. 1517–1523, 1987.

3. Y.I. Abramovich, N.K. Spencer, and A.Y. Gorokhov, "Resolving manifold ambiguities in direction-of-arrival estimation for nonuniform linear antenna arrays", *IEEE Trans. Sig. Proc.*, submitted 1996, accepted 1999.

4. C. Proukakis and A. Manikas, "Study of ambiguities of linear arrays", in *Proc. ICASSP-94*, Adelaide, 1994, vol. 4, pp. 549–552.

5. T.J. Shan, M. Wax, and T. Kailath, "On spatial smoothing for direction-of-arrival estimation of coherent signals", *IEEE Trans. Acoust. Sp. Sig. Proc.*, vol. 33, pp. 806–811, 1985.

6. Y.I. Abramovich and N.K. Spencer, "Generalised spatial smoothing using partial arrays for multiple target bearing estimation", in *Proc. ICASSP-98*, Seattle, 1998, vol. 4, pp. 2061–2064.

7. N.K. Spencer, "Nonlinear signal processing and nonuniform linear antenna array design for DOA estimation of coherent sources", in *Proc. NSIP-99*, Antalya, 1999, pp. ???–???

8. M.B. Sverdlik, *Optimal Discrete Signals*, Sovietsckoe Radio, Moscow, 1975, In Russian.

9. Y.I. Abramovich, N.K. Spencer, and S.J. Anderson, "Experimental trials on antenna array adaptive processing for surface-wave over-the-horizon radar", in *Proc. ASA-99*, Sydney, 1999, p. 53.

Performance of Simplex Signaling in Circular Trellis-Coded Modulation

Frank A. Alder, Jeffrey C. Dill
Electrical Engineering & Computer Science
Ohio University
Athens, Ohio 45701, USA

Alan R. Lindsey
US Air Force Research Laboratory
525 Brooks Road
Rome, New York 13441, USA

Abstract - *This paper presents measures of performance of simplex signaling in a circular trellis-coded modulation (CTCM) scheme. Background is given on both CTCM and simplex signaling. The CTCM system is shown to give substantial coding gain when compared to conventional BPSK. Performance is also shown to improve as trellis size increases.*

I. Introduction

Trellis-coded modulation (TCM) provides a means by which band-limited channels can reap the benefits of error control coding by combining coding and modulation into a single step. This coding/modulation step is accomplished by integrating a multi-level or multi-phase signaling constellation with a state-oriented encoder, such as a convolutional encoder. Where a binary block code might use a binary phase-shift keyed (BPSK) modulator to transmit each of the coded bits individually, a TCM scheme would choose one signal from the constellation to represent a number of the coded bits at one time. In TCM the tradeoff is decoder complexity for coding gain.

Recently, circular trellis-coded modulation (CTCM) has been proposed [2-10]. Also known as high-dimensional trellis-coded modulation (HDTCM), CTCM takes the basic concepts of TCM (such as signal partitioning) and applies them to achieve coding gain on a power-limited channel, such as a spread-spectrum channel, by using a high-dimensional signaling constellation. A CTCM system can be viewed as a block code with trellis structure in the sense that source data is encoded block-by-block independently. Additionally, CTCM satisfies a so-called "state constraint," which specifies that the starting state of a particular source data block must equal the ending state. This property alleviates the need to set tail information bits to zero to drive the encoder to the "all-zero" state, a necessary procedure in conventional TCM.

In [10], the assignment of transmission symbols from a simplex signaling constellation is investigated for use in CTCM, where a source alphabet size of four is emphasized. This paper takes these symbol assignments and determines performance criteria in the form of distance distributions and bounds on the bit-error performance of the system. It will be shown that the CTCM system provides substantial coding gain when compared to BPSK modulation.

II. Background

A. *Simplex Signaling*

For arbitrary dimensions $N \geq M-1$, M simplex signals exist for unit pulse energy if there exist M signals of $S=\{s_1, s_2, \cdots, s_M\}$ in N dimensions such that:

$$\left|s_i - s_j\right| = \sqrt{2M} \qquad i \neq j, 1 < i, j < M \tag{1}$$

where $s_i=[s_{i1}, s_{i2}, \cdots, s_{iN}]$, and $s_{ik} \in \{-1,0,1\}$, $1 \leq i \leq M$, $1 \leq k \leq N$. A simplex signal constellation can be loosely defined as a set of signals that are equidistant from each other and energy equivalent [11]. Simplex signals are not orthogonal, but they achieve the same error probability as an equally likely orthogonal signaling set while using less energy. Hence, simplex signaling is employed when transmission energy is constrained.

The three-dimensional simplex will be emphasized in this paper. It is desired to create simplex signals utilizing number of dimensions (usually much) greater than three. For example, one three-dimensional simplex signal might be

$$s=[1,0,-1,0,-1,0,0,0] \tag{2}$$

In this case, eight dimensions exist and dimensions 1, 3, and 5 are occupied. Additionally, shorthand notation can be used to describe simplex signals. Equation (2) can be denoted as [1,-3,-5], for example. This notation indicates non-zero pulses in dimensions one, three, and five, and their associated polarities. There is only one simplex that utilizes only dimensions 1, 3, and 5 and contains the signal in (2) as a member. Using shorthand notation, the matrix expression for this simplex is

$$S=\left[(1 \;\; 3 \;\; 5);(1 \;\; -3 \;\; -5);(-1 \;\; 3 \;\; 5);(-1 \;\; -3 \;\; 5)\right]^T. \tag{3}$$

Note that the negative of (3) and, furthermore, any simplex, is also a simplex. By using all 8 dimensions, many simplexes can be formed that contain (2). E.g. $S = [(1\ \ -3\ \ -5);(-1\ \ 4\ \ 6);(2\ \ 3\ \ -6);(-2\ \ -4\ \ 5)]^T$, which is just one of dozens of simplexes that contain (2) as a member.

B. Circular Trellis-Coded Modulation

The trellis-coded modulation scheme introduced in [2,3], the so-called circular trellis-coded modulation (CTCM), is the backbone of this research. A straightforward way to think of CTCM is as a block code with trellis structure. That is, individual source data blocks are mapped to a particular path through the trellis instead of an output data block. The mapping of the source data blocks to trellis paths is one-to-one. The CTCM scheme is characterized by the following four parameters: N (dimensionality of the transmitted signal space), n (size of the source alphabet), D (trellis depth), and B (input source symbol sequence length, or block length).

The CTCM system can be described by these four parameters in the form (N,n,D,B). The number of trellis states in CTCM is determined by $S = n^D$. Source alphabet size of n=4 is emphasized in this paper. Trellis depth refers to the number of transitions needed for a given state to reach any other state in the trellis. The block size denotes the size of the input source block as well as the number of transitions in a legal trellis path. The block size must be greater than or equal to the trellis depth plus one ($B \geq D + 1$).

One drawback of conventional TCM is that the decoder must know the starting state of the encoder before transmission. This is usually accomplished by "padding" the source data with additional zeros to force the encoder to an all-zero state before additional source data is encoded. CTCM alleviates this problem by forcing the starting and ending states of a particular block to be the same. This property of CTCM is known as the state constraint [2,7,8].

The state constraint is satisfied through proper design of the state table. A state table lists, for every state in the trellis, what the next state transition will be for any given input symbol. Since a 4-ary source is emphasized in this paper, the state table will have a number of rows equal to the number of states in the trellis, and four columns which correspond to each of the four source symbols. The design of the state table is achieved through the use of a Zech logarithm table, and is discussed in detail in [2], which also presents computer code capable of generating state tables for arbitrary source alphabet size n and trellis depth D. Only one state table exists for a given pair of n and D. An example of a state table is shown in Table I.

The transmission symbol table is a look-up table analogous to the state table. Both the state and transmission symbol tables contain entries for all trellis states and all possible source symbols. However, where the state table defines the next state transition given the current state and current input, the transmission symbol table describes what symbol is transmitted when transiting to that next state. Given the state table and the transmission symbol table, the trellis is completely described. A trellis diagram can then be constructed with all trellis states, their associated transitions, and the related symbols [2,3].

III. Simplex Signaling Performance For 16-State Trellis

With the problem of symbol assignment addressed in [4,11], the performance of CTCM using simplex signals will be explored in this section. Specifically, a 16-state CTCM trellis will be emphasized. Two methods will be employed in the performance analysis: distance distributions and bounds on the bit error rate of the system. Performance using various block sizes will be compared. The bounds on the bit error rate will also be compared to the performance of BPSK.

A. Distance Distributions

One method of measuring the performance of an CTCM system is known as a "distance distribution." A distance distribution is a histogram-type distribution that is created by measuring the distance from all the legal paths in the trellis to one reference path [3,5]. In general, there are m_i paths with distance d_i from the reference path. The distance value d_0 is usually zero and the distance from the reference path to itself is zero, hence the value of m_0 is one. The first non- zero multiplicity (other than m_0) occurs at what is called the minimum Euclidean distance, which is denoted by d_{min}.

TABLE I. CTCM State Table For 16-State Trellis (n=4, D=2).

Current State		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Next State	0 Input	1	3	4	5	6	7	8	9	10	11	12	13	14	15	16	2
	1 input	2	6	10	16	3	12	15	11	4	9	7	14	13	8	5	1
	2 input	4	7	1	8	12	3	5	14	2	13	6	11	9	16	15	10
	3 input	10	12	2	15	7	6	16	13	1	14	3	9	11	5	8	4

The reference path is typically chosen to be the "all-zero" path. The all-zero path is the path taken through the trellis for a source input of exclusively zero symbols. For a block size of B, a group of B consecutive zero source symbols maps to the all-zero path. Note that, for a source alphabet size of $n = 4$, the zero source symbol corresponds to two zero bits.

The parameter known as d_{free} refers to the free Euclidean distance of the trellis-coding scheme. This parameter is very important in determining performance gain of any CTCM trellis [2,5]. N_{free}, which denotes the multiplicity of codewords (legal trellis paths) with distance d_{free} from the reference path, is equally important.

For a particular trellis, d_{free} can be defined as [1]:

$$d_{free} = \min\left[\sum_n d^2(a_n, b_n)\right]^{1/2} \quad \forall\{a_n\} \neq \{b_n\} \tag{4}$$

where a_n and b_n denote two codewords and the function d^2 denotes the squared Euclidean distance between the two codewords. d_{free} can be defined in a more straightforward manner as the minimum Euclidean distance between two legal codewords that have the same starting state and ending state in an CTCM trellis [2]. For a given trellis and simplex signals with unit pulse amplitude, this can be expressed as:

$$d_{free} = \sqrt{8(D+1)} \tag{5}$$

where D denotes the depth of the trellis. Additionally, recall that the minimum block size is equal to the trellis depth plus one, i.e., $B_{min} = D + 1$.

In terms of the distance distribution, d_{free} is the largest value d_{min} can attain for a particular trellis. Obviously, it is desired to force d_{min} to reach d_{free}. Additionally, it is desired to minimize N_{free}, the multiplicity at distance d_{free}, and it is known that d_{min} dominates the error performance for high signal-to-noise ratios [3,5]. Therefore, reaching the value of d_{free} and minimizing the value of N_{free} for a particular trellis ensures optimal performance.

As discussed in [4,11], there are 24 possible transmission symbol tables that can be constructed for a CTCM trellis with a source alphabet size of $n = 4$. Using a transmitted signal space dimensionality of $N = 8$ and 10, all possible symbol assignments were determined by using the computer program in [4]. Distance distributions were then calculated for each of the 24 possible symbol assignments.

In [6], a 12-dimensional assignment was investigated. Table II lists values for N_{min} and d^2_{min} for this 12-dimensional assignment, as well as the two distributions (for $N = 8$ and 10) that reach d_{free} discussed in [4]. Distributions for block sizes of $B = 7$ and $B = 8$ were not included in [6].

Through investigation of Table II it can be seen that the 12-dimensional case reaches d_{free} at a block size of B=6, which is smaller than the block sizes required for either the 8 or 10-dimensional cases. However, when comparing the values of N_{free} for the three cases, the 12-dimensional value of N_{free} is significantly higher than the other two cases. It can then be concluded that the 8- and 10-dimensional assignments achieve better performance than the 12-dimensional assignment while using less bandwidth. In the 8-dimensional case, the bandwidth reduction is 1/3.

A. Bit-Error Performance

Bit-error probabilities cannot be computed for CTCM in a closed form. To obtain exact P_{be} curves, the system would have to be simulated. However, bounds on the bit-error probability can be computed, which will give a rough estimate of system performance. Additionally, the actual P_{be} curves will asymptotically approach the upper bound on P_{be} for high signal-to-noise ratios [3]. The upper bound on P_{be} was derived in [3] and the resulting equation is shown in (7):

$$P_{be} \leq \sum_{j=2}^{M} \frac{\Delta B_{1j}}{B \cdot k} Q\left(d_{1j} \cdot \sqrt{\frac{E_b}{N_o}}\right) \tag{6}$$

where B is the block size, k is the number of bits per symbol, ΔB_{1j} is the number of mismatched bits in the source sequences associated with codeword 1 and codeword j, d_{1k} is the Euclidean distance between codewords 1 and j, E_b/N_o is the signal to noise ratio, M is the number of legal codewords, and $Q(\cdot)$ represents the Gaussian Q-function. Equation (7) gives a performance benchmark for the system. It will be used to determine system performance for various block sizes and symbol assignments.

TABLE II. COMPARISON OF MINIMUM DISTANCE VALUES AND MULTIPLICITIES FOR 16- STATE TRELLIS.

Dim.	$B=3$		$B=4$		$B=5$		$B=6$		$B=7$		$B=8$	
	d^2_{min}	N_{min}	d^2_{min}	N_{min}	d^2_{min}	N_{min}	d^2_{min}	N_{min}	d^2_{min}	N_{min}	d^2_{min}	N_{min}
$N=8$	12	4	14	4	20	16	20	6	24	28	24	32
$N=10$	12	1	14	8	20	21	22	12	24	21	24	24
$N=12$	12	1	16	1	20	1	24	37	--	--	--	--

Fig. 1 shows the best P_{be} union bound curve (as determined by testing all 24 symbol assignments), for block sizes of $B = 3$ through 7, using $N = 8$ dimensions in signaling. The error probability improvement with increasing block size is easily seen from this plot. This is in agreement with Shannon's theorem, in that, the more data that is encoded at a time, the better the error performance will be.

Additionally, as expected, the error performance is asymptotically approaching a limit. That is, the improvement in the P_{be} union bound in going from $B = 3$ to $B = 4$ is far greater than the improvement shown in going from $B = 6$ to $B = 7$. The performance of BPSK is also shown in the plot as a reference point. For $B = 7$, the CTCM scheme provides a gain of approximately 5 dB over BPSK.

IV. SIMPLEX SIGNALING PERFORMANCE FOR 64-STATE TRELLIS

A. *Distance Distributions*

Distance distributions for the 64-state trellis were calculated in the same manner as the 16-state distributions discussed in Section III. From (6), it can be seen that the free Euclidean distance is defined as $d_{free} = \sqrt{32}$ for the 64-state trellis, an increase over the value of $d_{free} = \sqrt{24}$ for the 16-state case.

Dimensionality values of $N = 10$, 12, 16, 20, and 24 can be used in construction of the transmission symbol table for a 64-state trellis. Each of these five N values were used in the construction of transmission symbol tables and their corresponding distance distributions. As in the 16-state case, several "good" distributions existed for each combination of block size and dimensionality. Table III lists the values of d^2_{min} and N_{min} for the best distribution for each combination of N and B.

In examining Table III, the primary characteristic that is noticed is that there is no distribution, for any combination of N and B, that reaches the free Euclidean distance value of $d_{free} = \sqrt{32}$. It is not clear why this is the case, although there are a number of possible of explanations [4]. The most obvious explanation is that the block sizes tested have not increased to the point that d_{free} can be attained. That is, certain assignments may reach d_{free} for larger block sizes than those tested, say, B=20. However, this imposes a huge computational limitation, since the run-time for the B=10 cases is on the order of days.

B. *Bit-Error Performance*

The union bound on the probability of bit error for the 64-state trellis can be computed in the same manner as the 16-state trellis by utilizing (7). For comparison purposes, Fig. 2 shows P_{be} union bound curves for a block size of $B = 7$. The plot compares performance using 10-dimensional and 24-dimensional signaling for the 64-state trellis. Additionally, the union bound curve for the best 16-state case (from section III, $N = 8$) is shown, along with the BPSK performance curve.

Since the 64-state trellis gives a wide range in the number of dimensions that can be utilized in constructing the transmission symbol table, the two extremes of $N = 10$ and $N = 24$ were chosen to be included in Fig. 2. The plots are culled from the symbol assignments that yield the best distance distributions for the respective dimensionalities. Improvement over the 16-state trellis is easily noticed from the plot. Additionally, the 24-dimensional case offers improvement over the 10-dimensional case, with a gain of over 0.5 dB present at a block size of $B = 7$ and P_{be} value of 10^{-6}.

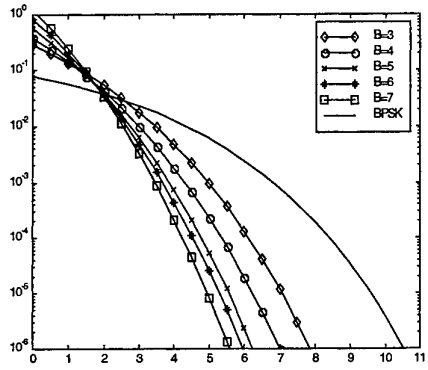


Fig. 1. Union bound on bit-error probability for 16-state trellis (B = 3 through 7), and BPSK.

TABLE III
COMPARISON OF MINIMUM DISTANCE VALUES AND MULTIPLICITIES FOR 64- STATE TRELLIS.

Block Size	$N=10$		$N=12$		$N=16$		$N=20$		$N=24$	
	d^2	m	d^2	m	d^2	m	d^2	m	d^2	m
$B=4$	16	9	16	4	16	4	16	4	20	32
$B=5$	20	11	20	25	22	35	22	10	22	10
$B=6$	20	3	24	21	24	3	24	3	24	3
$B=7$	24	7	28	21	28	14	30	21	30	14
$B=8$	28	24	30	16	28	8	30	16	30	16
$B=9$	28	18	30	18	28	9	30	18	30	18
$B=10$	28	10	30	20	28	10	30	20	30	20

Whether or not this 0.5 dB improvement is significant would be subject to the specific design problem. Since 24-dimensional signaling gives a 240% increase in bandwidth over the 10-dimensional case, the bandwidth/performance trade-off would need to be weighed heavily. However, since the improvement does exist, an increase in the dimensionality of the transmitted signal space is a viable route that can be explored to improve the bit-error rate of the system.

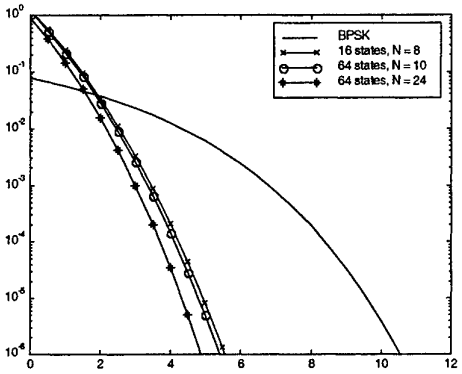


Fig. 2. BPSK performance and bit-error probability union bounds for $B = 7$ ($S = 16$ and 64, N = various).

The 64-state trellis demonstrates nearly 5.5 dB of gain over BPSK at $B = 7$. At larger block sizes, if the minimum distance value moves closer to d_{free}, the improvement over BPSK would increase further. Additionally, by increasing the size of the trellis (to 256 states, etc.), and thus the value of d_{free}, further improvement would be seen. However, the drawback of increasing the block size and the size of the trellis is that the decoding algorithm would also increase in complexity, if throughput is to be maintained [2].

V. Conclusions

This paper has examined the performance of simplex signaling in a circular trellis-coded modulation scheme. Both Euclidean distance distributions and bounds on bit-error performance are examined as performance benchmarks. Performance is shown to improve as various parameters (such as block size and trellis size) are increased.

References

[1] G. Ungerboeck, "Channel coding with multilevel/phase signals," *IEEE Transactions on Information Theory*, vol. 28, pp. 55-67, January 1982.

[2] Y.C. Lo, "Circular Trellis-Coded Modulation in Spread Spectrum Communications," *Ph.D. Dissertation*, School of Electrical Engineering and Computer Science, Russ College of Engineering and Technology, Ohio University, Athens, OH, August 1997.

[3] C. Chen, "The Performance Analysis and Decoding of High Dimensional Trellis-Coded Modulation for Spread Spectrum Communications," *Ph.D. Dissertation*, School of Electrical Engineering and Computer Science, Russ College of Engineering and Technology, Ohio University, Athens, OH, August 1997.

[4] F.A. Alder, "Symbol Assignment and Performance of Simplex Signaling in High Dimensional Trellis-Coded Modulation," *Master's Thesis*, School of Electrical Engineering and Computer Science, Russ College of Engineering and Technology, Ohio University, Athens, OH, August 1998.

[5] J.C. Dill, Y.C. Lo, and A.R. Lindsey, "Circular trellis-coded modulation with high-dimensional simplex signals," *MILCOM98 Conference*, Bedford, MA, October 1998.

[6] J.C. Dill, Y.C. Lo, and A.R. Lindsey, "Butterfly structure in trellis-coded modulation," *GLOBECOM98 Conference*, Sydney, Australia, November 1998.

[7] Y.C. Lo, J.C. Dill, C. Chen, and S.R. Lopez-Permouth, "The encoding of a new trellis-coded modulation with state permutation structure," *4th International Symposium on Communication Theory and Applications*, Ambleside, U.K., July 1997.

[8] Y.C. Lo, J.C. Dill, and A.R. Lindsey, "Circular trellis-coded modulation with permuted-state structure," *Defense Applications of Signal Processing*, Adelaide, Australia, August 1997.

[9] J.C. Dill, A.R. Lindsey, Y.C. Lo, and C. Chen, "A novel error-correcting codec for m-ary orthogonal modulations," *Asilomar Conference on Signals, Systems, and Computers*, Monterey, CA, November 1996.

[10] F.Ag Alder, J.C. Dill, and A.R. Lindsey, "Simplex symbol assignment in circular trellis-coded modulation," *MILCOM99 Conference* (accepted), Atlantic City, New Jersey, October 1999.

[11] J.M. Wozencraft and I.M. Jacobs, Principles of Communication Engineering. John Wiley and Sons, New York, NY, 1965.

Direction Finding Based on Spatial Time-Frequency Distribution Matrices

Moeness G. Amin
Department of Electrical and Computer Engineering,
Villanova University, Villanova, PA 19085
Phone: (610)519-7305 Fax: (610)519-4436
E-mail: moeness@ece.vill.edu

Spatial time-frequency distributions (STFDs) have been recently shown to be a powerful tool for solving direction finding and blind source separation problems for multi-sensor array receivers. These spatial distributions are the natural means to deal with source signals that are localizable in the time-frequency domain. This paper examines the eigenstructure of the spatial time-frequency distribution matrices. It is shown that improved estimates of the signal and noise subspaces are achieved by constructing the subspaces from the time-frequency signatures of the signal arrivals rather than from the data covariance matrices. This improvement is more evident in low signal-to-noise ratio (SNR) environment and in the cases of closely spaced sources. The paper considers the MUSIC technique to demonstrate the advantages of STFDs and uses it as grounds for comparison between time-frequency and conventional subspace estimates.

Key Words: time-frequency analysis; subspace analysis; time-frequency MUSIC; spatial time-frequency distributions; array signal processing

1. INTRODUCTION

Although the applications of the spatial time-frequency distributions to blind source separation and DOA problems using multiple antenna arrays in nonstationary environments have been introduced in [1,2], yet so far there has not been sufficient analysis that explains their offerings and justifies their performance. The aim of this paper is to examine the eigenstructure of the spatial time-frequency distribution matrices and provide statistical analysis of their respective signal and noise subspaces. The paper shows that the subspaces obtained from the STFDs are robust to both noise and angular separation of the waveforms incident on the array. This robustness is primarily due to spreading the noise power while localizing the source energy in the time-frequency domain. By forming the STFD matrices from

This work was supported by ONR under Grant #N00014-98-1-0176.

the points residing on the source time-frequency signatures, we in essence, increase the input signal to noise ratio, and hence improve subspace estimates.

This paper is organized as follows. Section 2 presents the signal model and gives a brief review of the definition and basic properties of the spatial time-frequency distributions. In Section 3, we consider nonstationary environment characterized by frequency-modulated (FM) source signals, and show the potential improvement in direction-of-arrival (DOA) estimation using STFDs. Section 4 examines the performance of the direction finding MUSIC technique based on the covariance and STFD noise subspace estimates.

2. BACKGROUND

2.1. Signal Model

In narrowband array processing, when n signals arrive at an m-element array, the linear data model

$$\mathbf{x}(t) = \mathbf{y}(t) + \mathbf{n}(t) = \mathbf{A}(\mathbf{\Theta})\mathbf{d}(t) + \mathbf{n}(t) \tag{1}$$

is commonly assumed, where the $m \times n$ spatial matrix $\mathbf{A}(\mathbf{\Theta}) = [\mathbf{a}(\theta_1)...\mathbf{a}(\theta_n)]$ represents the mixing matrix or the steering matrix, and $\mathbf{a}(\theta_i)$ are the steering vectors. Due to the mixture of the signals at each sensor, the elements of the $m \times 1$ data vector $\mathbf{x}(t)$ are multicomponent signals, whereas each source signals $d_i(t)$ of the $n \times 1$ signal vector $\mathbf{d}(t)$ are often a monocomponent signal. $\mathbf{n}(t)$ is an additive noise vector whose elements are modeled as stationary, spatially and temporally white, zero-mean complex random processes, independent of the source signals. That is,

$$E[\mathbf{n}(t+\tau)\mathbf{n}^H(t)] = \sigma\delta(\tau)\mathbf{I} \quad \text{and} \quad E[\mathbf{n}(t+\tau)\mathbf{n}^T(t)] = \mathbf{0} \quad \text{for any } \tau \tag{2}$$

where $\delta(\tau)$ is the Kronecker delta function, $\mathbf{I}$ denotes the identity matrix, σ is the noise power at each sensor, superscript ^{H} and ^{T} respectively denote conjugate transpose and transpose, and $E(\cdot)$ is the statistical expectation operator.

In equation (1), it is assumed that the number of sensors is larger than the number of sources, i.e., $m > n$. Further, matrix $\mathbf{A}$ is full column rank, which implies that the steering vectors corresponding to n different angles of arrival are linearly independent. We further assume that the correlation matrix

$$\mathbf{R_{xx}} = E[\mathbf{x}(t)\mathbf{x}^H(t)] \tag{3}$$

is nonsingular, and the observation period consists of N snapshots with $N > m$.

Under the above assumptions, the correlation matrix is given by

$$\mathbf{R_{xx}} = E[\mathbf{x}(t)\mathbf{x}^H(t)] = \mathbf{A}(\mathbf{\Theta})\mathbf{R_{dd}}\mathbf{A}^H(\mathbf{\Theta}) + \sigma\mathbf{I}, \tag{4}$$

where $\mathbf{R_{dd}} = E[\mathbf{d}(t)\mathbf{d}^H(t)]$ is the signal correlation matrix. For notational convenience, we drop the argument $\mathbf{\Theta}$ in equation (1) and simply use $\mathbf{A}$ instead of $\mathbf{A}(\mathbf{\Theta})$.

2.2. Spatial Time-Frequency Distributions

The spatial time-frequency distributions (STFDs) based on Cohen's class of time-frequency distribution were introduced in [1] and its applications to direction find-

ing and blind source separation have been discussed in [2] and [1], respectively. In this paper, we focus one key member of Cohen's class, namely the pseudo Wigner-Ville distribution (PWVD) and its respective spatial distribution. Only the time-frequency (t-f) points in the autoterm regions of PWVD are considered for STFD matrix construction. In these regions, it is assumed that the crossterms are negligible. The discrete form of pseudo Wigner-Ville distribution of a signal $x(t)$, using a rectangular window of odd length L, is given by

$$D_{xx}(t,f) = \sum_{\tau=-\frac{L-1}{2}}^{\frac{L-1}{2}} x(t+\tau)x^*(t-\tau)e^{-j4\pi f\tau}, \tag{5}$$

where * denotes complex conjugation. The spatial pseudo Wigner-Ville distribution (SPWVD) matrix is obtained by replacing $x(t)$ by the data snapshot vector $\mathbf{x}(t)$,

$$\mathbf{D_{xx}}(t,f) = \sum_{\tau=-\frac{L-1}{2}}^{\frac{L-1}{2}} \mathbf{x}(t+\tau)\mathbf{x}^H(t-\tau)e^{-j4\pi f\tau}. \tag{6}$$

Substitute (1) into (6), we obtain

$$\mathbf{D_{xx}}(t,f) = \mathbf{D_{yy}}(t,f) + 2\mathrm{Re}\left[\mathbf{D_{yn}}(t,f)\right] + \mathbf{D_{nn}}(t,f). \tag{7}$$

We note that $\mathbf{D_{xx}}(t,f)$, $\mathbf{D_{yy}}(t,f)$, $\mathbf{D_{yn}}(t,f)$, $\mathbf{D_{ny}}(t,f)$, and $\mathbf{D_{nn}}(t,f)$ are matrices of dimension $m \times m$, whereas the source TFD matrix $\mathbf{D_{dd}}(t,f)$ is of dimension $n \times n$. Under the uncorrelated signal and noise assumption and the zero-mean noise property, the expectation of the crossterm TFD matrices between the signal and noise vectors is zero, i.e., $E\left[\mathbf{D_{yn}}(t,f)\right] = E\left[\mathbf{D_{ny}}(t,f)\right] = \mathbf{0}$, and it follows

$$E\left[\mathbf{D_{xx}}(t,f)\right] = \mathbf{D_{yy}}(t,f) + E\left[\mathbf{D_{nn}}(t,f)\right] = \mathbf{A}\mathbf{D_{dd}}(t,f)\mathbf{A}^H + E\left[\mathbf{D_{nn}}(t,f)\right]. \tag{8}$$

For narrowband array signal processing applications, the mixing matrix $\mathbf{A}$ holds the spatial information and maps the auto- and cross-TFDs of the source signals into auto- and cross-TFDs of the data.

It is noted that relationship (8) holds true for every (t,f) points. In order to reduce the effect of noise and ensure the full column rank property of the STFD matrix, we consider multiple time-frequency points. This allows more information of the source signal t-f signatures to be included into their respective subspace formulation. Joint-diagonalization [3] and time-frequency averaging are the two main approaches that have been used for this purpose [1, 2, 4]. In this paper, we only consider averaging over multiple time-frequency points.

3. SUBSPACE ANALYSIS FOR FM SIGNALS

In this paper, we focus on frequency modulation (FM) signals, modeled as

$$\mathbf{d}(t) = [d_1(t), ..., d_n(t)]^T = [D_1 e^{j\psi_1(t)}, ..., D_n e^{j\psi_n(t)}]^T, \tag{9}$$

where D_i and $\psi_i(t)$ are the amplitude and phase of ith source signal. For each sampling time t, $d_i(t)$ has an instantaneous frequency $f_i(t) = \frac{1}{2\pi}\frac{d\psi_i(t)}{dt}$. To simplify the analysis, we assume that the FM signals are mutually uncorrelated over the observation period. That is,

$$\frac{1}{N}\sum_{k=1}^{N} d_i(k)d_j^*(k) = 0 \qquad \text{for } i \neq j,\ i,j = 1,...,n. \tag{10}$$

In this case, the signal correlation matrix in (4) is

$$\mathbf{R_{dd}} = diag\left[D_i^2,\ i = 1,2,...,n\right]$$

where $diag[\cdot]$ is the diagonal matrix formed with the elements of its vector valued arguments. The ith diagonal element of TFD matrix $\mathbf{D_{dd}}(t,f)$ in (8) is given by

$$D_{d_id_i}(t,f) = \sum_{\tau=-\frac{L-1}{2}}^{\frac{L-1}{2}} D_i^2 e^{j[\psi_i(t+\tau)-\psi_i(t-\tau)]-j4\pi f\tau}. \tag{11}$$

Assuming that the third-order derivative of the phase is negligible over the window length L, then $f_i = \frac{1}{2\pi}\frac{d\psi_i(t)}{dt}$, and $\psi_i(t+\tau) - \psi_i(t-\tau) - 4\pi f_i\tau = 0$. Accordingly,

$$D_{d_id_i}(t,f) = \sum_{\tau=-\frac{L-1}{2}}^{\frac{L-1}{2}} D_i^2 = LD_i^2. \tag{12}$$

Similarly, the noise STFD matrix $\mathbf{D_{nn}}(t,f)$ is

$$\mathbf{D_{nn}}(t,f) = \sum_{\tau=-\frac{L-1}{2}}^{\frac{L-1}{2}} \mathbf{n}(t+\tau)\mathbf{n}^H(t-\tau)e^{-j4\pi f\tau}. \tag{13}$$

Under the assumption of temporally and spatially white noise, the statistical expectation of $\mathbf{D_{nn}}(t,f)$ is given by

$$E\left[\mathbf{D_{nn}}(t,f)\right] = \sum_{\tau=-\frac{L-1}{2}}^{\frac{L-1}{2}} E\left[\mathbf{n}(t+\tau)\mathbf{n}^H(t-\tau)\right] e^{-j4\pi f\tau} = \sigma\mathbf{I}. \tag{14}$$

Therefore, when we select the time-frequency points along the t-f signature or the IF of an FM signal, the SNR in model (8) is LD_i^2/σ, which has an improved factor L over the one associated with model (4).

The pseudo Wigner-Ville distribution of each FM source has a constant value over the observation period, providing that we leave out the rising and falling power distributions at both ends of the data record. For convenience of analysis, we select those $N - L + 1$ t-f points of constant distribution value for each source signal. Therefore, the averaged STFD over the time-frequency signatures of n_o signals, i.e., $n_o(N - L + 1)$ t-f points, is given by

$$\hat{\mathbf{D}} = \frac{1}{n_o(N-L+1)}\sum_{q=1}^{n_o}\sum_{i=1}^{N-L+1} \mathbf{D_{xx}}(t_i, f_{q,i}), \tag{15}$$

where $f_{q,i}$ is the instantaneous frequency of the qth signal at the ith time sample. The expectation of the averaged STFD matrix is

$$\mathbf{D} = E\left[\hat{\mathbf{D}}\right] = \frac{1}{n_o}\sum_{p=1}^{n_o}\left[LD_p^2\mathbf{a}(\theta_p)\mathbf{a}^H(\theta_p) + \sigma\mathbf{I}\right] = \frac{L}{n_o}\mathbf{A}^o\mathbf{R}_{\mathbf{dd}}^o(\mathbf{A}^o)^H + \sigma\mathbf{I}, \quad (16)$$

where $\mathbf{R}_{\mathbf{dd}}^o$ and $\mathbf{A}^o$, respectively, represent the signal correlation matrix and the mixing matrix constructed by only considering n_o signals out of the total number of signal arrivals n.

It is clear from (16) that, when n_o signals are selected, the SNR improvement becomes $G = L/n_o$ (we assume $L > n_o$ throughout this paper). Therefore, from the SNR perspective, it is better to select (t, f) points that belong to individual signals, and to separately evaluate the respective STFD matrices. Accordingly, STFD-based direction finding is, in essence, a discriminatory technique in the sense that it does not require simultaneous localization and extraction of all unknown signals received by the array. With STFDs, direction finding can be performed using STFDs of a subclass of the impinging signals with specific time-frequency signatures. In this respect, the proposed direction finding technique acts as a spatial filter, removing all other signals from consideration and, subsequently, saves any downstream processing that is required to separate interference and signals of interest. It is also important to note that with the ability to construct the STFD matrix from one or few signal arrivals, the well known $m > n$ condition on source localization using arrays can be relaxed, i.e., we can perform direction finding or source separation with the number of array sensors smaller than the number of impinging signals [5]. From the angular resolution perspective, closed spaced sources with different t-f signatures can be resolved by constructing two separate STFDs, each corresponds to one source, and then proceed with subspace decomposition for each STFD matrix separately, followed by a appropriate source localization method (MUSIC, for example). The drawback of performing several direction finding using different STFD matrices is clearly the need for repeated computations of eigendecompositions and source localizations.

4. SIMULATIONS

The t-f MUSIC is introduced in [2], where the angles of arrival are estimated by locating the highest peaks of the spectrum provided by using the noise subspace of the STFD matrix, rather the covariance matrix, which is the case in conventional MUSIC.

The following example compares the performance of conventional and t-f MUSIC. Consider a uniform linear array of 8 sensors separated by half a wavelength. Two chirp signals emitted from two sources positioned at angle θ_1 and θ_2. The start and end frequencies of the chirp signal of the source at θ_1 are $\omega_{s1} = 0$ and $\omega_{e1} = \pi$, while the corresponding two frequencies for the signal of the other source at θ_2 are $\omega_{s2} = \pi$ and $\omega_{e2} = 0$, respectively. The noise used in this simulation is zero-mean, Gaussian distributed, and temporally white. The noise power, σ, is adjusted to give the desired $SNR = -10log(\sigma)$.

Fig. 1 displays the variance of the estimated DOA $\hat{\theta}_1$ versus SNR fro the case $(\theta_1, \theta_2) = (-10°, 10°)$. The curves in this figure show the theoretical and experimental results of the conventional MUSIC and t-f MUSIC (for L=33 and 129). The CRB is also shown in Fig. 1. Both impinging signals are selected when performing t-f MUSIC ($n_o = n = 2$). We assume that the number of signals is correctly estimated for each case. Simulation results are averaged over 100 independent trials of Monte Carlo experiments. The advantages of t-f MUSIC in low SNR cases are evident from this figure.

5. CONCLUSIONS

The advantages of STFD-based direction finding over traditional direction finding methods using data covariance matrices were demonstrated using the MUSIC algorithm. The t-f MUSIC technique outperforms the conventional MUSIC technique in the two situations of low SNR and closely spaced sources. Detailed performance analysis of DOA-based STFD is given in reference [6].

REFERENCES

1. A. Belouchrani and M. Amin, "Blind source separation based on time-frequency signal representation," *IEEE Trans. Signal Processing*, vol. 46, no. 11, pp. 2888-2898, Nov. 1998.
2. A. Belouchrani and M. Amin, "Time-frequency MUSIC," *IEEE Signal Processing Letters*, vol. 6, no. 5, pp. 109-110, May 1999.
3. G. H. Golub and C. F. Van Loan, *Matrix Computations*, 3rd edition. Baltimore, Maryland: Johns Hopkins University Press, 1996.
4. K. Sekihara, S. Nagarajan, D. Poeppel, and Y. Miyashita, "Time-frequency MEG-MUSIC algorithm," *IEEE Trans. Medical Imaging*, vol. 18, no. 1, pp.92-97, Jan. 1999.
5. M. G. Amin, "Spatial time-frequency distributions for direction finding and blind source separation," in *Proc. SPIE: Wavelet Applications IV*, vol.3723, pp.62-70, April 1999.
6. Y. Zhang, W. Mu, and M. G. Amin, "Subspace analysis of spatial time-frequency distribution matrices," submitted to *IEEE Trans. Signal Processing*, June 1999.

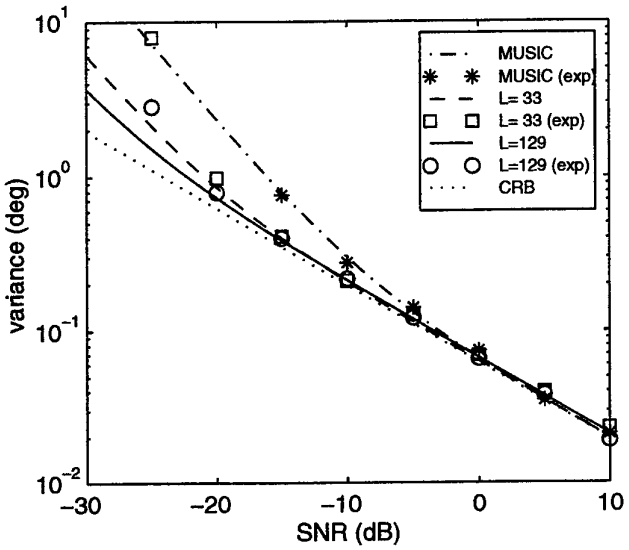


FIG. 1. Variance of DOA estimation vs. SNR.

Robust Adaptive Recovery of Spread-Spectrum Signals with Short Data Records [1]

Stella N. Batalama[†], Michael Medley [‡] and Dimitris A. Pados[†]

[†] *Department of Electrical Engineering, State University of New York at Buffalo, Buffalo, NY 14260; and* [‡] *Air Force Research Laboratory, IFGC, 525 Brooks Rd., Rome, NY 13441*

E-mail: batalama@eng.buffalo.edu; medleym@rl.af.mil; pados@eng.buffalo.edu

We consider the problem of adaptive reception of a multipath Direct-Sequence Spread-Spectrum (DS-SS) signal in the presence of unknown correlated SS interference and additive impulsive noise. The proposed SS receiver structure is comprised by a vector of adaptive chip-based Hampel non-linearities followed by an adaptive Auxiliary-Vector linear tap-weight filter. The non-linear receiver front-end adapts itself to the unknown prevailing noise environment providing robust performance over a wide range of underlying noise distributions. The adaptive Auxiliary-Vector linear tap-weight filter allows rapid SS interference suppression with a limited data record. Numerical and simulation studies offer comparisons with the conventional Minimum-Variance-Distortionless-Response (MVDR) SS receiver [6]-[9] as well as MVDR filtering preceded by vector adaptive chip-based non-linear processing [10].

Key Words: Spread spectrum communication; impulse noise; adaptive filters

CONTENTS

0. INTRODUCTION

Signal detection in the presence of impulsive channel noise has been considered extensively in the past (for example in [1]-[3] and references therein), while detection of a direct-sequence spread-spectrum (DS-SS) signal under similar channel conditions has been studied in [4], [5], and [10]. Receiver proposals in [4], [5] involve the use of either a conventional signature matched filter or a majority-vote receiver

[1] This work was supported by AFOSR under contract F49620-93-C-0063.

(hard-limiter non-linearity per chip followed by signature matched-filtering). In [4] it is reported that neither one of the above proposals is universally effective against the combination of SS interference and non-Gaussian impulsive noise. In [10], adaptive receivers are developed that are comprised by a vector of adaptive chip-based non-linearities followed by an adaptive linear tap-weight filter. The structures proposed in [10] tap the relative merits of both non-linear and linear signal processing and exhibit superior BER performance in the presence of combined impulsive and SS interference. In particular, the non-linear receiver front-end adapts itself to the unknown prevailing impulsive noise environment, while the adaptive linear tap-weight filter that follows the non-linearly processed chip samples combats effectively the SS interference. This article enhances our previous work in [10] in the following aspects. The receiver design objective is shifted to superior bit error rate (BER) performance under rapid short-data-record adaptation. In addition, the signal model is generalized to account for multipath signal reception and a new Hampel-type non-linear pre-processor is considered that encompasses the pre-processors considered in [10] as special cases.

1. SIGNAL MODEL

The baseband received signal is viewed as the aggregate of the multipath received SS signal of interest with signature code $\mathbf{S}_0$ of length L (if T is the symbol period and T_c is the chip period then $L = T/T_c$), $K-1$ multipath received DS-SS interferers with unknown signatures $\mathbf{S}_k$, $k = 1, \ldots, K-1$, and non-Gaussian (impulsive) interference. For notational simplicity and without loss of generality, we choose a synchronous signal set-up. We assume that the multipath spread is of the order of a few chip intervals, M, and since the signal is bandlimited to $B = 1/T_c$ the channel is modeled as a tap-delay line with $M+1$ taps spaced at chip intervals T_c. After conventional chip-matched filtering and sampling at the chip rate over a multipath extended symbol interval of $L+M$ chips, the $L+M$ data samples are organized in the form of a vector $\mathbf{r}$ given by

$$\mathbf{r} = \sum_{k=0}^{K-1} \sum_{m=0}^{M} c_{k,m} \sqrt{E_k} (b_k \mathbf{S}_{k,m} + b_k^- \mathbf{S}_{k,m}^- + b_k^+ \mathbf{S}_{k,m}^+) + \mathbf{n} \tag{1}$$

where, with respect to the k-th SS signal, E_k is the transmitted energy, b_k, b_k^-, and b_k^+ are the present, the previous, and the following transmitted bit, respectively, and $\{c_{k,m}\}$ are the coefficients of the frequency-selective slowly fading channel modeled as independent zero-mean complex Gaussian random variables that are assumed to remain constant over several symbol intervals. $\mathbf{S}_{k,m}$ represents the 0-padded by M, m-cyclic-shifted version of the signature of the k-th SS signal $\mathbf{S}_k$, $\mathbf{S}_{k,m}^-$ is the 0-filled $(L-m)$-left-shifted version of $\mathbf{S}_{k,0}$, and $\mathbf{S}_{k,m}^+$ is the 0-filled $(L-m)$-right-shifted version of $\mathbf{S}_{k,0}$. Finally, $\mathbf{n}$ represents additive complex non-Gaussian impulsive noise.

For conceptual and notational simplicity we may rewrite (1) as follows:

$$\mathbf{r} = \sqrt{E_0} b_0 \mathbf{w}_{\text{R-MF}}^{(0)} + \mathbf{I} + \mathbf{n} \tag{2}$$

where $\mathbf{w}_{\text{R-MF}}^{(0)} = \sum_{m=0}^{M} c_{0,m} \mathbf{S}_{0,m}$ is the effective (channel processed) signature of the SS signal of interest (signal-0) and $\mathbf{I}$ identifies comprehensively both the Inter-

Symbol and the SS interference present in (1). We use the subscript R-MF in our effective signature notation to make a direct association with the RAKE Matched-Filter receiver that is known to correlate the signature $\mathbf{S_0}$ with $M+1$ size-L shifted windows of the received signal (that correspond to the $M+1$ paths of the channel), appropriately weighted by the conjugated channel coefficients $c_{0,m}, m = 0, \ldots, M$. In our notation, this RAKE operation corresponds to linear filtering of the form $\mathbf{w}_{\text{R-MF}}^{(0)H}\mathbf{r}$, where H denotes the Hermitian operation.

2. RECEIVER ARCHITECTURE AND ALGORITHMIC DEVELOPMENTS

For the multipath signal model of the previous section, the general receiver structure under consideration is given in Fig. 1. The receiver consists of a non-linear front-end in the form of a vector of parametrized non-linearities $\mathbf{g}(\mathbf{r}\,;\cdot) : \mathcal{C}^{L+M} \longrightarrow \mathcal{C}^{L+M}$, followed by linear filter post-processing by an $L+M$ complex tap-weight filter $\mathbf{w}$. The non-linear pre-processor considered in this present work employs Hampel-type non-linearities $\mathbf{g}(\mathbf{r}\,;\alpha_1,\alpha_2,\alpha_3) = [\,g(r_1;\alpha_1,\alpha_2,\alpha_3)\,\cdots\,g(r_{L+M};\alpha_1,\alpha_2,\alpha_3)\,]^T$ where T denotes the transpose operation and

$$g(x;\alpha_1,\alpha_2,\alpha_3) \triangleq \begin{cases} x, & \text{if } |x| < \alpha_1, \quad 0 < \alpha_1 \\ \alpha_1 \frac{x}{|x|}, & \text{if } \alpha_1 \le |x| < \alpha_2, \quad 0 < \alpha_1 \le \alpha_2 \\ \frac{\alpha_3 - |x|}{\alpha_3 - \alpha_2}\alpha_1 \frac{x}{|x|}, & \text{if } \alpha_2 \le |x| \le \alpha_3, \quad 0 < \alpha_1 \le \alpha_2 \le \alpha_3 \\ 0, & \text{otherwise .} \end{cases} \tag{3}$$

In (3) x is a complex number and $|x|$ denotes the magnitude of x. The linear region of the Hampel non-linearity has the effect of passing the observations undistorted. The non-linear regions either completely reject (remove) or "correct" the observations. The latter is considered as an adjustment of the magnitude while maintaining the phase. The parameters α_1, α_2, and α_3 are positive cut-off parameters to be determined adaptively. The Hampel pre-processor is a generalization of the puncher and clipper pre-processors considered in [10].

In [10] the linear filter post-processor was chosen to be the Minimum-Variance-Distortionless-Response (MVDR) solution for the non-linearly processed data vectors. Adaptive SS interference suppression with MVDR post-processing has two shortcomings that we attempt to improve upon in this article. First, the adaptive optimization computational complexity may be prohibitive for mobile SS receivers due to the $(L+M)\times(L+M)$ autocorrelation matrix inversion operation. This is particularly true for systems with large spreading gain L. Second and most important, the data estimated adaptive implementation of the MVDR filter $\hat{\mathbf{w}}_{\text{MVDR}}^{(0)}(\text{N}) = [\mathbf{w}_{\text{R-MF}}^{(0)H}\hat{\mathbf{R}}_{\mathbf{g}}^{-1}(\text{N})\mathbf{w}_{\text{R-MF}}^{(0)}]^{-1}\,\hat{\mathbf{R}}_{\mathbf{g}}^{-1}(\text{N})\,\mathbf{w}_{\text{R-MF}}^{(0)}$, where $\hat{\mathbf{R}}_{\mathbf{g}}(\text{N}) = \frac{1}{N}\sum_{n=1}^{N}\mathbf{g}(\mathbf{r}_n)\mathbf{g}(\mathbf{r}_n)^H$ is the sample average estimate of the autocorrelation matrix over a data record of N Hampel processed input vectors, exhibits disappointing short data record performance. Data records of size many times the input vector dimension $L+M$ are necessary to approach satisfactorily the BER performance of the ideal $\mathbf{w}_{\text{MVDR}}^{(0)}$ filter, i.e. the filter that assumes perfectly known $\mathbf{R}_{\mathbf{g}}^{-1}$. In the following we address those two issues of reduced optimization complexity and superior small-sample-support performance in the context of what we call auxiliary-vector processing.

We consider the class of linear filter post-processors $\mathbf{w}$ in Fig. 1 that are "distortionless" in the $\mathbf{w}^{(0)}_{\text{R-MF}}$ vector direction of interest, i.e. $\mathbf{w}^H\mathbf{w}^{(0)}_{\text{R-MF}} = \|\mathbf{w}^{(0)}_{\text{R-MF}}\|$ so that no cancelation of the signal of interest takes place. This filter class is the set of all filters that can be written in the form

$$\mathbf{w}^{(0)}_{\text{AV}} = \mathbf{w}^{(0)}_{\|\text{R-MF}\|} - \mu^{\mathbf{g}}\mathbf{Q}^{\mathbf{g}} \tag{4}$$

where, for notational simplicity, $\mathbf{w}^{(0)}_{\|\text{R-MF}\|} \triangleq \frac{\mathbf{w}^{(0)}_{\text{R-MF}}}{\|\mathbf{w}^{(0)}_{\text{R-MF}}\|}$ denotes the normalized RAKE matched filter for the SS signal of interest 0, $\mu^{\mathbf{g}}$ is a complex scalar, and $\mathbf{Q}^{\mathbf{g}}$ is a vector in the $L+M$ complex space that is orthonormal with respect to $\mathbf{w}^{(0)}_{\|\text{R-MF}\|}$:

$$\mathbf{Q}^{\mathbf{g}H}\mathbf{w}^{(0)}_{\|\text{R-MF}\|} = 0 \quad \text{and} \quad \|\mathbf{Q}^{\mathbf{g}}\| = 1. \tag{5}$$

The superscript $\mathbf{g}$ that appears in (4) is intended to serve as a reminder that any specific choice for the scalar μ and the vector $\mathbf{Q}$ needs to account for the non-linear Hampel pre-processor $\mathbf{g}(\cdot)$ in Fig. 1. The receiver architecture that incorporates post-filtering by $\mathbf{w}^{(0)}_{\text{AV}}$ in (4) is shown in Fig. 2.

In contrast to minimum output variance optimization that leads to the optimum $\mathbf{w}^{(0)}_{\text{MVDR}}$ filter [10], we propose to choose an "auxiliary vector" $\mathbf{Q}^{\mathbf{g}}$ that satisfies the orthonormality constraint in (5) and maximizes the magnitude of the cross-correlation between points (a) and (b) of the receiver structure in Fig. 2. Standard, Lagrange multipliers derivation shows that this vector is

$$\mathbf{Q}^{\mathbf{g}} = \frac{\mathbf{R}_{\mathbf{g}}\mathbf{w}^{(0)}_{\|\text{R-MF}\|} - (\mathbf{w}^{(0)H}_{\|\text{R-MF}\|}\mathbf{R}_{\mathbf{g}}\mathbf{w}^{(0)}_{\|\text{R-MF}\|})\mathbf{w}^{(0)}_{\|\text{R-MF}\|}}{\left\|\mathbf{R}_{\mathbf{g}}\mathbf{w}^{(0)}_{\|\text{R-MF}\|} - (\mathbf{w}^{(0)H}_{\|\text{R-MF}\|}\mathbf{R}_{\mathbf{g}}\mathbf{w}^{(0)}_{\|\text{R-MF}\|})\mathbf{w}^{(0)}_{\|\text{R-MF}\|}\right\|}. \tag{6}$$

Then, the complex scalar weight $\mu^{\mathbf{g}}$ in the receiver structure of Fig. 2 is chosen to be the value that minimizes the Mean-Square (MS) error between points (a) and (c). Direct application of the Yule-Walker theorem shows that this MS-optimum value of $\mu^{\mathbf{g}}$ is

$$\mu^{\mathbf{g}} = \frac{\mathbf{Q}^{\mathbf{g}H}\mathbf{R}_{\mathbf{g}}\mathbf{w}^{(0)}_{\|\text{R-MF}\|}}{\mathbf{Q}^{\mathbf{g}H}\mathbf{R}_{\mathbf{g}}\mathbf{Q}^{\mathbf{g}}}. \tag{7}$$

This filter design approach can be generalized to cover processing with multiple auxiliary vectors of the form $\mathbf{w}^{(0)}_{\text{AV}} = \mathbf{w}^{(0)}_{\|\text{R-MF}\|} - \sum_{i=1}^{P}\mu^{\mathbf{g}}_i\mathbf{Q}^{\mathbf{g}}_i$, where each $\mathbf{Q}^{\mathbf{g}}_i$, $i = 1, \cdots, P$, is orthonormal with respect to $\mathbf{w}^{(0)}_{\|\text{R-MF}\|}$. The weighted auxiliary vectors are conditionally optimized in a sequential fashion as follows: $\mathbf{Q}^{\mathbf{g}}_1$ and $\mu^{\mathbf{g}}_1$ are chosen as before, in (6) and (7), respectively. Given $\mathbf{Q}^{\mathbf{g}}_1$ and $\mu^{\mathbf{g}}_1$, $\mathbf{Q}^{\mathbf{g}}_2$ is set to be the orthonormal to $\mathbf{w}^{(0)}_{\|\text{R-MF}\|}$ vector that maximizes the magnitude of the cross-correlation between $\left(\mathbf{w}^{(0)}_{\|\text{R-MF}\|} - \mu^{\mathbf{g}}_1\mathbf{Q}^{\mathbf{g}}_1\right)^H\mathbf{g}(\mathbf{r})$ and $\mathbf{Q}^{\mathbf{g}H}_2\mathbf{g}(\mathbf{r})$. Given $\mathbf{Q}^{\mathbf{g}}_1$, $\mu^{\mathbf{g}}_1$, and $\mathbf{Q}^{\mathbf{g}}_2$, the weight value $\mu^{\mathbf{g}}_2$ is chosen to minimize the MS error between $\left(\mathbf{w}^{(0)}_{\|\text{R-MF}\|} - \mu^{\mathbf{g}}_1\mathbf{Q}^{\mathbf{g}}_1\right)^H\mathbf{g}(\mathbf{r})$ and $\mu^{\mathbf{g}*}_2\mathbf{Q}^{\mathbf{g}H}_2\mathbf{g}(\mathbf{r})$. Inductively, we can obtain similar expressions for $\mathbf{Q}^{\mathbf{g}}_{p+1}$ and $\mu^{\mathbf{g}}_{p+1}$ given $\mathbf{Q}^{\mathbf{g}}_1, \mu^{\mathbf{g}}_1, \cdots, \mathbf{Q}^{\mathbf{g}}_p, \mu^{\mathbf{g}}_p$, $1 \le p < P$. The auxiliary vector generation procedure may stop when the cross-correlation magnitude $\left|\left(\mathbf{w}^{(0)}_{\|\text{R-MF}\|} - \sum_{i=1}^{p}\mu^{\mathbf{g}}_i\mathbf{Q}^{\mathbf{g}}_i\right)^H\mathbf{R}_{\mathbf{g}}\mathbf{Q}^{\mathbf{g}}_{p+1}\right|$ falls below a prespecified threshold value.

The auxiliary vector filter $\mathbf{w}_{\text{AV}}^{(0)}$ defined above has two major advantages in comparison with the $\mathbf{w}_{\text{MVDR}}^{(0)}$ that was used in [10]. The first advantage has to do with the pertinent optimization computational complexity. While both filters are a function of the RAKE matched filter $\mathbf{w}_{\text{R-MF}}^{(0)}$ and the Hampel pre-processed input data autocorrelation matrix $\mathbf{R}_{\mathbf{g}}$, no matrix inversion operation is required for the auxiliary-vector filter. The second and most important advantage has to do with the short data record behavior of the filter estimators $\hat{\mathbf{w}}_{\text{AV}}^{(0)}(\text{N})$ and $\hat{\mathbf{w}}_{\text{MVDR}}^{(0)}(\text{N})$ that are based on an $N-$point estimate of the autocorrelation matrix $\hat{\mathbf{R}}_{\mathbf{g}}(\text{N}) = \frac{1}{N}\sum_{n=1}^{N}\mathbf{g}(\mathbf{r}_n)\mathbf{g}(\mathbf{r}_n)^H$. The variance of $\hat{\mathbf{w}}_{\text{AV}}^{(0)}(\text{N})$ is significantly lower than the variance of $\hat{\mathbf{w}}_{\text{MVDR}}^{(0)}(\text{N})$ and this translates to superior short data record performance as seen in the next section. Of course, as $N \rightarrow \infty$, $\hat{\mathbf{w}}_{\text{AV}}^{(0)}(\text{N}) \rightarrow \mathbf{w}_{\text{AV}}^{(0)}$ and $\hat{\mathbf{w}}_{\text{MVDR}}^{(0)}(\text{N}) \rightarrow \mathbf{w}_{\text{MVDR}}^{(0)}$ with probability one, and in general $\mathbf{w}_{\text{AV}}^{(0)} \neq \mathbf{w}_{\text{MVDR}}^{(0)}$. So, this is a case of trading bias for lower variance.

To complete the algorithmic developments for a fully adaptive implementation of the DS-SS receiver in Fig. 2, we turn our attention to the Hampel cut-off parameters α_1, α_2, α_3. Adaptive cut-off parameter optimization can be pursued exactly as in [10], in the form of a decision driven, Minimum Bit-Error-Rate (MBER) stochastic approximation recursion.

3. NUMERICAL AND SIMULATION STUDIES

We examine DS-SS signal transmissions with speading gain $L = 63$ in the presence of 5 SS interfering signals and impulsive noise. The normalized synchronous signature cross-correlation of the interfering signals with the signal of interest is approximately 25% while the signature codes of the interferers are nearly orthogonal to each other. The communication channel is modeled as a multipath Rayleigh fading channel with 4 paths and zero mean complex Gaussian fading coefficients of variance 0.5 (i.e. $E\{|c_{k,m}|^2\} = 0.5$) for all paths $m = 0, \cdots, 3$ and all SS signals $k = 0, \cdots, 5$. The average total received interfering signal energies $E_k \sum_{m=0}^{3} E\{|c_{k,m}|^2\}$ are set equal to $9, 10, 11, 12$, and $13dB$ for $k = 1, 2, \cdots, 5$, respectively. The impulsive channel noise is modeled according to the familiar ϵ-mixture disturbance model $f_\epsilon(x) = (1-\epsilon)f_0(x) + \epsilon f_1(x)$ where $\epsilon \in [0, 1]$ accounts for the probability under which the noise is $f_1(\cdot)$ distributed. The nominal pdf $f_0(\cdot)$ is taken to be 0-mean complex Gaussian with variance $\sigma_0^2 = 1$. The "contaminating" pdf $f_1(\cdot)$ is assumed to be 0-mean complex Gaussian with variance $\sigma_1^2 = \gamma^2\sigma_0^2$ ($\gamma^2 = 1,000$) and ϵ is set equal to 0.2.

In Fig. 3 we compare the BER behavior of the conventional MVDR filter, the Hampel-MVDR filter, and the Hampel-AV (Auxiliary-Vector) filtering procedures developed herein. All cut-off parameter and filter estimates are based on a data record of 128 samples. The multipath fading channel is assumed to remain constant during adaptation and the BER induced by each receiver for each channel is averaged over 100 randomly drawn channels and 10 receiver realizations per channel. As seen in Fig. 3, the superiority of the Hampel-AV adaptive receivers is apparent.

To study the effect of the sample support size on the BER performance of the adaptive receivers under examination, we fix the average total received energy of the SS signal of interest at $E_0 \sum_{m=0}^{3} E\{|c_{0,m}|^2\} = 12dB$ and we repeat the studies of Fig. 3 as a function of the data record size. Fig. 4 demonstrates the superiority of AV post-filtering over the whole data support range of practical interest.

REFERENCES

1. A. D. Spaulding and D. Middleton, "Optimum reception in an impulsive interference environment-Part I: Coherent detection," *IEEE Trans. Commun.*, vol. 25, pp. 910-923, Sept. 1977.
2. A. D. Spaulding, "Locally optimum and suboptimum detector performance in a non-Gaussian interference environment," *IEEE Trans. Commun.*, vol. 33, pp. 509-517, June 1985.
3. A. B. Martinez, P. F. Swaszek, and J. B. Thomas, "Locally optimal detection in multivariate non-Gaussian noise," *IEEE Trans. Inform. Theory*, vol. 30, pp. 815-822, Nov. 1985.
4. B. Aazhang and H. V. Poor, "Performance of DS/SSMA communications in impulsive channels-Part II: Hard-limiting correlation receivers," *IEEE Trans. Commun.*, vol. 36, pp. 88-97, Jan. 1988.
5. B. Aazhang and H. V. Poor, "An analysis of nonlinear direct-sequence correlators," *IEEE Trans. Commun.*, vol. 37, pp. 723-731, July 1989.
6. U. Madhow and M. L. Honig, "MMSE interference suppression for direct-sequence spread-spectrum CDMA," *IEEE Trans. Commun.*, vol. 42, pp. 3178-3188, Dec. 1994.
7. C. N. Pateros and G. J. Saulnier, "An adaptive correlator receiver for direct-sequence spread-spectrum communication," *IEEE Trans. Commun.*, vol. 44, pp. 1543-1552, Nov. 1996.
8. M. L. Honig, U. Madhow, and S. Verdu,"Blind adaptive multiuser detection," *IEEE Trans. Inform. Theory*, vol. 41, pp. 944-960, July 1995.
9. A. Kansal, S. N. Batalama, and D. A. Pados, "Adaptive maximum SINR RAKE filtering for DS-CDMA multipath fading channels," *IEEE J. Select. Areas Commun.*, vol. 16, pp. 1765-1773, Dec. 1998.
10. S. N. Batalama, M. Medley, and I. N. Psaromiligkos, "Adaptive Robust Spread-Spectrum Receivers," *IEEE Trans. Commun.*, vol. 47, pp. 905-917, June 1999.

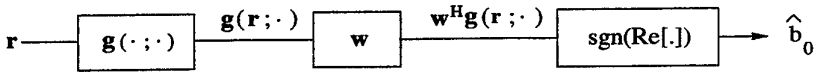


FIG. 1. General receiver structure.

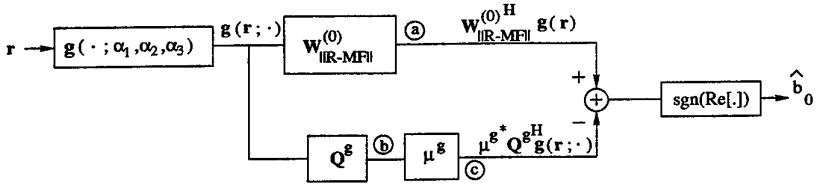


FIG. 2. Auxiliary-vector receiver structure.

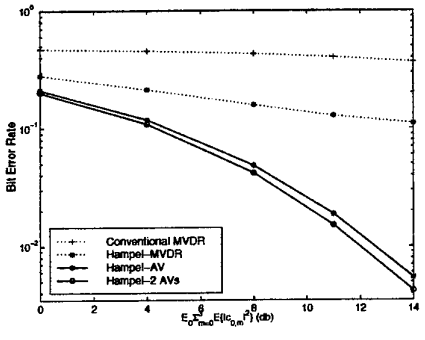


FIG. 3. Bit-Error-Rate versus total received energy for the SS signal of interest.

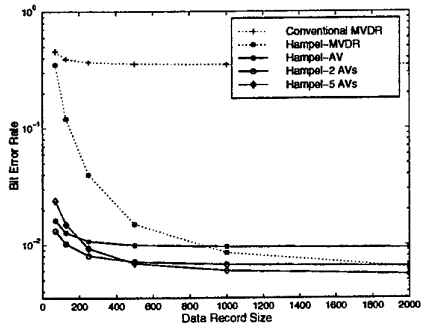


FIG. 4. Bit-Error-Rate versus sample support.

A Model for Shape and Texture Content-Based Image Compression

by
Capt. Robert J. Bonneau, Ph.D.
AFRL/SNRT

Abstract--- Currently a large area of research is being devoted to compressed domain content-based compression due to the JPEG-2000 and MPEG-7 requirements. Current compression quantization techniques are not geared to extract models of image structures as a part of the compression process due to the fact that the they are geared to attain maximum compression ratio for the lowest cost. Additionally the entropy encoding processes leave the compressed data unrecognizable due to its need to completely randomize the data for maximum compression. Recently there has been much work in wavelet and fractal methods for texture and shape segmentation as well as data compression. These methods contain implicit models for shape and texture coding as a natural part of the compression process. We thus develop a method for wavelet fractal compression which extracts and codes shape and texture primitives. This method makes use of the Mallat Gaussian derivative basis set and an implicit Markov shape and texture model.

I. Introduction

Recently there has been much interest in defining the relationship between wavelets and fractals [5]. We will first discuss a particular type of wavelet developed by Stephane Mallat which uses the Gaussian Derivative for a basis set. Secondly we will discuss how the properties of this wavelet can be used to enhance the block quantization process in wavelet fractal encoding. We will then show the parallels between he wavelet fractal block quantization process and a well known Wavelet Markov model and use the properties of the model to enhance the compression process. We will then describe how this Markov model can be used to integrate texture and shape into the compression process. Finally we will describe the image reproduction process and show an application of our compression method.

II. Gaussian Derivative Wavelet

The high g(x) and low h(x) pass elements of the Gaussian derivative wavelet transform are often cascaded in a filter bank structure as is shown in Figure 1. This methodology is a computationally efficient means of dividing a signal into organized set of frequency bands. One type of filter bank structure is the Mallat [14] multiresolution frequency bank or MRA for short. A particular implementation of this transform know as the dyadic MRA structure. This particular implementation of the MRA uses the Gaussian derivative[14] basis set as shown in . As we shall show later the Gaussian derivative has many desirable properties useful in signal and image analysis including an ability to accurately reveal boundaries and edges in signals and imagery minimal artifact production in image reproduction.The result is a an extremely effective means of detecting singularities (edges) in a signal. If we represent our gaussian $g_\sigma(x)$ Mallat [14] uses an elegant technique for two dimensional decomposition which lends itself to image compression and is the basis of our wavelet fractal method. The algorithm consists of first preprocessing an image with a multiscale wavelet decomposition as described by Mallat. In two dimensions the lowpass is performed as a separable transform as shown in Figure 3 and the highpass is done as the 1-D transform in X and Y dimensions independently.

$$Wf = (W_s^1 f(x,y), W_s^2 f(x,y))_{j \in Z} \tag{1}$$

The X and Y highpass functions are then combined to form modulus edge image and gradient edge image as is described in equations 2 and 3.

$$M_s f(x,y) = \sqrt{\left|W_s^1 f(x,y)\right|^2 + \left|W_s^2 f(x,y)\right|^2} \tag{2}$$

$$A_s f(x,y) = \arg(W_s^1 f(x,y) + iW_s^2 f(x,y)) \tag{3}$$

The difference between Mallat in his two dimensional biorthogonal transform and the ordinary orthogonal Harr two dimensional transform is that he does not subsample his images as in the 1 dimensional case and that he applies only one filter in the X and Y directions to compute a polar representation for modulus maxima for each of the high pass bands. Thus, Mallat has only two high pass bands which can be represented in terms of x and y or polar representation.

IV. Fractal Encoding

We now describe the process of fractal quantization. Classic fractal compression is described in terms of Harr basis sets and cross scale approximation which is in effect performed due to the averaging process in fractal compression. There are essentially three parameters needed for fractal reconstruction as was discussed in Jaquin and Fischers's techniques and indicated in equation 1

$$f(x) = Tf(x) = U_L f(x) + b = Q_L f(2x) + b \tag{4}$$

where f(x) is the image to be transformed and T is a contractive operator with unique fixed point $\bar{f}$. Encoding f means finding and operator T having a fixed point $\bar{f}$ approximately equal to f while decoding is equivalent to finding the fixed point $\bar{f}$ by iterating T starting with an image selected at random. U_L represents the transformation applied to the domain blocks which both grey levels and decimates; Q_L represents a simpler U_L that simply scales the already decimated domain blocks, and b is the intensity offset applied to the domain blocks. Note that f, b ε V where V is a discrete and finite dimensional space. Jaquin[10] describes this block matching between scales in terms of a Markov operator since in a pure fractal (Global IFS) [3] intensity and geometric matching occurs at all wavelet scales whereas in Jaquin's fractal approximation it occurs between two wavelet scales.

Instead of using the Harr basis set we insert the Mallat Gaussian derivative wavelet transform into the fractal compression. It is also interesting to note as shown in that the biorthogonal spline has a much more localized frequency response than the Harr basis set [5]. This fact is extremely important in the fractal reconstruction process since the Harr basis contains many high frequency artifacts in reproduced images due to its sharp spatial domain cutoff. The Gaussian derivative spline has a much smoother spatial cutoff and thus much less tendency to create artifacts in imagery. as is shown in the frequency plot below looking at all models of fractal encoding we realize that the trend is to reconstruct an object from its low frequency components to its high frequency components. In the wavelet fractal case this method approximates wavelet coefficients across scales. This model fits with the fractal method since low frequency corresponds to large scale. We also know that the Mallat multiresolution decomposition happens in dyadic scales which, in the Harr basis set case, corresponds to blocks which are dimensions are powers of 2 in size. Thus we shall keep with this framework for the Gaussian derivative basis set. Putting this in the context of the wavelet transform we re-write equation 4 by inserting equation 1 as:

$$W_{2^{j-1}} f(x,y) = Q_L W_{2^j} f(x,y) + b \tag{5}$$

Thus we build our reconstructed image from the low frequency or large scale images first and then eventually reconstruct the final image. Note that his process use the wavelet decomposition to explicitly separate scales by frequency. Thus for a given block size we only have information that is fits that particular scale.

V. Localized Texture Coding

In order to determine the intensity offset or b parameter from equation 5 between two scales we find the blocks that best match eachother between two scales In traditional fractal encoding we do this do an exhaustive LMS search of range to domain blocks and take the block with the lowest difference. If we look at wavelet theory however we find that such exhaustive block matching is unnecessary. The reason stems from observing the properties of a Markov wavelet model described with by Luettgen and Willsky[12]

To represent this Markov random field we define a given node in the quad tree structure as s, its children nodes as $s\alpha_{NW}, s\alpha_{NE}, s\alpha_{SE}, s\alpha_{SW}$ and its parent node as $s\ddot{\gamma}$ where $\bar{\gamma}$ shifts the wavelet coefficients from parent $s\ddot{\gamma}$ to child s. Now, defining a MRF on a $2^N \times 2^N$ lattice, a state at the mth level represents the values of the MRF at $16(2^{N-m}-1)$ points. This set of points is denoted as Γ_s and it is the union of 4 mutually exclusive subsets. In general we can divide Γ_s into four set sets of $4(2^{N-m(s)}-1)$ points in a similar fashion, and we denote these subsets as $\Gamma_{s,i}, i \in \{NW,NE,SE,SW\}$. Now if we have the

random variable Z representing the current state of any Γ_s at any stage of the tree then we insert our local IFS relationship as

$$Z_t = Q_L Z_T + b \tag{6}$$

defined in the fractal wavelet context as

$$Z_T \in W_{2^{j-1}} f(x,y) \qquad Z_t \in W_{2^j} f(x,y) \tag{7}$$

thus the basic probabalistic Markov relationship is defined as

$$p_{Z_t, t \in \Gamma_{s\alpha_i} | Z_T, T \in \Gamma_s}(Z_t, t \in \Gamma_{s\alpha} | Z_T, T \in \Gamma_s) = p_{Z_t, t \in \Gamma_{s\alpha_i} | Z_T, T \in \Gamma_{s,i}}(Z_t, t \in \Gamma_{s\alpha} | Z_T, T \in \Gamma_{s,i}) \tag{8}$$

This relationship defines regions of constant texture. For any region of constant texture the slope of the energy decay across scales remains fixed. This energy decay also known as the Lipschitz or Holder exponent, also closely related to fractal dimension [15], can be characterized in detail by the Markov model. If we select some α where $0 < \alpha < 1$ and the function f(x,y) is uniformly Lipschitz over an open set of reals if there exists a constant K such that for all points (x,y) of this open set

$$M_{2^j} f(x, y) \leq K(2^j)^{\alpha} \tag{9}$$

α represents slope of the decay function at any given spatial point in an image. The b parameter a linear approximation of α between any scales if it is computed using the localized markov approach since α seeks to characterized the decay across scales within cone of influence of the transform. This cone of influence resides within the spatial boundaries of the quadtree in our case because of the dyadic decomposition of the wavelet transform in this area.

VI. Localized Shape Coding

To find a mapping Q_L from shape to texture for a particular object we need an efficient mechanism for determining a geometric mapping. If we recall, range and domain blocks in traditional fractal encoding are obtained by subsampling the image and then matching each range block to every possible domain block in an image[10]. Needless to say this encoding process takes an extremely long time and is one of the major drawbacks of traditional fractal encoding.

To simplify the process of range to domain block matching thus finding the mapping for Q_L in equation 5 we classify the range and domain blocks by summing the blocks' gradient angle parameters since these gradient are accurate indications of energy and direction within each block. Jaquin used a similar procedure in his classification of blocks in traditional fractal encoding. by applying the centered operator to each block. Now with Mallat's wavelet decomposition energy direction is already indicated as a natural part of the process.Thus the operation of block classification thus becomes a lookup table procedure rather than an exhaustive matching process. Both range and domain block position in the image are stored. The block rotation value is also determined by applying the appropriate flip that makes the block gradient angles match most closely

In addition to block matching the shape contour information about objects in the scene can be also included in the compressed object by encoding around modulus maxima values. This also is a crude form of quantization around blocks with energy above a given threshold since energy with the Gaussian derivative is centered around the edges or zero crossings of the encoded of the encoded image. This process can be used to associate groups of blocks in conjunction with the localized texture metrics. Other well known shape grouping algorithms may also be employed such as chain coding.

VII. Object Quantization

Rate distortion in a wavelet fractal sense is handled by how many wavelet scales are used in the decomposition of the image.[7] The number of scales starts with the largest scale, and thus largest range block used. We go to the next lower scale and thus next smaller range block in a quadtree form if we do not find an adequate match between range to domain block. This process is now described in terms of classic rate distortion.

Traditional quantization deals with Q [17] as the set of admissible scalar frequency quantization choices and T as the complete spatial tree and S as the pruned spatial tree where $S \leq T$. In our model based approach instead of S our tree structure is pruned according to some model preconfigured model. We denote this tree model S_M which consists of the set of nodes which characterize our texture and shape region of interest.

Our model optimization problem seeks to solve the traditional rate distortion problem where a frequency quantizer $q \in Q$

$$\min_{q \in Q, S_M \leq T} D(q, S_M) \text{ subject to } R(q, S_M) \leq R_{budget} \tag{10}$$

where $D(q,S_M)$ and $R(q,S_M)$ represent the distortion and rate respectively associated with the frequency and spatial quantizer choices $q \in Q$ and $S_M \leq T$

As in conventional cost/optimization, this constrained optimization problem can be solved by unconstraining it via the Lagrange multiplier $\lambda \geq 0$, which quantifies the trade-off between rate and distortion and minimizing the unconstrained Lagrangian cost

$$J(q, S_M) = D(q, S_M) + \lambda R(q, S_M) \tag{11}$$

We now use the quantization process Q to define a simple model by the average holder exponent for the range of scales over which modulus maxima for a given object are computed. Since b is a linear approximation of the Holder exponent between two scales and we wish to define the wavelet transform across a range of scales we can approximate this α_S

Thus for a region of constant texture a measurement of α can be determined using the average value of b for a given range of encoding scales s denoted S {s:0 < s <N-1.} and this α_S is computed as:

$$\alpha_S \sim \frac{\sum_{s=0}^{N-1} b_s}{N-1} \tag{12}$$

Thus we segregate and prune regions of our image by restricting σ_{α_S} to be less than a certain maximum

$$\sigma_{\alpha_S} < max(\sigma_{\alpha_S}) \tag{13}$$

Because this quantization process naturally removes regions of low intensity around the modulus maxima and the modulus maxima define shape boundaries of objects our texture quantization methodology naturally finds the boundaries of objects and encodes them. Shapiro noticed that the natural decay of frequency between scales can be exploited to optimize the compression process by eliminating wavelet coefficients that fall below some threshold. For encoded texture and shape objects our procedure does not make this assumption but encodes the spectrum assuming our Markov model where spectral intensity can increase or decrease across scales. Our quantization method is applied within individual objects to optimize their overall quality. Variation of texture and shape definitions will result in different compression ratios depending on the size of the uniformity and size of the texture and shape region.

VIII. Decoding

Because our frequency mapping proceeds from low to high frequency our reconstruction process simply proceeds by iterating our compressed parameters on the low pass image. For each scale the iterative procedure forms the approximation of the next higher lowpass image and is then lowpass filtered and the process is repeated. This process thus removes all blocky artifacts in the image while still revealing the image features. The above technique leads to a direct reconstruction method as to in the following equation

j = J
while (j > 0)

$$S^d_{2^{j+1}} f = Q_L \frac{1}{\lambda_j} S^d_{2^j} f^*(\tilde{H}j+1, \tilde{H}j+1) + b \tag{14}$$

j = j - 1
endwhile

Thus in the compressed file our first step is to upsample compressed lowpass image and then use the stored fractal local iterated function parameters on it to reconstruct the original image building each new scale from the previous lowpass image. The image can be restored to any desired resolution simply by stopping the reconstruction process at a given scale. This iterative procedure is only order O(N) where N is the number of image blocks for all scales and thus can be performed real time with no special hardware. Thus this approach is a significant advance over the Mallat alternating projections reconstruction method because it is extremely efficient in its reconstruction speed.

IX. A Compression Application

We now define a compression application which will allow us to exploit the advantages of our methodology while still retaining good overall compression performance. It is well known that tranditional wavelet entropy encoding methods outperform fractal and wavelet fractal compression methods by pure PSNR to compression ratio performance. Thus we will only encode those regions of the image where detailed object analysis is required and leave the rest to be coded by traditional wavelet entropy methods[5], [15]. In our application we encode the region around Lenna's face with our encoding method . We maintain high quality around the facial region by compressing at a lower compression ratio around the face than in the rest of the image. We set our quantization procedure Q by which then specifies the objects that we encode. We quote two metrics for image PSNR quality. One for the region within Lenna and one for the overall image PSNR. Compression ratio is given for the overall image. Our results show that our overall image quality is less that that of traditional methods but at the higher compression ratios the image quality within the region of interest of our object surpasses that of the conventional methods. Thus by combining compression and recognition in one operation we can preserve image quality in those regions which have particular interest to the user and demphasize the background. This region of the image is now content addressable for recognition purposes in the compressed domain given that its compressed data (namely b values) were recognized by our quantization scheme Q such that they are the model Markov structure S_M.

Compress Ratio:	4:1	8:1	16:1	32:1	64:1
Bits/Pixel:	2.0	1.0	0.5	.25	.125
Object PSNR(db):	35.6	33.7	32.2	31.5	30.25
Image PSNR(db):	39.0	36.2	33.2	30.2	27.54
JPEG PSNR(db):	44.0	38.1	33.0	28.5	-
EZW PSNR(db):	-	39.6	36.3	33.17	30.23

Figure 1. Compression Ratio vs. PSNR on 512x512 Lenna 8bpp (preliminary)

X.Conclusion

The fractal-wavelet method offers a significant improvement over existing techniques because of its unified approach to image analysis and compression. The fractal wavelet method itself gives naturally higher compression and better reproductive quality than conventional DCT-based methods for specific regions of interest. By its wavelet frequency division process, it gives a more natural organization to existing fractal methods and allows more accurate block matching. As a result of its modulus maxima shape representation it gives a shape to texture content-based approach to compressed file organization. By its gradient based block matching technique it is significantly faster than existing wavelet-fractal compression methods.

XI. References

[1] Alexandru Bogdan, Image Coding Using Iterative Transformations with Applications to Image Communication, PhD Thesis, Columbia University, Feburary 1995.

[2] R. Bonneau, "The multiresolution transform and its application to image coding", SPIE/IEEE Conference on Visual Communication and Image Processing: Wavelets and Fractals, Orlando, March 1996.

[3] M. Barnsley, Fractals Everywhere 2nd Ed., Academic Press, New York, 1993.

[4] M. Barnsely and L. Hurd, *Fractal Image Compression*, A.K. Peters, Wellesley Mass., 1993.

[5] G. Davis, "Adaptive Self-Quantization of Wavelet Subtrees: A Wavelet-Based Theory of Fractal Image Compression", SPIE Conference on Mathematical Imaging: Wavelet Applications in Signal and Image Processing, San Diego, July 1995.

[6] I. Daubechies, "Ten lectures on wavelets," CBMS-NSF Series Appl. Math. SIAM, 1991.

[7] Y. Ficher, *Fractal Image Compression*, Springer Verlag, New York, 1995.

[8] P. Flandrin, "Wavelet analysis and synthesis of fractional Brownian motion", IEEE Transactions on Information Theory, vol. 38, no. 2, March 1992.

[9] W.H. Hwang and Stephane Mallat, "Characterization of Self-Similar Mulitfractals with Wavlet Maxima", NYU Technical Report 641, July 1993.

[10] A. Jacquin, A Fractal Theory of Iterated Markov Operators With Applications to Digital Image Coding, Georgia Institute of Technology PhD. Dissertation, August 1989.

[11] Joshi, R., Jafarkhani, H, Kasner, J., Fischer, T., Farvardin, N., Marcellin M., Bamberger, R., "Comparison of Different Mehtods of Clasification in Subband Coding of Images", IEEE Transactions on Image Processing, Vol 6. No 11, pp. 1473-1486, November 1997.

[12] Luettgen, M., Karl, W., Willsky, A., Tenny, R., "Multiscale Representations of Markov Random Fields", IEEE Transactions on Signal Processing, Vol 41, No 12, pp. 3377-3396, December 1993.

[13] S. Mallat and S. Zhong, "Characterization of signals from multiscale edges", IEEE Trans. on Patt. Anal. and Mach. Intell., Vol. 14, pp.710-732, July 1992.

[14] S. Mallat, "Multiresolution Approximations and Wavelet Orthonormal Bases of L2R", Transactions of the American Mathematicall Society, Volume 315, Number 1, September 1989.

[15] A. Pentland, "Fractal Based Description of Natural Scenes", IEEE Transactions on Pattern Analysis, and Machine Intelligence, Vol PAMI-6, No 6., November 1984.

[16] Ramchadndran, K, Orchard T., "An Investigation of Wavelet-Based Image Coding Using an Entropy -Constrained Quantization Framework", IEEE Transactions on Signal Processing, Vol 46, No 2, pp342-353, February 1998.

[17] J. Shapiro, "Embedded Image Coding Using Zerotrees of Wavelet Coefficients", IEEE Transactions on Signal Processing, Vol 41, No 12, December 1991.

Rotary-Wing Target Radar Signal Processing And Signature Simulator

Barry D. Bullard, Ph.D, P.E.
Patrick C. Dowdy

Georgia Institute of Technology
Georgia Tech Research Institute
Huntsville Research Operations Laboratory

Abstract

Utilizing coherent pulse Doppler radar waveforms and feature extraction signal processing techniques, a rotary –wing aircraft's rotor design such as main rotor configuration (single, twin tandem, twin coaxial, etc), number of blades (2, 3, 4, 5, etc), tail rotor blade count and configuration (cross, X, star, etc) can be determined. Such information can be used to assist in the classification and identification of the aircraft. This paper describes the development of a high fidelity coherent pulse Doppler radar time domain signature simulator for military rotary-wing aircraft targets. The simulator model's radar cross section (RCS) backscatter of the rotary-wing aircraft's airframe, main rotor blades, main hub section, and tail rotor blades as a function of time. The simulator also has models for simple clutter and noise, which can be added to the target return at any desired signal-to-clutter (S/C) and/or signal-to-noise (S/N) levels.

1. The Doppler Signature of a Rotary-Wing Aircraft

Figure 1 illustrates a typical military helicopter in forward flight. The main rotor blades are rotating in a clockwise direction as viewed from the top, with the tail rotor blades rotating in the plane of the drawing from top-to-bottom.

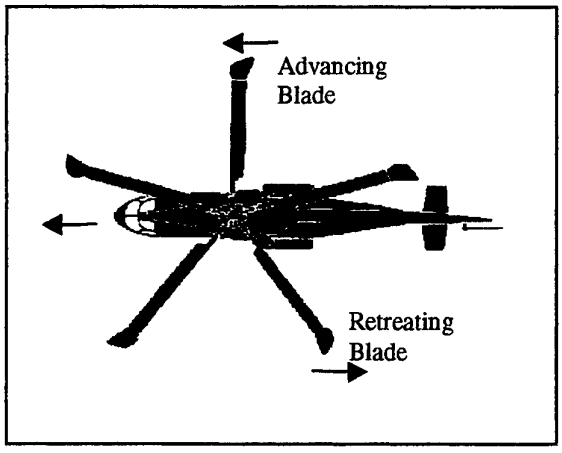


Figure 1. Helicopter in Forward Flight

If we assume that a pulse Doppler radar is viewing the aircraft of Figure 1 from the left along the flight path, the helicopter will exhibit several Doppler signatures to the radar processor. These will include positive Doppler shifts of the airframe, the advancing main and tail rotor blades, and the advancing elements of the hub region. Negative Doppler shifts will result from the retreating blades and retreating hub region scatters.

Figure No. 2 shows the complex magnitude time domain signature from the return of an actual military helicopter including noise and clutter. Figure 3 illustrates the power spectrum density of the signal (16384 point complex FFT). The recording pulse Doppler radar was aboard an instrumented aircraft in flight, thus the shifted clutter line. The helicopter was on an approaching profile in forward flight, resulting in additional Doppler shift of the airframe (i.e. skin). Centered around the helicopter airframe is the hub region returns and the advancing & retreating blade returns.

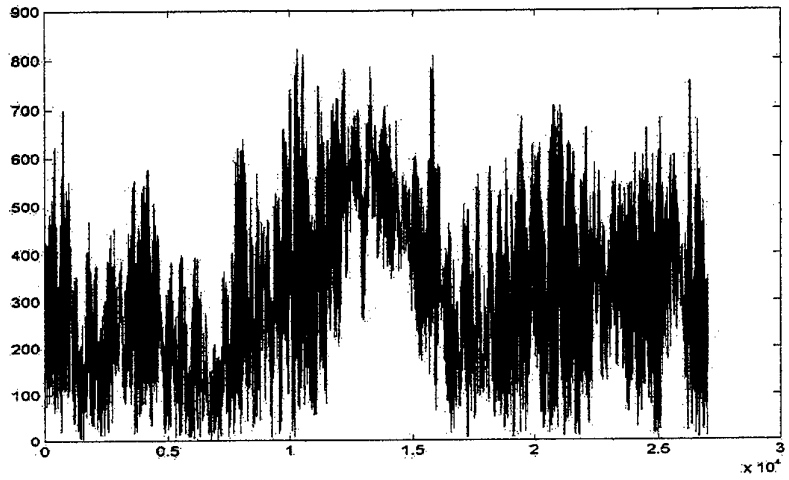


Figure 2. Magnitude Time Domain Data of RW Aircraft in Clutter and Noise

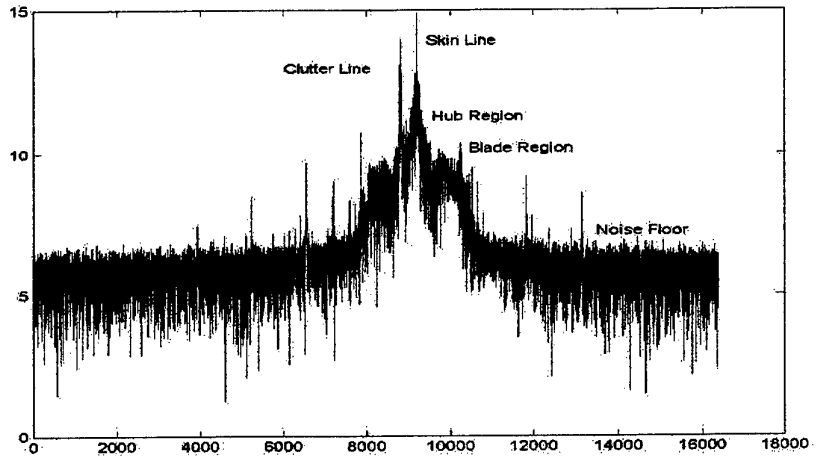


Figure 3. PSD of Time Domain Signal

Feature extraction of the radar signal can be achieved by selecting only the regions in the PSD that are of interest (i.e. skin, hub, blade, etc), and performing an inverse complex FFT to return the data to the time domain. Figure 4 illustrates such a process for the advancing blade flash region.

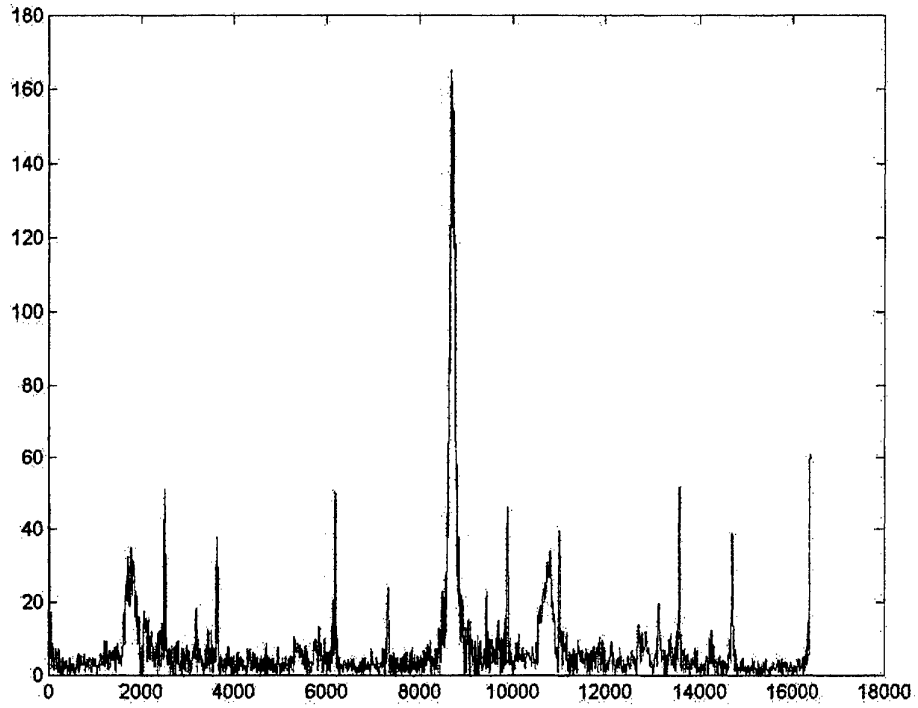


Figure 4. Time Domain - Advancing Blades

Analysis of Figure 4 reveals one main rotor blade "flash", two main rotor sweep-tip "flashes", and eight tail rotor "flashes". This information, combined with data from the retreating blade spectrum, can significantly characterize a given rotary-wing aircraft.

2. Rotary-Wing Signature Simulator

Figure 5 illustrates the coherent time-domain rotary-wing radar backscatter signature simulator developed at Georgia Tech Research Institute. Helo-Sim consists of models for:

1. advancing main rotor blades
2. advancing tail rotor blades
3. retreating main rotor blades
4. retreating tail rotor blades
5. hub signature (both advancing and retreating)
6. airframe (skin signature)
7. white noise
8. distributed clutter
9. radar waveforms

10. 3-D dynamic environment for the radar platform and rotary-wing target

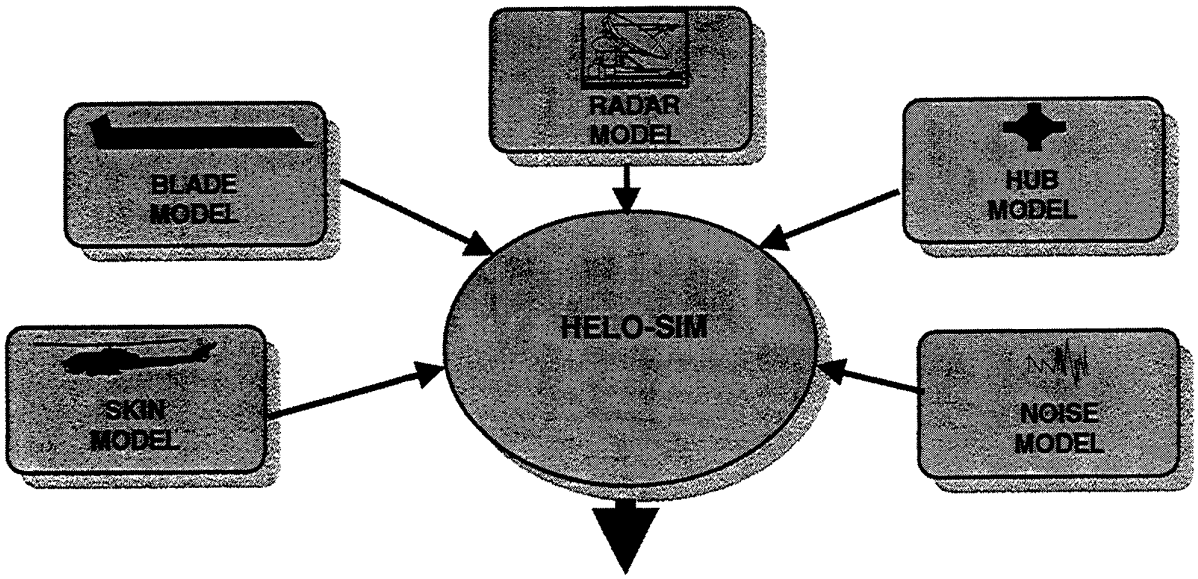


Figure 5. Pulse Doppler Radar Rotary-Wing Target Signature Simulator

The blade modeler generates returns based on the physical and dynamic properties of the main rotor and tail rotor blades of the aircraft. Variables include; blade length, major axis, minor axis, composite or all-metal, sweep-tip design of the main rotor blade, number of blades (main and tail), configuration of the tail rotors (cross, X, star, etc), and the rotation rate of the main and tail sections. The skin, noise, and hub models are based on probability estimators with user defined mean a variance settings. The radar model defines the radar wavelength and pulse repetition frequency.

Figure 6 is several complex time domain magnitude signatures generated by Helo-Sim for a hypothetical 4-blade main rotor/2-blade tail rotor helicopter. Subplots (a), (b), (c), and (d) illustrates the forward blade, hub, skin, and retreating blade signatures respectfully.

Figure 7 shows the composite complex time domain signature for the complete rotary-wing signature with added noise. Figure 8 illustrates the complex magnitude FFT for the signal of Figure 7. Review of Figure 7 indicates the helicopter's skin line, hub, and blade spectrums.

3. Summary

Researchers at Georgia Tech have developed a high fidelity pulse Doppler radar signature simulator for rotary-wing aircraft. Unique properties of a given helicopter, such as its main rotor blade design, tail rotor design, and hub structure can be characterized.

Simulated time domain signatures from the modeled helicopter can be presented to special radar processors for testing information to a non-cooperative target recognizer.

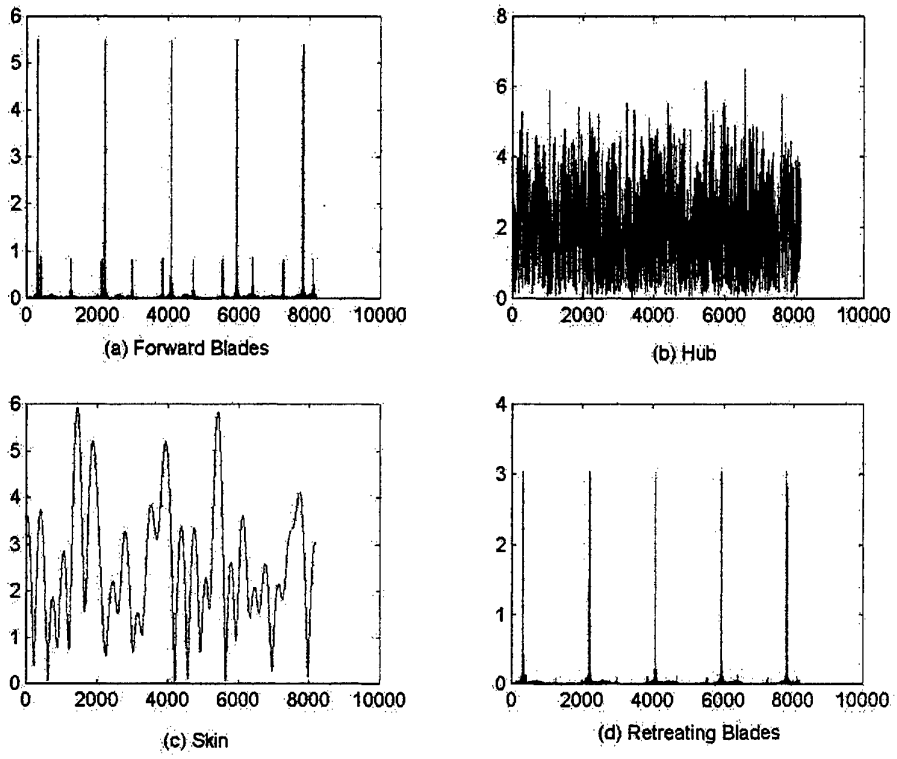


Figure 6. Helo-Sim Generated R/W Time Domain Signatures

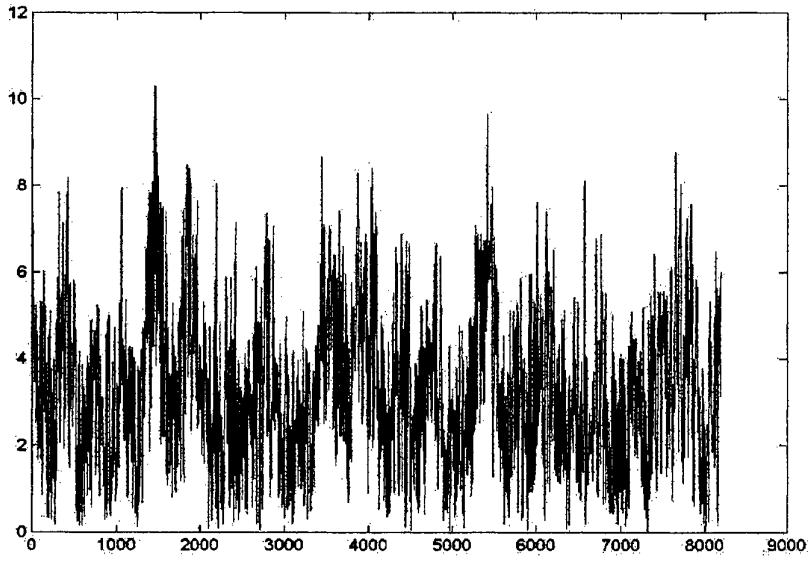


Figure 7. Composite R/W Time Domain Signature (Helo-Sim)

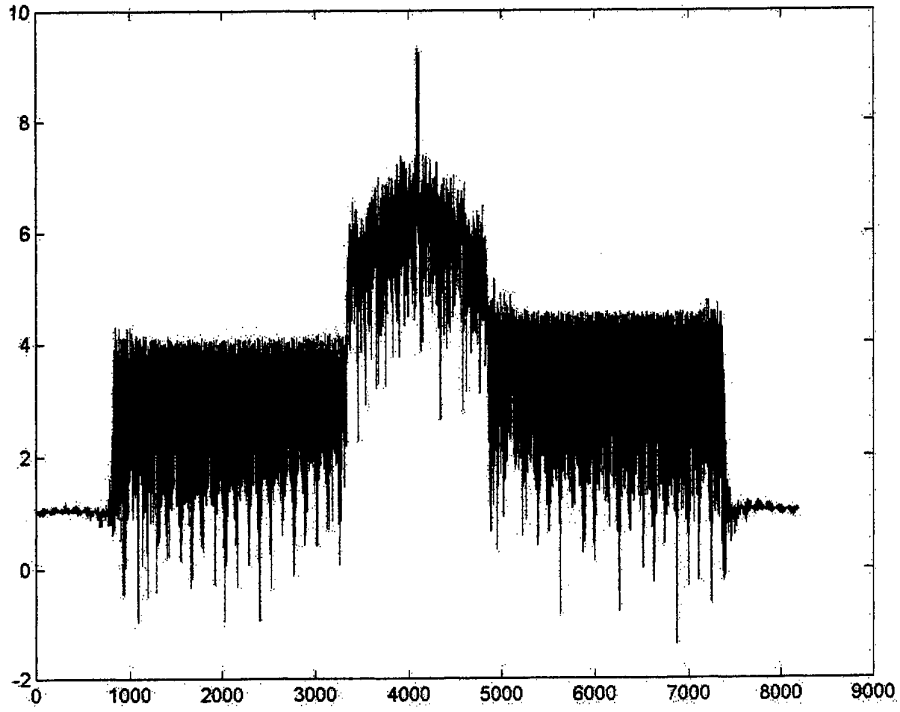


Figure 8. FFT Magnitude of Composite R/W Signature (Helo-Sim)

Comparison of Selected Features for Target Detection in Synthetic Aperture Radar Imagery

Tristrom Cooke* Nicholas J. Redding† Jim Schroeder‡ Jingxin Zhang§

Abstract

Several methods are available that capture the statistics of radar imagery. The best features, in the sense of man made target discrimination, are expected to be different for different types of natural background, and for different objects of interest such as vehicles. We demonstrate that discrimination of natural background and man made objects using low resolution Synthetic Aperture Radar imagery is possible using multiscale autoregressive (MAR), multiscale autoregressive moving average (MARMA) models, and singular value decomposition (SVD) methods. We use the model coefficients, moments of the model residual vectors, a subset of eigenvectors, and moments of the selected eigenvectors, as features for target discrimination. All the test imagery used here was 1.5 metre resolution.

Keywords: multiscale models, singular value decomposition, automated target detection, synthetic aperture radar, natural background.

1 Introduction

Characterising the natural background or clutter environment in synthetic aperture radar (SAR) imagery is an important step in developing better automated target detection (ATD) tools. The ability to discriminate one type of background from another can lead to the use of adaptive ATD algorithms which go beyond merely looking for radar bright objects. These models capture the variation in the statistical properties of a given region in an image. Natural backgrounds exhibit different statistics than man made objects, which we exploit as a means of discrimination. In this work we use Multiscale Autoregressive(MAR) models, Multiscale Autoregressive Moving Average (MARMA) models, and Singular Value Decomposition (SVD) techniques to derive features sensitive to the statistical differences between background clutter and man made objects such as vehicles.

2 Multiscale Modeling Methods

Multiscale methods have been applied to the problem of target detection and recognition in SAR imagery by Irving et al [1] and Subotic et al [2]. These methods differ from more traditional image analysis methods in that they operate on a sequence of related images, each being a view of the same scene at a different resolution, rather than a single high-resolution image. In our work, this sequence of images is organised as a hierarchical multiscale stack, where the resolution decreases from fine to coarse by a fixed factor progressively up through the levels of the stack. The models we impose on these stacks are generalisations of those used in ordinary time series. In this context, the level number (or scale index) plays the equivalent role to time.

*Centre for Sensor, Signal and Information Processing, SPRI Building, Mawson Lakes, SA, 5095, Australia. E-mail: tcooke@cssip.edu.au

†Defence Science and Technology Organisation, Salisbury, SA, Australia. E-mail: nick.redding@defence.dsto.gov.au

‡Centre for Sensor, Signal and Information Processing, SPRI Building, Mawson Lakes, SA, 5095, Australia. E-mail: schroede@cssip.edu.au

§Centre for Sensor, Signal and Information Processing, SPRI Building, Mawson Lakes, SA, 5095, Australia. E-mail: jingxz@cssip.edu.au

In [3] and [4], it was shown that multiscale autoregressive (MAR) models are effective in discriminating vehicle-sized objects from natural background in SAR imagery. Here we further quantify target discrimination performance using receiver operating characteristic (ROC) curves estimated from a larger data set. We also consider the wider class of multiscale autoregressive moving average (MARMA) processes. These are generalisations of ARMA models in time series analysis, and they contain MAR processes as a special case.

A multiscale stack is a sequence $\{X^{(m)}\}$ of images indexed by level number $m = 0, \ldots, M$, where M is a positive integer. Level 0 is the highest resolution image, and the resolution decreases by a factor of two in going from each level to its successor. The set of pixels at level m is

$$S^{(m)} = \{(i,j) : i = 0, \ldots, i_m - 1, j = 0, \ldots, j_m - 1\} .$$

The state of pixel (i,j) in $S^{(m)}$ is $X_{ij}^{(m)}$, a random variable which may be real or complex. Every pixel in every level except M has a parent pixel in the level above it. The parent of pixel (i,j) in level m is the pixel $(i \, div \, 2, j \, div \, 2)$ in level $m+1$, where div is the integer division operator (i.e. the remainder is discarded). The ancestor of (i,j) at level $m+k$ (provided $m+k \leq M$) is the pixel $(i \, div \, 2^k, j \, div \, 2^k)$. Every pixel in every level except the first has four children in the level below it. The children of (i,j) in $S^{(m)}$ are the four pixels $(2i, 2j)$, $(2i, 2j+1)$, $(2i+1, 2j)$ and $(2i+1, 2j+1)$, all in $S^{(m-1)}$. Note that $i_m = 2^{-k} i_0$ and $j_m = 2^{-k} j_0$, and that both i_0 and j_0 must be greater than or equal to 2^M.

The multiscale stacks used in this work were generated using coherent quadtree averaging and by use of a discrete wavelet transform; no significant differences between the two methods have been noted so far. The input to this process, a complex image of a region of interest at the highest resolution available, becomes level 0 of a complex multiscale stack $\{Z^{(m)}\}$, $m = 0, \ldots, M$. Each of the higher levels is generated from the level immediately below it with the resolution decreasing by a factor of two at each successive level. The state of each pixel at each level above the first is given by the arithmetic mean of the states of its children. So

$$Z_{ij}^{(m)} = \frac{1}{4} \left(Z_{2i,2j}^{(m-1)} + Z_{2i,2j+1}^{(m-1)} + Z_{2i+1,2j}^{(m-1)} + Z_{2i+1,2j+1}^{(m-1)} \right) .$$

Once the complex stack has been generated, each level is converted to decibels (log detection). This procedure produces a second multiscale stack $X = \{X^{(m)}\}$, $m = 0, \ldots, M$, which is the input to the model fitting and classification procedures.

A MAR process defined on a multiscale stack X satisfies

$$X_{ij}^{(m)} = \sum_{k=1}^{R(m)} a_k^{(m)} X_{i \, div \, 2^k, \, j \, div \, 2^k}^{(m+k)} + \xi_{ij}^{(m)} \quad , \tag{1}$$

where $R(m)$ is the order of the process for level m, $\xi_{ij}^{(m)}$ is zero mean noise, and the coefficients $a_1^{(m)}, \ldots, a_{R(m)}^{(m)}$ are real valued model parameters. An intercept term can be added to each level if required. The same coefficients apply to all pixels on the same level, but the coefficients for different levels need not be the same. For each level m, the $\xi_{ij}^{(m)}$ are independent and identically distributed, i.e. $\xi^{(m)} = \{\xi_{ij}^{(m)}\}$, $(i,j) \in S^{(m)}$, is a spatially stationary, strictly white noise field. We further assume that the noise fields of two different levels are independent.

To define a MARMA process on a multiscale stack X, we simply augment the above process like so:

$$X_{ij}^{(m)} = \sum_{k=1}^{R(m)} a_k^{(m)} X_{i \, div \, 2^k, \, j \, div \, 2^k}^{(m+k)} + \sum_{k=1}^{Q(m)} b_k^{(m)} \xi_{i \, div \, 2^k, \, j \, div \, 2^k}^{(m+k)} + \xi_{ij}^{(m)} \quad . \tag{2}$$

Here $b_1^{(m)}, \ldots, b_{Q(m)}^{(m)}$ are the coefficients of a multiscale moving average process.

The MAR process has a likelihood function which readily decomposes into a product of terms for each pixel and is amenable to computation. This makes likelihood approaches to model fitting and to classification straightforward, as demonstrated in [2]. By contrast, the likelihood function for MARMA is unwieldy. Iterative least squares techniques need to be used for model fitting, and likelihood based classification is problematic. We address this last issue by trying to use the fitted coefficients under the model as the input to the classification procedure, rather than trying to classify on the image data directly.

3 Singular Value Decomposition Method

An SVD feature approach, unproven and still in relative infancy, involves the analysis of high range resolution (HRR) profiles [10, 5] for ATD/R systems. Singular Value Decomposition (SVD) analysis of HRR data reveals that the range-space eigenvectors corresponding to the largest singular value accounts for more than 90 percent of target energy, thus it may be that the eigenvectors can be used as a target detection statistics. HRR analysis has so far concentrated on ATR so it is not clear what value may accrue to ATD algorithms based upon this approach. For a SAR image the analogy would be to use sequences of cross track pixels from a region of interest.

The motivation for interest in HRR SVD techniques within the USAF is an attempt to bridge the gap between MTI mode and SAR stripmap mode of wide area surveillance systems. MTI is good at detecting moving targets, but ineffective against stationary targets. SAR likewise is ineffective against moving targets but capable of accurately imaging stationary targets. It is possible that the SVD produces a useful feature space for SAR imagery not previously exploited.

A possible approach can be outlined as follows. Let X be a size N x M matrix containing a region of interest from a detected or complex SAR image. This would consist of N range bins or cross track pixels and M along track focussed strips. The SVD decomposition results in

$$U = Eigenvectors\, of\, [XX^T] = [u_1, u_2, ...u_N],$$

$$V = Eigenvectors\, of\, [X^T X] = [v_1, v_2, ...v_M],$$

$$\Lambda = diag[\lambda_{11}, \lambda_{22}, ...\lambda_{MM}].$$

The eigenvectors u_i correspond to the magnitude ordered eigenvalues λ_i, and represents energy from a target range profile (cross track pixels). Likewise, eigenvectors v_i span the along track subspace. It is predicted that the target energy is concentrated in eigenvectors u_1 and v_1. If the region of interest contains clutter only, it is predicted that the eigenvectors are evenly distributed in their components. A simple linear discriminant or other classifier can then be designed to classify a region as target or clutter only.

4 Other Selected Features/Discriminators

Along with the features previously discussed, numerous other features for magnitude only data such as maximum target intensity, Karhunen Loeve Transform (KLT) based methods and background distribution models (such as the U,V and W measures for K-distribution parameter estimation described in [8]) have also been considered. After extensive testing over a large and realistic database containing 53961 background and 22084 target samples, it was found that the best single feature found to date for target detection is the V measure for the estimation of the parameters of the K distribution.

A number of co-occurrence matrix features have already been described in Redding [9]. Lie [7] contains another set of co-occurrence based texture features, such as the 'busyness', Weber's contrast, 'average contrast' and 'homogeneity'. This paper also outlines an algorithm for efficiently calculating these measures.

There are many different ways in which an image can be decomposed into a weighted sum of orthonormal basis functions, where the weights can be used as features. Ghosal and McKee [6] for instance use a Zernike basis, while Rong and Bhanu [11] derive three features based on the means and moments of an image's coefficients in a Gabor basis.

A set of features that seems very promising for target detection is the SVD of the FFT. Although no firm physical explanation has been found for this, the first row of the first matrix produced by the SVD yields a reasonably good ROC curve as shown in Table 3. The best 5 other features (maximum intensity, one FFT coefficient, variance, skewness and a parameter derived from the KLT) gave a similarly good ROC. Choosing the best 8 features from all of the features tested (maximum intensity, 4 SVD of FFT coefficients, one FFT coefficient and the skewness) gave a FAR of about 7.8 percent for a PD of 90 percent.

Clearly, Automatic Target Detection can be considered as a form of target classification, wherein just two classes exist: Target only or Background Clutter only. It may be advantageous to approach ATD as a pattern classification problem and explore the use of modern classifiers such as Support Vector Machines (SVMs) or Neural Networks (NNs) to name just two possible techniques. The potential use of SVMs for target discrimination is summarised in [9].

5 Results

The multiscale models described in the previous section were each fitted to samples of "Background," "Grassland," "Woodland," and "Vehicles" extracted from 1.5m resolution SAR imagery. Additionally, eigenvectors from an SVD were computed from the magnitude imagery. The imagery was collected during trials of DSTO's Ingara airborne SAR. There were 53 samples of the background class, 256 samples of homogeneous grass, 256 samples of woodland, and 159 samples of the same vehicle from different orientations. Each sample was a 16×16 matrix of complex SAR data which was processed into a four level stack. Model fitting was accomplished by transforming each stack into a 256×4 data matrix suitable for use as input into the standard time series fitting routines in MatLab.

For this work, we only estimated the coefficients for level 0 of the multiscale stack. Four different models were fitted to the SAR data:

- a second order autoregression denoted MAR(2);
- a second order autoregression denoted MAR(3);
- a fourth order autoregression, MAR(4); and
- a second order autoregressive moving average denoted MARMA(2,2).

Table 1 summarises the performance of the multiscale techniques using the area under the ROC curves derived from coefficients and/or residual vector moments under several multiscale model cases. In most cases, the coefficients vary quite widely and the two populations overlap to varying extents. However, linear discrimination works (albeit with some error) for certain model sizes. Several model orders are invoked: MAR(2), MAR(3), MAR(4), and MARMA(2,2), over regions of interest (ROI) sized 8×8, 16×16, and 32×32. The complex multiscales are created either by complex quadtree averaging or by use of a "db1" wavelet transform; the results are similar. We caution, however, that the background clutter is reasonably homogeneous without additional cultural features, and that the targets are comparatively strong. In all cases the targets are multiple images of the same vehicle from different orientations.

Quadratic discriminant rules were also formulated from the training data for each of the multiscale models. Classifiers based upon these rules performed poorly relative to the linear discriminant rules. Examination of the scatterplots indicate a possible reason for this. The fitted coefficients of each class are quite widely dispersed with plenty of outliers. As a result, the sample covariance matrices needed for quadratic discriminators would be poor estimators of the underlying population covariance matrices. We are currently exploring the use of a Support Vector Machine (SVM) as an alternative to linear discriminant classifiers.

Typical results from use of SVD derived features (over 16×16 region size) for the target discrimination task are shown in Table 2. Specifically, area under the ROC curve using both the u_1 and v_1 eigenvector, or 32 coefficients total input into the linear discriminant is shown, and results from using just the moments of the u_1 and v_1 eigenvectors. It can be seen that using the variance of the first two eigenvectors produced a perfect ROC curve. However, recall that these results are obtained on "ideal" data in that the background clutter is homogeneous, the target signal strength is high. It can be seen from Table 3 that testing on low contrast data results in an obvious degradation in target discrimination performance.

For the features used in Table 3 which are based only on the magnitude of the radar return, a larger data set with more realistic contrast levels was available for testing. This data set was extracted from a large image

MAR Model	Area	Comments/Features
MAR(2)	.560	159 Vehicles/Grass, 8 x 8 Regions, MAR Coefficients
MAR(2)	.928	159 Vehicles/53 Backgrounds, 16 x 16 Regions, Intensity Data, Peak Freq Difference
MAR(2)	.922	159 Vehicles/53 Backgrounds, 16 x 16 Regions, Intensity Data, Min Freq Location
MAR(3)	.910	159 Vehicles/Grass, 16 x 16 Regions, MAR Coefficients
MAR(4)	.999	159 Vehicles/Grass, 16 x 16 Regions, 3 MAR Coefficients/3 Residual Vector Moments
MAR(4)	.990	159 Vehicles/Grass, 16 x 16 Regions, 4 Residual Vector Moments
MARMA(2,2)	.998	159 Vehicles/Grass, 32 x 32 Regions, MARMA Coefficients
MARMA(2,2)	.990	159 Vehicles/Grass, 16 x 16 Region, MARMA Coefficients/Variance of Residual Vector
MARMA(2,2)	.990	300 Vehicles/Grass, 16 x 16 Region, MARMA Coefficents/Variance of Residual Vector
MARMA(2,2)	.980	327 Vehicles/Grass, 16 x 16 Region, MARMA Coefficents/Variance of Residual Vector
MARMA(2,2)	.970	159 Vehicles/Woods, 16 x 16 Region, MARMA Coefficents/Variance of Residual Vector
MARMA(2,2)	.940	159 Vehicles/Woods, 16 x 16 Region, MARMA Coefficents/Variance of Residual Vector
MARMA(2,2)	.930	159 Vehicles/Woods, 16 x 16 Region, MARMA Coefficents/Variance of Residual Vector

Table 1: Estimated area under a ROC curve for linear discriminant applied to the features derived from multiscale models

SVD Vectors	Area	Comments/Features
u_1,v_1	.994	159 Vehicles/53 Backgrounds, 8 x 8 Region, u_1 and v_1 Used
u_1,v_1	.964	159 Vehicles/53 Backgrounds, 8 x 8 Region, Mean of u_1,v_1
u_1,v_1	.964	159 Vehicles/53 Backgrounds, 8 x 8 Region, Variance of u_1,v_1
u_1,v_1	.836	159 Vehicles/53 Backgrounds, 8 x 8 Region, Skewness of u_1,v_1
u_1,v_1	.533	159 Vehicles/53 Backgrounds, 8 x 8 Region, Kurtosis of u_1,v_1
u_1,v_1	.993	159 Vehicles/53 Backgrounds, 16 x 16 Region, u_1 and v_1 Used
u_1,v_1	.926	159 Vehicles/53 Backgrounds, 16 x 16 Region, Mean of u_1,v_1
u_1,v_1	.927	159 Vehicles/53 Backgrounds, 16 x 16 Region, Variance of u_1,v_1
u_1,v_1	.888	159 Vehicles/53 Backgrounds, 16 x 16 Region, Skewness of u_1,v_1
u_1,v_1	.720	159 Vehicles/53 Backgrounds, 16 x 16 Region, Kurtosis of u_1,v_1

Table 2: Estimated area under a ROC curve for linear discriminant applied to the features derived from an SVD

into which were inserted 512 real targets at a variety of contrasts. A prescreening algorithm was applied to this large image, which resulted in the extraction of 76045 64 × 64 images (although only an 8 × 8 subset of this, corresponding roughly to the target pixels, was used for the calculation of the features). 53961 of these contained background, while 22084 contained inserted targets. This larger data set is much more challenging than the complex data set, as can be seen by a comparison of the area under the ROC curves for the Maximum Intensity, which was 0.974 for the smaller complex data set, but 0.845 for the new data set.

6 Conclusions

Our results show that linear discrimination of three types of natural background in SAR imagery, "woodland," "background," and "grassland", from man made bright objects is possible using 16 × 16 samples when the classification is made using the fitted coefficients and residual vector moments under one of several multiscale models. The autoregressive component of "grassland" is first order on average, while that of "woodland" is second order. Adding a first order moving average component improves the classification accuracy beyond that attainable using pure multiscale autoregression. The data are very clean, thus more extensive testing of the MAR and MARMA method on low resolution SAR data would be necessary before conclusions can be made.

The SVD methods worked exceptionally well on the clean data, especially using the variance of just two eigenvectors as a two element feature vector. Unfortunately, the results degrade significantly when tested against lower contrast and higher clutter data. The SVD still produced better discrimination than most other individual features on this data.

SVD/K/FFT	Area	Comments/Features
K Parameters	.821	All Clutter and Contrasts, 8 x 8 Region, U Measure Used as feature
K Parameters	.851	All Clutter and Contrasts, 8 x 8 Region, V Measure Used as Feature
K Parameters	.795	All Clutter and Contrasts, 8 x 8 Region, W Measure Used as Feature
u_1	.726	All Clutter and Contrasts, 8 x 8 Region, Variance of u_1
v_1	.840	All Clutter and Contrasts, 8 x 8 Region, Variance of v_1
u_1	.841	All Clutter and Contrasts, 8 x 8 Region, Eigenvector u_1 Used as Feature
v_1	.841	All Clutter and Contrasts, 8 x 8 Region, Eigenvector v_1 Used as Feature
Not Applied	.845	All Clutter and Contrasts, 8 x 8 Region, Max Intensity Used as Feature
Not Applied	.974	786 Vehicles/4096 Backgrounds, 8 x 8 Region, Max Intensity Used as Feature
FFT	.778	All Clutter and Contrasts, 8 x 8 Region, Real FFT Coefficient Used
FFT	.926	786 Vehicles/4096 Backgrounds, 8 x 8 Region, Real FFT Coefficient Used
K/FFT/SVD	.901	All Clutter and Contrasts, 8 x 8 Region, 5 Best Features Combined
SVD/FFT	.915	All Clutter and Contrasts, 8 x 8 Region, 8 SVD of FFT Used as Feature
SVD-FFT/K	.967	786 Vehicles/4096 Backgrounds, 8 x 8 Region, 9 Best features Combined

Table 3: Estimated area under a ROC curve for linear discriminant applied to the features derived from an SVD, K Distribution Parameters, and FFT Coefficients Using a Range of Contrast and Clutter Ratios

7 Acknowledgements

The authors gratefully acknowledge the Defence Science and Technology Organisation for use of the Ingara SAR image data for this work completed under the contract JP129/Analysts' Detection Support System from DSTO.

References

[1] Irving W.W., Willsky A.S. and L.M. Novak (1995), *A multiresolution approach to discriminating targets from clutter in SAR imagery*, in Proceedings of the SPIE, pp. 272-299.

[2] Subotic N.S. et al (1997), *Multiresolution Detection of Coherent Radar Targets*, IEEE Transactions on Image Processing, Vol. 6, No. 1, pp. 21-34.

[3] Howard D.L. (1997), *Multiscale Autoregressive Models Discriminate Radar Images of Vehicle-Sized Objects from Natural Background*, Proceedings of the International Workshop on Image Analysis and Information Fusion, Adelaide, November 6-8, 1997, pp. 207-215.

[4] Schroeder, J.E., and Howard, D., *Multiscale Modeling in Complex Synthetic Aperture Radar Imagery for Manmade Object Detection*, Asilomar'98, November, 1998.

[5] V. Bhatnagar, A. K. Shaw, and R. W. Williams, *Improved Automatic Target Recognition Using Singular Value Decomposition*, Proc. ICASSP'98, pp. 2717-2720, Seattle, WA, May, 1998.

[6] S. Ghosal and D. McKee, *Target detection in foveal ATR systems*, Proceedings of CVPR, 1996.

[7] W. Lie, *An efficient threshold-evaluation algorithm for image segmentation based on spatial graylevel co-occurrences*, Signal Processing, Vol.33 p121-126, 1993.

[8] Redding, N.J. *Estimating the Parameters of the K Distribution in the Intensity Domain*, Defence Science and Technology Organisation Draft Report, 1998.

[9] N. J. Redding, *Design of the Analysts' Detection Support System for Broad Area Aerial Surveillance*, Defence Science and Technology Organisation Technical Report, DSTO-TR-0746, December, 1998

[10] A. W. Rihaczek, and S. J. Hershkowitz, *Radar Resolution and Complex-Image Analysis*, Artec House, 1998.

[11] S. Rong and B. Bhanu, *Modeling clutter and context for target detection in infrared images*, Proceedings of CVPR, 1996.

Turbo Channel Estimation and Equalization for Mobile Data Communications

Linda Davis[†] and Iain Collings[‡]

[†] *Dept. Electrical & Electronic Engineering*
University of Adelaide, AUSTRALIA
[‡] *School of Electrical & Information Engineering*
University of Sydney, AUSTRALIA

E-mail: linda@eleceng.adelaide.edu.au

In this paper, we consider turbo equalization for fast-fading frequency-selective channels when perfect channel state information is not available at the receiver. The receiver performs channel estimation/equalization and decoding in an iterative fashion using soft-decisions. Simulation studies of a turbo equalizer based on our recently proposed single stage joint channel estimation and equalization algorithm demonstrate the impressive performance of turbo equalization.

Key Words: communication systems, equalization, MAP estimation, iterative decoding

1. INTRODUCTION

Transmission channels of modern high data rate, high mobilility, digital communication systems often exhibit both fast-fading and intersymbol interference (ISI) characteristics. As a consequence, equalization can sometimes seem an overwhelmingly difficult task. Nevertheless, new turbo equalizers have been developed, and have proved to be very effective in compensating for both the fast-fading and the ISI effects in these channels.

Turbo processing originally arose in the context of channel coding for error correction [2] in additive white Gaussian noise (AWGN) channels. The term *turbo* refers to the fact that the data is processed multiple times in a feedback arrangement before a final digital decision is made. Recently, turbo processing has been demonstrated in the context of equalization for coded data transmitted over fast-fading ISI channels [3, 4, 5]. In each of these works, perfect channel state information has been assumed to be available to the equalizer.

In the absence of perfect channel state information, traditional (non-turbo) adaptive equalizers use least mean squares (LMS) or recursive least squares (RLS) algorithms to directly update linear FIR equalizer coefficients [6, 7]. Explicit estimation of the channel is avoided, but these equalizers are unable to adequately compensate for the fast-fading environment.

An alternative approach is to separately estimate the channel, and provide the estimates on-line to an equalizer. Channel estimators which learn and exploit the statistics of the channel with Kalman filtering and prediction have been demonstrated to have improved performance in the fast-fading frequency-selective environment [8, 9, 10]. For flat-fading channels (i.e. not frequency-selective, no ISI), this approach may be realized using pilot symbols and interpolation [11].

Recently, a more integrated approach to channel estimation and equalization has evolved, known as per-survivor processing (PSP) [12]. Based on maximum likelihood sequence estimation (MLSE), channel estimates are formed using Kalman filters along each of the surviving hypothesis paths in the MLSE trellis. Thus there is a different channel estimate for each state in the trellis, and the result is *joint* channel and data estimation. PSP has been demonstrated to outperform conventional techniques in fast-fading ISI channels [12], and reduced complexity algorithms have been proposed [13, 14, 15, 16]. Unfortunately, the PSP approach is not suited to the turbo structure because MLSE entails hard decisions.

The maximum *a posteriori* (MAP) algorithm [17, 18] has received much attention as an alternative to MLSE for equalization. This is because it provides optimal soft decisions in contrast to the hard decision MLSE approach. When the *a posteriori* probabilities are retained as soft decisions, the algorithm is often referred to as the APP algorithm [19]. Unlike MLSE, there is no concept of a surviving path to each state, and therefore where joint equalization and channel estimation is to be achieved, channel estimates must be based only on the trellis state. Our recent work [1, 20] provides a generalized framework for achieving this by expanding the trellis state-space. In flat-fading channels, the algorithm reduces to that described in [21, 22].

This paper focusses on incorporating our APP equalizer [1, 20] (with joint channel estimation) into a turbo processing structure. The receiver performs channel estimation/equalization and decoding in an iterative fashion using soft decisions. Simulation studies of a turbo equalizer based on our recently proposed single stage joint channel estimation and equalization algorithm demonstrate the impressive performance of turbo equalization.

2. TURBO EQUALIZATION

In this section we consider the motivation for turbo equalization and review turbo processing principles in the context of *equalization* for coded data transmitted over fast-fading ISI channels. We then describe a turbo equalizer configuration incorporating our APP equalizer (with joint channel and data estimation) as well as an alternative using a single channel estimate.

2.1. The Optimal Receiver

It is well known that the optimal receiver for any communication system takes all available information (on the channel, modulation format, coding, *a priori* knowledge of data source statistics, for example) into account in finding a joint estimate of all the unknown parameters, which includes the transmitted data. Even if it were possible to design such a complicated receiver, the computational burden would more than likely be prohibitive. Furthermore, modularity is important in any practical system. Thus, from both design and implementation considerations,

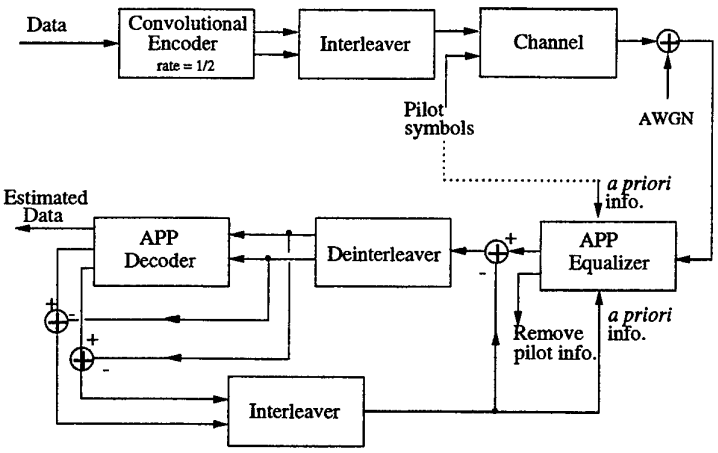


FIG. 1. Transmission system with turbo equalization.

it is desirable to separate the receiver into a cascade of processing subsystems (or stages).

From an information theoretic point of view, hard decisions at the output of any processing stage result in a loss of information. Soft decisions reflect confidence in the decision, and this information may be further utilized by subsequent processing stages. However, even if soft decisions are passed between all stages, the overall processing may be far from optimal. This is because later stages benefit from information derived at earlier stages, but not vice versa. This has motivated iterative (i.e. turbo) processing at the receiver where soft decisions from later stages (e.g. decoding) are fed back as *a priori* information to earlier stages (e.g. equalization) to refine decisions.

2.2. Turbo Equalization

For illustration of turbo equalization in this paper, we consider the simplest configuration with two stages, i.e. an equalization stage and a decoding stage. The concepts are easily extended for turbo equalization with further stages (corresponding to outer coding or turbo coding, for example).

Fig. 1 shows the complete transmission system. The information available to the receiver consists of the channel output (in the form of a set of sufficient statistics from the matched filter at the receiver front-end), and any *a priori* information on the transmitted symbols (e.g. pilot symbols, zero-tailing).

The equalizer (first stage) takes the channel output and forms soft estimates of the transmitted symbols. The decoder (second stage) then uses these soft estimates and knowledge of the coding algorithm to form soft estimates of the original data, and in doing so, forms a refined estimate of the transmitted symbols. These estimates are then passed back to the equalizer as *a priori* transition probabilities for the next iteration of equalization and decoding on the same data block. This feedback (iteration) process repeats, either until the data estimates converge, or until a processing delay limit is reached.

Important in the implementation of turbo processing is the concept of extrinsic information for each stage. This is the information about the data estimates, which is generated from the processing at that particular stage. In subsequent turbo iterations, this extrinsic information must not be included in the input to that same stage of processing (hence the two subtractions of information at each iteration in Fig. 1).

Formally, the information about the n-th transmitted symbol at the output of stage d may be written in terms of log-likelihood ratios as

$$L_{d,n} = L^a_{d,n} + L^e_{d,n} \tag{1}$$

where $L^a_{d,n}$ is the *a priori* information (consisting of information available to the receiver and extrinsic information contributed by other stages of processing), and $L^e_{d,n}$ is the extrinsic information for stage d.

The relationship (1) holds only if the soft decisions input to the stage (represented by $L^a_{d,n}$) are independent over time n. For the case of a channel with memory (due to ISI or correlation in the fading), this will not be true as information at the output of the equalizer will be correlated. Further, by the very nature of coding, the refined estimates of the transmitted symbols will also be correlated. Fortunately, these correlations dominate symbols which are closely located, and therefore may be overcome by interleaving the data between coding and transmission through the channel as shown in Fig. 1.

2.3. Joint Channel Estimation & Equalization

In the absence of perfect channel state information at the receiver, the APP equalizer forming the first stage of the turbo configuration (as shown in Fig. 1) can be implemented in one of two ways. Firstly, a separate channel estimator may be used to provide channel state information to the APP algorithm. Secondly, a more sophisticated joint channel estimation and equalization may be achieved using an APP equalizer with expanded state-space as described in [1]. The latter configuration is the focus of this paper.

The operation of the APP algorithm may be represented by a state trellis [7]. It accepts *a priori* symbol probabilities (soft inputs) and produces *a posteriori* symbol probabilities via the forward-backward recursions [19]. At each time n, and for each state transition ij in the trellis, the following expression needs to be evaluated:

$$b_{n,ij} = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\{-\frac{1}{2\sigma^2}|z_n - \hat{z}_{n,ij}|^2\} \tag{2}$$

where z_n is the output of the receiver matched filter, σ^2 is the observation noise variance, and

$$\hat{z}_{n,ij} = \sum_{\ell=0}^{L-1} \tilde{x}_{n-\ell,ij}\, f_{n,\ell} \tag{3}$$

where L is the length of the channel, $\{\tilde{x}_{n,ij}\}$ are the data hypothesis values represented by the transition ij for which the calculation is being made, and $f_{n,\ell}$ is the time-varying channel impulse response.

If the channel, $f_{n,\ell}$, is known exactly, or estimates, $\hat{f}_{n,\ell}$, are provided to (3) by a separate channel estimation algorithm, then the state trellis for the channel

has $Q^{(L-1)}$ states, where Q is the size of the modulation symbol set (e.g. $Q = 8$ for 8-PSK), and L is the number of taps in the FIR channel model (i.e. channel memory). The key to achieving joint channel estimation and equalization with a MAP receiver is to expand the state space of the trellis to $Q^{(P+L-1)}$ states, where P is the number of additional hypotheses. This expansion ensures that there is enough extra memory in the trellis to estimate the channel $\hat{f}_{n,\ell,i}$ for each trellis state i in (3). Minimum mean square error (MMSE) estimators can then be used with these additional hypotheses to provide a different channel estimate for each state in the trellis [1]. The channel estimate takes the form:

$$\hat{f}_{n,\ell,i} = \sum_{p=1}^{P} h_{\ell,i,p}\, z_{n-p} \tag{4}$$

where $h_{\ell,i,p}$ are the MMSE channel coefficients corresponding to the hypothesis of state i.

Since channel estimation is performed as part of the receiver signal processing, the receiver front-end is non-coherent. When phase-shift keying (PSK) is used for the transmitted symbols, pilot symbols are required to resolve the phase ambiguity within the equalizer. These pilot symbols are handled seamlessly by the APP algorithm. However, they incur a penalty in SNR and reduce the effective bandwidth used to transmit the data. (Note that the expanded state equalizer [1] is quite general in that it is not restricted to PSK modulation, and can be modified for differentially encoded transmission [20]).

3. SIMULATIONS

The performance of a turbo equalizer incorporating our APP equalizer (with joint channel and data estimation) is demonstrated here by simulation of a binary PSK system. At the transmitter, the message is convolutionally encoded with rate 1/2 (octal generators (133, 171), code memory 6). Following interleaving, pilot symbols are added (1 : 8) to the symbol sequence transmitted over a fast-fading channel. Here we consider both a flat, and a two-path (symbol period T spaced) ISI fast-fading Rayleigh channel with Doppler spreading described by Jakes' spectrum. The normalized fading rate, $f_d T = 0.5$, in both cases. The APP equalizer is symbol spaced, and uses $P = 3$ for channel prediction of the flat channel, whilst in the two-path case, a larger prediction order (here $P = 6$) is necessary to adequately estimate the ISI. In both simulations, up to five turbo iterations are carried out by the receiver. Fig. 2 shows the impressive BER performance gains achievable with turbo iterations (numbered 1 to 5). For reference, the BER performance of a non-iterative equalizer/decoder with perfect channel state information (CSI) is also shown. Missing data points are associated with low BER, and are the subject of ongoing (longer) simulations.

REFERENCES

1. L. M. Davis, I. B. Collings, and P. Hoeher, "Joint MAP equalization and channel estimation for frequency-selective fast-fading channels," in *Proc. of IEEE Globecom'98 Communication Theory Mini-Conference*, (Sydney, Australia), pp. 53–58, Nov. 1998.

2. C. Berrou, A. Glavieux, and P. Thitimajshima, "Near Shannon limit error-correcting coding and decoding : Turbo-codes," in *Proc. of IEEE Int. Conf. on Communications (ICC'93)*, (Geneva, Switzerland), pp. 1064–1070, May 1993.

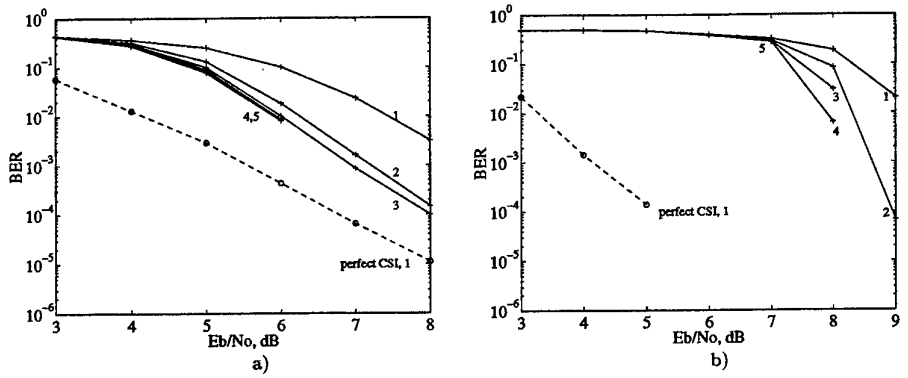


FIG. 2. Performance of turbo equalizer in a) flat (no ISI), and; b) 2 path T spaced (ISI) fast-fading ($f_d T = 0.05$) channels

3. C. Douillard, M. Jezequel, C. Berrou, A. Picart, P. Didier, and A. Glavieux, "Iterative correction of intersymbol interference: Turbo-equalization," *European Trans. on Telecommunications (ETT)*, vol. 6, pp. 507–511, September/October 1995.
4. G. Bauch, H. Khorram, and J. Hagenauer, "Iterative equalization and decoding in mobile communication systems," in *Proc. of European Personal Mobile Communications Conference (EPMCC'97)*, (Bonn, Germany), pp. 307–312, Sept./Oct. 1997.
5. D. Raphaeli and Y. Zarai, "Combined turbo equalization and turbo decoding," in *Proc. of Int. Symp. on Turbo Codes and Related Topics, (Brest, France)*, pp. 180–183, Sept. 1997.
6. S. U. H. Qureshi, "Adaptive equalization," *Proceedings of IEEE*, vol. 73, pp. 1349–1387, Sept. 1985.
7. J. Proakis, *Digital Communications*. McGraw Hill, 3rd ed., 1995.
8. R. A. Iltis, "Joint estimation for PN code delay and multipath using the extended Kalman filter," *IEEE Trans. on Communications*, vol. 38, pp. 1677–1685, Oct. 1990.
9. M. K. Tsatsanis, G. B. Giannakis, and G. Zhou, "Estimation and equalization of fading channels with random coefficients," *Signal Processing*, vol. 53, pp. 211–229, Sept. 1996.
10. L. M. Davis, I. B. Collings, and R. J. Evans, "Coupled estimators for equalization of fast-fading mobile channels," *IEEE Trans. on Communications*, vol. 46, pp. 1262–1265, Oct. 1998.
11. J. K. Cavers, "An analysis of pilot symbol assisted modulation for Rayleigh fading channels," *IEEE Trans. on Vehicular Technology*, vol. 40, pp. 686–693, Nov. 1991.
12. R. Raheli, A. Polydoros, and C.-K. Tzou, "Per-survivor processing: A general approach to MLSE in uncertain environments," *IEEE Trans. on Communications*, vol. 43, pp. 354–364, Feb 1995.
13. J. H. Lodge and M. L. Moher, "Maximum likelihood sequence estimation of CPM signals transmitted over Rayleigh flat-fading channels," *IEEE Trans. on Communications*, vol. 38, pp. 787–794, June 1990.
14. Q. Dai and E. Shwedyk, "Detection of bandlimited signals over frequency selective Rayleigh fading channels," *IEEE Trans. on Communications*, vol. 42, pp. 941–950, Apr. 1994.
15. X. Yu and S. Pasupathy, "Innovations-based MLSE for Rayleigh fading channels," *IEEE Trans. on Communications*, vol. 43, pp. 1534–1544, February/March/April 1995.
16. M. E. Rollins and S. J. Simmons, "Simplified per-survivor Kalman processing in fast-fading frequency selective channels," *IEEE Trans. on Communications*, vol. 45, pp. 544–553, May 1997.
17. R. W. Chang and J. C. Hancock, "On receiver structures for channels having memory," *IEEE Trans. on Information Theory*, vol. 12, pp. 463–468, Oct. 1966.
18. K. Abend and B. D. Fritchman, "Statistical detection for communication channels with intersymbol interference," *Proceedings of IEEE*, vol. 58, pp. 779–785, May 1970.
19. L. R. Bahl, J. Cocke, F. Jelenik, and J. Raviv, "Optimal decoding of linear codes for minimizing symbol error rate," *IEEE Trans. on Information Theory*, vol. 20, pp. 284–287, Mar. 1974.
20. L. M. Davis and I. B. Collings, "On the benefits of pilot-aided MAP receivers over differential PSK," in *Proc. of IEEE Wireless Communications and Networking Conference (WCNC'99)*, (New Orleans, LA, USA), Sept. 1999.
21. M. J. Gertsman and J. H. Lodge, "Symbol-by-symbol MAP demodulation of CPM and PSK signals on Rayleigh flat-fading channels," *IEEE Trans. on Communications*, vol. 45, pp. 788–799, July 1997.
22. P. Hoeher and J. Lodge, "Iterative decoding / demodulation for coded DPSK systems," in *Proc. of IEEE Globecom'98*, vol. 1, (Sydney, Australia), pp. 598–603, Nov. 1998.

IDEA — A Signals Analysis Package for ELINT Processing and Extraction

Stephen D. Elton and Jennifer A.L. Newman

Electronic Warfare Division, Defence Science and Technology Organisation, Department of Defence, P.O. Box 1500, Salisbury, SA 5108, Australia

The purpose of this article is to review the functionality of a radar signals analysis package that has been developed for the processing and extraction of ELINT information. At the same time we highlight some of the technical challenges that face a radar intercept receiver in collecting and analysing the signals that make up the modern radar signal environment.

Key Words: IDEA, ELINT, radar signal analysis, pulse trains, deinterleaving.

1. INTRODUCTION

Electronic intelligence, or ELINT, involves the collection of signals intelligence through the interception of non-communications electromagnetic emissions. Signals of interest may originate from a variety of man-made electromagnetic sources including radars, transponders, data links and identification friend or foe (IFF) [9]. However, our interest here will be limited to radar-based ELINT. Because of the wideband nature of the modern ELINT receiver, a radar signal of interest (SOI) will often be recorded in the presence of other, possibly many, interference signals and the SOI must be isolated through *deinterleaving* processing before detailed parametric information can be extracted from the signal. In the context of radar signal analysis, deinterleaving refers to the process of isolating signals from a time interleaved record of pulsed radar emissions. In addition to basic intelligence gathering which involves collecting, analysing and locating the sources of the intercepted radar emissions, strategic ELINT operations are used to build radar emitter libraries or data bases. These libraries are then employed in tactical missions by a radar intercept receiver for emitter identification purposes, including threat warning.

In this article we provide an overview of the functionality of a signals analysis package dedicated to ELINT processing and extraction and in so doing, highlight some of the technical challenges that face a radar intercept receiver in processing the signals associated with the modern radar signal environment. IDEA, the Interactive Deinterleaver for ELINT Analysis is a software package for laboratory use that incorporates a number of tools that would be familiar to an experienced ELINT analyst, but also includes several novel signal processing methods that have been developed by researchers at the Defence Science and Technology Organisation.

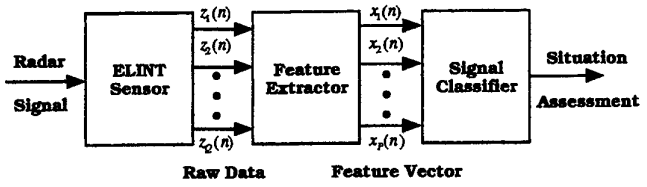


FIG. 1. A generic ELINT signal classification system.

2. ELINT PROCESSING AND EXTRACTION

In the case of pulsed radar emissions, ELINT processing and extraction is generally carried out by analysing a series of digital *pulse descriptor words* (PDWs) that is created by the radar intercept receiver. A PDW may be viewed as a *feature vector* that characterises each pulse recorded by the receiver in a compact fashion. A generic ELINT system that illustrates this concept is shown in Fig. 1. Typical signal descriptors include pulse angle of arrival (AOA), radio frequency (RF), pulse width (PW), pulse amplitude (PA), and pulse time of arrival (TOA), together with *intra-pulse* features such as within-pulse phase and frequency modulation. From these parameters one can deduce the pulse repetition interval (PRI) of a signal which corresponds to the time difference of arrival (TDOA) between two successive pulses, as well as the scan period between consecutive radar sweeps.

Pulse AOA is an important signal discriminator for isolating a SOI since it cannot be influenced by a radar emitter in a deceptive manner. Other parameters, such as RF and PRI, can be and frequently are modified on a pulse-by-pulse basis by a modern radar source, either as a means of providing radar emitter performance enhancement, or as a method of defeating the intercept receiver. Significantly, pulse-by-pulse signal agility has the potential to confuse an intercept receiver and may lead to signal *fragmentation*. This can be a particularly serious problem in automated processing systems and may result in multiple radar emitters being reported by an intercept receiver, rather than a single emitter with parameter agility. A major challenge to radar signal intercept analysis is to add an additional layer of *information* processing to the analysis sequence to correct for signal fragmentation.

With the advent of precision digital receiver technology, electronic *fingerprinting* can, in principle, provide a means of assisting with the deinterleaving process and involves the extraction of a unique signature from a radar signal. However, signal-to-noise-ratio considerations do come into play with these methods and may inhibit the performance of some of the techniques currently under investigation [5].

3. THE IDEA SIGNALS ANALYSIS PACKAGE

In this section we provide an overview of the functionality of IDEA and concentrate on recent enhancements to the package. An earlier review appeared in [2].

3.1. Background

IDEA is a computer-based signals analysis package written in C++ and designed to run on Unix workstations. As IDEA's name would suggest, there is a strong emphasis on *interactivity* by the human operator and this is intrinsic to IDEA's design philosophy. The interactivity is realised through a range of visualisation and

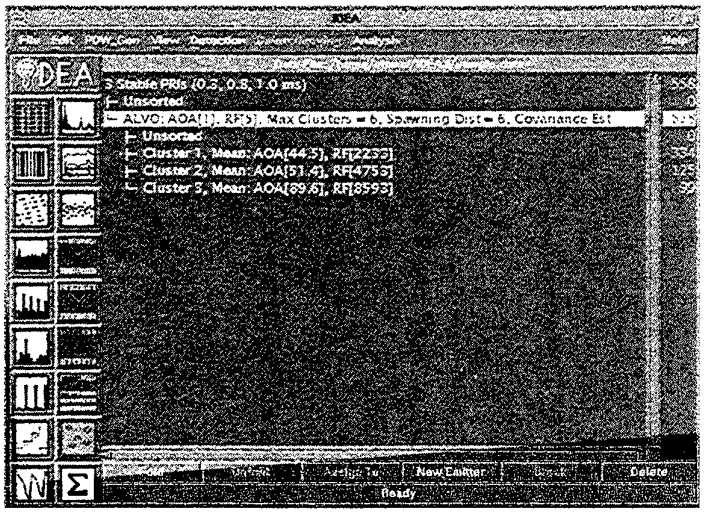


FIG. 2. The main processing window of IDEA.

analysis tools that are made available to the operator via a graphical user interface. A number of data viewers and signal detector windows also update automatically as a new tool is invoked, but retain the previous view (plotted in a different colour) so that the effect of the new tool can be assessed visually. In addition to the signal analysis tools, IDEA contains an in-built signal generation capability, allowing the user to generate a range of synthetic time interleaved signals of varying complexity.

3.2. IDEA's Tree Structure

The main processing window of IDEA is shown in Fig. 2. In this example a data file has been loaded and a Deinterleaving tool employed to process the PDWs created for the signals that make up the data record. The hierarchical *tree structure* shown in Fig. 2 has been built up through analysis and represents a unique and powerful approach to deinterleaving processing that is one of IDEA's main strengths. Each node of the tree is labelled according to the analysis technique that was used to create it, including input parameters that are employed by an analysis tool. Once a SOI or portion of a SOI has been isolated through processing and a node created for it, the pulses of that node may be operated on further by a number of other tools; the intent being to reveal additional parametric information on the signal. Nodes may also be grouped together within the IDEA tree structure and this is a particularly useful facility when looking for evidence of signal fragmentation.

3.3. Processing Tools

3.3.1. Data Viewers

The data View tools make up the first set of tools that are likely to be used by an ELINT analyst and include: As Text viewers, TOA Raster plots, parameter Histograms, Parameter vs. Time plots, and Parameter vs. Parameter viewers. The Histograms may be used to help isolate a SOI by setting up a Parametric Filter within IDEA, which is essentially a parameter-based band-pass filter. The Parameter vs. Parameter viewers on the other hand, could be used to cue a cluster-based

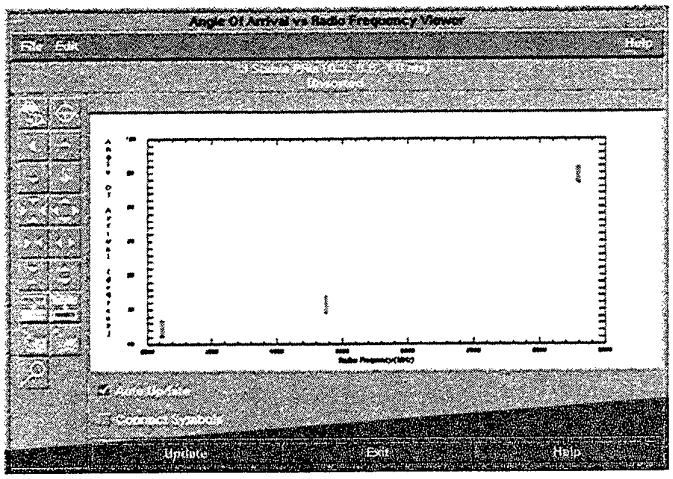


FIG. 3. AOA vs. RF data viewer.

deinterleaving tool. In Fig. 3 we show an AOA vs. RF plot which clearly illustrates the presence of three radar sources and suggests that the Cluster Analysis tool, also within IDEA, could readily deinterleave the three signals. The results of cluster analysis for this particular data set are shown in the tree structure of IDEA's main processing window in Fig. 2, where the nth feature vector, $\boldsymbol{x}(n)$, for cluster processing has been defined in terms of the RF and AOA measurements via

$$\boldsymbol{x}(n) = [\mathrm{rf}(n)\ \mathrm{aoa}(n)]'. \tag{1}$$

3.3.2. Signal Detectors

The signal Detection tools provide an alternative means of recognising the presence of a particular type of signal embedded in an interleaved data record and include TDOA histograms of various forms. The TDOA histograms may be viewed as autocorrelation devices and reveal information on the PRIs of interleaved signals, e.g., see [9]. This information could then be used to prime a pulse TOA-based *sequence search* algorithm. The Primed Sequence Search tool within IDEA is another method of isolating a SOI and searches for repetitive patterns in the arrival time sequence recorded for a signal [6].

3.3.3. Deinterleaving Tools

The Deinterleaving tools supported by IDEA include Parametric Filters and a Primed Sequence Search algorithm as previously stated. The tools also include an Auto Sequence Search technique [7], the essence of which is arguably the workhorse deinterleaving method for a large number of fielded radar intercept receivers.

Cluster Analysis makes up the final Deinterleaving tool. In the context of ELINT processing and extraction, cluster analysis seeks to assign a sequence of N feature vectors $\{\boldsymbol{x}(0), \boldsymbol{x}(1), \ldots, \boldsymbol{x}(N-1)\}$ derived from a buffer of N radar pulses, where $\boldsymbol{x}(n) \in \mathbb{R}^P$, to a finite number of source classes by searching for structure in a multi-dimensional feature space. IDEA contains an *unsupervised* neural network algorithm for cluster analysis that is also self-evolving in that the network adapts

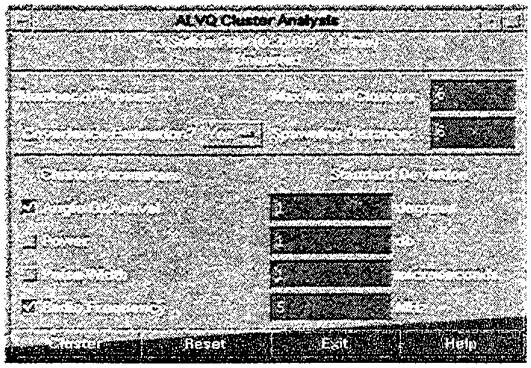


FIG. 4. The IDEA cluster analysis dialogue box.

itself to the number of radar emitters or emitter modes detected in a data buffer. An early version of the method has been reported in [3] and the dialogue box for IDEA's current implementation of the algorithm is shown in Fig. 4.

Despite the strength of cluster analysis as a deinterleaving tool, it can be vulnerable to emitter agility. Consider, for example, a radar source that hops between three different RFs while its other parameters remain constant. A cluster-based deinterleaving technique might report the intercepted signal as three separate emitters. One possible solution to help remedy this situation is to re-group those pulses from the three clusters as a trial merge within IDEA and compute a TDOA histogram for the pooled pulses. The resulting TDOA histogram could then be compared with the TDOA histogram computed for each of the individual clusters prior to merging the data. If the TDOA histogram for the pooled pulses is more "ordered" compared to that obtained for each of the non-pooled pulse TDOA histograms, the merge should be allowed and the signals re-associated. This type of processing is essentially a visual implementation of the entropy-based scheme outlined in [1].

3.3.4. Signal Analysers

The IDEA Analysis tools operate on a SOI once it has been isolated from other signals. The tools currently include a histogram-based method for revealing the clock period that is used to generate the PRI sequence of a discrete jittered signal [9] and recursive estimators for the mean PRI and TOA phase parameters of a signal. To illustrate the latter, consider the following model-based approach to pulse TOA analysis in which the nth pulse TOA, t_n, for a signal is modelled as

$$t_n = t_\phi + nT + \alpha_n, \quad n = 0, 1, \ldots, N-1, \tag{2}$$

where t_ϕ and T are constant parameters denoting the TOA phase and mean PRI respectively, and α_n is a Gaussian jitter random variable with distribution $\mathcal{N}(0, \sigma_\alpha^2)$. Least squares estimators for t_ϕ and T have been reported in [4] and implemented in IDEA. Results for the least squares estimator of T are shown in Fig. 5.

4. CONCLUDING REMARKS

The interactive nature of IDEA emphasises the important role that a human operator can play in the laboratory-based analysis of ELINT data. It also highlights

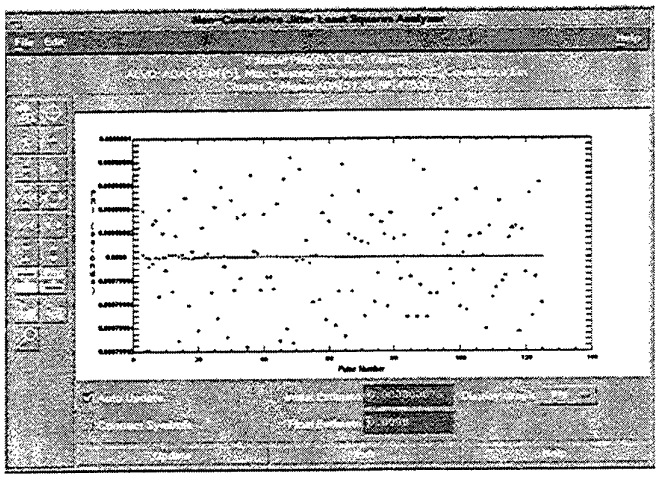


FIG. 5. Recursive least squares estimation of the mean PRI, T, of a signal.

the types of skills that can be brought to bear by an experienced analyst, particularly as they relate to visual pattern recognition and a synergistic approach to ELINT problem solving. To translate these skills into an autonomous system where signal intercept and analysis is carried out automatically, e.g., see [8], represents one of the major technical challenges facing the ELINT community today.

ACKNOWLEDGMENT

The authors would like to acknowledge the significant contributions made by a number of colleagues over the years in the development of the IDEA signals analysis package.

REFERENCES

1. Agg, D.A., "The Development of an ESM Data Processing Scheme Based on Pattern Recognition Techniques," *IEE Colloquium on "Electronic Warfare Systems,"* (Digest No. 9), pp. 5/1–4, 1991.
2. Balin, M., "IDEA - Interactive Deinterleaver for ELINT Analysis," *Proceedings of the International Conference on Signal Processing Applications and Technology,* Vol. 1, pp. 852–856, 1996.
3. Elton, S.D., "An Adaptive Learning Vector Quantisation Algorithm for Cluster Analysis and Radar Pulse Train Deinterleaving," *Australian Journal of Intelligent Information Processing Systems,* Vol. 5, No. 1, pp. 29–33, 1998.
4. Gray, D.A., Slocumb, B.J. and Elton, S.D., "Parameter Estimation for Periodic Discrete Event Processes," *Proceedings of the 1994 International Conference on Acoustics, Speech and Signal Processing,* Vol. 4, pp. 93-96, 1994.
5. Herskovitz, D., "The Other SIGINT/ELINT," *Journal of Electronic Defense,* Vol. 19, No. 4, April, pp. 35–41, 1996.
6. Mardia, H.K., "New Techniques for the Deinterleaving of Repetitive Sequences," *IEE Proceedings F. Radar and Signal Processing,* Vol. 136, No. 4, August, pp. 149–154, 1989.
7. Noone, G. and Balin, M., *An Improved Sequence Search Deinterleaving Algorithm,* unpublished internal report, 1994.
8. Sciortino, Jr., J.C., "Autonomous ESM Systems," *Naval Engineers Journal,* November, pp. 73–84, 1997.
9. Wiley, R.G., *Electronic Intelligence: The Analysis of Radar Signals, Second Edition,* Artech House: Norwood, MA, 1993.

Wigner-Ville analysis of HF radar measurements of a surrogate theatre ballistic missile

Gordon J. Frazer and Stuart J. Anderson
Surveillance Systems Division, DSTO, PO Box 1500, Salisbury, SA 5108, AUSTRALIA

The paper describes results obtained by applying the Wigner-Ville distribution to high frequency line-of-sight radar measurements of a surrogate theatre ballistic missile during powered flight. The Wigner-Ville distribution is able to discriminate between target and interfering transient events and to accurately determine the instantaneous Doppler law of the target. It is a useful pre-processing step for determining the instantaneous received signal level and the coherent processing loss due to target acceleration.

Key Words: time-frequency, high frequency radar, theatre ballistic missile

1. INTRODUCTION

In this paper we describe the results obtained by applying the Wigner-Ville distribution (WVD) to measurements obtained from a high frequency (HF) line-of-sight radar. The measurements were of an accelerating surrogate theatre ballistic missile (TBM) target during powered flight immediately following launch. We have shown that the WVD is able to discriminate between the accelerating target and a interfering transient event which produces a similar Doppler spectrum to the target and hence a potentially confusing display at the radar output. The WVD is also able to accurately determine the instantaneous Doppler law of the target, at a temporal resolution comparable with the radar sweep (or pulse) duration. Knowledge of the instantaneous Doppler law is applied to the problem of determining the coherent processing loss due to target acceleration, and to determining the instantaneous received signal level. For an example of time frequency distributions applied to HF skywave data (not of accelerating targets) see [1].

The paper is set out as follows. We provide background to the data used in section 2. Next we provide a signal model and justify the selection of the WVD as the one of many potential time-frequency distributions which may have been applied. We present our results in section 4 and provide our conclusions in section 5.

2. BACKGROUND

During September 1997 four surrogate TBMs were launched from a site in northern West Australia to test a variety of TBM launch detection sensors. One of the

sensor suite was a bistatic HF line-of-sight radar. The radar was positioned tens of kilometres from the launch site and it operated at a carrier frequency of approximately 25MHz. The radar waveform was linear frequency modulated continuous wave (LFMCW) and the repetition frequency (WRF) was 50Hz. A set of coherent measurements were collected, each of 5s duration and there was a short inter-dwell gap between each coherent measurement interval. The choice of waveform and low WRF meant that Doppler measurements of the high velocity target were ambiguous for most of the flight. The long coherent integration time (CIT) increased radar sensitivity although the target acceleration decreased the coherent processing gain achieved and limited the accuracy of velocity measurements.

3. SELECTION OF THE WIGNER-VILLE DISTRIBUTION

The data consisted of complex time series which correspond to sequences of coherent measurements from the particular azimuth range cells which contained the target. Preliminary analysis suggested that the following signal model would be appropriate.

$$z(t) = A\,\mathrm{e}^{\jmath 2\pi[f_0 t + \frac{\beta}{2}t^2]} + c(t) + n(t) \tag{1}$$

for $\{t : 0 < t < T\}$ where A is the complex amplitude, f_0, β are the linear FM parameters, $c(t)$ represents clutter, $n(t)$ represents noise and T is the coherent integration time (CIT) of the radar. In the case of a bistatic HF line of sight radar, $c(t)$ includes contributions such as; the direct signal from the transmitter, range sidelobes from the direct signal, additional targets, say from a booster stage in a multi-stage rocket and meteor and ionospheric scatter. In general, both the clutter and noise are unknown, although we assume the relative energy is such that

$$\frac{|A|^2 T}{\left[\int_0^T |c(t)|^2\,\mathrm{d}\,t + \int_0^T |n(t)|^2\,\mathrm{d}\,t\right]} \gg 1 \tag{2}$$

and thereby consider $z(t)$ as deterministic with unknown parameters, unknown clutter and background noise, and high signal to clutter plus noise energy ratio. The objective is to determine f_0 and β as part of the task of establishing the accelerating target dynamics.

Next we discuss the suitability of using the Wigner-Ville distribution (WVD) to determine f_0 and β. We consider the case of continuous time signals for convenience in the subsequent derivations, although the results extend to discrete time sequences in a straightforward manner. Complete expositions on the WVD and instantaneous frequency estimation are given in [2–4] and the references therein.

3.1. Definition of the Wigner-Ville distribution

The WVD of a deterministic signal $z(t)$ is defined as

$$\mathrm{W}_z(t, f) = \int_{-\infty}^{\infty} z(t + \frac{\tau}{2}) z^*(t - \frac{\tau}{2})\,\mathrm{e}^{-\jmath 2\pi f \tau}\,\mathrm{d}\,\tau \tag{3}$$

here $*$ denotes conjugation. $x(t)$ is a continuous time analytic signal, and

$$z(t) = x(t) + \jmath \mathbf{H}[x(t)] \tag{4}$$

for a real signal $x(t)$ and the Hilbert transform operator $\mathbf{H}$. Let $Z(f)$ be the Fourier transform of $z(t)$ and $X(f)$ the Fourier transform of $x(t)$. For $z(t)$ to be analytic then

$$Z(f) = \begin{cases} 2X(f) & f > 0 \\ X(f) & f = 0 \\ 0 & f < 0 \end{cases} \tag{5}$$

The requirement that $z(t)$ in (3) be analytic is motivated by:

1. Recognition that the quadratic form in (3) introduces interaction features, or cross terms, between individual additive components of $z(t)$. Enforcing that $z(t)$ be analytic at least eliminates cross terms generated by interaction between the symmetric positive and negative frequency components of a real signal $x(t)$.
2. Additional computational savings realisable when restricting the domain of $\mathrm{W}(t, f)$ to the half plane $f > 0$.

In many radar applications $z(t)$ is the output of quadrature demodulation in the radar receiver. The signal is complex but not necessarily zero for $f < 0$. In this case $z(t)$ can be used directly in (3) without enforcing (5) provided $\mathrm{W}(t, f)$ is evaluated over the full t, f plane.

3.2. WVD of a single linear FM signal

Consider the WVD for a single component linear FM signal. Let

$$z(t) = A\, e^{j2\pi[f_o t + \frac{\beta}{2} t^2]} \quad \text{for } t \in (-\infty, \infty) \tag{6}$$

with complex amplitude A, initial frequency f_0 and chirp rate β, then it can be shown that

$$\mathrm{W}(t, f) = |A|^2 \cdot \delta(f - [f_0 + \beta t]) \tag{7}$$

The WVD of a single component complex linear FM signal shows ideal concentration in the time frequency plane along the instantaneous frequency (IF) law of the signal. The instantaneous frequency law, $f_i(t) = f_0 + \beta t$, can be determined by

$$f_i(t) = \underset{f}{\operatorname{argmax}} \left[\mathrm{W}(t, f)\right] \quad \forall\, t \tag{8}$$

For signals which are not purely deterministic, such as that in (1), the instantaneous frequency law may be estimated using (8) to give an IF estimate $\hat{f}_i(t)$.

3.3. WVD of multi-component signals

The WVD can become uninterpretable when there are more than a modest number of component signals with comparable energy. Consider the two additive component signal

$$z(t) = Ap(t) + Bq(t) \tag{9}$$

with complex scalars A and B and the component energy constraint

$$\int_{-\infty}^{\infty} |p(t)|^2 \, \mathrm{d}\, t \approx \int_{-\infty}^{\infty} |q(t)|^2 \, \mathrm{d}\, t \tag{10}$$

The WVD of $z(t)$ may be written in terms of the signal components $p(t)$ and $q(t)$

$$\begin{aligned} \mathrm{W}_z(t,f) &= |A|^2\,\mathrm{W}_p(t,f) + AB^*\,\mathrm{W}_{pq}(t,f) \\ &\quad + BA^*\,\mathrm{W}_{qp}(t,f) + |B|^2\,\mathrm{W}_q(t,f) \end{aligned} \tag{11}$$

The WVD of the sum of the two components $p(t)$ and $q(t)$ is the sum of the individual WVD of each component W_p and W_q (auto-terms) plus the cross WVD between components W_{pq} and W_{qp}. The individual auto and cross WVD are scaled by the appropriate products of the scalars A and B.

Given the constraint (10), and assuming now that one is primarily interested in determining the IF law $f_i^p(t)$ of component $p(t)$, the relative interpretability of $\mathrm{W}_z(t,f)$ will be determined by the scalars A and B. If $A \gg B$ then

$$\mathrm{W}_z(t,f) \approx |A|^2\,\mathrm{W}_p(t,f) \tag{12}$$

and $f_i^p(t)$ will be well estimated by (8). For $A \approx B$ all terms in (11) will contribute to W_z making interpretation and estimation based on W_z difficult. With $A \ll B$

$$\begin{aligned} |A|^2\,\mathrm{W}_p(t,f) \ll AB^*\,\mathrm{W}_{pq}(t,f) + BA^*\,\mathrm{W}_{qp}(t,f) \\ + |B|^2\,\mathrm{W}_q(t,f) \end{aligned} \tag{13}$$

and W_z will be of little use in determining $f_i^p(t)$.

3.4. WVD of transient signals

Consider a single component signal which is non-zero for only a small fraction of the radar CIT. Let $s(t)$ be an arbitrary finite energy signal with $\frac{\delta t}{2} \ll T$ and define $z(t)$ as

$$z(t) = \begin{cases} s(t) & t_c - \frac{\delta t}{2} < t < t_c + \frac{\delta t}{2} \\ 0 & \text{otherwise} \end{cases} \tag{14}$$

with $\frac{\delta t}{2} < t_c < T - \frac{\delta t}{2}$. Spectrum analysis procedures which assume stationarity which are applied to $z(t)$ will not capture the transient nature of this signal. Any stationary spectrum estimate will appear broadband, of the order $\frac{1}{\delta t}$ Hz bandwidth or greater, depending on $s(t)$, and may be confused with the spectrum of signals such as that in (1).

The WVD has time and frequency marginal properties which ensure that the time and frequency support of $z(t)$ are captured in the distribution. The time marginal property $|z(t)|^2 = \int_{-\infty}^{\infty} \mathrm{W}_z(t,f)\,\mathrm{d}f$ indicates that for $z(t)$ in (14) $\mathrm{W}_z \neq 0$ only in the interval $t_c - \frac{\delta t}{2} < t < t_c + \frac{\delta t}{2}$. Hence, W_z captures the transient nature of $z(t)$.

4. RESULTS

We have applied the WVD in the following ways; to assist with discriminating between accelerating and transient targets or scatterers, to determine the target instantaneous Doppler law, and to determine the instantaneous received target energy law and the processing loss caused by the target acceleration.

4.1. Accelerating target v. transient

Figures 1 and 2 (left) show the range–Doppler (RD) maps generated for two separate beam steer directions measured during the same CIT. In figure 1 the accelerating target is visible as a large smear in Doppler at range cell 7. In figure 2 (left) the transient meteor scatterer is visible at range cell 19, which is also smeared in the Doppler domain. We seek improved discrimination between the accelerating and the transient scatterers. Figures 2 (right) and 3 (left) show the WVD computed from the time series corresponding to the mentioned range cells. The instantaneous Doppler law of the accelerating target is visible and so is the transient behaviour of the meteor scatterer. The WVD is able to discriminate clearly between these two types of scatterers.

4.2. Instantaneous Doppler law

The instantaneous Doppler law has been extracted from the peak of the WVD shown in figure 2 (right) according to (8). CITs prior to and later than this measurement interval have also been analysed using the WVD, with the instantaneous Doppler law again estimated from the peak according to (8). Appropriate smoothing of the estimates using polynomial models reduces estimate variance, allows Doppler law prediction into the inter-dwell intervals and can be easily integrated to produce phase law estimates. The sequence of unsmoothed Doppler estimates shown in figure 3 (right) covers an interval of approximately 10 CITs. Clearly, it is possible to determine accurate instantaneous Doppler sequences, with temporal resolution of approximately the sweep duration, as compared with the CIT for conventional Doppler processing (20ms v. 5.12s). Note that the target second stage motor ignited at approximately 18s and that accurate time of ignition can be determined from the instantaneous Doppler law.

4.3. Instantaneous energy law

Knowledge of the instantaneous Doppler law can also be used to construct a demodulation reference signal, $s(t)$. This signal has unity amplitude and instantaneous frequency law which is the conjugate of the estimated instantaneous Doppler law of the target, i.e. the instantaneous frequency law of $z(t)$. $s(t)$ can be used to demodulate $z(t)$ giving the approximately stationary time series $z'(t)$.

$$z'(t) = z(t) \cdot s(t) \tag{15}$$

The instantaneous energy of the demodulated time series is

$$\hat{p}(t) = \mathbf{G}|z'(t)|^2 \tag{16}$$

where $\mathbf{G}$ is some smoothing operator. The instantaneous energy is shown in figure 4 (left), which shows three different levels of local smoothing, (i.e. different $\mathbf{G}$).

4.4. Processing loss due to target acceleration

The processing loss due to target acceleration compared with a comparable target of constant velocity can be determined. One contrasts standard Doppler processing applied to the time series $z(t)$ and to the demodulated version $z'(t)$. It can be

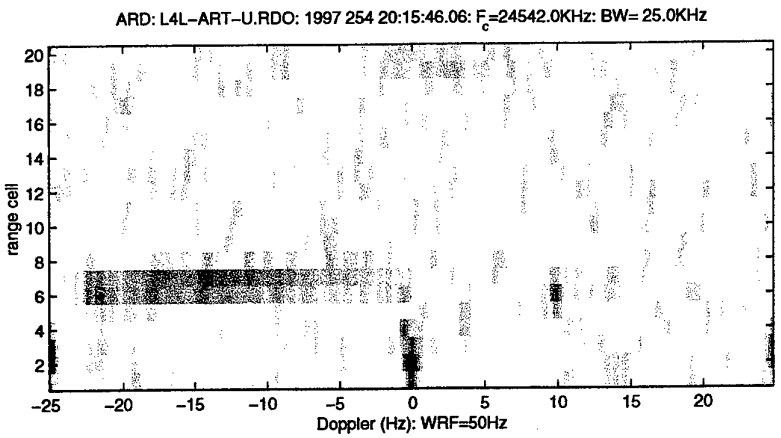


FIG. 1. Range-Doppler map showing the accelerating target smeared in Doppler in range cells 6 and 7. The direct wave from the transmitter and ground clutter is visible surrounding 0Hz Doppler and centered in range cell 2. The coasting spent first stage of the two stage TBM can be seen at range cell 6 with 10Hz Doppler.

seen from figure 4 (right) that the processing loss is approximately 10dB for this particular CIT.

5. CONCLUSIONS

The Wigner-Ville distribution has been applied to HF line-of-sight radar measurements of a TBM launch. This approach has assisted with discriminating between the accelerating target and transient meteor scatterers. It has provided estimates of the Doppler law of the target at a temporal resolution of approximately 20ms compared with standard processing which had a temporal resolution of 5.12s. It has also allowed determination of the instantaneous energy law of the target and has provided an estimate of the processing loss when standard Doppler processing is applied to a rapidly accelerating target.

REFERENCES

1. S. J. Anderson and S. E. Godfrey, "Time-frequency signatures of transient phenomena observed with HF skywave radar," in *Proceedings of the International Conference Modern Radar '94*, (Kiev, Ukraine), 1994.
2. L. Cohen, "Time-frequency distributions — A review," *Proceedings of the IEEE*, vol. 77, pp. 941–981, July 1989.
3. B. Boashash, *Methods and Applications of Time-Frequency Signal Analysis*. Melbourne, Australia: Longman Cheshire, 1991. ISBN No. 582 86873 4.
4. B. Boashash, "Interpreting and estimating the instantaneous frequency of a signal-Part I: Fundamentals," *Proceedings of the IEEE*, vol. 80, pp. 520–538, April 1992.

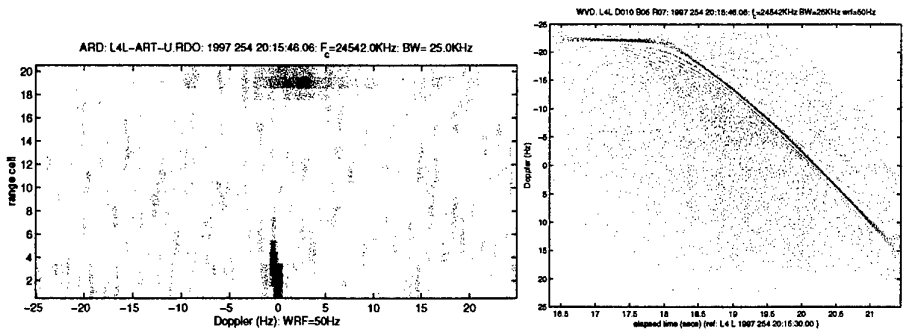


FIG. 2. (Left) Range-Doppler map showing the transient meteor scatterer in range cell 19. It is difficult to discriminate between this smeared feature and the smeared accelerating target shown in the previous figure. (Right) WVD of the time series containing the accelerating target from range cell 7 in figure 1. Second stage ignition occurred at 18s.

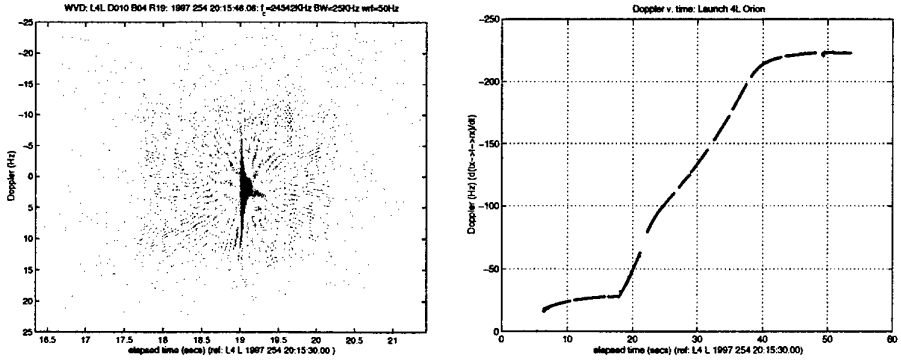


FIG. 3. (Left) WVD of the time series containing the transient meteor scatterer from range cell 19 in figure 2 (left). (Right) Doppler v. elapsed time. Computed using the WVD. The gaps are due to missing sweeps during the radar interdwell gap and some lost instantaneous Doppler detections at either end of individual radar CIT. The plot is a sequence of point measurements, one per radar sweep, and not a continuous line.

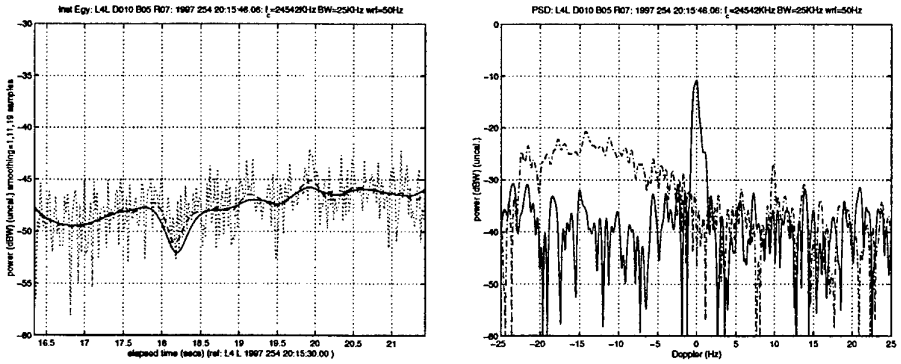


FIG. 4. (Left) Instantaneous energy law v. elapsed time for the interval 16.5s to 21.5s after launch. The three curves correspond to 1 (..), 11 (- -) and 19 (-) sample zero phase moving average smoothing. (Right) The Doppler spectrum computed over the full CIT for the original time series (-.) and for the demodulated time series (-). The processing loss caused by assuming a constant velocity target is approximately 10dB in this case.

Colored Noise Matched Filtering with Unknown Covariance

J.S. Goldstein[1], J.R. Guerci[1], S. Huang[1] and I.S. Reed[2]

[1] *SAIC, 4001 N. Fairfax Dr., Suite 600, Arlington, VA 22203-1616 USA*
[2] *University of Southern California, Los Angeles, CA 90089-2565 USA*

This paper is concerned with colored noise matched filtering when the noise covariance is unknown *a priori*. Both principal components analysis and canonical correlation analysis require knowledge of the noise covariance matrix and its inverse. Furthermore, neither of these methods yield an optimal representation of the noise subspace for rank one detection problems even when the covariance is known. The multistage Wiener filter is shown to be optimal when the noise covariance is unknown in the sense that it determines the best noise subspace representation and reduced-rank approximation as a function of rank.

Key Words: rank reduction, signal representation, signal compression.

1. INTRODUCTION

Classical detection problems in radar, sonar and communications determine the presence or absence of a target signal observed in noise. It is common to assume that all signals are independent Gaussian random processes as a starting point in the analysis. Under these conditions, the solution is found by analyzing the target (or desired) signal and the noise statistics. The target analysis is achieved through the use of prior knowledge under the hypothesis that the target is present. This step may utilize a steering vector, a matched field processor or correlation information such as a CDMA code. The noise covariance can not be assumed to be known in practical problems, and the noise statistics must be estimated. This paper is concerned primarily with the estimation of the noise statistics and the impact that it has on the detection problem.

Define an N-dimensional signal vector $\mathbf{s}$, which can be considered to be a radar steering vector, the output from a matched field processing routine for sonar or a correlation vector used in communications. Let the N-dimensional test vector $\mathbf{x}$ be the observed vector being considered for target presence or absence. If the $N \times N$ noise covariance matrix $\mathbf{R}$ is known *a priori*, then a popular constant false alarm

rate (CFAR) test is given by

$$\Lambda = \frac{|\mathbf{s}^H \mathbf{R}^{-1} \mathbf{x}|^2}{\mathbf{s}^H \mathbf{R}^{-1} \mathbf{s}} \underset{H_0}{\overset{H_1}{\gtrless}} \eta, \tag{1}$$

where H_1 denotes the target present hypothesis, H_0 is the null hypothesis and η is a threshold determined by the false alarm rate [1-3].

The numerator of the test statistic in (1) is the output power of a colored noise matched filter $\mathbf{v}^H \mathbf{z}$, where the whitened signal vector is $\mathbf{v} = \mathbf{R}^{-1/2}\mathbf{s}$ and the whitened data vector is $\mathbf{z} = \mathbf{R}^{-1/2}\mathbf{x}$. The test then compares the ratio of the colored noise matched filter's output power and the output noise power to a threshold. The expected value of the numerator and the denominator are identical when a target is absent in the known covariance case, and the mean value of the logarithm of the left-hand side of (1) is 0 dB. The mean value of this test statistic increases when the target is present as a function of the signal-to-interference plus noise ratio (SINR). Note that the colored noise matched filter output is identical to the output of a Wiener filter $\mathbf{g} = \mathbf{R}^{-1}\mathbf{s}$: $y = \mathbf{v}^H \mathbf{z} = \mathbf{g}^H \mathbf{x}$.

The estimation of the noise covariance is considered in Sect. 2, along with discussions pertaining to sample support, complexity and rank reduction. An analysis of the colored noise matched filter is then studied in Sect. 3, where the multistage Wiener filter is derived from an optimization of the noise subspace representation as a function of rank. Simulation results are also depicted in Sect. 3 to demonstrate algorithm performance. The summary is presented in Sect. 4.

2. NOISE COVARIANCE ESTIMATION

A maximum likelihood estimate is often used to replace the covariance matrix when the true covariance matrix is not known. Replacing the matrix inversion in (1) with this estimate results in a normalized form of the celebrated sample matrix inversion (SMI) algorithm [4].

Many attributes of the SMI algorithm may not be obvious to the casual observer. For example, the number of samples required for the statistics to converge in an SINR sense is at least $2N$ independent and identically distributed (*iid*) snapshots. Also the computational complexity grows as $\mathrm{O}(N^3)$ due to the matrix inversion requirement in the CFAR test, where $\mathrm{O}(\cdot)$ denotes the highest order of the underlying polynomial and at least $2N$ samples are used in the covariance estimation.

The detection of small signals in competing noise requires that the signal space be enlarged to obtain a subspace where the signal and noise can be separated with sufficient resolution. This fact has led to adaptive processing in two and more dimensions using, for example, spatial, temporal, Doppler and polarization degrees of freedom. An impact of increasing both the number of dimensions in the signal space (for signal discrimination) and increasing the bandwidth of each dimension in that space (for resolution) is that the parameter N quickly becomes very large.

The critical problem which occurs as N increases is satisfying the requirement for at least $2N$ *iid* samples for covariance estimation. This is particularly difficult in the radar and sonar problems, where the samples are taken from disparate spatial locations. In other words, the training region for estimating the noise statistics is

in the range domain. Thus the sensor must estimate the noise statistics in the test cell using data which is increasingly distant as the parameter N increases.

Rank reduction is a method to potentially reduce the sample support and computational complexity requirements of these large dimensional problems. The goal here is to find a low rank subspace that can accurately represent the noise present in the test cell. If the rank is reduced from N to $M < N$, then the sample support requirement and the computational complexity can both be reduced accordingly. Some approaches have been proposed to solve this problem from a statistical framework, however they really presuppose the estimation of the noise covariance matrix.

The principal components inverse (PCI) and eigencanceler algorithms utilize a low rank estimate formed by those eigenvectors of the SMI covariance matrix which correspond with the largest eigenvalues [5-7]. The cross-spectral metric (CSM) uses a low rank estimate composed of those eigenvectors of the SMI covariance matrix which actually maximize the SINR (this solution has been shown to be generally different than that obtained by PCI and the eigencanceler) [8-11]. One key attribute of the CSM method is that it solves the optimal signal representation problem (for the noise process) with an eigenvector basis. Therefore optimal compression of the noise subspace is obtained by the CSM relative to an eigenvector basis. The CSM achieves this capability due to its target signal dependency, a feature lacking in the PCI and eigencanceler algorithms.

Unfortunately the damage has already been done in these cases since the averaging to obtain the original covariance matrix (or an eigenvector basis for the space spanned by its columns) needs to be accomplished first. If N is reasonably large, then it quickly becomes unreasonable to assume that there is a region of space large enough to support the $2N$ or more samples which are *iid* with the noise present in the test cell. That is, the clutter and interference are not capable of retaining the properties of stationarity and homogeneity over extended regions. Note that computing the eigendecomposition is equivalent to computing the inverse from an information perspective, and that computing the inverse is equivalent to having solved the full-rank problem. Thus, the PCI, eigencanceler and CSM methods can all be considered as attempts at reducing the rank to improve performance after having to compute the original full-rank solution. While it is true that the "larger" eigenvectors may be estimated with lower sample support, they generally span a suboptimal subspace and cannot compensate for problems such as a lack of homogeneity.

Finally, it is mentioned that canonical correlation analysis (CCA) [12-14] is not of use in classical rank one signal detection problems. While one could use canonical coordinates with the SMI covariance matrix, the solution degenerates to the direct solution of the full-rank Wiener filter. CCA therefore falls into the category of requiring the full-rank solution while not even motivating rank reduction for this problem.

3. COLORED NOISE MATCHED FILTERING

What is really desired is a method to obtain the normalized colored noise matched filter without knowledge of either the true or SMI noise covariance matrices. In addition, it is desireable to obtain this result while simultaneously only using those training samples which are most correlated with the noise in the test cell. Therefore

it is necessary to obtain a low rank subspace which is optimized in the sense that it provides the optimal representation of the noise in the test cell using the noise in the training data. Note that this is an information theoretic solution which must be solved before the full-rank problem is formulated. Now consider explicitly solving this optimal representation problem without knowledge of the noise covariance matrix, its inverse or its eigenstructure.

The first step in the optimization process is to recognize that the white noise matched filter, $\mathbf{s}$ (assumed for convenience to be a unit vector), is the best solution without prior knowledge of the noise covariance. All of the remaining information (and coloring) that is not accounted for in the white noise matched filter "lives" in its orthogonal complement $\mathbf{s}^{\perp}$. Define the first basis vector to be $\mathbf{s}$.

Consider the implementation of a full-rank Wiener filter in $\mathbf{s}^{\perp}$ to estimate the colored noise projected into the rank one subspace $\mathbf{s}$. This yields a partitioned-form processor called a generalized sidelobe canceler [11], and is well known to be equivalent with the original colored noise matched filter for the target signal with known noise covariance. While this result is useful, it cannot be the solution currently sought since it requires knowledge of the noise covariance.

Another approach is to pick the optimal rank one subspace in $\mathbf{s}^{\perp}$ for estimating the colored noise that "leaked" through the filter $\mathbf{s}$, without knowledge of the noise covariance. This solution is easily verified to be the white noise matched filter $\mathbf{r}_{x_0 d_0}$, which is the cross-correlation vector between the first stage white noise matched filter output $d_0 = \mathbf{s}^H \mathbf{x}$ and the data vector $\mathbf{x}_0 = \mathbf{B}\mathbf{x}$ in $\mathbf{s}^{\perp}$ (where $\mathbf{B}$ is the $(N-1) \times N$ projection matrix into $\mathbf{s}^{\perp}$, termed a blocking matrix). Define the next basis generating vector as the unit vector $\mathbf{h}_1 = \mathbf{r}_{x_0 d_0} / ||\mathbf{r}_{x_0 d_0}||$. Also define the scalar Wiener filter g_1 as the optimal linear filter for estimating d_0 from $d_1 = \mathbf{h}_1^H \mathbf{x}_0$. The first stage subspace estimation error is then given by $e_1 = d_0 - g_1^* d_1$.

Next a recursion is developed which optimizes the remaining basis selection for $i = 1, 2, \ldots, N-1$. Define $\mathbf{B}_i$ as the projection operator into $\mathbf{h}_i^{\perp}$, which yields the data vector $\mathbf{x}_i = \mathbf{B}_i \mathbf{x}_{i-1}$. Then the cross-correlation vector $\mathbf{r}_{x_i e_i}$ maximizes the noise residual in $\mathbf{h}_i^{\perp}$ without knowledge of the noise covariance matrix. The next basis generating vector is then chosen to be the vector $\mathbf{h}_{i+1} = \mathbf{r}_{x_i e_i} / ||\mathbf{r}_{x_i e_i}||$. Now define the scalar $d_{i+1} = \mathbf{h}_{i+1}^H \mathbf{x}_i$. The Wiener filter $\mathbf{g}_{i+1}$, of dimension $i+1$, is the optimal linear filter which estimates d_0 from the vector $\vec{\mathbf{d}}_{i+1} = \left[\begin{array}{ccc} d_1 & \cdots & d_{i+1} \end{array} \right]^T$. The error at each stage is then given by $e_i = d_0 - \mathbf{g}_i^H \vec{\mathbf{d}}_i$. This stage-wise maximization of the colored noise residual combines a whitening innovations with a correlation operator that converges to the colored noise matched filter.

The algorithm described thus far determines an optimal basis which has been constructed to provide the best estimate of the residual noise at each stage. Unfortunately the required Wiener filter at stage i is a vector of length i. A simplificaton is achieved, however, through the realization that the output can be mapped to the input by the following relation:

$$||\mathbf{r}_{x_i e_i}|| = ||\mathbf{r}_{x_i d_i}||. \tag{2}$$

The result of (2) unifies the decoupled stages of the innovations procedure while simultaneously decomposing the vector Wiener filter into a nested multistage Wiener filter [15-18] composed only of scalar weights w_i (see Fig. 1). The ratio of the out-

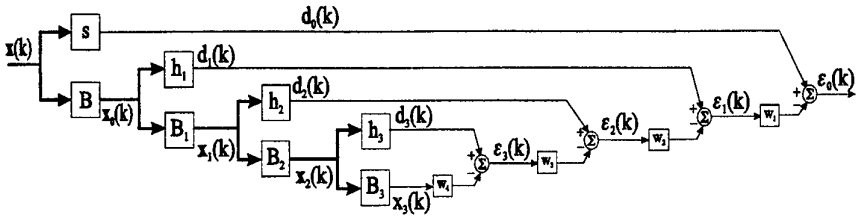


FIG. 1. The multistage Wiener filter for N=5.

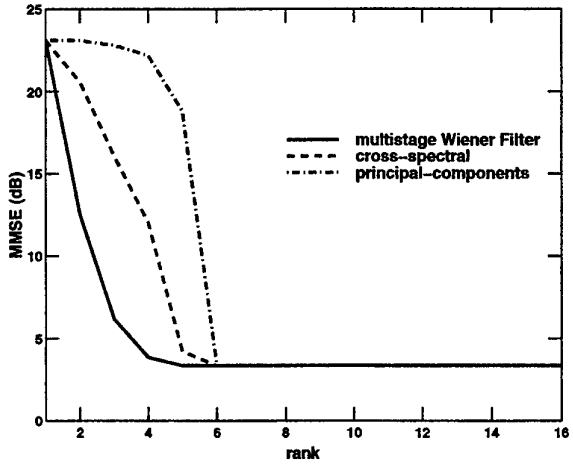


FIG. 2. The relative performance of reduced-rank processors.

put power of this filter to the output noise power yields an identical CFAR test to (1):

$$\Lambda = \frac{|\varepsilon_0|^2}{\pi_0} = \frac{|\mathbf{s}^H \mathbf{R}^{-1}\mathbf{x}|^2}{\mathbf{s}^H \mathbf{R}^{-1}\mathbf{s}} \underset{H_0}{\overset{H_1}{\gtrless}} \eta, \tag{3}$$

where $\pi_0 = \mathbf{E}\left[|\varepsilon_0|H_0|^2\right]$. This filter demonstrates the following properties: 1) The colored noise matched filter is exactly obtained without a matrix inversion; 2) The multistage Wiener filter generates a nonunitary diagonalization of the covariance where the diagonal elements are mean-square error values; 3) The most correlated signal energy is represented in the fewest spectral coefficients of this decomposition; and 4) Optimal signal compression is obtained by truncating the multistage Wiener filter.

An example is now considered to demonstrate the relative performance of the multistage Wiener filter, the cross-spectral metric and the principal components algorithms. Consider a 16 element array with half-wavelength spacing. There are 5 jammers present whose directions of arrival are -60°, -30°, -17°, 14° and 34°, and whose signal-to-noise ratios in dB are 30, 32, 27, 30 and 29, respectively. The minimum mean square error, which is also the output power of the colored noise matched filter, is 3.3440 dB while the white noise matched filter output power is 23.1077 dB. The convergence as a function of rank for these reduced-rank processors then explicitly evaluates the relative signal representation and compression capa-

bility of the algorithms. The results, depicted in Fig. 2, demonstrate the superior performance of the multistage Wiener filter.

4. SUMMARY

The multistage Wiener filter is derived from a new optimization procedure. This procedure directly demonstrates that the filter simultaneously performs whitening via an innovations process and correlation on the whitened data in a stagewise manner. This result is interpreted as a colored noise matched filter implementation and more clearly explains the optimal properties of the filter structure.

REFERENCES

1. Brennan, L.E. and Reed, I.S. (1973). Theory of adaptive radar. *IEEE Trans. Aerosp. Electron. Syst.*, 9:237–251, March 1973.
2. Chen, W.S. and Reed, I.S. (1991). A new CFAR detection test for radar. *Digital Signal Proc.*, 4:198–214, October 1991.
3. Robey, F.C., Fuhrmann, D.R., Kelly, E.J. and Nitzberg, R. (1992). A CFAR adaptive matched filter detector. *IEEE Trans. Aerosp. Electron. Syst.*, 28(1):208–216, January 1992.
4. Reed, I.S., Mallett, J.D. and Brennan, L.E. (1974). Rapid convergence rate in adaptive arrays. *IEEE Trans. Aerosp. Electron. Syst.*, AES-10(6):853–863, November 1974.
5. Kumaresan, R. and Tufts, D.W. (1980). Data adaptive principal component signal processing. In *Proc. 19th IEEE Conf. Decision and Control*, pages 949–954, Albuquerque, NM, December 1980.
6. Kirsteins, I.P. and Tufts, D.W. (1994). Adaptive detection using a low rank approximation to a data matrix. *IEEE Trans. Aerosp. Electron. Syst.*, 30(1):55–67, January 1994.
7. Haimovich, A.M. and Bar Ness, Y. (1991). An eigenanalysis interference canceler. *IEEE Trans. Signal Processing*, 39(1):76–84, January 1991.
8. Goldstein, J.S., Williams, D.B., Mersereau, R.M. and Holder, E.J. (1994). Inter-space and intra-space transformations for sensor array processing. In *Proc. 28th Asilomar Conf. Signals, Syst. Comput.*, volume I, pages 638–642, Pacific Grove, CA, November 1994.
9. Goldstein, J.S., Kogon, S.M., Reed, I.S., Williams, D.B. and Holder, E.J. (1995). Partially adaptive radar signal processing: The cross-spectral approach. In *Proc. 29th Asilomar Conf. Signals, Syst. Comput.*, volume 2, pages 1383–1387, Pacific Grove, CA, November 1995.
10. Goldstein, J.S. and Reed, I.S. (1997). Reduced rank adaptive filtering. *IEEE Trans. Signal Processing*, 45(2):492–496, February 1997.
11. Goldstein, J.S. and Reed, I.S. (1997). Theory of partially adaptive radar. *IEEE Trans. Aerosp. Electron. Syst.*, 33(4):1309–1325, October 1997.
12. Hotelling, H. (1936). Relations between two sets of variates. *Biometrika*, 28:321–377, 1936.
13. Anderson, T.W. (1958). *An Introduction to Multivariate Statistical Analysis.* John Wiley & Sons, New York, NY, 1958.
14. Scharf, L.L. and Thomas, J.K. (1998). Wiener filters in canonical coordinates for transform coding, filtering, and quantizing. *IEEE Trans. Signal Processing*, 46(3):647–654, March 1998.
15. Goldstein, J.S. and Reed, I.S. (1996). Multidimensional Wiener Filtering Using a Nested Chain of Orthogonal Scalar Wiener Filters. Technical Report CSI-96-12-04, University of Southern California, December 1996.
16. Goldstein, J.S. and Reed, I.S. (1997). A New Method of Wiener Filtering and its Application to Interference Mitigation for Communications. In *Proc. IEEE MILCOM*, volume 3, pages 1087–1091, Monterey, CA, November 1997.
17. Goldstein, J.S., Reed, I.S.,Scharf, L.L. and Tague, J.A. (1997). A low-complexity implementation of adaptive Wiener filters. In *Proc. 31th Asilomar Conf. Signals, Syst. Comput.*, volume 1, pages 770–774, Pacific Grove, CA, November 1997.
18. Goldstein, J.S., Reed, I.S. and Scharf, L.L. (1998). A multistage representation of the Wiener filter based on orthogonal projections. *IEEE Trans. Information Theory*, 44(7):2943–2959, November 1998.

Applications of Probabilistic Least Squares Tracking to Pulse Train Deinterleaving

Douglas A. Gray[1,2] and Mark L. Krieg[1,3]

[1] CRC for Sensor Signal and Information Processing
[2] EEE Dept, University of Adelaide
[3] Defence Science and Technology Organisation, Australia

ABSTRACT

The Probabilistic Least Squares Tracking (PLST) algorithm is a recursive way of estimating both the states and associations of mixture models and a Kalman predictor/filter version of this algorithm is considered. The problem of deinterleaving superimposed stable pulse trains from multiple radars can be formulated as a mixture process in both the state transition and measurement models. Using a simple switched version of the PLST Kalman filter the states and associations can be estimated. The estimated associations are the indicator variables identifying which pulse is associated with each pulse train and is the information required to deinterleave the various pulse trains.

1. Introduction

Consider M independent sources which are dynamically varying in time according to a linear state space representation and a set of measurements where at each time instant the measurement comes from only one of the M sources, i.e., the measurements can be considered samples of a mixture of the M sources. The Probabilisitc Multi-Hypothesis Tracking (PMHT) algorithm of Streit and Luginbuhl [1,2] allows the estimation of both the states and the measurement to model associations by a clever application of the EM algorithm. An alternative is to use a least squares approach; and a Probabilistic Least Squares Tracking (PLST) method has recently been proposed by Krieg and Gray [3,4,5] and compared with PMHT.

An important electronic warfare problem in the identification of transmitted radar signals is the deinterleaving problem - this is the problem in which the superimposition of a series of pulses, each series being termed a pulse train, from different radars are received by a single receiver and the times of arrival of each pulse are then accurately measured. The problem of separating the constituent different pulse trains given just the time of arrival of each pulse is known as the deinterleaving problem. The times of arrival (TOA's) of stable pulse trains (i.e., pulse trains with constant pulse repetition intervals (PRI)) can be formulated as a state space representation [6,7,8] and some approaches to the deinterleaving problem using this approach have been proposed [8,9].

In this paper we modify previous formulations of this problem and of PLST by introducing a mixtures model into the state transition equation that is coupled to the mixtures model for the measurements. This coupling is through the associations that are explicit to the PLST algorithm and provide a simple way for the practical implementation of recursive PLS Kalman filters.
The results of the application of this approach to a simple three pulse train deinterleaving problem is given.

2. Kalman Predictor and Filter Estimation - PLS Approach

Consider a mixture model, where, at time t_k, the measurements z_k are generated by one of M linear models, i.e.,

$$z_k = \begin{cases} H_k^1 x_k^1 + w_k^1 \\ H_k^2 x_k^2 + w_k^2 \\ \mathbf{M} \\ H_k^M x_k^M + w_k^M \end{cases}$$

depending on which model generates the k-th data point. For each linear model, the time varying H_k^ρ are assumed known and the w_k^ρ represent independent noise processes, each with a noise covariance R_k^ρ. The unknown states, x_k^ρ, are to be estimated and vary according to

$$x_{k+1}^{(\rho)} = F_k^{(\rho)} x_k^{(\rho)} + v_k^{(\rho)} \qquad \text{where} \qquad E\{v_k^{(\rho)} v_l^{(\rho)T}\} = Q_k^{(\rho)} \delta_{kl}$$

In the PLS approach a set of the assignments, α_k^ρ, to be estimated from the data are introduced - the α_k^ρ, are zero or one depending on whether the measurement z_k at time t_k was generated by the ρ^{th} source. The α_k^ρ can be interpreted as the probability that the measurement at time t_k originated from the ρ^{th} source. The problem of estimating both the unknown states $x_k^{(\rho)}$ and the weights α_k^ρ has been termed probabilistic least squares (PLS) [3], and a batch algorithm for this was presented in [4] and a recursive least squares was derived in [5].

The Kalman filter derivation of the PLS approach is to minimise the following expression w.r.t the unknown states and associations.

$$J_k = J_{k-1} + \sum_{\rho=1}^{M} \{\alpha_k^\rho\}^2 \varepsilon_k^{\rho T} R_k^{\rho^{-1}} \varepsilon_k^\rho + \sum_{\rho=1}^{M} \xi_k^{\rho T} Q_k^{\rho^{-1}} \xi_k^\rho$$

where

$$\varepsilon_k^\rho = z_k - H_k^\rho x_k^\rho. \quad \text{and} \quad \xi_k^\rho = x_k^\rho - F_{k-1}^\rho x_{k-1}^\rho$$

subject to the constraint

$$\sum_{\nu=1}^{M} \alpha_k^{(\nu)} = 1 \qquad \forall \quad k.$$

The recursive solutions for the PLS Kalman predictor, $\hat{x}_{k+1/k}^\rho$, and the PLS Kalman filtered output, $\hat{x}_{k/k}^\rho$, are then given by

$$\hat{x}_{k+1/k}^\rho = F_k^\rho \hat{x}_{k/k}^\rho$$

and

$$\hat{x}_{k+1/k+1}^\rho = \hat{x}_{k+1/k}^\rho + K_{k+1}^\rho \{z_{k+1} - H_{k+1}^\rho \hat{x}_{k+1/k}^\rho\}$$

where the Kalman gain, K_{k+1}^ρ, is given by

$$K_{k+1}^\rho = \{\hat{\alpha}_{k+1}^\rho\}^2 P_{k+1/k+1}^\rho H_{k+1}^{\rho T} R_{k+1}^{\rho\ -1}$$

The recursion being finally completed by the following update equations for the error covariances of the predicted and filtered states.

$$P^{\rho}_{k+1/k} = F^{\rho}_{k} P^{\rho}_{k/k} F^{\rho T}_{k} + Q^{\rho}_{k}$$

and

$$P^{\rho}_{k+1/k+1}{}^{-1} = P^{\rho}_{k+1/k}{}^{-1} + \{\hat{\alpha}^{\rho}_{k+1}\}^2 H^{\rho T}_{k+1} R^{\rho -1}_{k+1} H^{\rho}_{k+1}$$

Similarly to previous approaches the estimated associations are given by

$$\hat{\alpha}^{\rho}_{k+1} = \frac{\left(\underline{e}^{\rho T}_{k+1} R^{\rho}_{k+1}{}^{-1} \underline{e}^{\rho}_{k+1}\right)^{-1}}{\sum_{\mu=1}^{M} \left(\underline{e}^{\mu T}_{k+1} R^{\mu}_{k+1}{}^{-1} \underline{e}^{\mu}_{k+1}\right)^{-1}}$$

As is their wont, the equations are nonlinear but coupled in such a manner that allows, at each time instant, an iterative, but computationally demanding, solution.

3. The Deinterleaving of Stable Pulse Trains

The deinterleaving problem, discussed earlier, may be formulated as a state space problem, by defining the state of the i-th stable pulse train, $\underline{x}^{\rho}_{k}$ by

$$\underline{x}^{\rho}_{k} = \begin{bmatrix} t^{\rho}_{k} \\ T^{\rho}_{k} \end{bmatrix}$$

where $t^{(\rho)}_{k}$ is the precise time of last occurrence of a pulse from the ρ^{th} pulse train and T^{ρ}_{k} is the pulse repetition interval which is allowed to be time varying, but with a very small variance [6,7,8,9].

To allow for the fact that k is the received pulse number index rather than time, a switched transition matrix of the form

$$F^{\rho}_{k} = \begin{cases} F_1 = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} & \textit{if the pulse at time } t_{k+1} \textit{ originates from the } \rho^{th} \textit{source} \\ F_2 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} & \textit{otherwise} \end{cases}$$

with the corresponding measurement covariances

$$Q^{\rho}_{k} = \begin{cases} Q_1 = \begin{bmatrix} \sigma^2_{jitter} & \\ 0 & \sigma^2_{pri} \end{bmatrix} & \textit{if the pulse at time } t_{k+1} \textit{ originates from the } \rho^{th} \textit{source} \\ Q_2 = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix} & \textit{otherwise} \end{cases}$$

The measurement matrix is time and pulse train invariant and is given by $H^{\rho}_{k} = [1 \quad 0]$

The PLS method can be applied in either a recursive Kalman filter or batch mode. Here we consider the recursive mode described above whereby given $\{\hat{x}^{\rho}_{k|k}, \hat{\alpha}^{\rho}_{k}\}$ we use the Kalman filter to get $\hat{x}^{\rho}_{k+1|k+1}$ and directly calculate $\hat{\alpha}^{\rho}_{k+1}$. These two steps are coupled through the error terms and must be done iteratively for each k. This necessitates some initialisation which is detailed below. The switched nature of the transition matrix can be handled by hard limiting the associations and dynamically using them to determine which model for the transition matrix should be implemented.

The iteration at time instant t_{k+1} may be summarised as

(1) Initialise the associations using prediction errors $\varepsilon^{\rho}_{k+1}(0) = z_{k+1} - H^{\rho}_{k+1} F_1 \hat{x}^{\rho}_{k|k}$

(2) Calculate the $\hat{\alpha}^{\rho}_{k+1}(0)$'s.

(3) For $n = 1, 2, \ldots$ till convergence

(3.1) Update the covariances and gains according to

$$P^{\rho}_{k+1|k}(n) = F^{\rho}_{k} P^{\rho}_{k|k}(n_{conv}) F^{\rho T}_{k} + Q^{\rho}_{k} \text{ where } F^{\rho}_{k}, Q^{\rho}_{k} \text{ are determined by } \hat{\alpha}^{\rho}_{k+1}(n).$$

$$K^{\rho}_{k+1}(n) = \{\hat{\alpha}^{\rho}_{k+1}(n-1)\}^2 P^{\rho}_{k+1|k}(n) H^{\rho T}_{k+1} (R^{\rho}_{k+1} + \{\hat{\alpha}^{\rho}_{k+1}(n-1)\}^2 H^{\rho}_{k+1} P^{\rho}_{k+1|k}(n) H^{\rho T}_{k+1})^{-1}$$

$$P^{\rho}_{k+1|k+1}(n) = P^{\rho}_{k+1|k}(n) - K^{\rho}_{k+1}(n) H^{\rho}_{k+1} P^{\rho}_{k+1|k}(n)$$

(3.2) Update the state estimates according to

$$\hat{x}^{\rho}_{k+1|k+1}(n) = F^{\rho}_{k} \hat{x}^{\rho}_{k|k}(n_{conv}) + K^{\rho}_{k+1}(n)\{z_{k+1} - H^{\rho}_{k+1} F^{\rho}_{k} \hat{x}^{\rho}_{k|k}(n_{conv})\}$$

(3.3) Recalculate the association estimates according to

$$\varepsilon^{\rho}_{k+1}(n) = z_{k+1} - H^{\rho}_{k+1} \hat{x}^{\rho}_{k+1|k+1}(n) \text{ and consequently } \hat{\alpha}^{\rho}_{k+1}(n)$$

(3.4) Form "hard assignments" by

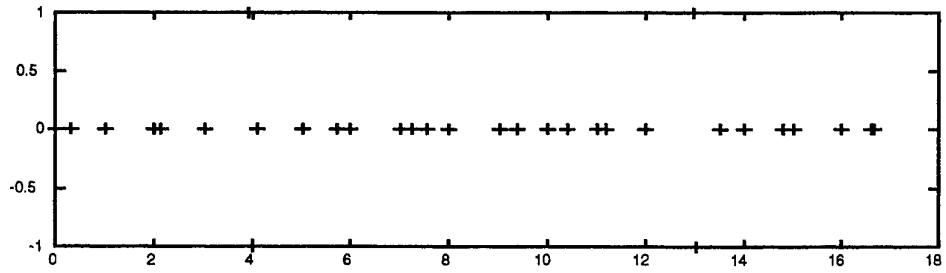
$$\text{find } \nu = \arg\max_{\rho} \{\hat{\alpha}^{\rho}_{k+1}(n)\} \text{ and set } \hat{\alpha}^{\rho}_{k+1} = \delta_{\rho\nu}.$$

(4) end $n = n_{conv}$

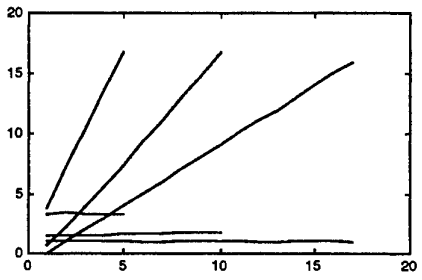
An example application of this are three stable pulse trains with PRIs and start times of $[1,\sqrt{2}+0.4,\pi][$ and $[0,0.3,4.1]$ respectively. The initialisation of the algorithm used the actual measured times of the first three pulses and the PRIs are randomly set within 2.5% of the exact PRI. The tight restriction on the intialisation of the PRIs and the fact that the number of pulse trains is assumed known need further investigation but could be overcome by first preprocessing the data with some histogramming technique. The covariances were initialised by setting $P_{111}^{\rho} = \begin{bmatrix} 5 & 0 \\ 0 & .01 \end{bmatrix}$ which is ten times the assumed model noise covariance matrix. Note the use of a high variance on the time of last occurence of the ρ–th pulse and the small variance on the PRIs.

At each time update of the Kalman filter the iteration described above was repeated 10 times. The times of arrival were jittered by adding Gaussian noise of standard deviation of 0.01 i.e., 1% of the smallest PRI, and 32 data samples were used.

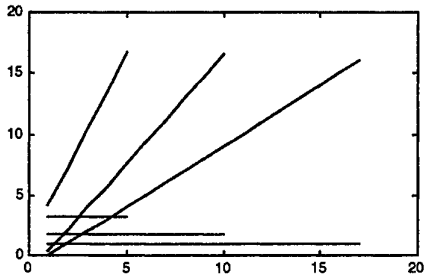
Plotted below is a sample run showing the difference between the exact and the estimated associations using the PLST Kalman filter just running forward in time. Two errors were made at times around 4 and 13.



Plotted below is a sample run of the Kalman filter estimates of the TOAs of the three pulse trains and their PRIs when (a) the exact associations are known and (b) when these are estimated as outlined above. Close agreement was obtained.



(a) Associations known exactly



(b) Associations estimated

Results averaged over an ensemble of realisations indicated that on average about 3 errors per 32 points resulted. At the same time the corresponding errors in the state estimates were very small indicating that deinterleaving rather than parameter estimation was important for this example. However for other examples the converse situation can occur.

4. Summary

The problem of deinterleaving pulse trains can be formulated in state space using mixture models for both the measurement and state transition equations. These are coupled through the unknown set of associations. A simple modification of the PLST algorithm allows this to be recursively solved using a Kalman filter. Smoothing can readily be incorporated by considering a batch process and could show a significant decrease in the errors.

Simulations indicate that, provided good initialisation can be obtained, the approach does show some promise although many issues associated with initialisation need resolution.

5. References

[1] R L Streit and T E Luginbuhl *Probabilistic Multi-Hypothesis Tracking* Technical Report TR-10,428, NUWC Newport RI 1995.

[2]R L Streit and T E Luginbuhl *A Probabilistic Multi-Hypothesis Tracking Algorithm without Enumeration and Pruning* Proceedings of the Sixth Joint Service Data Fusion Symposium, Laurel, Maryland, 14-18 June 1993, 1015-1024.

[3] M.L. Krieg and D.A. Gray *Comparison of probabilistic least squares and probabilistic multi-hypothesis tracking algorithms for multi-sensor tracking* Proc. IEEE Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP-97), Vol. 1, Munich Germany, 515-518.

[4].D. A. Gray, M.L. Krieg and M. R. Peake *Estimation of the Parameters of Mixed Processes by Least Squares Optimisation* Proceedings of Fourth International Conference on Optimization; Techniques and Applications (ICOTA 98) Vol 1, Perth Australia, July 1998, 891-898.

[5] D A Gray and M L Krieg *Recursive Least Squares and Kalman Filtering Approaches to Tracking the Parameters of Mixture Models* Proceedings of Workshop commun GdR ISIS (GT1) and NUWC "Approches Probabilistes Multipistes pour l'Extraction Multipistes Paris Nov 1998 paper 5.

[6] E T Kofler and C T Leondes New *Approach to the Pulse Train Deinterleaving Problem* Internation Journal on Systems Science 20(12), 2663-2671, 1989

[7] D.A. Gray, B.L. Slocumb and S.D. Elton, *Parameter Estimation for Periodic Discrete Event Processes* Proceedings Int Conf on Acoustics Speech and Signal Processing, vol 4, April 1994, Adelaide, Australia, 93-96.

[8] J B Moore and V Krishnamurthy *Deinterleaving Pulse trains using Discrete-Time Stochastic Dynamic-Linear Models* IEEE Trans on Signal Processing, vol 42, No11, Nov 1994 3092-3103.

[9] B J Slocumb and E W Kamen *The Pulse train PDA Analysis and Deinterleaving Filter* SPIE vol 3068 296-307.

Array Element Localisation Using Simulated Annealing

Michael V. Greening
Defence Science and Technology Organisation,
Salisbury Site, MOD Building 79, P.O. Box 1500,
Salisbury, S.A., 5108, Australia

Abstract

Array processing techniques such as beamforming or matched field processing require accurate knowledge of the location of individual elements in the array. For horizontal arrays laid on the ocean floor, relative arrival times measured across the array from nearby implosive sources are often used to aid in estimating the sensor positions. However, the inverse problem of determining the sensor positions from the relative arrival times is both nonunique and ill-conditioned. This paper shows how simulated annealing can be used to solve this inverse problem. Synthetic studies show that relative sensor locations can be exactly found while tests with real data show an improvement in array gain comparable to the theoretical limit obtained from a perfectly known array.

Introduction

Remotely deployed systems often contain horizontal or vertical arrays mounted on the ocean floor and used to acoustically monitor areas of the ocean. One problem with remotely deployed systems is accurately determining the sensor positions in the array. Conventional beamforming is often considered to require sensor position estimates accurate to within $\lambda/10$ where λ is the wavelength of the signal measured.[1] More advanced array processing techniques such as adaptive beamforming or matched field processing require even more accurate estimates of the sensor positions.

One technique often used to help estimate the sensor positions in remote systems is to employ transient sources near the array and measure the arrival times of the signals across the array. If the location of the sources and the travel times to the sensors are known, then the location of all the sensors in the array can be unambiguously determined using triangulation from three sources. However, the source locations are often only known approximately and the travel times from source to sensor are often unknown with the arrival times at any sensor only known relative to the arrival times at other sensors. Unknown source positions can allow for a rotation or translation of the combined system of sensors and sources without

a change in the arrival times. Also, a given set of relative arrival times for a single source can be exactly reproduced by decreasing the source range while increasing the curvature of the array. Thus, the inverse problem of estimating the sensor positions from relative arrival times and with unknown source locations is both nonunique and ill-conditioned. The problem involves searching a multidimensional space of estimated sensor and source positions to minimize the error between the measured and predicted relative arrival times. One technique designed specifically to search ill-conditioned, multidimensional spaces is called simulated annealing.[2] This paper will show how to apply this technique to the problem of locating the sensors in a remotely deployed system. Both synthetic and real data will be examined and some recommendations on the number and location of sources will be given.

I Methodology

This section shows how to apply simulated annealing to the problem of locating a horizontal array mounted on the ocean floor. For the real data, a set of light bulbs[3] imploded near mid-depth was used as transient sources and the relative arrival times of the direct arrival and surface reflection were measured across the array. The problem then is to use simulated annealing to find a set of source and sensor locations which will reproduce the relative arrival times.

Simulated annealing involves a series of iterations in which the unknown parameters (ie. source and sensor locations) are perturbed. For each iteration, the relative arrival times of the direct arrival and surface reflection are calculated for the modelled parameters. The modelled arrival times are then compared with the measured arrival times and the total time error E is given as an estimate of the goodness of fit of the modelled source and sensor positions to their true values. For successive iterations, the change in error ΔE is calculated. If the error has decreased ($\Delta E < 0$), the new parameter configuration is accepted. If the error has increased ($\Delta E > 0$), the new configuration has a probability P of being accepted with the probability being drawn from the Boltzmann distribution:

$$P(\Delta E) = \exp(-\Delta E/T), \tag{1}$$

where T is a controlling parameter analoguous to temperature in the physical process of annealing. Accepting some perturbations which increase E allows the algorithm to escape from local suboptimal minima in the search space. Decreasing T with successive iterations decreases the probability of accepting an increase in error, and the algorithm eventually converges to a solution which should approximate the global minimum.

Two factors involved in developing an efficient and effective simulated annealing algorithm are the method of decreasing the temperature T, and the method of perturbing the unknown parameters. A starting temperature T_0 was chosen which allows at least 90% of all perturbations to be accepted. A number of perturbations

η are then performed before decreasing the temperature according to $T_{j+1} = \alpha T_j$, where $\alpha < 1$. The process is terminated when further temperature steps do not result in a lower error. The values of η and α to use depend on the difficulty of the inversion. Increasing both η and α should decrease the final error but also increases the number of iterations and time required. An estimate of η and α can be obtained by using synthetic data and chosing η and α large enough that the final error is zero or the resulting sensor locations are accurate to within an acceptable tolerance. With real data, η and α can be initialized to the values obtained from the synthetic study and then increased until a further increase in their values does not decrease the final error.

The method of perturbing the parameters can have a major effect on the efficiency of simulated annealing. Changing only one parameter at a time allows the algorithm to converge for a sensitive parameter while continuing to search for less sensitive parameters. Changing multiple parameters in one perturbation allows for quicker convergence when coupled parameters are involved and also allows for easier jumping between local minima. For our problem, the unknown parameters are the source and sensor locations along with the bottom depth, which was assumed to be known only within 20 m for the purposes of the study. For every perturbation a source, sensor or bottom depth is randomly chosen to be changed. If a source is picked, the position of only a single source is changed for a given perturbation. If a sensor is picked, either a single sensor, multiple sensors or the entire array can be changed in the following manner. The entire array can be changed by shifting it horizontally or by rotating it about some angle. A single sensor can be changed by changing its distance or bearing from the previous sensor without moving other sensors, providing that the separation between pairs of sensors does not exceed the length of cable joining them. Multiple sensors can also be changed by moving all sensors before or after the sensor picked as above by the same change given to that sensor. Finally, when changing a parameter, it is changed in one of two ways. Either a new value is picked within a Gaussian distribution centered on the current value or a new random value is chosen from the entire allowable range for that parameter.

After the simulated annealing algorithm stops, a gradient descent algorithm was applied using the result of the simulated annealing as the initial estimate. This is used to ensure that the absolute minimum of the current trough is found and can help to speed up simulated annealing after a suitably cool temperature is reached.

II Results

The simulated annealing algorithm for array element localisation was tested on both real and synthetic data. The real data was collected at the RDS-2 trial in November 1998 in 110 m water in the Timor Sea off the northern coast of Australia. The data examined was collected on the ULITE array deployed by the Space and Naval Warfare Systems Center (SPAWAR), San Diego, CA., USA.

The planned deployment of the ULITE array and light bulbs is shown in Fig. 1. The array consists of three arms of 32 sensors tied in the center and with each arm containing a slight curvature to break the left/right symmetry from the arm. The planned light bulb locations were at mid-depth with three light bulbs on each side of each arm, approximately 100 m distance from the arm. Using Fig. 1 to generate synthetic data, the simulated annealing algorithm was tested with the following uncertainties: the center of the array was assumed to be known to within 100 m; the bearing or range to any sensor was unrestricted except that the range between a pair of sensors could not be greater than the length of cable separating the pair; the horizontal position of a light bulb was assumed to be known within 100 m; the depth of a light bulb was assumed to be known only within 20 m; the water depth was assumed to be known within 20 m. With the above uncertainties, if the relative arrival times were known exactly (ie. not digitised), then the relative array shape and light bulb positions were found exactly. Absolute positions could shift or rotate as long as all light bulb positions stayed within the allowed uncertainty. If the relative times were only known within a digitisation sample, then the relative position of any sensor could shift from its true relative position by as much as the distance travelled by sound within the time of the digitisation sample. Increasing the number of light bulbs decreases the positional shift introduced by the digitisation. The relative arrival times of the direct arrival and bottom reflection for the light bulb at (100 m, 300 m) is shown in Fig. 2.

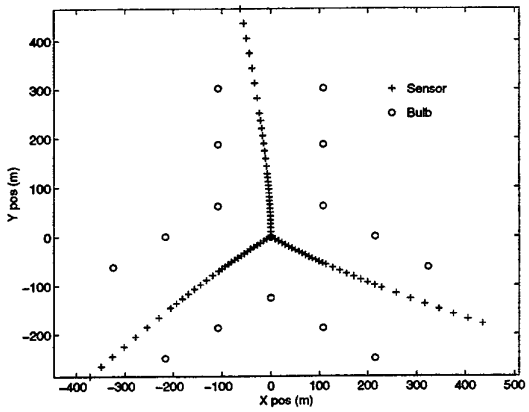


Figure 1: Trial layout plan

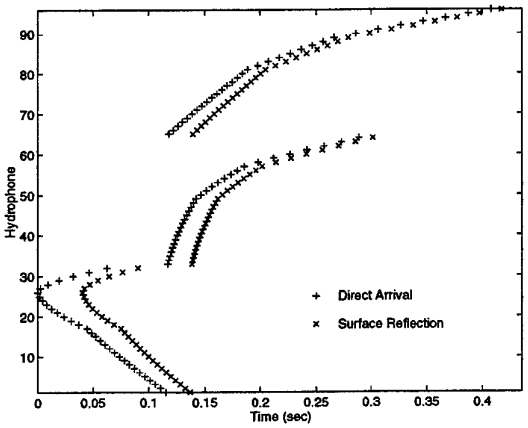


Figure 2: Relative arrival times

The estimated shape of the ULITE array is shown in Fig. 3 along with the shape estimated by SPAWAR. Although ground truth is unavailable, the similarity of shape provides some confidence that the correct shape was obtained. The SPAWAR estimate assumed fixed light bulb locations at mid-depth and at the recorded GPS positions. The simulated annealing technique used at DSTO allowed uncertainties of 100 m in the horizontal location, and 20 m in the depth of a light bulb. This makes the technique more robust to errors in estimated light bulb positions and allows for drift in the light bulb as it is lowered to depth. For individual arms of the array, the relative shapes estimated by DSTO and SPAWAR are very similar as

shown in Fig. 4. Although the relative shapes of individual arms are very similar, the relative shapes of the entire array show a difference of a 2° rotation in the direction of the northward pointing arm relative to the other two arms. The reason for the difference between the two estimated shapes is believed to be caused by the location of the light bulbs, which were not as tightly concentrated along the arms of the array as in the plan. Consider a light bulb which is near endfire to one arm and near broadside to another. A shift in the light bulb position can cause a large change in the relative arrival times across the broadside array but very little change across the endfire array. Thus, having many light bulbs near endfire of one arm can cause a relative shift in the heading between two arms of the array with little difference in the relative arrival times. Determining the number and location of light bulbs required to accurately estimate an array shape is array dependent. A study of a single, nearly linear, array of 200 m length showed that four light bulbs with two along one side, one along the other side and one near endfire always provided solutions accurate to within the digitisation rate, provided that the light bulbs were within 200 m of the array.

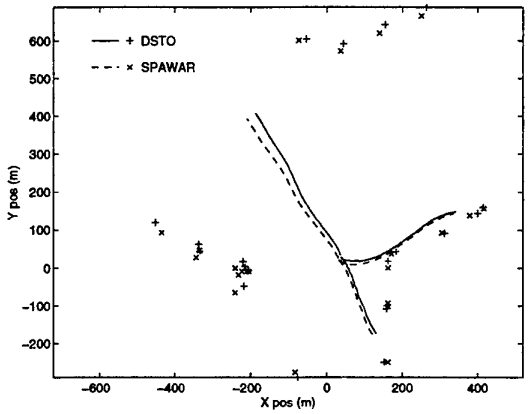


Figure 3: Estimated trial layout

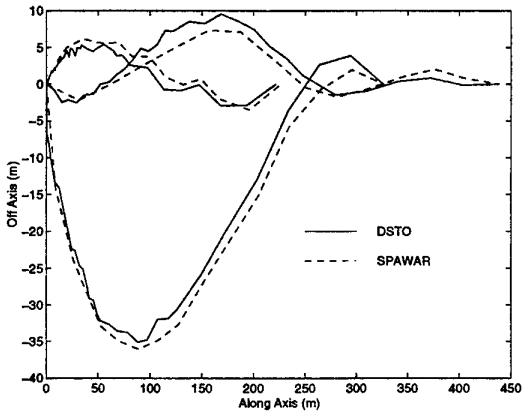


Figure 4: Estimated shapes of array arms

The errors between the measured and modelled arrival times are shown in Fig. 5. Each horizontal line shows the error in arrival times for both the direct arrival and surface reflection from all light bulbs. Nearly all errors fall within the digitisation sample size of 0.002 sec, providing further confidence that the true array shape is well approximated. Sensor 32 is the outermost sensor on the northward pointing arm of the array and was connected to a surface buoy. It is believed that the buoy was causing this sensor to move and thus, an accurate estimate of its position could not be found, and it contains large errors in the arrival time estimates. This was also found by SPAWAR. This sensor is not plotted in Fig. 3.

A final indication of the accuracy of the estimated array shape can be obtained by finding the array gain provided when using the estimated shape in beamforming. Beamforming was performed using the eastward pointing arm on an 80 Hz tonal target at 100° relative to North. Only the 16 sensors that were cut for 24 Hz are used. The beamformed output is shown in Fig. 6 and indicates an approximate

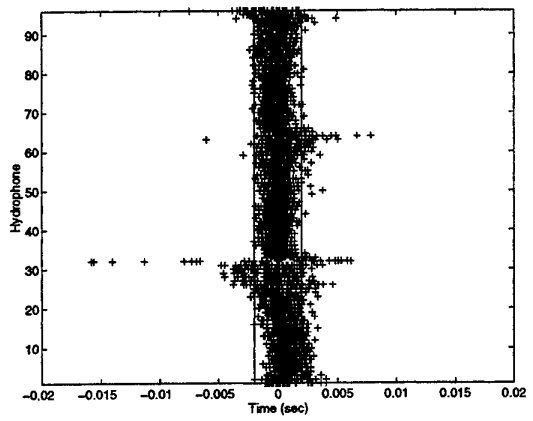


Figure 5: Arrival time errors

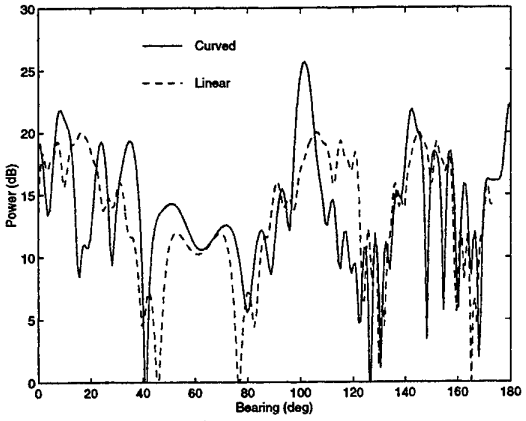


Figure 6: Beamformed output of east arm

6 dB gain when the estimated array shape is used compared to a linear array. This matches the theoretical loss for a signal arriving on the curved array but processed assuming a straight array. Again, this is a strong indication that the true array shape is well represented.

III Summary

This paper has shown how simulated annealing can be used to accurately perform array element localisation on remotely deployed systems. Synthetic studies have shown that the relative sensor positions can be determined within the accuracy defined by the digitisation rate if sufficient light bulbs are employed nearby along the array and at endfire to the array. Tests with real data agreed well with an independently performed localisation estimate, and also improved the response of conventional beamforming to approximately the theoretical limit.

Acknowledgments

The author would like to gratefully acknowledge the Space and Naval Warfare Systems Center in San Diego, CA, USA for the use of the ULITE data and for the collaboration in comparing array element localisation results.

References

1. B.D.Steinberg, *Principles of Aperture and Array System Design*, Wiley, New York, 1976.
2. S.Kirkpatrick, C.D.Gelatt, and M.Vecchi, "Optimization by Simulated Annealing," Science **220**, 671-680, (1983).
3. G.J.Heard, M.McDonald, N.R.Chapman, and L.Jaschke, "Underwater light bulb implosions: a useful acoustic source," Oceans 97. MTS/IEEE. Conference Proceedings. IEEE Part 2, vol 2, 755-762, (1997).

A multicriteria shortest path algorithm or plotting a safe path through a field of sensors

Christina Hallam*, K.J. Harrison** and J.A. Ward**

*Maritime Operations Division, HMAS Stirling, WA 6958, AUSTRALIA and
**Department of Mathematics and Statistics, Murdoch University, WA 6150, AUSTRALIA

We describe an algorithm for finding Pareto-optimal paths in a multicriteria shortest path problem. We use the algorithm to find approximate solutions to the problem of guiding a mobile object such as a submarine from one location to another, through a field of sensors at known positions, within a fixed time period and with minimum probability of detection.

1. INTRODUCTION

Our goal is to guide a submarine from one location to another, through a field of sensors at known positions, within a fixed time period and with minimum probability of detection. The techniques presented here can be used in other applications - for example, to find approximately optimal flight paths for aircraft [7], or to plan paths requiring obstacle avoidance in robotics [6]. The mathematical problem that underpins our approach is a discrete one involving undirected graphs, and we first discuss that.

2. A MULTICRITERIA OPTIMISATION PROBLEM

We are given a directed graph $\mathcal{G} = [\mathcal{V}, \mathcal{E}]$, where $\mathcal{V}$ and $\mathcal{E}$ are the finite sets of vertices and edges respectively, and specified *start vertex* s and *target vertex* t. We allow more than one edge between a pair of vertices, so that we will be able to incorporate multiple speed options in our application. We are also given k non-negative functions $f_1, f_2, ..., f_k$ defined on $\mathcal{E}$, where for each $j = 1, 2, ..., k$, the number $f_j(e)$ represents a cost associated with the edge $e \in \mathcal{E}$, such as a measurement of the time required to traverse e or of the danger of being detected whilst on e.

A *path* from vertex v to vertex v' in $\mathcal{G}$ is usually described as a sequence of edges $p = (e_1, e_2, ..., e_{r-1})$, where e_i is an edge joining v_i to v_{i+1} for $i = 1, 2, ..., r-1$, and $v_1 = v$ and $v_r = v'$. We only consider paths that start at s, and so we say that a path is a path to the vertex v if it is a path from s to v. We assume that there is at least one path to every vertex in $\mathcal{V}$, and include here the 'trivial path to s' containing only the vertex s as this is needed to initiate our algorithm.

The classical *shortest path problem* arises when there is only one cost function, and the task is to find a path p from s to t with minimal cost $c_1(p) = \sum_{e \in p} f_1(e)$,

where the sum is taken over all those edges e that form p. There are many efficient polynomial time algorithms, such as Dijkstra's algorithm, for doing this.

In the *multicriteria shortest path problem* (where $k > 1$), we let $c_j(p) = \sum_{e \in p} f_j(e)$ for each j, and $\mathbf{C}(p) = (c_1(p), c_2(p), ..., c_k(p))$ be the *cost vector* associated with the path p. There is a natural partial order on cost vectors defined by $\mathbf{C}(p) \leq \mathbf{C}(p')$ if and only if $c_j(p) \leq c_j(p')$ for $j = 1, ..., k$. We say that a path p to a vertex v *dominates* another path p' to the same vertex if $\mathbf{C}(p) \leq \mathbf{C}(p')$. As there may be no single dominant path to t, the problem is one of finding *Pareto-optimal* paths, which are those that are not dominated by any other path; that is, those corresponding to minimal cost vectors. Each such path is optimal in that no improvement can be achieved in one cost function, without worsening at least one of the others.

The task of determining all Pareto-optimal paths is computationally complex. Typically, ad hoc single Pareto-optimal paths are found using methods such as the weighted sum approach, the ε-constraint method and goal programming. Genetic algorithms have also been used, but their effectiveness is closely tied to the method used to assign fitness [1]. Recent literature also describes algorithms that simultaneously find all Pareto-optimal paths [2],[3],[5].

In practice there may be more than one path to a specified vertex corresponding to a given cost vector $\mathbf{C}(p)$, and we say that all such paths are *equivalent.* For many applications it is sufficient to determine just one Pareto-optimal path to t for each minimal cost vector. Tung and Chew [8],[9] give such an algorithm that does this for simple graphs. In the next section we describe modifications to this algorithm that improve its efficiency and make it appropriate for non-simple graphs.

3. THE PARETO ALGORITHM

The basic Pareto algorithm determines one Pareto-optimal path to t for each minimal cost vector. It iteratively generates paths from the start vertex s. At each iteration it selects a path $p = (e_1, e_2, ..., e_{r-1})$, the *test path*, to some vertex v', and analyses all one-edge extensions $[p : e] = (e_1, e_2, ..., e_{r-1}, e)$, where e is an edge joining v' to a neighbouring vertex v. It uses pruning criteria to reject an extension as *inadmissable* if it is either dominated by a known path to v, or if it can be shown that each possible completion of $[p : e]$ to a path to t is equivalent to, or dominated by, one already found. The latter is decided using a heuristic that provides a lower bound for the cost of such completions.

The algorithm uses a vertex labelling scheme to keep track of possible test paths at each iteration, and selects from amongst these on the basis of a selection function that incorporates the heuristic.

3.1. Components of the algorithm

We briefly discuss the important selection function and pruning criteria before describing the Pareto-algorithm.

3.1.1. Selection function

Although only one selection function is required for the basic Pareto-algorithm, we introduce a family of such functions that may be used in a later modification of the algorithm.

A selection function assigns a numerical value to each potential test path. To see how, we first let $\alpha = (\alpha_1, \alpha_2, ..., \alpha_k)$ denote a vector of non-negative weights such that $\mathrm{tr}\,\alpha = 1$, where the trace tr of a vector is the sum of its components. The special weight vectors $\delta^{(j)}$, where $j = 1, ..., k$, for which the j^{th} coordinate is 1 and the others all equal 0, and *positive* weight vectors α, for which each α_j is positive, will be important.

For each α and each path p, we let $W_\alpha(p) = \alpha_1 c_1(p) + ... + \alpha_k c_k(p)$, and for each vertex v, $h_\alpha(v) = \min_p W_\alpha(p)$, where the minimum is taken over all paths from v to t. (In particular, $W_{\delta(j)} = c_j$ for each j.) Note that the weights reflect the relative influence of the cost functions only if the cost functions f_j have similar numerical ranges. The values $h_\alpha(v)$ are easily computed by applying Dijkstra's algorithm to the simple graph $\mathcal{G}' = [\mathcal{V}, \mathcal{E}']$, obtained by putting an edge between two vertices v' and v whenever there is at least one edge between them in $\mathcal{G}$, and weighting that edge by the minimum of $W_\alpha(e)$, where the minimum is taken over all edges between v' and v in $\mathcal{G}$. We let p_α denote the path constructed to t for which $W_\alpha(p_\alpha) = h_\alpha(s)$. We can assume that p_α is Pareto-optimal, although this will require us to modify Dijkstra's algorithm if some of the α_j equal 0.

For each positive weight vector α, we define the selection value $S_\alpha(p)$ of a path p to a vertex v to be $S_\alpha(p) = W_\alpha(p) + h_\alpha(v)$. It gives a lower bound on the corresponding weighted sum of costs of paths to t that are extensions of p. The Pareto algorithm uses such a selection function, choosing a test path at each iteration from amongst those known paths with minimum selection value. It is easy to check that the positivity of α ensures that a selected test path cannot be dominated by any other path to the same vertex, and so must be Pareto-optimal.

3.1.2. Pruning criteria

Clearly the number of potential test paths may grow exponentially with the number of vertices and edges in $\mathcal{G}$, and so it is essential to eliminate unnecessary paths as soon as possible.

For each vertex v and integer $1 \leq j \leq k$, let $q_j(v) = \min_p c_j(p)$, where the minimum is taken over all paths from v to the target vertex t; that is, $q_j(v) = h_{\delta(j)}(v)$. Furthermore, let $\mathbf{Q}(v) = (q_1(v), q_2(v), ..., q_k(v))$. Then an extension $[p : e]$ of the current test path p to v' to vertex v is *rejected,* in the sense that it is not added to the set of possible test paths, if either

(P1) $\mathbf{C}([p : e]) \geq \mathbf{C}(p')$ for some known path p' to v, or

(P2) $\mathbf{C}([p : e]) + \mathbf{Q}(v) \geq \mathbf{C}(p')$, where p' is a known path to t.

The first condition identifies extensions to v that are dominated by a known path to v. The second identifies extensions that can at best be completed to paths that are equivalent to or dominated by known ones. This is a valid rejection criterion because we only want one path from each equivalence class of Pareto-optimal paths.

Let Γ_0 be the set of Pareto-optimal paths constructed to t when calculating $\mathbf{Q}(t)$ and $h_\alpha(s)$ for the chosen α. We can use Γ_0 to seed the pruning process. As Dijkstra's algorithm is quadratic, it may be worthwhile generating a reasonably large set Γ_0 by including paths p_α for several choices of α.

3.2. The steps of the Pareto algorithm

We briefly describe the essential steps of the Pareto algorithm as follows:

A1 Fix a selection function S_α (most commonly $\alpha = (1/k, 1/k, ..., 1/k)$), and use Dijkstra's algorithm to calculate the corresponding values $h_\alpha(v)$, and $\mathbf{Q}(v)$, for each vertex v. Generate additional paths in Γ_0 if desired.

A2 Let Σ denote the set of possible test paths, and Γ the set of known Pareto-optimal paths to t, at the current iteration The algorithm is initialised by setting Σ equal to the trivial path (that is, the path with no edge) and $\Gamma = \Gamma_0$.

A3 Pick a test path p by finding the path in Σ that has the minimum selection value. If there are no such paths, stop. Remove p from Σ.

A4 For each one-edge extension $[p : e]$ of p,

(i) reject $[p : e]$ if it satisfies either (P1) or (P2), with $p' \in \Gamma$, or

(ii) if $[p : e]$ is an admissible path to t, add $[p : e]$ to Γ, and go to step A5, or

(iii) if $[p : e]$ is an admissible path to a vertex other than t, calculate its selection value and add $[p : e]$ to Σ.

Go to step A3.

A5 For each path p' in Σ to vertex v, reject p' if $\mathbf{C}(p') + \mathbf{Q}(v) \geq \mathbf{C}([p : e])$. Return to step A4.

When the algorithm stops, Γ contains a complete set of inequivalent Pareto-optimal paths, one for each cost vector. They have been found in increasing order of the selection values.

It is worth mentioning the labelling scheme that keeps track of the paths that are generated. Each vertex $v \neq s$ is given a finite sequence of labels $\theta^1(v), \theta^2(v), ...$, where each label represents a path to v, and where $\theta^n(v)$ is assigned to v the n^{th} time it is used to produce an admissible one-edge extension of a test path. The labels are assigned iteratively in the following sense. Suppose that the vertex v is used for the n^{th} time when an edge e from v' to v is added to a test path p to produce an admissible extension $[p: e]$, and suppose that $\theta^i(v')$ is the label corresponding to p. Then we set $\theta^n(v) = [e, i, \mathbf{C}([p : e])]$. The third component $\mathbf{C}([p : e])$ is not used to specify a path, but just to record the cost of $[p : e]$. The scheme is initialised by setting $\theta^1(s) = [-, -, \mathbf{0}]$.

3.3. Locating special Pareto-optimal paths

We describe modifications to the Pareto algorithm that may be used when it is sufficient to find Pareto-optimal paths to t that satisfy certain cost restrictions, as will be the case in our application.

Suppose firstly that we wish to locate only Pareto-optimal paths p' that satisfy prescribed bounds on some or all of their costs. For example, we may require $c_m(p') \leq \tau_m$ for some m's, where the τ_m give the bounds. Then the pruning criteria should be modified by adding: a one-edge extension $[p : e]$ of the test path p to vertex v is *rejected* if

(P3) $c_m([p : e]) + q_m(v) \geq \tau_m$ for the relevant m.

If such bounds are given for every m then $W_\alpha(p') \leq \alpha_1\tau_1 + ... + \alpha_k\tau_k$ for every weight vector $\alpha = (\alpha_1, ..., \alpha_k)$. This suggests an additional pruning criterion: that $[p : e]$ be pruned if

(P4) $S_\alpha([p:e]) \geq \alpha_1\tau_1 + ... + \alpha_k\tau_k$,where α is the index of the selection function in use.

In our application, we will seek a Pareto-optimal path to t of least danger for which the time cost is at most some prescribed amount. Such a path can be found efficiently by updating the right hand side of (P3) and (P4) on the basis of the least dangerous known Pareto-optimal path which does not exceed the time limit.

4. APPLICATION TO SUBMARINE TRANSIT PATHS

As stated earlier, the aim is to direct a submarine from one location to another, through a field of sensors at known positions, within a fixed time period and with minimum probability of detection. If we make a number of simplifying assumptions, we can use the graph searching techniques described in previous sections to produce a near-optimal path.

4.1. Discrete formulation

For simplicity, we assume that the problem is two dimensional, in the sense that the object is constrained to move in a bounded two dimensional region called the *transit region.* We place a grid on the transit region so that the start and goal locations correspond to grid points, and assume that the grid size is sufficiently large (for a submarine, no less than $1km$) that we may ignore such complications as 'turning circles'. We also assume that the object is constrained to move in a straight line at constant speed from one grid point to a neighbouring grid point, and that only a finite number of speeds are permitted.

4.1.1. The graph and objective functions

The grid points form the vertices of a graph $\mathcal{G}$. There may be more than one edge joining a pair of adjacent vertices, one for each of the permitted speeds of the object. So it is useful to denote an edge by an ordered triple (v, v', z) if it connects vertices v and v' and the object is travelling z units per second between them.

Associated with each edge are costs corresponding to *danger* or the probability of detection by the sensors when traversing the edge, and *time* to cover that distance. Both depend on the speed of the object and the grid size. The first cost function $f_1(v, v', z)$ simply measures the time taken to traverse the edge (v, v', z).

To measure the danger, we need to know the probability in a fixed time period of detection at each point. To simplify things, we assume that the probability $P_m(r, z)$ of detection in the fixed time period by the m^{th} sensor depends only on the distance r of the object from the sensor and on its speed z, that it is independent of the probability of detection at any other sensor, and does not change with time. In practice, values of $P_m(r, z)$ are computed for a discrete set of distances and speeds using a complicated model that may also take into account such things as type and depth of sensor, type of submarine and local topography.

Under these assumptions, the probability of *not being detected* at position $\mathbf{x}$ in the fixed time period by one of M sensors is $N(\mathbf{x}, z) = \prod_{m=1}^{M}(1 - P_m(r_m, z))$, where r_m is the object's distance from the m^{th} sensor. We can use any one of a number of methods to determine an average probability $N_{(v,v',z)}$ of non-detection in the fixed time period when travelling along the edge (v, v', z), and from this we can calculate an approximate probability of non-detection whilst on (v, v', z). The

probability of non-detection on a path p is then the product of the probabilities of non-detection on the edges that make up p. Since we want an additive cost function, we take logarithms. So the second cost function, which we have loosely called the danger, can be defined as $f_2(v,v',z) = f_1(v,v',z)\left|\ln\left(1-N_{(v,v',z)}\right)\right|$. Since $1-N_{(v,v',z)} \in [0,1]$ for each edge, we minimise the probability of detection by minimising f_2.

4.2. Solving the discrete problem

To find the Pareto-optimal path p^* that satisfies a time constraint and gives the lowest probability of detection, we use a version of the algorithm that incorporates the modifications described above. We assume that τ, the maximum time allowed to move from the start to the goal location, is achievable but less than the minimum time required for the least dangerous paths from s to t; that is, $q_1(s) \leq \tau < \min\{c_1(p) : c_2(p) = q_2(s)\}$.

We add a step to the Pareto-algorithm which enables us to choose a selection function $S_\alpha = S_{(\alpha,1-\alpha)}$ that leads to the desired solution more quickly. As a guide to how to do this, we observe that $c_1(p_\alpha)$ decreases as α increases. So we may do a binary search on the parameter α, using Dijkstra's algorithm, to identify an approximately optimal value of α, in the sense that the corresponding path p_α has time cost at most τ and lowest possible danger amongst all those paths constructed. This value of α determines the selection function S_α used in the Pareto algorithm.

4.3. Experimental results

Efficiency of computation may be improved by adopting a heirachical or multi-resolution approach. An approximate path $p^{(1)}$ is obtained using a course grid on the transit region. This grid is then restricted to a subregion that includes $p^{(1)}$, and refined. In this way, a sequence $(p^{(n)})$ of approximate paths is determined. As n increases, the submarine is given increased flexibility to change direction and speed. Efficiency can also be improved using some of the more sophisticated data management techniques available when coding the algorithm.

The attached figure shows approximately optimal paths obtained for one sensor geometry with different time constraints, under the assumption that the probability model for each sensor is the same and that only two speeds were possible. Both paths are shown on the danger-map corresponding to the lower speed, and the darker sections of each path correspond to higher speeds. A heirachical approach was used on a transit region of 62km $\times$ 62km, starting with a grid with spacing 3.26km and ending after the second iteration with a grid with spacing 1.63km. The longest CPU time for path construction was less than 1 second.

This technique can be used to deal with situations in which three dimensional motion is allowed, and directionally biassed sensors used. However, the appropriate probabilities must be available and there will be a heavier computational burden. We could also add additional constraints (pruning criteria) to allow for such requirements as the need for a diesel submarine to recharge its batteries near the surface, thus increasing its probability of detection, after extended periods submerged.

Figure 2: Approximate safe paths with time contraints of (1) 500min (ii) 428min

REFERENCES

1. Peter J. Bentley and Jonathan P. Wakefield, *An analysis of multiobjective optimisation with genetic algorithms*, preprint.
2. J.C.N. Climaco and E.Q.V. Martin, *A bicriterion shortest path algorithm*, Eur. J. Op. Res., 11 (1982), 399-404.
3. P. Hansen, *Bicriterion path problems*, in G. Fandel and T. Gal, editors, Multiple criteria decision making: theory and applications, Lecture Notes in Economics and Mathematical Systems 177, Springer-Verlag, 1980, 109-127.
4. D.B.Johnson, *Efficient algorithms for shortest paths in sparse networks,* J.of ACM, 24, 1977, 1-13.
5. E.Q.V. Martins, *On a multicriteria shortest path problem*, European J. Op. Res., 16, 1984, 236-245.
6. Joseph S.B. Mitchell and Christos H. Papadimitriou, *The weighted region problem: finding shortest paths through a weighted planar subdivision*, J. Assoc. Computing Machinery, 38 (1), 1991, 18-73.
7. Mario C. Selvestral and Simon Goss, *Optimal flight paths for aircraft: a little knowledge goes a long way*, Industrial and Engineering Applications of Artificial Intelligence and Expert Systems, IEA/AIE-95, 8^{th} International Conference, 1995.
8. Chi Tung Tung and Kim Lin Chew, *A bicriterion pareto-optimal path algorithm*, Asia-Pacific J. Op. Res., 5, 1988, 166-172.
9. Chi Tung Tung and Kim Lin Chew, *A multicriteria pareto-optimal path algorithm*, European J. Op. Res., 62, 1992, 203-209.

QUANTUM ALGORITHMS

K. J. Harrison
Department of Mathematics and Statistics, Murdoch University, WA 6150, AUSTRALIA

We outline the basic framework of quantum computing, and describe some key quantum algorithms and their applications. We also identify the crucial role of the Fourier transform in some of these algorithms. It is reasonable to expect that new, more powerful algorithms can be found which further exploit the Fourier transform in its various forms.

1. ORIGINS

Simulating quantum mechanical processes on conventional computers is generally computationally infeasible. Such a simulation typically involves an exponential slowdown in time compared to the evolution itself, since the amount of information needed to describe the evolving quantum system in classical terms grows exponentially with time. However in 1981 Feynman [7] suggested that it ought to be possible to turn this around and treat it as opportunity rather than an obstacle. He argued that by regarding the measurements obtained from experiments carried out on certain types of quantum mechanical devices as the results of complex computations, it might be possible to perform certain computational tasks beyond the reach of any conceivable conventional computer.

This vision of Feynman has stimulated research in various directions, including, of course, the search for ways of building devices that can function usefully as quantum computers. But is this really worth doing? Would quantum computers be significantly more powerful than conventional computers? Benioff [1] showed as early as 1980 that any computation that could be done by a conventional computer could, in principle, also be done by a quantum computer. Within the next ten or so years formal models of quantum computing had been developed and a number of contrived problems that could be solved more efficiently on quantum computers had been discovered [6], [2] [12].

However the first real breakthrough came in 1994 when Shor [13] presented an efficient quantum algorithm for factorization. This generated a great deal of excitement because no such conventional algorithm is known. It is widely suspected (but not proven) that none exists and that factorization is NP-hard. Indeed the very difficulty of factorizing large numbers is the key to the security of many commercially used encryption schemes. These will be rendered useless, of course, if the implementation of Shor's algorithm on quantum computers becomes a reality.

The next big step was taken in 1996 by Grover [8], [9], who presented a quantum search algorithm which gave a quadratic improvement over existing search methods. He has since adapted the key idea of his search algorithm to develop efficient quantum algorithms for solving problems not immediately related to searching [10].

In recent years efforts have been made to unify known quantum algorithms within various frameworks. For example, in 1995 Kitaev [11] recognized the crucial role of the Fourier transform of various finite Abelian groups, and exploited this in developing other algorithms for solving group-related problems. On the other hand, in a very recent paper [3] interesting and potentially powerful links were made between quantum algorithms and multi-particle interferometer experiments.

2. MODELS FOR COMPUTATION

All models of 'reasonable' conventional computers are equivalent, in the sense that any universal machine can simulate any other machine with at most a 'polynomial' slowdown. It suits our purpose here to describe conventional computers as devices that store binary digits (*bits*) in locations known as *registers*. These bits can be read (measured) at any time and they can be used as the inputs for Boolean functions. The outputs of these functions can then be stored in the registers, overwriting the existing values if necessary. In this setting a conventional algorithm is merely a sequence of Boolean functions, together with the appropriate bookkeeping that tracks the locations of the inputs and outputs of these functions. These functions are usually Boolean *gates* which act on just a few bits at a time. Such an algorithm can be represented graphically as a Boolean network or circuit. We apply an algorithm by prescribing the *initial state* of the registers (i.e. a binary string describing the values of the stored bits), and then applying the functions in the prescribed order. The *output* of the algorithm is just the final state of the registers.

Many commonly used Boolean functions (such as AND) are not $1-1$. So the algorithms described above are not *reversible*, in the sense that we cannot always recover the initial state from the final one. However by including extra bits (in registers commonly called *scratch pads*), we can assume that we are working with $1-1$ functions and our algorithms are reversible. To see this, let $B^{(n)}$ denote the set of all Boolean n-strings $x = x_1x_2x_3\ldots x_n$, where $x_i \in B = \{0,1\}$ for each i, and suppose that $f : B^{(n)} \to B^{(m)}$. Then the function $F : B^{(n+m)} \to B^{(n+m)}$ defined by $F(x,y) = (x, y \oplus f(x))$, where $\oplus$ denotes the bitwise *exclusive or (XOR)*, is $1-1$. Since the evolution of isolated quantum systems is reversible, models of reversible computing are more adaptable to quantum computation.

Whereas the fundamental unit of conventional information is the *bit*, the corresponding unit of quantum information is the *quantum bit* or *qubit*. The quantum state of single qubit system is a unit vector $|\psi\rangle$ in a two-dimensional complex inner product space, which we denote by $\mathcal{B}$ or $\mathcal{B}^{(1)}$. The space $\mathcal{B}$ has a distinguished orthonormal basis whose elements are labelled $|0\rangle$ and $|1\rangle$, and so the state $|\psi\rangle$ of the qubit can be expressed as a linear combination

$$|\psi\rangle = a\,|0\rangle + b\,|1\rangle\,, \text{ where } a, b \in \mathbf{C} \text{ and } |a|^2 + |b|^2 = 1. \qquad (1)$$

A *measurement* of $|\psi\rangle$ can be regarded as a projection onto one or other of the basis vectors $|0\rangle$ or $|1\rangle$. The outcome is not deterministic. In fact we obtain the result $|0\rangle$ with probability $|a|^2$ and $|1\rangle$ with probability $|b|^2$.

The state of an n-qubit quantum system is a unit vector $|\psi\rangle$ in a complex inner product space of dimension 2^n, which we denote by $\mathcal{B}^{(n)}$. This can be regarded as a tensor product of n copies of $\mathcal{B}$. The distinguished basis of the quantum state space $\mathcal{B}^{(n)}$ consists of all states in which each qubit has a definite value, either $|0\rangle$ or $|1\rangle$. These states are known as the *classical states* of the system, and can be labelled as binary n-strings $|x\rangle = |x_1 x_2 \dots x_n\rangle$, where each $x_j \in B$. Thus if $B^{(n)}$ denotes the set of all binary n-strings we can express $|\psi\rangle$ in the form

$$|\psi\rangle = \sum_{x \in B^{(n)}} a_x |x\rangle \,, \text{ where each } a_x \in \mathbf{C} \text{ and } \sum_{x \in B^{(n)}} |a_x|^2 = 1. \qquad (2)$$

If we measure all n qubits of the system in parallel, the state of the system becomes one of the classical states, and the probability of obtaining any one such state $|x\rangle$ is $|a_x|^2$.

A *quantum algorithm* can be described in the following simple manner. We start with the n qubits in a classical initial state such as $|000\dots0\rangle$ and then apply a unitary transformation. This is usually a product of standard *quantum gates* that act on just a few qubits at a time. The output of the computation is then obtained by measuring some or all of the qubits.

The probabilistic nature of the output of a quantum algorithm is an important difference from conventional computing. Unless the final state is one of the classical states rather than a superposition, repetitions of a quantum algorithm can produce different results.

Conventional computers can store and rotate vectors, and can simulate the quantum measuring process of projecting onto mutually orthogonal axes. So conventional computers can do anything quantum computers can do. The difference is in the speed and storage requirements. For example, merely to represent on a conventional computer a typical state $|\psi\rangle$ of an n-qubit quantum system, we need to store the 2^n coefficients a_x as in (2).

3. ALGORITHMS

The first quantum algorithms were given by Deutsch [4] [5] [6], and were designed to determine whether a given Boolean function possesses certain *global* properties, (i.e. joint properties of all the function values). The aim is to use a minimum number of function evaluations. By concentrating on global properties the algorithms are attempting to exploit the *parallelism* inherent in quantum mechanics.

3.0.1. Deutsch's algorithm

The algorithm now known as Deutsch's algorithm has evolved since its first appearance in 1985. In its most recent and powerful form we are given a function $f : B^{(n)} \to B$, and are told that f is either *constant* or *balanced*, that is, the values of f are either all the same or there are equally many zeroes and ones, (2^{n-1} of each, in fact). The problem is to decide whether f is constant or balanced. Any conventional algorithm may require as many as $2^{n-1} + 1$ function evaluations, but

Deutsch's algorithm requires only one application of $\mathcal{U}_f$, the quantum version of f, and $O(n)$ applications of standard quantum gates and register measurements.

The advantage of this quantum algorithm over conventional methods is lost, however, if we allow errors into our computations. If we are satisfied to guess whether f is balanced or constant on the basis of a number of function evaluations, then we can guess correctly with probability $1-\varepsilon$ using $O(\log(1/\varepsilon))$ evaluations. Since we can have an exponentially good probability of success with just polynomially many trials, the problem cannot really be regarded as hard.

3.0.2. Simon's algorithm

In 1993 Bernstein and Vazarini [2] gave a variation of Deutsch's problem which *is* hard conventionally and which showed for the first time that quantum computation is significantly more powerful than conventional computation. Simon [12] soon followed with a simpler example. Simon's algorithm has turned out to be quite significant since it established a pattern that has been subsequently used and generalized.

In Simon's problem we are given a Boolean function $f : B^{(n)} \to B^{(n)}$ which we are told is $2-1$ and has unknown period ξ, i.e. $f(x) = f(y)$ if and only if $y = x \oplus \xi$, $x, y \in B^{(n)}$. The problem is to determine the period ξ. Since f has 2^{n-1} distinct values, if we try to solve the problem simply by searching through the values of f we may need as many as $2^{n-1}+1$ values before we find a match (and hence the period ξ). A similar exponential number of evaluations is needed in order to guarantee any given probability of success if the observations are noisy. However the quantum algorithm presented by Simon achieves a given probability of success in determining ξ with $O(n)$ observations, even in the presence of noise.

3.0.3. Shor's algorithm

Shor's algorithm [13] is a method for factorizing a given positive integer N. It does this by solving an equivalent problem, that of finding the *order* of any number y coprime to N. (This is the least positive integer r for which $y^r \equiv 1 \mod N$.) The algorithm has essentially the same formal structure as Simon's algorithm, but it uses a quantum version of the discrete Fourier transform (DFT). A key ingredient of Shor's algorithm is the use of an efficient quantum circuit for evaluating the DFT for $\mathcal{Z}_N$, the additive group of integers $\mod N$. The quantum Fourier transform (QFT) for $\mathcal{Z}_N$ requires $O(\log N)^2)$ steps, which is an exponential improvement on both the standard method $\left(O(N^2)\right)$ and the fast Fourier transform $(O(N \log N))$.

3.0.4. Grover's algorithm

In its original form Grover's algorithm gave a method of speeding the identification of a particular object in a large data base. If the data base has size N, then it takes $N/2$ look-ups to be guaranteed a 50% chance of finding the one we want. Grover [8] presented a quantum algorithm which is almost certain to succeed after $O(\sqrt{N})$ 'observations'.

In the quantum setting of Grover's algorithm, the unknown object is represented as a 'marked' classical state ω in $B^{(n)}$, where $N = 2^n$. The algorithm starts by preparing the balanced superposition $s = 2^{-n/2} \sum_{x \in B^{(n)}} |x\rangle$. If we measure the system now, the chance of obtaining ω is merely $1/N$. The key to the algorithm

is a unitary operator $\mathcal{U}$ which rotates this state towards ω. This operator $\mathcal{U}$ is the product of two reflections. The first of these is the quantum equivalent of the characteristic function of ω, and the second reflection is associated with the superposition s. After $\pi\sqrt{N}/4$ applications of $\mathcal{U}$, the state of the system is very close to ω, and so a measurement will now almost certainly produce ω.

Since Grover's algorithm gives a quadratic improvement over conventional search methods, the question naturally arises as to whether quantum computers have the potential to do even better. In other words, is there a quantum search algorithm that requires significantly less than $\pi\sqrt{N}/4$ observations? It turns out that no such algorithm exists. In order to distinguish all possible values of the marked state, it is necessary to have at least $\sqrt{N/2}$ observations [14]. Grover's algorithm exceeds this bound by less than 12%.

This algorithm has since been adapted to develop efficient quantum algorithms for estimating means and standard deviations of statistical distributions [10].

4. FOURIER TRANSFORMS

It has recently been recognized [3],[11] that all of these key quantum algorithms, except for Grover's, solve problems that can be expressed in group theoretic terms as examples of finding 'hidden subgroups' of a finite Abelian group. In each case the algorithm begins by preparing a superposition of states using the quantum version of the appropriate Fourier transform. For the algorithms of Deutsch and Simon, the underlying group is $\mathcal{Z}_2^n$ and the Fourier transform is a Walsh-Hadamard transformation. For Shor's algorithm $\mathcal{Z}_N$ is the underlying group, and in this case we use the quantum equivalent of the standard discrete Fourier transform. The point of the Fourier transform is that it useful in recognizing periodicity. The role of periodicity is plainly visible in Shor's problem, but it has also been shown [3] to be present in disguised forms in the problems of Deutsch and Simon. In [11] Kitaev devised new quantum algorithms for computing stabilizers of actions of finite, Abelian groups. These too were based on quantum versions of the appropriate Fourier transform, and for which he also gave efficient algorithms.

As noted in [3], we can expect to find other problems associated with the subgroup structure of groups that turn out to be amenable to efficient quantum computation.

REFERENCES

1. Benioff, P. *Quantum mechanical Hamiltonian models of Turing machines*, J. Stat. Phys. **29** (1982) 515.
2. Bernstein E. & Vazirani U. *Quantum complexity theory*, Proc. 25th ACM Symp. on the Theory of Computing (1993) 11-20.
3. Cleve R., Ekert A, Henderson L., Macchiavello C., & Mosca M. *On quantum algorithms*, (1999) lanl e-print 990361
4. Deutsch, D.. *Quantum theory, the Church-Turing principle and the universal quantum computer*, Proc. Roy. Soc. London Ser. **A, 400** (1985) 97.
5. Deutsch, D.. *Quantum computational networks*, Proc. Roy. Soc. London Ser. **A, 425** (1989) 73.
6. Deutsch, D. & Jozsa, R. *Rapid solution of problems by quantum computation*, Proc. Roy. Soc. London Ser. **A, 439** (1992) 553-558.
7. Feynman, R.P. *Simulating physics with computers*, Int. J of Theoretical Physics, **21** (1982), 467-488.

8. Grover L.K. *A fast quantum mechanical algorithm for database search*, Proc. 28th Annual Symposium on the Theory of Computing, (1996) 212-219.

9. Grover L.K. *Quantum mechanics helps in searching for a needle in a haystack*, Phys. Rev. Letters **76**(2), (1996) 325-328.

10. Grover L.K. *A framework for fast quantum mechanical algorithms*, Proc. 30th Annual Symp. on the Theory of Computing (1998), lanl e-print 9711043.

11. Kitaev, A.Yu. *Quantum measurements and the abelian stabilizer problem*, lanl e-print 9511026.

12. Simon, D. *On the power of quantum computation*, Proc. 35th Annual Symp. on the Fundamentals of Computer Science (1994), 116-123.

13. Shor. P.W. *Algorithms for quantum computation: discrete log and factoring*, Proceedings of the 35th Annual Symp. on the Fundamentals of Computer Science (1994) 124.

14. Zalka C., *Grover's Quantum Searching Algorithm is Optimal*, lanl e-print 9711070

Object Enhancement in Time-Frequency Scans of Communications Environments

John Hefferan

Communications Division, DSTO

In this work we apply a nonlinear morphological filter with a one-dimensional structure element to the task of improving the "visibility" of several signals of interest (SOI). Two dimensional time-frequency scans representing a series of wideband snapshots of a communications environment are analysed. The morphological filter approach is compared with conventional thresholding and with no filtering. It is found that the morphological filter can improve the SOI visibility at specified levels of environment noise. Signals considered include frequency hopping and time co-incident multi-tone transmissions.

Key Words: Morphological filters; time-frequency scans

0. INTRODUCTION

The analysis of how spectral information varies over time has been the source of much research in recent times (see [1] for a recent review). Much of the work has been aimed at developing transforms which may be applied to time domain signals and thereby produce a set of characteristic features in the time-frequency domain. However, in wideband scans of communications environments, signals of interest are often represented as an increase in energy in a single frequency bin in the spectral FFT "snapshot". This coarse resolution, necessary to capture wideband environments, means that employing conventional transforms provides little benefit in detecting the presence of a SOI when compared with the observation of the level of spectral energy at specific frequencies and times.

Representation of these wideband scans as a two dimensional rasterised image presents visual observers with, essentially, a task in image recognition. The task becomes one of detecting the presence or absence of a SOI based on the carrier level representation observed in the scan. Further observations such as co-incident transmissions or transmissions related in time or frequency may also further characterise a particular SOI.

In this work we have applied morphological filter techniques to the task of improving the "visibility" of several signals of interest present in two-dimensional time-frequency scans of a communication environment.

1. MORPHOLOGICAL FILTERS

Morphological filters are well known in the field of image processing [2,3,4]. Essentially, they are a class of nonlinear filters capable of enhancing the features and boundaries present in noise contaminated objects.

As noted in the literature [2] a nonlinear closure filter can be defined for any linear space (eg: R^2 and R^3). Closing of a set X by B is defined as

$$(X \oplus B) \ominus B \tag{1}$$

Here, X represents the image and B is the set called the "structure element". The operators $\oplus$ and θ are respectively, the Minkowski sum and difference operators [2]. The operator $\oplus$ is also called the dilation of X by B, ie:

$$X \oplus B = \{h \mid [(B')_h \cap X] \neq \varnothing\} \tag{2}$$

Here, B' is the reflection of set B, ie

$$B' = \{-b : b \in B\} \tag{3}$$

Dilation of X by B is thus the set of all non-zero intersections between set X and all h translations of the reflection of the structure element, B. Erosion of X by B may be defined by

$$X \ominus B = \{h \mid (B)_h \subseteq X\} \tag{4}$$

Erosion of X by B is the set of all h translations of the structure element B which are contained within set X. We can see from (1) therefore, that closure can be thought of as the dilation of set X by B followed by the erosion of the result by B.

Usually, the structure element B, is a sphere (in R^3) or disc (in R^2). Hence, applying the closure filter is analagous to rolling or sliding the sphere or disc around the outside of the object for dilation, then around the inside of the dilated object for erosion, thereby altering the "morphology" of the original object. Closure also has the property of idempotence, which means that applying the closure operation more than once has no further effect on the result.

2. A MORPHOLOGICAL FILTER FOR NARROWBAND AND FREQUENCY HOPPING SIGNALS

In the case of two dimensional time-frequency scans of a communications environment, spectral signatures of various emissions can be viewed as individual noise contaminated objects which vary as a function of time. For narrowband emissions, with an appropriate selection of FFT size, the object can be contained entirely within one "spectral bin". When this is done there is no (or negligible) correlation between adjacent frequency bin components at any given time t. Hence, we can define a simplified structure element as the one dimensional column vector B, of dimension $M \times 1$. In this work, M is defined as the "window-length" of the morphological filter.

For the case of M = 3 we have;

$$B = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}. \tag{5}$$

The application of a closure filter to a 2D time-frequency scan using this structure element involves sliding B along M adjacent time samples in the scan and performing a dilation. This is then followed by a scan with an erosion operation, as defined in (4) above, again using B. In this present work the closure operation on the scan has been augmented by a number of other operations.

Firstly, a data reduction step is carried out where the noise floor of the scan is estimated and then an appropriate threshold level set. The threshold level (T), below which data is removed, is set at

$$T = mean(X) + 2\sigma \tag{6}$$

Here σ is the standard deviation of the distribution of background noise levels in the scan. The effect of thresholding is to remove much of the low-level background noise in the scan and leave only candidate

objects of level greater than 2 standard deviations above the noise floor mean. Following the data reduction by thresholding, a modified dilation operation is applied.

In this work the dilation is preceeded by setting

$$(X \cap B)^{\prime} = \max(X \cap B) \tag{7}$$

for each *h* translation of B applied to X. This effectively sets each element of the intersection equal to the maximum in that intersection. The dilation is then applied to the $(X \cap B)''$ term. The purpose of this operation is to reduce the effect of short duration fades in any narrowband object present in the image.

During the erosion phase, the application of B with M=3 effectively removes objects of less than three time divisions duration. This "decluttering" of the image is designed to remove short duration spurious interferers and higher level background impulsive noise present in the scan.

3. THE COMMUNICATIONS SYSTEM

Simulated communication scans of different emission types have been generated and applied to one of three filters. The emission types considered were:

1. A single Narrowband tone. The tone ($A\sin(2\pi f_c t)$) commences at time t_1 and ends at time t_2. Here, $0 < \{t_1,t_2\} < T$ where T is the time of the final spectral estimation of the environment. The f_c term is the carrier frequency of the emission;

2. Several adjacent narrowband tones, co-incident in time. Here, the tones are defined by $A\sin(2\pi(f_c + nf_d)t)$ where f_d is the separation between adjacent tones and n is an element of $\{n: |n| = 0,1,2,3,4,\ldots\}$. Once again the tones start at t_1 and t_2 where $0 < \{t_1,t_2\} < T$ as above;

3. Frequency hopping narrowband emissions. A series of tones with randomly changing frequency steps were generated.

Each of these emissions was contaminated with Additive White Gaussian Noise (AWGN) of varying levels and then filtered using one of the filters described below.

Scans of the received signals were constructed by taking successive spectra of the time domain signal and assembling the spectra into a two dimensional (time-frequency) file. In this case the files were $N_f = 512$ frequency bins wide by $N_t = 155$ successive spectral samples.

The filters compared in this work were:

1. No filter. Here, the noise contaminated scan is passed through unchanged;

2. Threshold Filter. The noise floor is estimated and a threshold level set as in (6) above. Scan data above the threshold level is passed unchanged whereas data below the threshold is filtered out;

3. Modified Morphological Closure Filter. A modified form of the morphological filter implementing the closure operation with a structure element as in (5) was implemented. The dimension of the structure element (M) was able to be varied.

The "quality" of the filtered scans was then estimated and compared over different noise levels and with different filter types used in the filtering stage. The performance of the filters was measured in several ways.
Firstly, the filtered scan (X_f) was compared with a scan of the original transmission (S). The sum squared error between the two scans was calculated at differing noise levels.

$$SquaredError = \frac{1}{N}\sum_{i=1}^{N}\left(x_{fi} - s_i\right)^2 \quad (8)$$

Here, $N = N_f \times N_t$, the total number of data elements in each scan and $\{x_i \in X\}$, $\{s_i \in S\}$.

Secondly, an attempt has been made to estimate the "visibility" of the signal of interest (SOI) in the filtered scan. Here, visibility is also relevant as it indicates the ease with which specific SOI could be detected by visual inspection of a rasterised representation of the 2D time-frequency scan. The visibility of the SOI (V_{SOI}) can be estimated by calculating the ratio of the SOI average intensity after filtering, *mean*(SOI_o), to the level of the background noise floor after filtering. This can be expressed as;

$$V_{SOI} = mean(SOI_o) / mean(X_f) \quad (9)$$

4. RESULTS

The squared error obtained when each of the filters is used with received scans at different noise levels is shown graphically in Figure 1 for the case of the Frequency Hopping SOI. The noise level is measured as the Carrier to Noise ratio (CNR) in dB. From this graph it can be seen that, as expected, the threshold filter substantially reduces the square error compared with no filter as the noise level increases. The further reduction in squared error obtained when the morphological filter is used is due to the filtering of *all* clutter less than M=3 raster periods in length.

SOI visibility for M = 2 and M = 3 is depicted in Figures 2 and 3 respectively. In Figure 3 it can be seen that the morphological filter with M = 3 gives very high visibility at lower background noise floor levels. However, as the noise increases, the morphological filter performance degrades giving, at best, a 4dB improvement in CNR over threshold filtering. The degradation in the morphological filter performance is due to the presence of large bursts of noise and the increasing number of fades in the SOI. The filter will remove fragments of the SOI when the fades produce excessive segmentation.

5. CONCLUSION

The modified morphological filter developed in this work has been shown to reduce the error in received time-frequency scans of an AWGN channel with several different SOI. For a given value of background noise, the morphological filter is seen to perform better than conventional thresholding of the input scan and also when no filtering is used.

For the signals tested, the M=3 morphological filter has been shown to increase the SOI_o/noise floor ratio providing approximately a 4dB gain in CNR over threshold filtering. This would effectively extend the detectable range of a SOI to an automated or manual detection system.

A potential application of this filter includes that of a "front-end" for a signal classifier system. The ability of the filter to increase the visibility of frequency hopping signals could also be developed.

REFERENCES

1. Qian,S and Chen,D., "Joint Time-Frequency Analysis", *IEEE Signal Processing Magazine*, March 1999, pp 52-67.

2. Serra,J.,*Image Analysis and Mathematical Morphology,* Academic Press,New York, 1982.

3. Gonzalez,RC and Woods,RE., *Digital Image Processing*, Addisson Wesley, 1993, §8.4.

4. Thomson,CM and Shure,L., *Image Processing Toolbox for Use with Matlab*, The Math Works Inc., 1993.

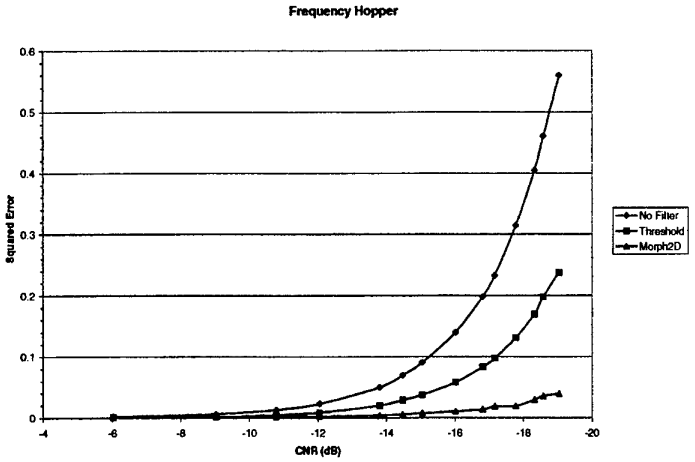


FIG. 1. Squared Error vs. CNR (SOI = FH)

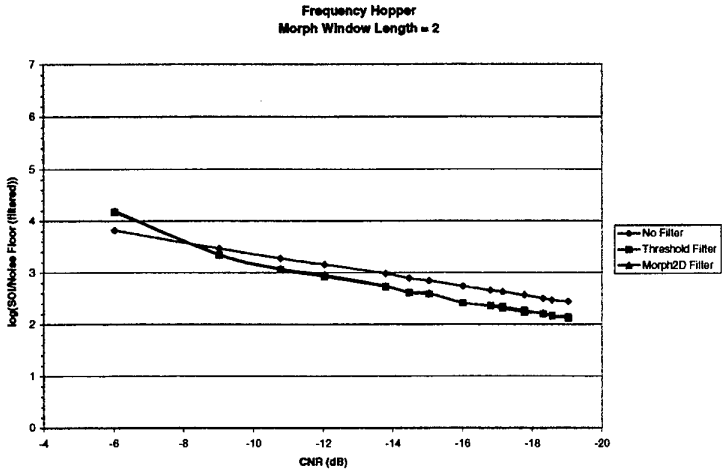


FIG. 2. SOI Visibility – FH – Morph Window Length = 2

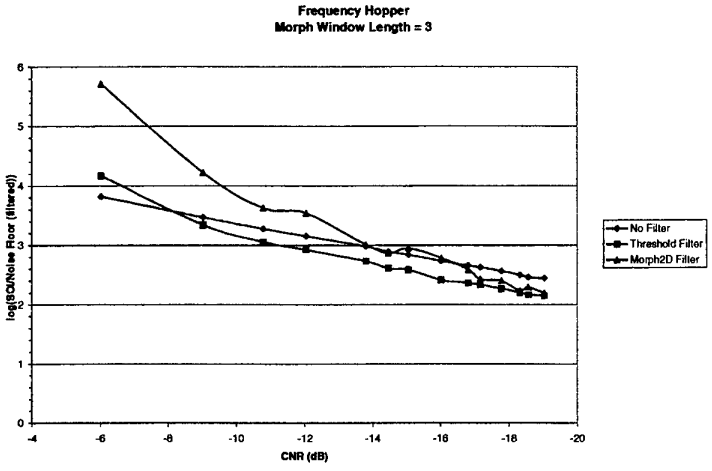


FIG. 3. SOI Visibility – FH – Morph Window Length = 3

A Speech Segmentation Algorithm with Application to Speaker Identification.

By: O.P. Kenny, R.C. Price, and J. Willmore.

Radar Target Identification Group, and Information Management and Fusion Group, Defence Science and Technology Organisation, P.O. Box 1500, Salisbury, 5108, Australia.

ABSTRACT.
Speaker identification has become an increasingly useful tool in areas such as forensics and for military applications. Commonly used speaker identification systems are based on statistical based models to represent speech [1]. A speaker is chosen by its likelihood function for a given statistical model. In practical situations, it is common for more than one speaker to be present.

The focus of this paper is to determine if there is an advantage in segmenting speakers before forming the log-likelihood scores for test speech utterances. Our aim is to segment the speakers without adversely increasing the processing required for speaker identification. Consequently, this constraint was reflected in the implementation of the segmentation algorithm.

Experimental results, based on marked data, have shown the performance of the segmentation algorithm varies depending on the set of speakers. It has been found that segmentation performance is estimated to be between 70 and 95 percent. An improvement to the performance of speaker identification was also observed, after segmentation had been performed.

1.0 Introduction.
Speaker identification has been the subject of research from the mid seventies through to the present day. Early speaker identification approaches used long-term averages of acoustic features such as the pitch [2], [3]. Long term averaging smoothed out phonetic variations leaving the speaker's vocal tract shape. This approach requires integration over long periods of speech, of the order of 20s. Other approaches compared individual phonetic sounds that compose the utterance [4], [5]. This comparison gave a measure of the distance from one speaker to another rather than the textual information. These methods used hidden Markov models for phonetic structure detection. Another commonly used approach is based on neural networks [6], [7]. This uses a closed set of speaker training data to form decision boundaries for the speakers.

We used two speaker identification approaches a statistical model and a neural net-based approach [8]. The statistical model system uses a Gaussian mixture model to represent the statistical information for the speaker's speech. Training data is used to form the parameter set of the model. Speaker identification is then achieved by obtaining log-likelihood scores for test utterances for a given speaker model. The speaker chosen is the one that gives the maximum log-likelihood score.

The focus of this paper is to find a computationally efficient algorithm to separate a set of speakers before identification. The speakers are segmented into two groups and used to form log-likelihood scores against each speaker model. The identified speaker is chosen to be the maximum in the class and the furthest away from the background speakers.

Speaker separation has appeared in the literature. Chen, Brown, and Bovey [9] developed an algorithm for speech segmentation for the work environment. This algorithm used pre-captured speech samples of potential speakers and found use as an audio indexing system.

Cohen and Lapidus [10] investigated the problem of unsupervised speech segmentation in telephone conversions. In this case, no a priori speaker information was assumed. This algorithm accepted dual telephone speech data, detected events of simultaneous speakers, and then segmented the speech assigning each to a group. Takagi and Itahashi [11] clustered Japanese utterances using seven prosodic feature vectors. Sugiyama, Murakami and Watanabe [12] investigated speech segmentation and clustering for an unknown number of multiple signal sources based on ergodic hidden Markov models where each speaker corresponded to a single signal source. Other segmentation algorithms that have been reported include those in [13] and [14].

The outline of this paper first reviews speaker identification based on statistical models. This is followed by the description of the speaker segmentation algorithm. Finally, experiments were then used to determine the effectiveness of this algorithm and to determine if it provides an improvement to speaker identification.

2.0 Review of the Speaker Identification based on Statistical Models.

Common speaker identification systems are based on statistical models. The speech first undergoes preprocessing including speech activity detection, and conversion to its corresponding cepstral coefficients. These coefficients describe predominate physiological factors that distinguishes one person's voice from another [1].

Statistical models are formed using Gaussian mixture models (GMMs) these are weighted summations of Gaussian functions each having their own mean and variance. These weightings, means, and variances form the parameter set for a particular speaker. The parameter set can be estimated using the expectation maximization algorithm or a form of vector quantisation from a source of training data.

The probability of a feature vector, $\vec{x}_t$, given a parameter set, λ, is given by,

$$p(\vec{x}_t|\lambda)=\sum_{i=1}^{N} p_i b_i(\vec{x}_t) \tag{1}$$

where

$$b_i(\vec{x})=\frac{1}{(2\pi)^{D/2}|\Sigma_i|^{1/2}}\exp\left\{-\frac{1}{2}(\vec{x}-\vec{\mu}_i)'\Sigma_i^{-1}(\vec{x}-\vec{\mu}_i)\right\}. \tag{2}$$

The parameter set for the GMM is denoted as $\lambda=\{p_i,\vec{\mu}_i,\Sigma_i\}$ for $i=1,\ldots M$ with the constraint that $\sum_{i=1}^{M} p_i=1$. It is assumed that the time observations for the feature vectors, $\vec{x}_t$, are statistically independent of each other. Consequently, the probability for a series of observations for a given model is,

$$p(X|\lambda)=\prod_{t=1}^{T} p(\vec{x}_t|\lambda). \tag{3}$$

Speaker identification is achieved by choosing the maximum probable model given the observation vectors,

$$\hat{S}=\arg\max_{1\le k\le S}\Pr(\lambda_k|X). \tag{4}$$

Bayes theorem allows us to rearrange Eqn (4) to give,

$$\hat{S}=\arg\max_{1\le k\le S}\frac{p(X|\lambda_k)\Pr(\lambda_k)}{p(X)}. \tag{5}$$

Eqn (5) can be reduced from the fact that certain probabilities equate to unity or have no direct influence on the maxima. Consequently, speaker identification is obtained by choosing the maximum likelihood of the observation vectors given a model formulized as,

$$\hat{S} = \arg\max_{1 \le k \le S} p(X|\lambda_k). \tag{6}$$

A further simplication is made from statistical independence and use of logarithms to yield,

$$\hat{S} = \arg\max_{1 \le k \le S} \sum_{t=1}^{T} \log p(\vec{x}_t|\lambda_k). \tag{7}$$

Intuitively, if there are observations that do not belong to the true speakers this effects the log-likelihood scores and may force a bias unto the result.

3.0 Segmentation Algorithm.

The paradigm for the segmentation algorithm is to partition the speech waveform into separate segments. Each of these segments belongs to a speaker, and is consequently referred to as utterance units. The locations of these utterance units are denoted by their start and stop positions in the speech waveform. The statistical parameters for the target speakers were already known. The target speaker statistical models initiate the partitioning of the utterance units into separate groups. Each group, corresponding to a speaker, form observation vectors for each person in the conversation. Normalized log-likelihood scores are obtained for the target speakers against each group. The most likely speaker has the maximum likelihood score and furthermost distance from the background speaker. A more detailed discussion of the paradigm's steps now follow.

In some practical situations, such as telephone conversions, not only is there more then one speaker but other types of signal as well. These signals include dial tone frequency modulation, faxes, or modems. It is necessary to first remove those signals that are not speech. The details of this processing is not discussed in detail here, but the end result of this initial processing produces speech data that has been converted to a 16 bit linear format.

Once in the proper format a speech activity detector is used to produce indexing to locate utterance units. The utterance units are extracted from the speech with use of an energy envelope detector. The time duration used in this detector was of the order of three second long, and formed continuous speech energy samples.

Generally, a region where the signal has a large energy concentration corresponds to "voiced" speech. It is these regions that contain physiological information about the speaker and are used to train statistical models. To determine the utterance location the signal energy must exceed a threshold. This threshold value is dynamically set in the sense that it adapts to environmental noise changes. To determine the threshold value an estimate of the signal to noise ratio is obtained by sorting the energy vector from lowest to highest value. Signal and noise energy estimates are obtained from the lower and upper quarters of the sorted energy vector. The value of the threshold is formulated by an equation, being a function of SNR, and takes into account the absence of speech signals.

To this point, the resulting speech segments are regions where the signal energy exceeds the calculated threshold and belongs to either speaker. However, for better performance it is desirable to have as many feature vectors assigned to a segment as possible. To achieve longer time duration segments the length and distance between segments are noted. The desire is to concatenate segment to form longer segments. If the distance between the two adjacent segments is less than the syllabic rate then the two segments are concatenated. This procedure is followed for the entire signal with the resulting utterance unit lengths being several words long. On completion of the concatenation process, the utterance units of smaller length are discarded to minimize the variance of the speaker log-likelihood estimates.

The utterance units are normalized and used to generate cepstral coefficients of time duration lengths of 20ms. The coefficients are formed from an auto-correlation function and taking the inverse Fourier transform of the logarithm of the spectrum. The first ceptral coefficient is discarded since it corresponds to the power of the signal

and does not convey information about the speech process. Finally, the mean of the coefficients is subtracted and the result is passed on to form speaker log-likelihood score.

Log-likelihood scores are obtained using Gaussian mixture models whose parameters are for the target speakers. The result is an array of log-likelihood scores against each target speaker and utterance units. Each element of the array is obtained by,

$$l_{ik} = \sum_{t=1}^{L} \log p\left(\vec{x}_t^i | \lambda_k\right). \tag{8}$$

Where $\vec{x}_t^i$ denotes the cepstral coefficient for the i^{th} utterance unit and t^{th} time observation. The relationship between the target speakers is used to determine which group the utterance unit belongs. The log-likelihood array can be viewed as a set of segmentation feature vectors for a set of observations. The feature vector are formed from the log-likelihood score for each target speaker [15], and the time observations correspond to the utterance units. The task now is given these feature vectors, for a set of observations, to separate them into two separate groups.

From the set of feature vectors two vectors are chosen that give the maximum separation distance between them assuming each belongs to each speaker. An initial projection vector is formulated using these two feature vectors. This projection vector is constructed in such a way that the common information is subtracted. The resultant projection operator places the feature vectors, for each time observation, into positive and negatives half spaces that separate the speakers.

Iteration is used to re-estimate the projection vector. The projected observations are sorted and percentiles of the feature vectors are grouped. The percentile is increased after each iteration. This process chooses feature vectors that best describe the group. A common feature for each group is extracted from the principle component of the singular value decomposition and the corresponding projection vector reconstructed. After the completion of the iteration the resulting projection vector is used to segment each speaker onto a half space.

Once the utterance units have been grouped then the overall log-likelihood for each group is obtained by the summation of the individual log-likelihoods for each utterance unit. Speaker identification can then be made.

4.0 Experimental results.

In this section an example is given to demonstrate the segmentation algorithm and to show the affect that mixed speakers has on log-likelihood scores. The second experiment performed evaluates the effectiveness speaker segmentation before performing speaker identification. The conclusions drawn from this experiment were based on several conversations.

Figure 1 shows a scatter plot for a two-person conversation illustrating each speaker has been separated. Figure 2 presents the result of the log-likelihood scores for the mixed speakers. This figure also gives the log-likelihood scores after the two speakers have been separated. The true speaker was speaker number 1. Before splitting the maximum log-likelihood corresponded to speaker number two, and after splitting it corresponded to the correct speaker. The intent of this plot was to demonstrate the fact that, in some cases, a speaker can bias the true log-likelihood scores. Preliminary results have shown there is an improvement in speaker identification using splitting in the testing procedure.

5.0 Conclusion.

This paper has reported on the hypothesis that speaker separation improves the performance of speaker identification. The situation investigated was for the two-speaker case. Utilizing the information in the log-likelihood scores of the utterance unit made it possible to separate the speakers. Regrouping and adding the log-likelihood scores results in a set of scores for each speaker. We then evaluated speaker identification to determine weather splitting improves the performance of the overall speaker identification.

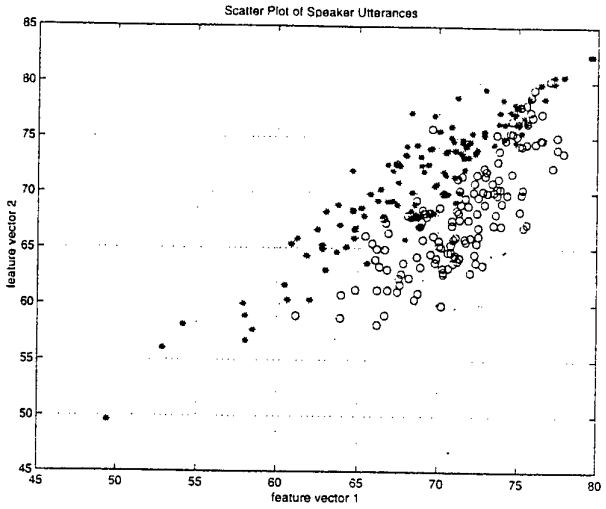


Figure 1 Scatter plot for the Segmentation of the Speakers.

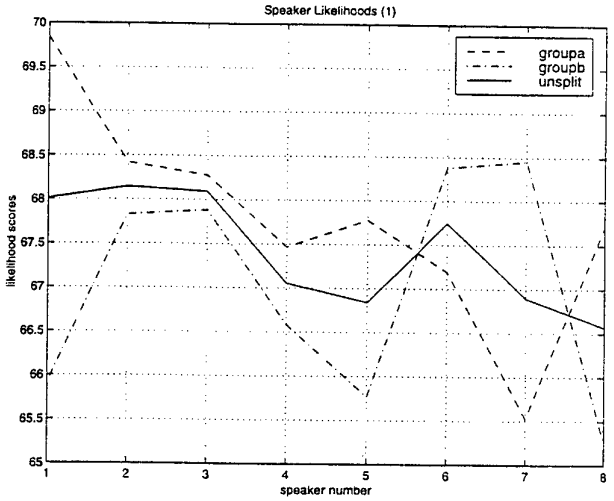


Figure 2 Plot of the log-likelihood scores.

References.

[1] "Robust Text-Independent Speaker Identification Using Gaussian Mixture Speaker Models", Douglas A. Reynolds and Richard C. Rose, *IEEE Transactions on Speech and Audio Processing*, Vol. 3, No. 1, January, 1995.

[2] "Talker Recognition by longtime averaged spectrum", S. Furui, F. Itakura, and S. Saito, *Electon., Commun. In Japan*, Vol. 55-A, No. 10, pp. 54-61, 1972.

[3] "Long-term Feature Averaging for Speaker Recognition", J. Markel, B. Oshika, and A. Gray Jr., *IEEE Transactions on Acoustics, Speech, and Signal Processing*, Vol. ASSP-25, pp. 330-337, August, 1977.

[4] "A Text-independent Speaker Recognition Method Robust against Utterance Variations", T. Matsui and S. Furui, *Proceedings of the IEEE ICASSP*, 1991, pp. 377-380.

[5] "Free-text Speaker Identification over long Distance Telephone Channel using Hypothesized phonetic Segmentation" *Proceedings of the IEEE ICASSP*, 1992, pp. II. 177-180.

[6] "Voice Identification using Nearest-neighbor distance Measure", A. Higgins, L. Bahler, and J. Porter, *Proceedings of the IEEE ICASSP*, April 1993, pp. II-375-378.

[7] "Radial basis functions networks for speaker recognition", J. Oglesby and J. Mason, *Proceedings of the IEEE ICASSP*, May 1991, pp. 387-390.

[8] "Implementation of a Speaker Identification System", William Roberts, Jonthan Willmore and Richard Price, Information Technology Division, Defence Science and Technology Organisation, 1998.

[9] "Speech Segmentation in Working Environments", X. Chen, P. Brown, and J. Bovey, *IEE Colloquium on Prospects for Spoken Language Technology*, p. 6/1-4, May, 1997.

[10] "Unsupervised Speaker Segmentation in Telephone Conversions", Arnon Cohen and Vladimir Lapidus, 1996 *Transactions of the IEEE*, pp. 102-105, 1996.

[11] "Prosodic Pattern of Utterance units in Japanese Spoken Dialogs", K. Takagi and S. Itahashi, *ICSLP 94*, Vol. 1, pp. 143-146, September 1994.

[12] "Speech Segmentation and Clustering Based on Speaker Features", M. Sugiyama, J. Murakam and H. Watanbe, *Proceedings of ICASSP*, Vol. 2, 1993, pp395-398.

[13] "Speech Segmentation and Clustering Problem Based on An Unknown-Multiple N Signal Source Model- An Application to Segmented Speech Clustering Based on Speaker Features", Masahide Sugiyama, *Systems and Computers in Japan*, Vol. 25, No. 9, 1994.

[14] "An Unsupervised, Sequential Learning Algorithm for the Segmentation of Speech Waveforms with Multiple Speakers", Man-Humg Siu, George Yu and Herbert Gish, *Proceedings of ICASSP*, Vol. 2, pp. 189-192.

[15] "Speaker Recognition System", Peter Chatalin, Ph.D. Thesis, Queensland University of Technology, 1997.

Finite Dimensional Algorithm for Optimal Scheduling of Hidden Markov Models Sensors. [1]

Vikram Krishnamurthy[†]

[†] *Department of Electrical and Electronic Engineering*
University of Melbourne, Parkville, Victoria 3052, Australia
vikram@ee.mu.oz.au

1. INTRODUCTION

There are several signal processing applications where a variety of sensors are available for measuring a given process, however physical and computational constraints may impose the requirement that at each time instant, one is able to use only one out of a possible total of M sensors. There is also growing interest in flexible sensors such as multi-mode radar which can be configured to operate in one of many modes for each measurement. In such cases, one has to make the decision: *Which sensor (or mode of operation) should be chosen at each time instant to provide the next measurement.* It may also happen that one can associate with each type of measurement a per unit-of-time measurement cost, reflecting the fact that some measurements are more costly or difficult to make than others, although they may contain more useful or reliable information. The problem of optimally choosing which one of the M sensor observations to pick at each time instant is called the *sensor scheduling problem.* The resulting time sequence which at each instant specifies the best sensor to choose is termed the *sensor schedule sequence.*

Several papers have studied the sensor scheduling problem for systems with linear Gaussian dynamics where linear measurements in Gaussian noise are available at a number of sensors (see [1] for the continuous-time problem and [7] for the discrete-time problem). For such linear Gaussian systems, if the cost function to be minimized is the state error covariance (or some other quadratic function of the state), then the solution has a nice form: the optimal sensor schedule sequence can be determined a priori and is *independent* of the measurement data (see [1], [7] for details). This is not surprising; since the Kalman filter covariance is independent of the observation sequence.

In this paper we study the discrete-time sensor scheduling problem when the underlying process is a finite state Markov chain that is observed in white noise. The signal model is as follows: At each time instant, observations of a Markov chain in white noise are made at M different sensors. However, only one sensor observation can be chosen at each time instant. The aim is to devise an algorithm that optimally picks which single sensor to use at each time instant, in order to minimize a given cost function. We will show that *unlike the linear Gaussian case, the optimal sensor schedule in the HMM case is data dependent.* This means that past observations together with past choices of which observation to pick influence which observation to choose at present.

In our recent work [4], we formulated the HMM sensor scheduling problem and presented an infinite dimensional dynamic programming functional recursion for its solution. An approximate algorithm was then presented in [4] which was based on discretizing the dynamic programming recursion to a finite grid.

The main contribution of this paper is to present an optimal **finite dimensional** solution to the HMM sensor scheduling problem. Indeed we show that the solution to the dynamic programming equation is piecewise linear and convex. An algorithm is given for computing these piecewise linear segments. The finite dimensional scheduling algorithms presented in this paper are similar to those recently used in the operations research (see [6] for a tutorial survey) and in robot navigation systems [3] for the optimal control of Partially observed Markov Decision Processes (POMDP). However, our problem has the added complexity that the cost function is quadratic function of the information state – whereas the standard POMDP problem consists of a cost that is linear in the information state. We show that by a novel change of coordinates, the problem can be re-expressed as a standard Hidden Markov Model control problem and optimally solved using similar algorithms to those use for solving POMDPs.

2. SIGNAL MODEL AND PROBLEM FORMULATION

Let $k = 0, 1, \ldots$ denote discrete time. Assume X_k is an S-state Markov chain with state space $\{e_1, \ldots, e_S\}$. Here e_i denotes the S-dimensional unit vector with 1 in the i-th position and zeros elsewhere. This choice of using unit vectors to represent the state space considerably simplifies our subsequent notation. Define the $S \times S$ transition probability matrix A as

$$A = [a_{ji}]_{S \times S} \text{ where } a_{ji} = P(X_k = e_i | X_{k-1} = e_j), \qquad i, j \in \{1, \ldots, S\}.$$

Denote the initial probability vector pi_0 of the Markov chain as

$$\pi_0 = [\pi_0(i)]_{S \times 1} \text{ where } \pi_0(i) = P(X_0 = i), \;\; i \in \{1, \ldots, S\}.$$

2.1. Sensor Scheduling Problem

Assume there are L noisy sensors available which can be used to give measurements of X_k. At each time instant k, we are allowed to pick only one of the L possible sensor measurements. Motivated by the physical and computational constraints alluded to in the introduction, we assume that having picked this sensor, we are not allowed to look at any of the other $L-1$ observations at time k.

Let $u_k \in \{1, \ldots L\}$ denote the sensor picked at time k. The observation measured by this sensor is denoted as $y_k(u_k)$. Suppose at time k, we picked the lth sensor, i.e. $u_k = l$, where $l \in \{1, \ldots, L\}$. We assume that the measurement $y_k(l)$ of the l-th sensor belongs to a known finite set of symbols $O_1(l), O_2(l), \ldots, O_{M_l}(l)$. That is the l-th sensor can yield one of M_l possible measurement values at a given time instant. For $u_k \in \{1, \ldots, L\}$, denote the symbol probabilities as $b_i(u_k = l, y_k(u_k) = O_m(l)) = P(y_k(u_k) = O_m(u_k) | X_k = e_i, u_k = l)$, $i = 1, 2, \ldots, S$. These represent the probability that an output $O_m(l)$ is obtained given that the state of the Markov chain is e_i and that the lth sensor is chosen. The symbol probabilities are assumed

known. Finally define the symbol probability matrix

$$B(u_k, O_m(u_k)) = \text{diag}\begin{bmatrix} b_1(u_k, O_m(u_k)) \\ \vdots \\ b_S(u_k, O_m(u_k)) \end{bmatrix}$$

Finally, for notational convenience let

$$\phi = (A, B(l, O_m(l))), \quad l = 1, \ldots, L, \quad m_l = 1, \ldots, M_l \tag{1}$$

denote the entire parameter vector which comprises of the transition probability matrix and all the symbol probabilities of all the L sensors.

Let $Y_k = \{u_1, u_2, \ldots, u_k, y_1(u_1), y_2(u_2), \ldots, y_k(u_k)\}$ so that Y_k represents the information available at time k upon which to base estimates and sensor scheduling decisions. The sensor scheduling and estimation problem proceeds in three stages for each $k = 0, 1, \ldots, N-1$, where N is a fixed positive integer

1) Scheduling: Based on Y_k we generate $u_{k+1} = \mu_{k+1}(Y_k)$ which determines which sensor is to be used at the next time step.

2) Observation: We then observe $y_{k+1}(u_{k+1})$ where u_{k+1} is the sensor selected in the previous stage.

3) Estimation: After observing $y_{k+1}(u_{k+1})$ we generate our best estimate $\hat{X}_{k+1}$ of the state of the Markov chain X_{k+1} as $\hat{X}_{k+1} = \mathbb{E}\{X_{k+1}|Y_{k+1}\}$. Here $\hat{X}_k$ denotes the Hidden Markov Model filtered state estimate at time $k+1$ defined as

$$\hat{X}_{k+1} = \mathbb{E}\{X_{k+1} \mid Y_{k+1}\} = \sum_{i=1}^{S} e_i \, P(X_{k+1} = e_i \mid Y_{k+1}), \qquad \hat{X}_0 = \pi_0.$$

Note that the state estimate $\hat{X}_{k+1}$ is dependent on the scheduling sequence of sensors picked from time 1 to $k+1$, i.e. $u_1, \ldots, u_{k+1}$ (since it depends on Y_{k+1}).

With these steps in mind, we define the sensor scheduling sequence

$$\mu = \{\mu_1, \mu_2, \ldots, \mu_N\}$$

and say that the scheduling sequences are admissible if μ_{k+1} maps Y_k to $\{1, \ldots, M\}$. Note that μ is a sequence of functions.

We assume the following cost is associated with estimation errors and with the particular sensor schedule chosen. If based on the observation at time k, the decision is made at time k to choose the l-th sensor, $l \in \{1, \ldots, L\}$ at time $k+1$, i.e. $u_{k+1} = l$, the instantaneous cost incurred at time k is

$$(X_k - \hat{X}_k)' R_k(l)(X_k - \hat{X}_k) + c_k(X_k, l)\}. \tag{2}$$

Here $R_k(l)$, $l = 1, 2 \ldots, L$ are known positive definite weighing matrices and

Our aim is to find the optimal sensor schedule to minimize the total accumulated cost J_μ from time 1 to N over the set of admissible control laws:

$$J_\mu = \mathbb{E}\{\sum_{k=0}^{N-1} (X_k - \hat{X}_k)' R_k(u_{k+1})(X_k - \hat{X}_k) + \sum_{k=0}^{N-1} c_k(X_k, u_{k+1})\} + \mathbb{E}\{(X_N - \hat{X}_N)' R_N (X_N - \hat{X}_N)\} \tag{3}$$

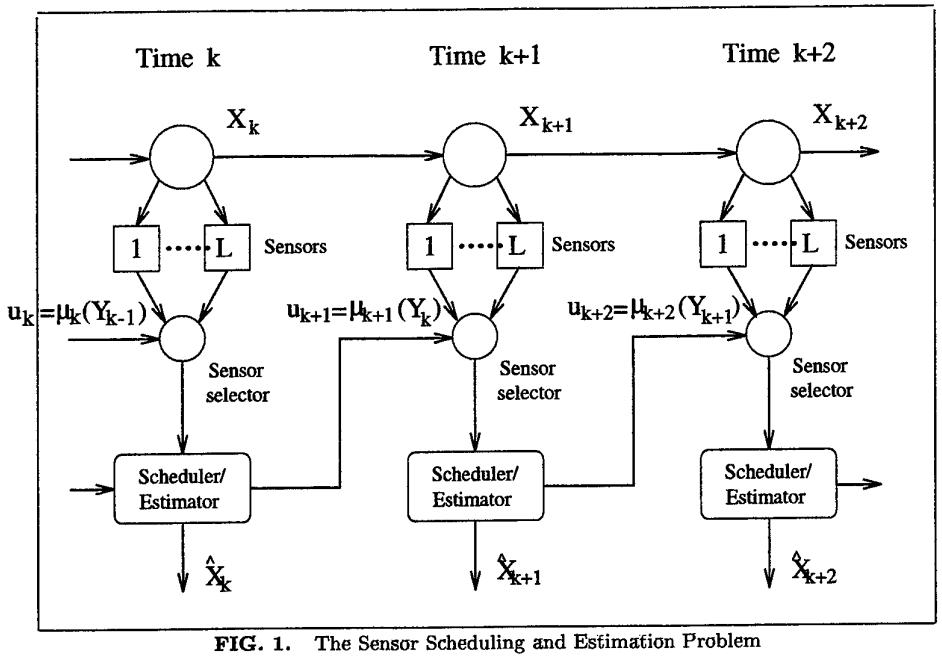


FIG. 1. The Sensor Scheduling and Estimation Problem

where $u_{k+1} = \mu_{k+1}(Y_k)$.

The above objective (3) can be interpreted as follows: The minimization of the first summation results in the optimal sensor schedule that minimizes the weighted mean square error in the state estimate of the Markov chain state X_k. In particular if R_k is set as the identity matrix for all time k, minimization of 3) yields the optimal sensor $u_k \in \{1, 2, \ldots, L\}$ to pick at each time instant k, $k = 1, 2, \ldots, N$ to yield the minimum mean square error state estimate of the Markov chain. The weight terms $R_k(l)$ allow different sensors $l \in \{1, 2, \ldots, L\}$ to be be weighed differently. The time index in R_k allows us to weigh the state estimate errors over time.

The second summation term reflects the cost involved in using a sensor (i.e. the unit time sensor charge) when the the Markov chain is in a particular state. The final term is the terminal cost at time N.

2.2. Information State Formulation

As it stands, the above HMM sensor scheduling problem is a partially observed infinite horizon stochastic control problem. As is standard with such stochastic control problems – in this section, we convert the partially observed stochastic control problem to a fully observed stochastic control problem defined in terms of the *information state* [5].

The information state at time k, which we will denote by π_k—column vector of dimension S, is merely the conditional filtered density of the Markov chain X_k given the observation history Y_k. That is $\pi_k(i) = P(X_k = e_i | Y_k)$, $i = 1, 2, \ldots, S$. Also because we have assumed that X_k is a unit vector $\in \{e_1, \ldots, e_N\}$, we straightforwardly have $\pi_k = \hat{X}_k$. (This is one of the notational advantages of depicting the state space by unit vectors).

The information state π_k is a sufficient statistic to describe the current state of a HMM (see [2] and [5]. The information state update is computed straightforwardly by the HMM state filter (also known as the "forward algorithm")

$$\pi_{k+1} = \frac{B(u_{k+1}, y_{k+1}(u_{k+1}))\, A'\, \pi_k}{\mathbf{1}_S' B(u_{k+1}, y_{k+1}(u_{k+1}))\, A'\, \pi_k} \tag{4}$$

where $\mathbf{1}_S$ represents an S-dimensional vector of ones.

Let $\mathcal{P}_\phi$ denote the set of all information states π that can be achieved by the sensors given the parameter vector ϕ defined in (1). That is,

$$\mathcal{P}_\phi = \left\{\pi \in \mathbb{R}^S : \mathbf{1}_S'\pi = 1, \quad 0 \le \pi(i) \le 1 \text{ for all } i \in \{1 \dots, S\}\right\} \tag{5}$$

Using the smoothing property of conditional expectation, the cost functional of (3) can be rewritten in the form

$$J_\mu = \mathbb{E}\{C_N(\pi_N) + \sum_{k=0}^{N-1} C_k(\pi_k, \mu_{k+1}(\pi_k))\} \tag{6}$$

$$C_N(\pi_N) = \sum_{i=1}^{S} (e_i - \hat{X}_N)' R_N (e_i - \hat{X}_N)\, \pi_N(i),$$

$$C_k(\pi_k, u_{k+1}) = \sum_{i=1}^{S} \left[(e_i - \hat{X}_k)' R_k(u_{k+1})(e_i - \hat{X}_k) + c_k(e_i, u_{k+1})\right]\, \pi_k(i)$$

$$C_0(\pi, u) = \sum_{i=1}^{S} [c_0(e_i, u) + (e_i - \pi_0)' R_0(u_1)(e_i - \pi_0)]\, \pi_0(i).$$

Substituting for $\hat{X}_k = \pi_k$ in in the above equations, after some algebraic manipulations we can write C_k as a quadratic function of the information π as follows:

$$C_k(\pi_k, u_{k+1}) = -\pi_k'\, R_k(u_{k+1})\, \pi_k + g_k'(u_{k+1})\pi_k \tag{7}$$

where $g_k(u)$ denotes the S dimensional vector with elements

$$g_k(e_i, u_{k+1}) = R_k(e_i, e_i, u_{k+1}) + c_k(e_i, u_{k+1}), \qquad i = 1, 2 \dots, S. \tag{8}$$

We now have a fully observed control problem in terms of the information state π: Find an admissible control law, μ, which minimizes the cost functional of (6), subject to the state evolution equation of (4).

2.3. Examples and Applications

1. **Optimal Filtering versus Prediction** Consider the tracking problem of measuring the state of a target from radar derived measurements. Assume that the

target's coordinates evolve according to a finite state Markov chain X_k with known transition probability matrix A. Assume that at each time instant k we have two choices:
(i) $u_k = 1$: Obtain a radar derived noisy measurement of the target position $y_k(u_k = 1)$. Assume that the noise density and hence the symbol probability matrix $B(u_k = 1, y_k)$ is known. After observing the targets position $y_k(u_k = 1)$, we compute the best *filtered* estimate of the target's position $\hat{X}_k$ by using the HMM filter. Let $c(X_k, 1)$ denote the cost of using the radar when the target's true position is X_k. For example, the cost $c(X_k, 1)$ would typically be large when the target X_k is close to the radar tracker.
(ii) $u_k = 2$: Do not observe the target state. This is equivalent to choosing $B(u_k = 2, y_k) = I$, as the observation y_k then contains no information of the state of the Markov chain X_k. Without using the radar for observing the target, we can only compute the best *predicted* estimate of the target via a Hidden Markov model state predictor. Let $c(X_k, 2)$ denote the cost of not using the radar.

In addition to the cost of using the radar $c(X_k, u_k)$, we also incorporate into our cost function the mean-square estimation error of the target's coordinates. Suppose our aim is to chose at each time between $u_k = 1$ (obtaining a radar derived observation and using a HMM filter) versus $u_k = 2$ (not making a measurement and using a HMM predictor) to minimize the cost function in (3). Then the problem is identical to the sensor scheduling problem posed above.

2. **Optimal Quantization Problem**: Given an S state Markov chain X_k observed in noise, consider the following joint source coding and estimation problem which seeks to compute the optimal tradeoff between quantization bits, channel transmission cost and state reconstruction. Suppose at each time k, one has to choose between the following L possibilities: Quantize the observation to l bits and transmit these l bits over the channel at a transmission cost of $c(l, X_k), l = 1, \ldots, L$. The receiver, seeks to recover the best estimate of X_k, i.e. the minimum mean square state estimate subject to the channel transmission cost $c(l, X_k)$. The cost function is then identical to (3). The sensor scheduling problem then yields the optimal answer to how many of bits one must quantize the observations at time k to minimize the transmission and reconstruction cost.

REFERENCES

1. M. Athans. On the determination of optimal costly measurement strategies for linear stochastic systems. *Automatica*, 8:397–412, 1972.
2. D.P. Bertsekas. *Dynamic Programming and Optimal Control*, volume 1 and 2. Athena Scientific, Belmont, Massachusetts, 1995.
3. A. R. Cassandra. *Exact and Approximate Algorithms for Partially Observed Markov Decision Process*. PhD thesis, Brown University, 1998.
4. J.S. Evans and V. Krishnamurthy. Optimal sensor scheduling for hidden Markov models. In *IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP98*, pages 2161–2164, Seattle, May 1998.
5. P.R. Kumar and P. Varaiya. *Stochastic systems - Estimation, Identification and Adaptive Control*. Prentice-Hall, New Jersey, 1986.
6. W.S. Lovejoy. A survey of algorithmic methods for partially observed Markov decision processes. *Annals of Operations Research*, 28:47–66, 1991.
7. L. Meier, J. Perschon, and R.M. Dressler. Optimal control of measurement systems. *IEEE Transactions on Automatic Control*, 12(5):528–536, October 1967.

Advisers With Attitude for Situation Awareness

Dr. Dale A. Lambert

Information Technology Division
Defence Science and Technology Organisation, Australia

ABSTRACT

This paper proposes intelligent virtual advisers and animated scenes in virtual realities as an appropriate technological aid for Situation Awareness. It then outlines the ATTITUDE agent architecture as a basis for building intelligent virtual advisers.

1. Introduction

Endsley ([1]) defines Situation Awareness (SA) as follows.

> "Situation awareness is the perception of the elements in the environment within a volume of time and space, the comprehension of their meaning, and the projection of their status in the near future."

As perception, comprehension and projection characterise mental attributes, SA is understood as a mental phenomenon, and in the absence of anthropomorphism, is understood to be about human minds. So viewed, SA is not a computer system or a screen, it is a state of human awareness.

Technologies and technological aids are often introduced to enhance the state of human awareness, and so the advancement of SA is partly about psychology, partly about technology, and partly about the integration of the two. The concept of a Common Reference Picture[1] (CRP), is often cited as the technological solution to SA. I am less than enthused by the CRP recommendation and in [2] cast a critical eye over each of its common, reference and picture components. The term picture is typically used to refer to a (possibly fused) track display. As a general technological aid to SA, track displays place considerable cognitive burden on the commander. This paper proposes a radical alternative to track displays as the primary technological aid for SA.

2. Virtual Advisers

In a command and control setting, SA is acquired by being informed about what is going on in the world. In our daily lives we are likewise informed of what is going on through the news services, with print, radio, television and the Internet serving as the dominant forms. The last two afford the advantage of being able to supplement the text, photographs and sounds of the first two with a dynamic imaging capability. Television news broadcasts typically involve news presenters, weather presenters, sports presenters, reporters, expert interviews, diagrams, graphs and video footage. The various individuals are assembled as advisers to the viewer and the visual footage is engaged

[1] Common Operating Picture in the US, Joint Operating Picture in the UK

wherever possible to refine the mental images that the viewer otherwise forms from the spoken and written words. News broadcasts provide the viewer with SA by carefully assembling these various components to tell a story.

Military commanders within command and control centres should perhaps acquire SA through facsimiles of news broadcasts, and in some sense, existing staff briefs to these commanders are a comparatively limited attempt to do just that. But while these proposed military news broadcasts would offer potential advantages over traditional briefings, they would still be impeded by some drawbacks.

1. The news broadcast approach to SA is people based and therefore lacks portability. Platform based radar operators or fighter controllers do not have the luxury of a team of advisers to provide them with news broadcasts relevant to their SA.
2. The news broadcast approach to SA engages visual footage wherever possible, but in the military context, the information available is often confined to terse military messages without direct visual footage.
3. Being people based, the news broadcast approach to SA is typically a one off performance by the various advisers. Ideally the viewer should be able to observe the news when it is convenient for them to do so, and be able to rewind and replay parts of the presentation.
4. The news broadcast approach to SA delivers information, but prevents the viewer from interactively conversing with the advisers to further enhance their SA.

In response to these difficulties I offer the following suggestions.

1. The news broadcast should be deliverable by software in addition to people. This facilitates greater portability.
2. Where appropriate, 2 and 3 dimensional virtual reality animations of military messages should be constructed to provide animated movie footage of the events being described.
3. The advisers in conventional news broadcasts should be replaced by virtual advisers in the software. They will provide a story telling counterpart in the software.
4. The virtual advisers should be intelligent. They should be repositories of particular expertise and be capable of interacting with the user about their area of expertise. The aim is to deliver virtual people who interact with the user, their environment, and one another, to meet the user's SA needs.

3. ATTITUDE

The remainder of this paper concerns itself with how we construct intelligent virtual advisers, independently of their animated form and their animated environments. In particular, it promotes as a solution, the ATTITUDE software product being developed at the Defence Science and Technology Organisation. The predecessor to ATTITUDE was designed as an intelligent controller for an AEW phased array radar, and so it is worth noting that the virtual advisers made available to a commander need not necessarily exist locally. They could perform various functions, be distributed across various platforms in the battlespace, and be made accessible to the commander through networking and an avatar form.

To understand why ATTITUDE delivers a framework for developing virtual people, it is necessary to appreciate its motivation. The current computer science paradigm began in the 1940s with a communicative gulf, with a human user flush with conceptualisations at one extremity and the computer as a complex electronic switching device at the other. The current computer science paradigm has sought to bridge this gulf by dragging the computer closer to the user *by embedding human conceptualisation within the machine and then interfacing those conceptualisations to the*

user as if primitive thereafter. Thus we have seen the familiar progression of machine languages, assembly languages, floating point arithmetic, higher level languages, graphical user interfaces, and speech processing systems. If we continue to pursue this paradigm, then at the automation limit we would interact with the computer as if it were another user, and we would predict and explain its behaviour in a similar manner to how we predict and explain human behaviour.

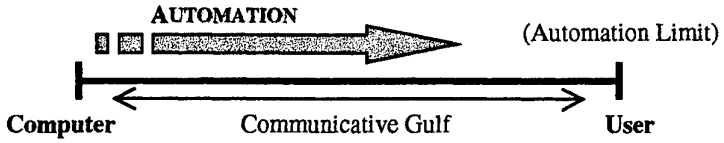


Figure 1: Automation Paradigm

Humans predict and explain the behaviour of other humans by ascribing mental attitudes to them, such as beliefs, desires, expectations, fears, hopes, *et cetera*, and when expressing these and other mental attitudes, the syntax of the expression always assumes the form

<subject> <attitude> that <propositional expression>

The following examples illustrate

Fred believes that the sky is blue Tom expects that it will rain Mary hopes that Tom is insightful

Expressions having this syntactic form are called propositional attitude expressions and the beliefs *et cetera* that they denote are termed propositional attitudes. In a propositional attitude expression: the subject, e.g. Fred, expresses which individual has the propositional attitude; the propositional expression, e.g. the sky is blue, expresses some assertion about the world; and the attitude, e.g. believes, expresses the kind of response the subject has toward the proposition. With subtle modification, propositional attitude observations such as

Fred believes that the sky is blue

can be transformed into propositional attitude instructions like

Fred believe that the sky is blue.

The latter is an instruction, commanding software agent Fred to believe that the sky is blue. A mechanism of this form allows us to not only predict and explain software behaviour at the automation limit, but to also *program* software behaviour at the automation limit. The use of propositional attitude instructions as primitive programming instructions I call attitude programming and ATTITUDE is so named because it practices attitude programming. ATTITUDE therefore satisfies two quite different motivations for virtual advisers: one emanating from a refinement of the current computer science paradigm, and the other coming from a strategy for enhancing SA.

4. Individuals With Attitude

Interacting with the World

In individual ATTITUDE agents, the world is described by the propositional expressions within propositional attitude instructions. The world of every ATTITUDE agent is understood as a world of facts in which, in their most primitive form, the facts are expressed as atomic formulae having the syntactic form (<relation> $<term_1>$ ... $<term_k>$). Each term represents an object and the relation identifies some relationship between those objects, e.g. (taller Clinton Howard). In Classical First Order Logic, the terms denoting objects are recursively grounded in symbols, variables and functions. In ATTITUDE, terms are expressions recursively formed from symbols, variables, functions, Booleans, integers, reals, subexpressions, indexed expressions, schedules, events,

scenarios, individuals and groups. As a consequence, ATTITUDE individuals can form very sophisticated structures to describe the world, including recursively self-referencing structures[1].

Individual ATTITUDE agents operate by being embedded in the world. To facilitate this interaction they are often equipped with sensors and effectors that allow them to observe and alter the world. An individual's conception of the world will often be limited by what they can experience, and this could range from a world of radar tracks through to a sophisticated virtual reality world. The individual's conception of the world is realised by the relations and objects that it is able to consider, and it is the sensors and effectors role to manage the interaction between events in the world and the propositional expressions within the individual. Attitudes exist to facilitate control of sensors and effectors.

Interacting with One Another

Individual ATTITUDE agents also interact socially. This is facilitated by the role of subjects within propositional attitude instructions. The instruction Fred believe that (blue sky) is an instruction to individual Fred to believe that the sky is blue, and so when it is executed by a second agent Tom, it causes an instruction to believe that the sky is blue to be sent from Tom to Fred. Individuals are able to communicate beliefs in this way. Expectations, anticipations and desires can also be exchanged between individuals.

Subjects can also include groups of individuals. Groups are simple sets of individuals and Boolean algebra operators are provided to include and preclude individuals from group membership. Groups can also be formed through queries. Issuing the query ?who ask if believe that (blue sky) will store in variable ?who the group of individuals who believe that the sky is blue. The interaction of individuals enables the collective to arrive at outcomes which none of the participating individuals might arrive at alone.

Determining Behaviour

The behaviour of each individual is determined by the propositional attitude instructions that it invokes in response to social and environmental cues. Each individual is designed to exhibit certain routine behaviours that are applicable to its domain of expertise. These are coded in routines, which comprise an atomic proposition goal and a state transition network of propositional attitude instructions. Routine execution involves navigating control through the transition network of instructions, with each instruction succeeding or failing. The routine is designed so that the atomic proposition goal will be satisfied if a successful path of execution can be found from the start node to a terminal node of the network. Routines provide a procedural approach to knowledge representation [5].

In addition, the believe attitude accommodates Horn clause beliefs for reasoning. Instructions such as Fred believe that (son ?x ?y) if (& (male ?x) (parent ?y ?x)) then make Fred believe the propositional expression (son ?x ?y) if (& (male ?x) (parent ?y ?x)). Each individual is able to engage their inference engine to reason about rules of this sort. Beliefs provide a declarative approach to knowledge representation [5].

[1] In mathematical terms this means that a set theoretic meta-theory for the model theory [3] of an ATTITUDE ontology may need to be a non-well founded set theory [4].

5. Automated Awareness

To succeed as an SA adviser, an ATTITUDE individual must be capable of assembling its own awareness of the world. When engaging the world, we rarely attend to individual facts in the world.

Figure 2: Animated Event and Scenario

In assessing the typical mental snapshot picture animated in the left of Figure 2, we are inclined to represent it as a set of facts, perhaps as

E_1 = { (at loc11 (blue_asset tgt3)), (at loc42 (faker tgt7)),
(at loc42 (low_altitude tgt7)), (at loc42 (approaching tgt7 tgt3))}.

Sets of facts are termed events in ATTITUDE and Boolean algebra combinations of events are formed as scenarios. The right of Figure 2 illustrates a scenario formed from two events. The term situation is used collectively for events and scenarios. Situation awareness is about being aware of situations, not facts or objects *per se*, and all inference within ATTITUDE is conducted relative to situations. Among other things, this allows ATTITUDE individuals to perform "what if" and knowingly counterfactual reasoning, by reasoning about possible events with known scenarios. [6] expands upon the simplified details presented here.

The author recommends a five-step approach in devising an ATTITUDE individual.

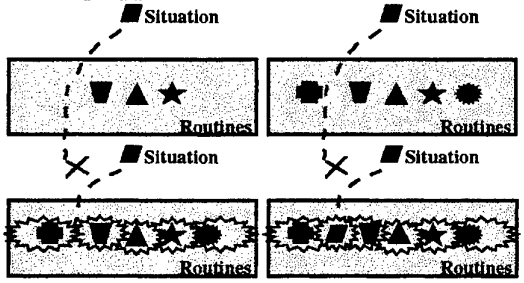


Figure 3: Automating Awareness

Step 1 involves deciding the type of *objects and relations* that lie within the scope of the individual's expertise. This is best done formally, so that the logical dependencies between those object and relation types are well defined. The individual will not understand facts that cannot be expressed in those terms. Step 2 involves deciding the *foreseeable events* that are of interest to the individual. These events may comprise a lattice structure if fragments of events of interest are also of interest. For each event of interest, one or more routines are defined to identify when those events have occurred and to identify the individual's involvement with those events. The situations accommodated by step 2 will comprise all the situations which can be generated from the events of interest *and* which can be serviced collectively by the routines designed to service the events comprising it.

As depicted in the top left of Figure 3, routines for the foreseeable events may not suffice as the unforeseeable might occur. To guard against this, step 3 advocates further expanding the set of

routines so that they include a *topologically well separated set* for which the space of possibilities is approximately covered. The inclusion of routines that roughly accommodate any situation increases the likelihood of any given situation being satisfactory handled by the individual, but as the top right of Figure 3 shows, it provides no guarantee. Consequently, as illustrated in the bottom left of Figure 3, step 4 is to allow the routines to operate with *tolerance*, so that each can be successfully executed with events similar to those intended. This secures completeness, in the sense that the individual will always respond to any given situation, though possibly with a degraded level of performance for unforeseen situations. Two approaches to tolerance have been developed for ATTITUDE, one based upon fuzzy inference ([7]) and the other upon Bayesian inference ([8]). Only the latter has been implemented to date. It involves the association of conditional probability tables with beliefs so that the internal structures produced by the inference engine can be interpreted as Bayesian networks, thereby enabling the probability of the query, including conditional queries, to be computed.

The fifth and final step is to accept that, while the routines of step 4 will always be able to deal with any situation that can arise, their manner of dealing with some situations will be less than satisfactory. Consequently step 5 is to incorporate routine *adaptation* capabilities, as depicted in the bottom right of Figure 3. A case-based approach for ATTITUDE was initiated in [9]. It applied weakest precondition semantics to the execution traces of ATTITUDE routines to learn the conditions under which each routine is likely to succeed, and then applied the Viterbi algorithm to select the routine that is most likely to achieve a nominated goal in the current situation. Another approach, based upon routine generation through genetic algorithms, is about to commence.

6. References

[1] Endsley , M. R. (1995). "Toward a Theory of Situation Awareness in Dynamic Systems", Human Factors Vol. 37 No. 1 pp. 32 – 64.

[2] Lambert, D. A. (1999). "Moving Pictures: Forecasting Situation Awareness", DSTO Research Report *in preparation*. DSTO Electronics and Surveillance Research Laboratory.

[3] Chang, C. C. and H. J. Keisler (1977). Model Theory (2nd. Ed.) North-Holland Publishing Company, Amsterdam.

[4] Aczel, P. (1988). Non-Well-Founded Sets, CSLI Lecture Notes No. 14. Center for the Study of Language and Information, Stanford.

[5] Winograd, T. (1975). "Frame Representations And The Declarative/Procedural Controversy", Readings In Knowledge Representation pp. 357 - 370, Edited R. J. Brachman and H. J. Levesque. Morgan Kaufmann Publishers Inc., Los Altos, California.

[6] Lambert, D. A. (1999). "Assessing Situations", In Proceedings of 1999 Information, Decision and Control, 503 – 508. IEEE.

[7] Lambert, D. A. and M. G. Relbe. (1998). "Reasoning With Tolerance", Proceedings of the IEEE Second International Conference on Knowledge-based Intelligent Electronic Systems. Vol. 3 pp. 418 – 427, Edited L.C. Jain and R. K. Jain. IEEE Inc., NJ.

[8] Fabian, I. and Lambert D. A. (1998). "First-Order Bayesian Reasoning", In Advanced Topics in Artificial Intelligence, 11th Australian Joint Artificial Intelligence, AI'98, 131 – 142. Springer-Verlag.

[9] Rutten, M. (1998). "Automated Situation Assessment Refinement", Masters dissertation in Mathematical Signal and Information Processing. The University of Adelaide.

Quadtree Focusing for UWB SAR

J. H. McClellan, S-M. Oh and M. C. Cobb [1]
School of ECE, Georgia Tech, Atlanta, GA 30332-0250

E-mail: jim.mcclellan@ece.gatech.edu

A common focusing algorithm for Ultra-Wideband, Wide-Angle (UWBWA) SAR systems is the delay-sum backprojector which coherently sums the SAR data. It is very robust to perturbations of the radar platform, but does require a huge amount of computation. In this paper, the Quadtree Backprojection algorithm, which approximates the backprojector, is presented and its characteristics are analyzed. The Quadtree algorithm has $\log_2 N$ stages, so it is possible to measure the focusing performance of each intermediate stage and perform early detection of targets.

1. INTRODUCTION

Various SAR focusing algorithms have been developed for low frequency, wideband radar, because low frequency (less than 1 GHz) signals have much better foliage penetration (FOPEN) and ground penetration (GPEN) capability, making them well-suited to detecting camouflaged targets or buried targets (e.g., land mines). At these low frequencies, the requirement of fine cross-range resolution demands a very long synthetic aperture. In turn, this leads to a very large integration angle, which increases the possibility that severe motion errors occur during the SAR data collection process. As a result, fast imaging algorithms based on the FFT are handicapped by non-uniform spatial sampling of the collected data.

2. DESCRIPTION OF QUADTREE ALGORITHM

The impulse response of the UWBWA SAR data collection process occupies a hyperbolic contour in the space-time domain, because the energy which was originally concentrated in a single point target in the space domain (x, z) spreads out over a hyperbola in the space-time domain (x, t). For each point (x, z) in the image, the delay-sum backprojection algorithm coherently sums the collected SAR data

[1] Prepared through collaborative participation in the Advanced Sensors Consortium sponsored by the U.S. Army Research Laboratory under Cooperative Agreement DAAL01-96-2-0001.

along this hyperbola in the space-time domain:

$$I(x,z) = \sum_{\alpha=0}^{N-1} D(\alpha, T_{x,z,\alpha}) \tag{1}$$

where $I(x,z)$ is the output image, $D(\alpha,t)$ is the radar data recorded at the α^{th} aperture point, and $T_{x,z,\alpha}$ is the time delay corresponding to the distance between the image position (x,z) and the α^{th} aperture point. When the computed time delay lies between signal samples, a bilinear interpolation is done to get the value for the summation (1). Backprojection is like a matched filter, but it is a space-variant computation since the shape of the hyperbola changes with range.

There are two important advantages of the the delay-sum backprojection algorithm: (1) simple motion compensation, and (2) localized processing artifacts. The simplicity of the computation which relies on a distance calculation means that there is no requirement to have a regularly spaced array, as required in FFT-based algorithms. However, there might be a need to change the relative weighting when doing the coherent summation to avoid "hot spots" in the final image.

The most notable disadvantage of backprojection is its computational complexity (order N^3, for $N \times N$ pixels, N sensors) as opposed to FFT-based algorithms such as the ω-k method (order $N^2 \log N$).

2.1. Quadtree Algorithm

A Quadtree is a general way of representing an algorithm, or data structure, in a hierarchical tree structure. The Quadtree for UWBWA SAR Processing was first introduced in [1] where a divide and conquer decomposition of the Delay-sum Backprojector was described. The radix-2 Quadtree algorithm has two features:

- Instead of coherently adding the data from all sensors, the summation is done two sensors at a time. As a result, we must form "virtual sensors" from two neighboring sensors at every iteration stage.
- Instead of computing the exact values of the image at all points in the ground patch at full resolution, use a multi-resolution scheme to approximate the final image. In this case, a ground patch is divided iteratively into a 2×2 set of subpatches at each stage.

At each iteration stage, the generation of new aperture points and new sub-images (sub-patches) can be described as a "parent-child" structure that is typical of tree-structured recursive algorithms (Fig. 1).

2.2. Interpolation and Beamforming

The core computation of the Quadtree image former is a two-sensor delay-and-sum beamformer, where the delay is recalculated at each iteration based on the distance between the new ground-patch centers and either the parent or child aperture positions. In the parent nodes, the radar data is aligned for the distance between the parent aperture positions and the parent ground patch center. When the beamforming is done, its output data must be aligned for the distance between the child aperture positions and the child ground-patch centers. These distance calculations are time-consuming because they involve square roots which then generate indices into the parent data. More than half of the total operations in the

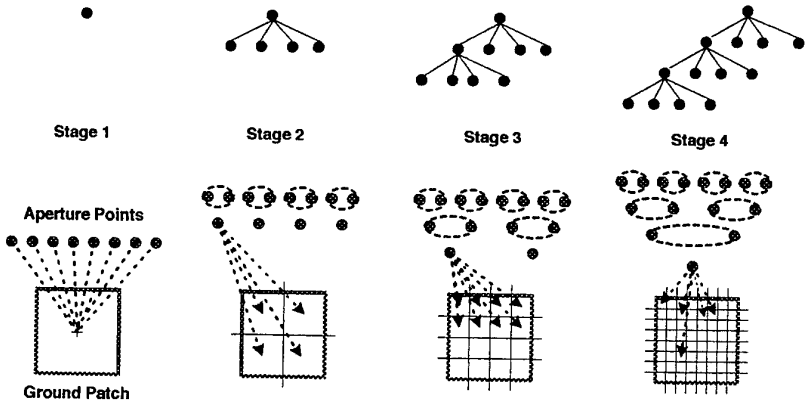


FIG. 1. Tree structure of the quad-tree algorithm.

Quadtree are involved in calculating indices. In general, several parent data values are selected by the indexing, and bilinear interpolation is done as part of the two-point beamforming.

3. OPERATION COUNT

The original Quadtree [1] was a radix-2 algorithm, but we have recently extended the algorithm to a mixed-radix form [2]. In this section, we summarize the number of operations ($\times$, $\pm$, $\sqrt{\cdot}$) required in the mixed-radix Quadtree algorithm. We make the following definitions:

L = number of apertures combined at each stage
$K \times K$ array of new image centers generated at each stage
$N_{p\ell}$ = number of Parent apertures at the ℓ-th iteration
$N_{c\ell}$ = number of Child apertures at the ℓ-th iteration
$N_{tc\ell}$ = number of Time (range) samples in a child node
$M_{p\ell}$ = number of Parent ground patches at the ℓ-th iteration
$M_{c\ell}$ = number Child ground patches at the ℓ-th iteration

We assume that $L = K$, so the aperture dividing factor equals the image patch expansion. Finally, we assume that we start with an equal number of apertures and image patch samples $N = K^p$. The relationships between these numbers is:

$$N_{c\ell} = \frac{N_{p\ell}}{K} = K^{p-\ell} \quad M_{c\ell} = M_{p\ell}K = K^{\ell} \quad N_{tc\ell} \approx K^{p-\ell}$$

The following four steps are analyzed:

1. Distance from each child aperture to the child ground-patch centers requires multiply, add and square roots to compute $N_c\ M_c^2$ distances at each stage:

$$\sum_{\ell=1}^{p} N_{c\ell} M_{c\ell}^2 = \sum_{\ell=1}^{p} K^{p-\ell} K^{2\ell}$$
$$= \frac{K}{K-1}(K^{2p} - K^p) \approx \frac{K^{2p+1}}{K-1}$$

2. The distance from Parent apertures to Child range bins which is done $N_{tc}N_pM_c^2$ times at each stage:

$$\sum_{\ell=1}^{p} N_{t\ell}N_{c\ell}M_{c\ell}^2K = K\sum_{\ell=1}^{p} K^{2p} = pK^{2p+1}$$

3. Interpolation which only needs multiplications and additions and is done $N_{tc}N_pM_c^2$ times at each stage:

$$\sum_{\ell=1}^{p} N_{t\ell}N_{p\ell}M_{c\ell}^2 = pK^{2p+1}$$

4. Coherent Summation which is done $N_{c\ell}(K-1)$ times at each stage:

$$\sum_{\ell=1}^{p} N_{c\ell}(K-1) = \sum_{\ell=1}^{p}(K-1)K^{p-\ell} = K^p - 1 \approx K^p$$

Thus, the total number of computations is:

$$\frac{K^{2p+1}}{K-1} + 2pK^{2p+1} + K^p - 1 \approx N^2(\frac{K}{K-1} + 2K\log_K N) \qquad (p = \log_K N)$$

4. MULTIRESOLUTION ENERGY MEASURE

The internal data structure of the Quadtree algorithm can be written as $v_s[r, a, m, n]$ where s is the stage number, r is the range bin number, a is the aperture point index, and (m, n) are the coordinates of the ground patch center position. In the Quadtree, the raw data is a function of only (r, a), but is gradually localized into the (x, z) coordinates as the iterations progress. The (r, a) dimensions contract and the (m, n) dimensions expand at each stage, representing finer sampling of the image patch.

A multiresolution measurement scheme can be constructed by summing the squared data within each subimage patch to obtain an *Energy Distribution Function* defined versus (m, n):

$$E_s[m, n] = \sum_r \sum_a |v_s[r, a, m, n]|^2$$

The s-th stage has 2^{s-1} image centers, so the domain for (m, n) grows with s. Figure 2 shows a set of images of $E_s(m, n)$ at each stage of a radix-2 QuadTree. In this figure we can see large signal regions in the early stages of the image become focused in the later stages. As the number of subimage centers grows, the regions with higher energy shows the locations where targets are likely to be found. A new target detection algorithm has been based on this focusing sequence [5].

5. HARDWARE IMPLEMENTATION

In surveillance applications the UWBWA SAR might be fielded in an unmanned airborne platform, so the constraints of power and size will drive the implementation. Our approach is to explore the use of a Context Switching Reconfigurable

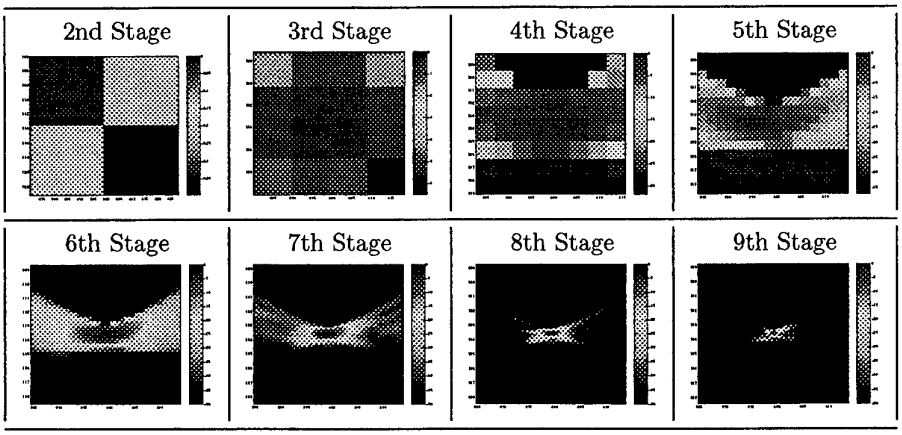


FIG. 2. Energy Measure for all levels of the Quadtree (first stage is the raw data).

Computer (CSRC) being developed by Lockheed-Sanders under the DARPA Adaptive Computing Systems program.

A CSRC has multiple-layers of logic, called contexts, which can be switched on a single clock cycle. Each context implements a particular function (multiply-add, DSP operation, I/O, etc.). As an algorithm executes, it switches context to accelerate a given function. The device allows for data to be shared between contexts and also has RAM to implement local data storage. We are developing only a single node to benchmark the acceleration of the distance and index calculations for the inner loops of the Quadtree.

5.1. Jump Level

Our approach to accelerating the Quadtree is run several stages of the algorithm and then to switch to another algorithm like ω-k or backprojection. If we do most of the Quadtree iterations then the FFTs for the ω-k algorithm should be small enough to be run on individual processors.

It is important to realize that the Quadtree algorithm produces data in aperture and range, not in image coordinates. If a final high-resolution image is desired there are two choices. The Quadtree algorithm can be carried out to the trivial case of one aperture, one pixel per image section, and one range cell per image section.

As a second option the aperture-range data in each subimage can be focused using either delay-sum or ω-k backprojection. The number of iterations of the Quadtree algorithm employed before final focusing is referred to as the *jump level.* In the hardware effort by Georgia Tech, Sanders, and ARL to create a real-time UWB SAR imager, delay-sum backprojection is used to create a final high-resolution image from Quadtree data. This technique has proven to be faster and more accurate than the Quadtree algorithm alone [3].

The squared distance calculations in the Quadtree algorithm can be formulated as two real adds per range cell, with significant overhead per subimage-aperture combination. This is very effective for the larger subimages, but the overhead

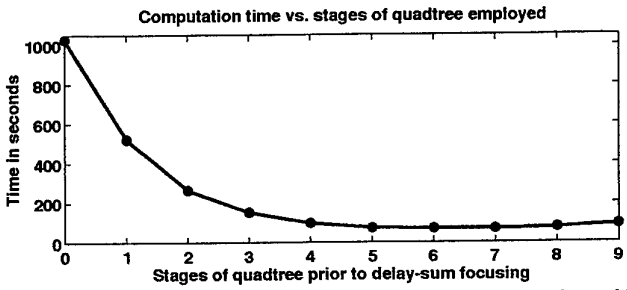


FIG. 3. Computation time as a function of the number of quadtree levels used before jumping to a backprojector to finish. In this case, the computation starts with an array of 2048 apertures, spaced by 11.25 cm. The image size is 1536 (downrange) × 512 (crossrange). Downrange × crossrange resolution is 3.75 cm × 11.25 cm.

becomes burdensome for the smaller subimages in the final few Quadtree stages. Using this formulation of the Quadtree algorithm, and a delay-sum backprojector specifically coded for fast imaging of Quadtree data, computation times were compared for different jump levels from the Quadtree to the delay-sum backprojector. One result, in this case showing that the Quadtree should be used to reduce a 2048 sensor array to no fewer than 32 apertures, is shown in Fig. 3.

Each subimage can also be focused at later stages using the ω-k backprojector, but the ω-k algorithm requires sensors to be evenly spaced along a linear array. If the sensors are necessarily spaced along a wandering path, as in airborne synthetic arrays, the Quadtree can be used to synthesize an evenly spaced virtual array. The optimal approach to this depends on the characteristics of the flight path, but results show that several stages of the Quadtree algorithm should be used to gradually approach a linear array. Then the ω-k backprojector is expected to provide the best solution to high-resolution focusing of Quadtree data.

REFERENCES

1. M. Rofheart and J. McCorkle, "An Order $N^2 \log(N)$ Back Projection Algorithm for Focusing Wide Angle Wide Bandwidth Arbitrary-Motion Synthetic Aperture Radar" *Proc. SPIE Radar Sensor Technology Conference,* vol. 2747, April 1996 , pp. 25–36.
2. S-M. Oh, J. H. McClellan, and M. C. Cobb, "Multi-Resolution Mixed Radix QuadTree SAR Focusing Algorithms," *Proc. 3rd Annual FedLabs Symposium,* Advanced Sensors, pp. 139–144.
3. M. C. Cobb, J. H. McClellan, S-M. Oh, M. Falco, J. Sichina, "Real-Time Image Formation Effort Using Quadtree Backprojection and Reconfigurable Processing," *Proc. 3rd Annual FedLabs Symposium,* Advanced Sensors, pp. 133–138.
4. L. Nguyen, R. Kapoor, D. Wong, and J. Sichina, "Ultra-Wideband Target Discrimination Utilizing an Advanced Feature Set," *Proc. SPIE,* vol. 3370, Orlando, FL, 1998.
5. L. M. Kaplan, S-M. Oh, J. H. McClellan, R. Murenzi, and K. R. Namuduri, "Target Detection During Image Formation for UWB Radar," *Proc. SPIE,* vol. 3810, Denver, CO, 1999.
6. R. Rau, *Postprocessing Tools for Ultra-Wideband SAR Images,* Ph.D. Thesis, Georgia Institute of Technology, December 1998.

Smoothing of Space-Time Power Minimization Based Preprocessor for GPS [1]

Wilbur L. Myrick[†], Michael D. Zoltowski[†], and J. Scott Goldstein[‡]

[†]*School of Electrical Engineering, Purdue University, West Lafayette, IN 47907-1285*

[‡]*SAIC 4001 N. Fairfax Drive, Suite 400, Arlington, VA 22203*

E-mail: wlm@ecn.purdue.edu, mikedz@ecn.purdue.edu, and sgoldstein@trg1.saic.com

An objective function that measures the deviation from smoothness for a space-time preprocessor anti-jam filter is developed. A linear combination of noise eigenvectors is formed to produce the desired space-time weights for minimizing GPS signal distortion based on minimization of the objective function. It is demonstrated that the smoothness of the spectrum characterizing the space-time preprocessor across angle and frequency depends on the choice of the space-time delay in the objective function.

Key Words: GPS; anti-jam filter; preprocessor; power minimization; smoothing

0. INTRODUCTION

GPS is known to provide significant force enhancement capability. This force enhancement capability has been demonstrated in every U.S. military operation since (and including) the Gulf War, but with this capability is a concern about the vulnerability of the GPS signal to jamming. The jamming threat is serious because of the physical design of the GPS system. The received power from the GPS satellites is approximately -157 dBW. Many jammers available on the arms market today either already cover the GPS frequencies, or can be modified to do so. A space-time preprocessing filter prior to the GPS correlators is one of several proposed methods for suppressing jammers. However, this type of filter also induces some distortion of the desired GPS signal.

It is known that tapped-delay line preprocessing of a spread spectrum signal introduces distortion of the desired GPS signal. Characterization of this type of distortion in time-only preprocessing has been previously studied in [2]. We here consider the space-time extension of such an interference suppression algorithm

[1]This research was supported by the Air Force Office of Scientific Research under grant no. F49620-97-1-0275.

proposed in [4] to effectively null both wideband and narrowband jammers while minimizing GPS signal distortion.

1. POWER MINIMIZATION BASED JOINT SPACE-TIME PREPROCESSOR

In the joint processing approach, each sample value input to the GPS receiver is formed from a linear combination of samples across both space and time. The space-time weights are realized through a tapped-delay line behind each digitized baseband antenna, as shown in Figure 1. The output of the preprocessor is then fed to a standard digital GPS receiver. The goal of the preprocessor is to suppress jammers as best as possible while simultaneously passing as many undistorted GPS signals as possible. Note that the anti-jam space-time filter will not be optimized for any one GPS satellite signal in terms of maximizing the SINR. The advantage of this approach is that the anti-jam space-time filter remains a separate component so that a standard digital GPS receiver may be employed.

The criterion for determining the optimal set of space-time weights is premised on the fact that the respective power levels of the desired GPS signals are significantly below the noise floor, as well as below the respective power levels of the potential jammers. The goal then is to drive the power of the preprocessor output down to the noise floor. This approach serves to place point nulls at the respective angle-frequency coordinates of strong narrowband interferers and spatial nulls in the respective directions of broadband interferers.

In order for the GPS receiver to provide accurate navigation information, it is necessary to track the signals from at least four different GPS satellites. Given the parallax error associated with GPS satellites at near-horizon relative to the aircraft, it is generally desirable to track the respective signals from a larger number of GPS satellites, e.g., twelve. It is desired then that the preprocessor "pass" unaltered as many GPS signals as possible. Thus, the magnitude of the multidimensional Fourier transform of the space-time weights should be as flat (smooth) as possible in the spectrum as a function of frequency and angular dimensions. The goal then is achieve a desired smoothness while simultaneously nulling both wideband and narrowband interferers. This motivates the minimization of an objective function that measures the deviation from smoothness.

1.1. Theoretical Development

A space-time power minimization based preprocessor for GPS was proposed in [4] for anti-jam protection. The output power of the space-time preprocessor is minimized under the constraint that the value of the first tap of the tapped-delay-line behind the reference element be unity. This is not necessarily the optimum constraint to reduce GPS signal distortion while nulling out the wideband and narrowband jammers. An alternative approach is taken here in which the space-time weights are expressed as a linear combination of the noise eigenvectors of the space-time correlation matrix, thereby insuring the desired nulling of jammers. The coefficients for linearly combining the noise eigenvectors are determined as those which minimize an objective function that measures the deviation from smoothness.

The $NM \times NM$ space-time correlation matrix is denoted $\mathbf{K}$, where M is the number of antennas and N is the number of taps per antenna. $\mathbf{K}$ can be expressed

as

$$\mathbf{K} = \mathbf{K}_s + \sigma_n{}^2\mathbf{I} \tag{1}$$

where $\sigma_n{}^2$ is the power of the noise per tap per antenna assumed without loss of generality to be both temporally and spatially white. $\mathbf{K}_s$ is the noise-free space-time correlation matrix. Since the GPS signals are at least 16 dB below the noise floor prior to the correlation with any one satellite's PN code, the contributions of the GPS signals to $\mathbf{K}_s$ are here considered negligible. We here assume that the number of jammers is such that not all degrees of freedom are consumed for jammer cancellation purposes. In this case, $\mathbf{K}_s$ is not full rank so that it can be formed from the $K < NM$ eigenvectors of $\mathbf{K}$ associated with the K largest eigenvalues.

An $NM \times (NM - K)$ matrix $\mathbf{E}_N$ is formed from the $NM - K$ eigenvectors of $\mathbf{K}$ associated with the smallest eigenvalue (of multiplicity $NM - K$) equal to the noise power, $\sigma_n{}^2$; these are the noise eigenvectors.

$$\mathbf{E}_N = [\mathbf{e}_{K+1}, \mathbf{e}_{K+2}, ..., \mathbf{e}_{NM}]. \tag{2}$$

The 2D FFT of the j-th noise eigenvector is expressed in terms of a spatial frequency, μ, and an angular digital frequency, ω, which are defined as follows. First, for sake of simplicity, we here assume a linear array of identical antennas equispaced by $d = \lambda/2$ along a line, where λ is the wavelength associated with the L1 frequency. In this scenario, the spatial frequency is defined as

$$\mu = 2\pi\frac{d}{\lambda}\sin\theta = \pi\sin\theta$$

where θ is the angle-of-arrival (AOA). The angular digital frequency is defined as $\omega = 2\pi\frac{\Delta f}{F_s}$ where Δf is the frequency offset relative to the L1 frequency and F_s is the sampling rate.

With these definitions, the 2D FFT of the j-th noise eigenvector may be expressed as

$$s_j(\mu,\omega) = \mathbf{f}^H(\mu,\omega)\mathbf{e}_j \tag{3}$$

where $\mathbf{f}(\mu,\omega) = \mathbf{f}_M(\mu)\otimes\mathbf{f}_N(\omega)$. Now $\mathbf{f}_M(\mu)$ and $\mathbf{f}_N(\omega)$ are respectively defined as

$$\mathbf{f}_M(\mu) = [1, e^{j\mu}, e^{j2\mu}, ..., e^{j(M-1)\mu}]^T \tag{4}$$

$$\mathbf{f}_N(\omega) = [1, e^{j\omega}, e^{j2\omega}, ..., e^{j(N-1)\omega}]^T. \tag{5}$$

Any of the noise eigenvectors when viewed as a space-time weight vector places a null at the angle-frequency coordinate (μ_i, ω_i) of the i-th narrowband jammer

$$\mathbf{f}^H(\mu_i,\omega_i)\mathbf{e}_j = 0 \tag{6}$$

and a spatial null in the direction of the ℓ-th wideband jammer

$$\mathbf{f}^H(\mu_\ell,\omega)\mathbf{e}_j = 0. \tag{7}$$

We desire to find a linear combination of these noise eigenvectors such that the 2D FFT spectrum is as smooth as possible. We seek therefore to minimize the objective function

$$\frac{1}{4\pi^2}\int_{-\pi}^{\pi}\int_{-\pi}^{\pi}|e^{jm\mu}e^{jn\omega}-\sum_{j=K+1}^{NM}z_j{}^*s_j(\mu,\omega)|^2d\mu d\omega \tag{8}$$

with respect to $\mathbf{z}^* = [z^*_{K+1}, z^*_{K+2}, ..., z^*_{NM}]^T$. To minimize (8) let us form a vector

$$\mathbf{s}(\mu,\omega) = [s_{K+1}(\mu,\omega), ..., s_{NM}(\mu,\omega)]^T. \tag{9}$$

We can now redefine our problem as minimizing

$$\frac{1}{4\pi^2}\int_{-\pi}^{\pi}\int_{-\pi}^{\pi}\left|e^{jm\mu}e^{jn\omega}-\mathbf{z}^H\mathbf{s}(\mu,\omega)\right|^2d\mu d\omega. \tag{10}$$

We now will utilize the complex vector minimization methods described in [1]. Using the notation from [1], we are solving for $\mathbf{z}$ such that

$$\nabla_{\mathbf{z}^*}\left\{\frac{1}{4\pi^2}\int_{-\pi}^{\pi}\int_{-\pi}^{\pi}\left|e^{jm\mu}e^{jn\omega}-\mathbf{z}^H\mathbf{s}(\mu,\omega)\right|^2d\mu d\omega\right\}=0. \tag{11}$$

Applying complex vector differentiation yields

$$\underbrace{\left\{\int_{-\pi}^{\pi}\int_{-\pi}^{\pi}\mathbf{s}(\mu,\omega)\mathbf{s}^H(\mu,\omega)d\mu d\omega\right\}}_{(NM-K)\times(NM-K)}\mathbf{z}=\underbrace{\int_{-\pi}^{\pi}\int_{-\pi}^{\pi}e^{-jm\mu}e^{-jn\omega}\mathbf{s}(\mu,\omega)d\mu d\omega}_{(NM-K)\times 1}. \tag{12}$$

Due to the limits of integration, the left hand side of (12) reduces to the Identity matrix and (12) simplifies to

$$\mathbf{z} = [\boldsymbol{\delta}^H_{n,m}\mathbf{e}_{K+1}, ..., \boldsymbol{\delta}^H_{n,m}\mathbf{e}_{NM}]^T \tag{13}$$

where $\boldsymbol{\delta}^H_{n,m} = \boldsymbol{\delta}_n \otimes \boldsymbol{\delta}_m$ with $\boldsymbol{\delta}_n = [\delta(n), ..., \delta(n-(N-1))]^T$ and $\boldsymbol{\delta}_m = [\delta(m), ..., \delta(m-(M-1))]^T$ where $\delta(\mathrm{n})$ is the Kronecker Delta function. This implies that the solution of (8) is based on selecting the same component of each noise eigenvector where the selected component depends upon the values of m and n. Now (13) is used to linearly combine the noise eigenvectors to form the space-time weight vector $\mathbf{h}$ where

$$\mathbf{h} = \mathbf{E}_N\mathbf{z}^*. \tag{14}$$

Any space-time delay factor $e^{jm\mu}e^{jn\omega}$ may be chosen for the objective function. The following simulations illustrate that choosing the space-time delay as $m = \frac{M-1}{2}$ and $n = \frac{N-1}{2}$ maximizes the desired smoothness of the 2D FFT associated with $\mathbf{h}$.

2. SIMULATIONS

Two scenarios are presented to illustrate the importance of choosing the best space-time delay. The simulations employ an $M = 7$ element equi-spaced linear array with $N = 7$ taps at each antenna as depicted in Figure 1. Table 1 summarizes

the values used in both scenarios. The narrowband jammers have different frequency offsets relative to the L1 frequency. Since we are assuming a $20MHz$ receiver bandwidth at each antenna, the noise floor was determined to be at $\sim 130dBW$ after bandpass filtering at each antenna.

The first scenario utilizes a space-time delay of $m = 0$ and $n = 0$ while the second scenario utilizes a space-time delay of $m = 3$ and $n = 3$. Both scenarios deal with maximizing the smoothness associated with the objective function to generate $\mathbf{h}$. Figure 2 illustrates the smoothness of the 2D FFT spectrum of $\mathbf{h}$ associated with the first scenario while Figure 3 illustrates the smoothness associated with the second scenario. Notice the contrast in smoothness associated with each simulation. While both scenarios provide the desired nulls for both the narrowband and wideband jammers, the second scenario provides the smoothest 2D spectrum across space and frequency of all possible space-time delays.

TABLE 1
Simulation Paremeters

Jammer Type	SNR	*AOA*	Bandwidth
Wideband	-100 dBW	$-60°$	20 MHz
Jammer Type	SNR	*AOA*	Frequency(rel.to L1)
Narrowband	-110 dBW	$-40°$	-8 MHz
Narrowband	-100 dBW	$-20°$	-5 MHz
Narrowband	-110 dBW	$0°$	1 MHz
Narrowband	-105 dBW	$20°$	5 MHz
Narrowband	-105 dBW	$40°$	8 MHz

3. CONCLUSION

An objective function that measures smoothness for a space-time preprocessor was presented. By maximizing the smoothness of the objective function, a space-time weight vector is formed from a linear combination of noise eigenvectors. It was shown that choosing a specific space-time delay in the objective function exhibited the desired smoothness across space and frequency, thereby minimizing GPS signal distortion.

REFERENCES

1. D. H. Brandwood, "A complex gradient operator and its application in adaptive array theory," *IEE Proc.,Parts F and H,* vol. 130, pp. 11-16, Feb. 1983

2. J. W. Ketchum and J. G. Proakis, "Adaptive Algorithms for Estimating and Suppressing Narrow-band Interference in PN Spread-Spectrum Systems," *IEEE Trans. on Communications,* May 1982, vol. COM-30, pp. 1169-1177.

3. A. S. Gecan and M. D. Zoltowski, "Power Minimization Techniques for GPS Null Steering Antennas," *Institute of Navigation (ION) Conference,* Palm Springs, CA, 13-15 Sept. 1995.

4. W. L. Myrick, M. D. Zoltowski and J. Scott Goldstein, "Anti-Jam Space-Time Preprocessor for GPS Based on Multistage Nested Wiener Filter," *Milcom '99.*

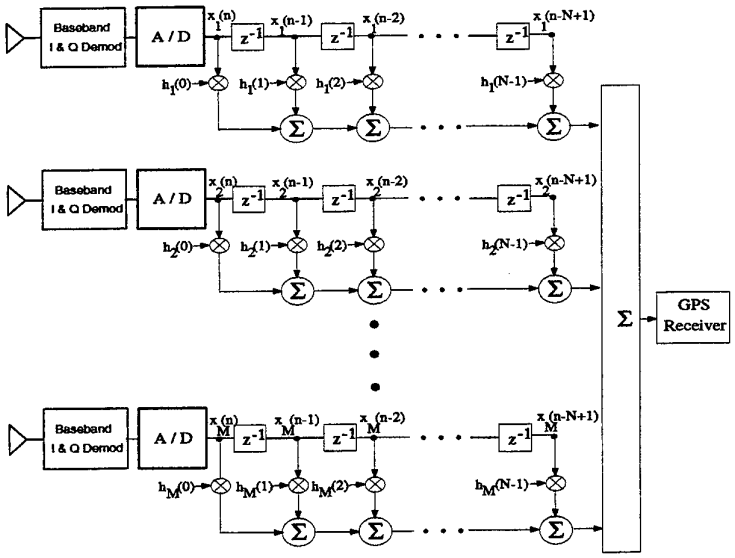


FIG. 1. Power minimization based joint space-time preprocessor.

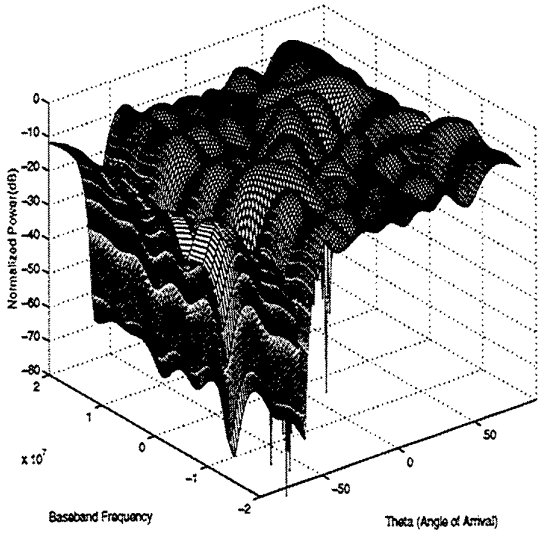
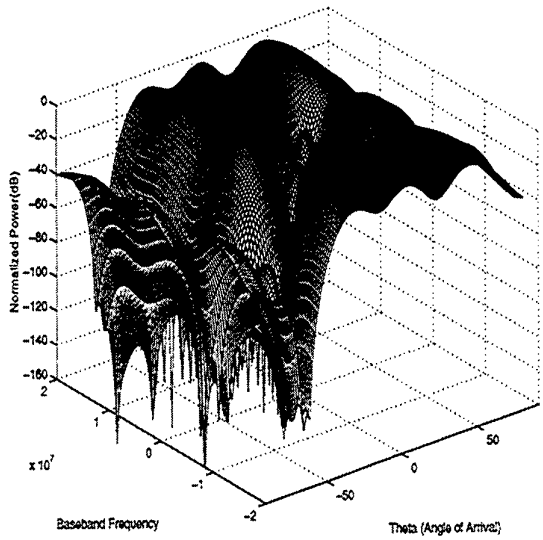


FIG. 2. 2D FFT mesh plot for $\mathbf{h}$ $(m = 0, n = 0)$.



A Hidden Markov Model Classifier for Anti-tank Guided Missiles*

C. Nilubol[1], F. Mujica[1,2], R. Murenzi[2], R. M. Mersereau[1], and M. J. T. Smith[1,2]

Center for Signal and Image Processing[1]
School of Electrical and Computer Engineering
Georgia Institute of Technology

Center for Theoretical Studies of Physical Systems[2]
Physics Department
Clark Atlanta University

July 2, 1999

Abstract

In this paper, we investigate the use of Hidden Markov Models for detection and recognition of Anti-Tank Guided Missiles (ATGMs). ATGMs produce thermal energy that can be received by infrared sensors. The problem is to recognize the ATGM thermal signals in IR sensor data as quickly as possible. The algorithm presented in this paper employs temporal processing followed by a Hidden Markov Model classifier. The performance of the approach is measured experimentally on a set of synthetically generated IR images with embedded ATGM thermal signatures.

1 Introduction

Anti-Tank Guided Missiles (ATGMs) pose a dangerous threat to tank crews. ATGM detection methods attempt to provide warning to the crew so that evasive action can be taken or countermeasures can be launched to neutralize the threat. Working with available on-board IR sensors, these algorithms must be able to detect ATGM firings as early in the game as possible to allow time for appropriate response. ATGM thermal signatures, as measured using high quality sensors, have a well-defined characteristic that can be used to discriminate threats from non-threats (detection) as well as discriminate one type of ATGM from another (classification).

Examples of ATGM signals (intensity vs. time) are plotted in Figure 1. The first one, corresponding to the HOT missile, has two "bright" peaks at the beginning of the signature. These are a consequence of the two stage motor of this missile. The second signature characterizes the MILAN missile. It has a single stage motor, which results in a profile with a single peak. Finally, the TOW missile is propelled by a two-stage motor with a .05 second boost and a .1 second sustain stage. ATGM signatures as seen through on-board IR sensors have low signal to noise ratios (SNRs), typically resulting in either poor detection rates or unreasonably high false alarm rates. Examples of this are shown in Figure 2, which depicts the noisy signatures obtained from the IR sensor for HOT, MILAN, and TOW ATGMs. The SNRs make classification a challenging problem. In addition, noise in the background can make detection difficult as well. Figure 3 shows three examples of time histories for randomly selected background pixels. As can be seen, these pixels have signatures that can easily be confused with those of an ATGM.

The goal of an ATGM warning system is to detect and classify a missile threat as soon as possible so that appropriate measures can be taken. Given that typical ATGM flight times are generally only slightly more than 10 seconds, it is desirable to perform detection in one second or less. As it turns out, the onboard IR sensor used for detection has limited spatial resolution. This in conjunction with the missile's linear trajectory being pointed directly toward the sensor results in the signature information residing in a single pixel position. That is, each pixel

*This work was sponsored in part by US-Army CECOM under contract DAAB07-94-C-M756.

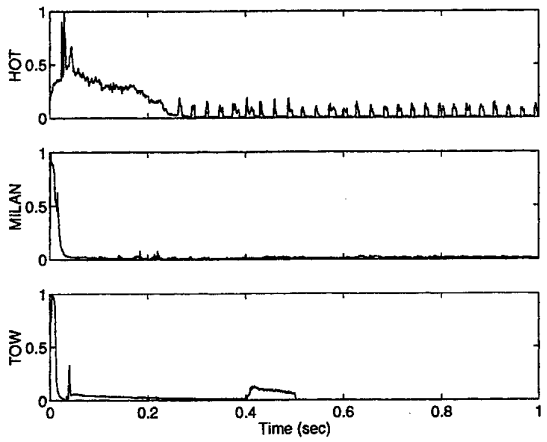


Figure 1: Examples of ideal ATGM sequences.

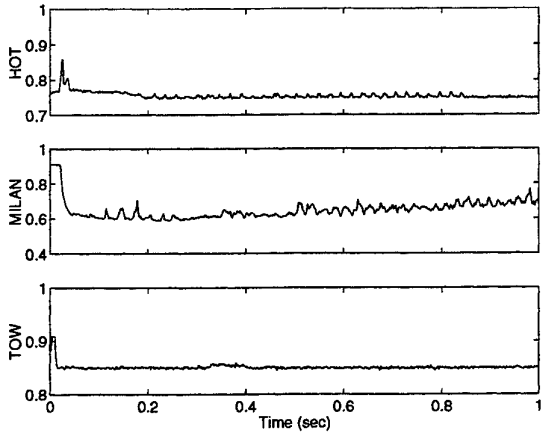


Figure 2: Examples of ATGM sequences as observed through on board IR sensors.

position (as a function of time) represents either noise or a potential ATGM signature. Consequently, we examine the 1-D time histories associated with the pixel positions in an attempt to detect and classify ATGM signatures.

Owing to constraints on computational complexity, we approach the detection problem in a hierarchical fashion, which allows the processing to focus on the image regions that most likely represent an ATGM threat. In the first stage of the approach, we employ a simple temporal processing routine to remove obvious noise/clutter pixel positions from further consideration. This first stage effectively converts the 3-D IR data into a set of 1-D candidate signals.

The temporal processing procedure consists of first filtering along the temporal dimension generally on a per pixel basis. Each pixel in the filtered frame is compared with a spatially weighted threshold. If the pixel value exceeds the threshold, the pixel location is marked as an ATGM candidate location and passed on for addition testing. All pixel positions failing this threshold test are removed from further consideration.

One can also perform the temporal processing hierarchically by dividing the image frames into contiguous sub-blocks, computing the average for the sub-blocks, and then applying the temporal filtering to these block averages. Temporally filtered blocks whose pixels exceed a spatially weighted threshold are then subdivided into smaller blocks with repeated application of the temporal filtering. Hierarchical processing of this type can be effective when ATGM signal variations straddle two or more pixel positions.

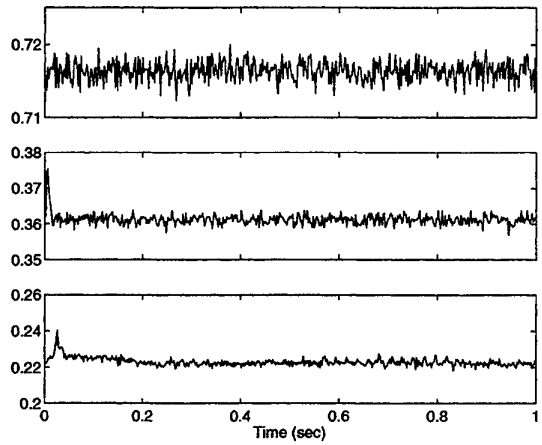


Figure 3: Examples of background noise signals.

2 HMMs for ATGM Detection and Classification

With the IR images now converted into a set of 1-D time signals, the task at hand is to determine if any of these signals constitutes a threat and, if so, determine the type of threat. This is an M-class classification problem, where $M-1$ of the classes correspond to ATGMs, and the remaining class constitutes noise signatures. To address this classification problem, we investigated the use of Hidden Markov models (HMMs), as this technique has proven to be effective in speech recognition and target recognition problem areas [1, 2, 3]. For this application, we view the mechanics associated with the ATGM engines as a process that produces a heat signature. The heat signature is observed by the IR sensor, and is the only observation available to us. We assume that the engine process can be modeled as Markov, that is, represented reasonably well by a finite state machine where at every time instance a transition is made between states $q^0, q^1, \ldots, q^T$, and an observation sequence $\mathbf{O} = \overline{O}_1, \overline{O}_2, \cdots, \overline{O}_T$, is generated according to a probability density function associated with that state.

The basic assumption of the HMM is that the successive observation samples produced by the same state are assumed to be independent of each other and the time t, and that the probability of a state transition depends only on the previous state [4]. Therefore, the joint probability of the state sequence $\mathbf{q}$ being generated by the Markov model and the observation sequence $\mathbf{O}$ being generated by that state sequence can be calculated by the product

$$P(\mathbf{O}, \mathbf{q}|p, A, B) = p_{q^0} \prod_{t=1}^{T} a_{q^{t-1}q^t} b_{q^t}(\overline{O}_t),$$

where $p_i = P(q^0 = i)$ is the probability of the initial state, $A = [a_{ij}]$ is the transition probability matrix, and $b_{q^t}(\overline{O}_t)$ is the probability density of the observation $\overline{O}_t$ at state q^t.

2.1 Estimation of the HMM parameters

To employ HMMs for the ATGM classification problem, the parameters are chosen to optimize a specific criterion, given the training data. The general approach is to use Maximum Likelihood Estimation (MLE), which attempts to maximize the likelihood of the training data given the model of the correct class. MLE is usually achieved by using the Expectation-Maximization (EM) algorithm, an example of which is the Baum-Welch algorithm [4].

Assume that the quantities available are the observations $\mathbf{O}$, the latest estimates of the unknown probabilities in the model λ, and an assumed initial-state distribution $p_i, i = 1, \cdots, S,$. The procedure sets up a reestimation formula that is at least as good as the previous initial estimate. Baum first proved the solution to this problem is to maximize an auxiliary function that leads to an increase in the overall likelihood of a new set of HMM parameters. The auxiliary function then can be expanded and separated into components based on individual parameters. These components lead to closed form solutions for the reestimation of the model set parameters. To compute the likelihood efficiently,

we use a recursion involving forward and backward (FB) probabilities. The procedure underlying the computation of FB probabilities is as follows.

1. Forward probability:
   For $t = 1, \cdots, T;\ \ i = 1, \cdots, S$, the ***forward probability***, $\alpha_t(i)$, which represents the probability of the partial observation sequence $O_1, O_2, \cdots, O_t$ and state S_i at time t, given the model λ, is computed using:

   (a) Initialization:

$$\alpha_1(i) = p_i b_i(o_1), \qquad 1 \lesssim i \lesssim N.$$

   (b) Induction:

$$\alpha_{t+1}(j) = \left[\sum_{i=1}^{N} \alpha_t(i) a_{ij}\right] b_j(O_{t+1}), \qquad 1 \lesssim t \lesssim T-1, 1 \lesssim j \lesssim N.$$

   (c) Termination:

$$P(\mathbf{O}|\lambda) = \sum_{i=1}^{N} \alpha_T(i).$$

2. Backward Probability:
   For $t = T, T-1, \cdots, 1;\ \ i = 1, \cdots, S$, the ***backward probability***, $\beta_t(i)$ which represents the probability of the partial observation sequence $O_{t+1}, O_{t+2}, \cdots, O_T$ and state S_i at time t, given the model λ, can be computed by:

   (a) Initialization:

$$\beta_T(i) = 1, \qquad 1 \lesssim i \lesssim N.$$

   (b) Induction:

$$\beta_t(i) = \sum_{j=1}^{N} a_{ij} b_j(O_{t+1}) \beta_{t+1}(j), \qquad t = T-1, T-2, \cdots, 1, 1 \lesssim i \lesssim N.$$

The reestimation formulas for the coefficients of the mixture density, i.e., c_{jm}, μ_{jk}, and Γ_{jk}, are

$$\begin{aligned}
\bar{c}_{jk} &= \frac{\sum_{t=1}^{T} \gamma_t(j,k)}{\sum_{t=1}^{T} \sum_{k=1}^{M} \gamma_t(j,k)} \\
\bar{\mu}_{jk} &= \frac{\sum_{t=1}^{T} \gamma_t(j,k) \cdot O_t}{\sum_{t=1}^{T} \gamma_t(j,k)} \\
\bar{\Gamma}_{jk} &= \frac{\sum_{t=1}^{T} \gamma_t(j,k) \cdot (O_t - \mu_{jk}) \cdot (O_t - \mu_{jk})'}{\sum_{t=1}^{T} \gamma_t(j,k)}
\end{aligned}$$

where $\gamma_t(j,k)$ is the probability of being in state j at time t with the k^{th} mixture component accounting for O_t, i.e.,

$$\gamma_t(j,k) = \left[\frac{\alpha_t(j)\beta_t(j)}{\sum_{j=1}^{N} \alpha_t(j)\beta_t(j)}\right] \left[\frac{c_{jk}\Psi(O_t, \mu_{jk}, \Gamma_{jk})}{\sum_{m=1}^{M} c_{jk}\Psi(O_t, \mu_{jk}, \Gamma_{jk})}\right],$$

and Ψ is a Gaussian density, with mean vector μ_{jm} and covariance matrix Γ_{jm} for the m^{th} mixture component in state j. The mixture gains c_{jm} satisfy the stochastic constraint

$$\begin{aligned}
\sum_{m=1}^{M} c_{jm} &= 1, \qquad 1 \lesssim j \lesssim N, \\
c_{jm} &\gtrsim 0, \qquad 1 \lesssim j \lesssim N, 1 \lesssim m \lesssim M,
\end{aligned}$$

so that the PDF is properly normalized.

The recognition of an unknown firing sequence involves computing the likelihood of the observation sequence given each HMM model. The likelihood computation takes the most likely state sequence through the model for the final score. For this we use the Viterbi algorithm, its main advantage being that it considers computation of disjoint paths separately.

3 Experimental Results and Discussions

To assess the utility of HMMs for ATGM signature detection and classification, a set of synthetic data was generated using software from GTRI [5]. The imaging sensor was modeled with the GTSENSE package. A 3D representation of the White Sands missile range was used to obtain the background imagery, which was rendered using GTSCENE. The ATGM signatures are based on the Meppen ATGM live tests conducted in May 1998.

The simulated ATGM warning system was configured to be mounted on a stationary platform and contained four IR sensors. Each sensor had a field of view of 90 and 14.25 degrees in the horizontal and vertical orientations respectively and operated at a rate of 500 frames per second. A 360 degree field of view was obtained by pointing each sensor to a different cardinal direction. The focal plane array had 256×32 pixels and was obtained with a uniformly spaced sampling pattern that mapped the world onto a flat plane.

The experiments performed involve 1) detection of the ATGM signatures against a set of randomly selected background pixel signals; and 2) a classification test among 3 target types: HOT, MILAN, and TOW.

1. Detection

   Background pixel signals were chosen from synthetic sensor images. These signals were obtained by randomly choosing background pixels located outside the neighborhood of the real signature from each training image. The set of background signals is divided into subsets of 200 training sequences and 100 testing sequences. Each sequence contains 100 samples. Experiments are performed using 25 samples/sequence obtained by decimation and also by using the full undecimated sequences directly to determine if complexity could be reduced without loss in detection performance. Three simple features were used, all of which were derived by temporal filtering (with different coefficient values). Table 1 shows the detection performance of the baseline HMM at 25 and 100 samples per sequence as a confusion matrix that tabulates the correct detection and false alarms.

2. Classification

   Here we investigate the capability of the HMM to distinguish among three target types: HOT, MILAN, and TOW. The training and testing in this stage repeats the same procedure described in the previous part. A set of training sequences was selected from the available pool of sequences, and the remaining sequences were used for testing. This process was repeated 200 times using different subsets of the sequence pool. A small amount of noise was added to the set at each run. An average was taken over the 200 runs to assess the performance. Table 2 shows the classification performance of the baseline HMM at 25 and 100 samples per sequence as a confusion matrix that tabulates the correct and incorrect classifications.

It can be seen from Table 1 and Table 2 that HMMs are highly capable of both detecting and classifying the ATGM sequences and the background noise signals when the original sequences are subsampled by a factor of 5 and 20 (100 and 25 samples per sequence, respectively). The detection/classification of each system is 99%/91.3% for a 100-sample system, and 100%/88.7% for the 25-sample system, that we investigated.

In summary, ATGM detection and classification is a relatively new problem that has arisen from the recent introduction of low-cost anti-tank weaponry. The development of an onboard early ATGM warning system is expected to have a significant impact on the safety of tank crews. At this point in time, no no such system exist that provide reliable detection and threat classification, operating under the above-mentioned sensor quality limitations and response time constraints. The work reported in this paper represents a first step toward providing a solution to the early ATGM warning system problem and might be used as a benchmark for future work. Our conclusion is that HMMs provide a promising approach for rapid threat detection. The next aspect of the problem we plan to study is detection and discrimination in the presence of gun flashes, which can cause false alarms.

(a)		
Percent Classified as		
	Targets	Unknown
Targets	100	0
Unknown	0	100
total percent detection = 100		

(b)		
Percent Classified as		
	Targets	Unknown
Targets	98	2
Unknown	0	100
total percent detection = 99		

Table 1: The effect of the number of the observation length on the detection performance evaluated on the testing data set: *(a)* T = 25 samples and *(b)* T = 100 samples.

(a)			
Percent Classified as			
	HOT	MILAN	TOW
HOT	77	2	21
MILAN	5	90	5
TOW	1	0	99
total percent recognition = 88.7			

(b)			
Percent Classified as			
	HOT	MILAN	TOW
HOT	84	1	15
MILAN	6	92	2
TOW	2	0	98
total percent recognition = 91.3			

Table 2: The effect of the number of the observation length on the classification performance evaluated on the testing data set: *(a)* T = 25 samples and *(b)* T = 100 samples.

4 Acknowledgement

The authors would like to thank John Stewart and Russ Stanton of GTRI for providing data and relevant software packages used in our experiments.

References

[1] C. Nilubol, Q. H. Pham, R. M. Mersereau, M. J. T. Smith, M. A. Clements, "Translational and rotational invariant hidden Markov model for automatic target recognition," *Proc. of SPIE*, vol. 3357 pp. 179-185, 1998

[2] C. Nilubol, R. M. Mersereau, M. J. T. Smith, "Solving SAR Automatic Target Recognition Problems Using Hidden Markov Modeling," *Proc. of the 3rd Annual Fedlab Symposium*, pp. 325-330, Feb. 1999.

[3] C. Nilubol, R. M. Mersereau, and M. J. T. Smith, "An Improved hidden Markov model classifier for SAR images," *To appear in the Proc. of SPIE*, 1999

[4] L. Rabiner, "A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition," *Proc. IEEE*, vol 77, pp. 257-286, Feb. 1989.

[5] R. Stanton, et. al., "Preliminary and partial report concerning ATGM missile signatures," *Quarterly Report by Georgia Tech Research Institute*, April, 1999.

Investigation of Target Detection in HF Skywave Radar using Thomson's Multiple-Window Method

J. Praschifka and C.F. Doudle

Surveillance Systems Division, DSTO, PO Box 1500, Salisbury, SA 5108, AUSTRALIA

E-mail: justin.praschifka@dsto.defence.gov.au

Thomson's Multiple-Window (MW) method is applied to the problem of target detection in HF skywave radar. The MW method makes minimal assumptions about the noise environment, is specifically designed for short data segments and has high resolution. These properties suggest that the MW method may be useful in dealing with some of the challenges of target detection in skywave radar. The performance of a simple MW detector and a conventional CFAR detector are compared using data from the Jindalee skywave radar.

1. INTRODUCTION

This paper investigates the application of Thomson's Multiple Window (MW) method to the problem of target detection in HF skywave radar. The motivation is to overcome performance limitations of conventional target detectors in dealing with inhomogeneous noise environments. The performance of conventional and MW-based detectors are analysed in some simple scenarios, using both synthetic and real data. Conclusions are drawn as to the feasibility of the MW detector and areas for further research are identified.

The paper is organised as follows. In section 2 we provide background to the problem at hand. In section 3 we briefly describe the MW approach to detection and illustrate its performance by comparing it with a conventional Constant False Alarm Rate (CFAR) detector using a simple data model. In section 4 we carry out a similar performance comparison by injecting sysnthetic targets into real noise data from the Jindalee radar. Conclusions are presented in section 5.

2. BACKGROUND

HF skywave radars use ionospheric refraction to detect and track targets over vast coverages and at ranges of up to 3000 km. A fundamental design requirement for a target detector in skywave radar is that it perform effectively in an inhomogeneous noise environment which can include radar system noise, surface clutter returns, meteor returns, atmospherics and various forms of radio-frequency interference.

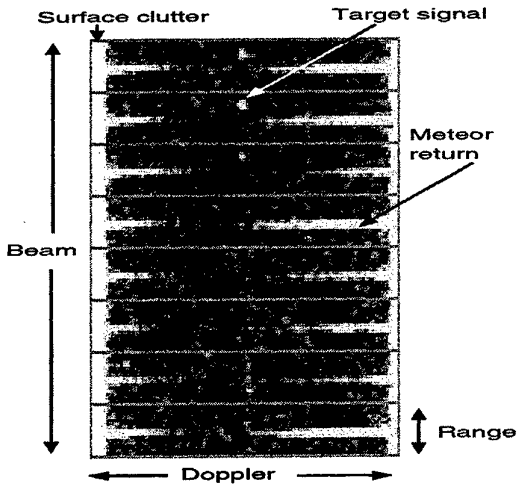


FIG. 1. *A dwell of raw Jindalee data, showing signal power on a grey-scale ARD display. The left and right hand edges of the display correspond to 0Hz Doppler shift.*

For aircraft surveillance the radar revisit time is typically of the order of tens of seconds and must generally be assumed greater than the correlation time of the noise. This means that for each revisit the target detector is required to provide CFAR candidate detections to the tracking system, given minimal prior information about the noise environment. Moreover, the number of available data samples per revisit is generally relatively small.

The Jindalee skywave radar transmits a repetitive, linear sweep, frequency-modulated continuous-wave signal and processes the received signal using an approximate, 3-dimensional matched filter for azimuth, range and Doppler. The processing includes pre-processing to excise transients due to atmospherics and meteor returns, data windowing to control sidelobe leakage and CFAR processing to estimate local noise power statistics using samples from a neighbourhood of each Azimuth-Range-Doppler (ARD) cell (see [1] and references therein for a discussion of CFAR processing algorithms). An example dwell of data, in raw ARD format, is shown in figure 1. The advantages of this conventional approach are low to moderate computational load and good performance over a moderate range of operating conditions. However, because CFAR processing requires a neighbourhood of ARD space, there is the potential for bias due to the inhomogeneity of the noise statistics. This can result in a non-uniform false alarm rate and degraded probability of detection, both of which can degrade tracking performance.

Thomson's Multiple-Window (MW) method[2, 3, 4] provides a relatively new method for CFAR detection of harmonic lines in unknown Gaussian noise. Importantly, the MW method makes minimal assumptions about the noise environment, is specifically designed for short data segments and has high resolution. In essence, the MW method uses a series of orthogonal data windows to generate multiple, independent realisations of the noise process in the frequency domain, based on a single coherent radar dwell. This means that the MW method can estimate local noise statistics without recourse to neighbouring ARD samples, thereby avoiding the potential bias of conventional CFAR processing.

3. THE MULTIPLE-WINDOW METHOD

We now briefly describe the MW method of target detection, contrasting it with conventional CFAR processing. Detailed expositions of the MW method can be found in [2, 3, 4]. For simplicity we apply the MW method in the sweep/Doppler domain only, and this is done independently in each Azimuth-Range (AR) cell. We also assume that there is at most one target per AR cell, though the results are valid for multiple targets, provided that the targets are well separated in Doppler.

For ideal targets the received discrete-time signal in any given AR cell, prior to Doppler processing, can be modelled as a harmonic line in coloured noise:

$$x[n] = \mu \exp(j2\pi f_0 n) + z[n]; \quad n = 0, \ldots, N-1 \tag{1}$$

where the constants f_0 and μ are the Doppler frequency and complex amplitude of the target, respectively, and N is the number of sweeps within a coherent radar dwell (typically $\lesssim 128$). The noise $z[n]$ is assumed to be a stationary, zero-mean, complex, Gaussian random process with power spectral density $S(f)$. The target detection problem is defined in terms of the binary hypothesis test $\mathrm{H}_0 : \mu = 0$ versus $\mathrm{H}_1 : \mu \neq 0$, to which the usual Neyman-Pearson criterion is then applied.

Doppler processing in skywave radar conventionally employs data windowing to avoid sidelobe leakage from very strong surface clutter returns. In this paper we use as an example the minimum 4-sample Blackman-Harris window. In contrast, the MW method employs an orthonormal set of windows $\{\boldsymbol{v}_k\}_{k=0}^{N-1}$, each of which satisfies the $N \times N$ Toeplitz matrix eigenvalue problem [2]

$$\boldsymbol{\mathcal{D}}\boldsymbol{v}_k = \lambda_k \boldsymbol{v}_k, \quad [\boldsymbol{\mathcal{D}}]_{mn} = \frac{\sin\{2\pi W(m-n)\}}{\pi(m-n)} \tag{2}$$

For any specified analysis bandwidth W the multiple windows have a maximal energy concentration property within the band $f \in (-W, W)$, which is measured by the eigenvalues $\{\lambda_k\}$. This allows the windows to be arranged in descending order of fractional energy concentration, $1 > \lambda_0 > \lambda_1 > \cdots > \lambda_{N-1}$. Although sidelobe leakage decreases with increasing W, this must be traded against reduced frequency resolution.

For any chosen W only the first K windows are retained, the remainder being discarded due to poor sidelobe performance. Thomson[3] suggests the choice $K = 2NW - 1$ or $K = 2NW - 3$ to minimise sidelobe leakage. For $W = 4/N$ the first multiple window has sidelobe performance similar to that of the abovementioned Blackman-Harris window, so this value of W will be used for comparison purposes in this section.

For a single, arbitrary Doppler window $w[n]$ it is straightforward to show using (1) that the windowed Discrete Fourier Transform (DFT) of $x[n]$ at f_0, denoted by $y_w(f_0)$, is a complex Gaussian random variable distributed as

$$y_w(f_0) \sim \mathcal{CN}(\mu\sqrt{G}, S(f_0)). \tag{3}$$

where G is the net processing gain of the window, $G = (\sum_{n=0}^{N-1} w[n])^2$ and we assume, without loss of generality, that all windows are normalised to unit incoherent power gain, $\sum_{n=0}^{N-1} w[n]^2 = 1$. It is important to note that, in deriving (3), it is also assumed that $S(f)$ is slowly varying within the analysis band $|f - f_0| < W$.

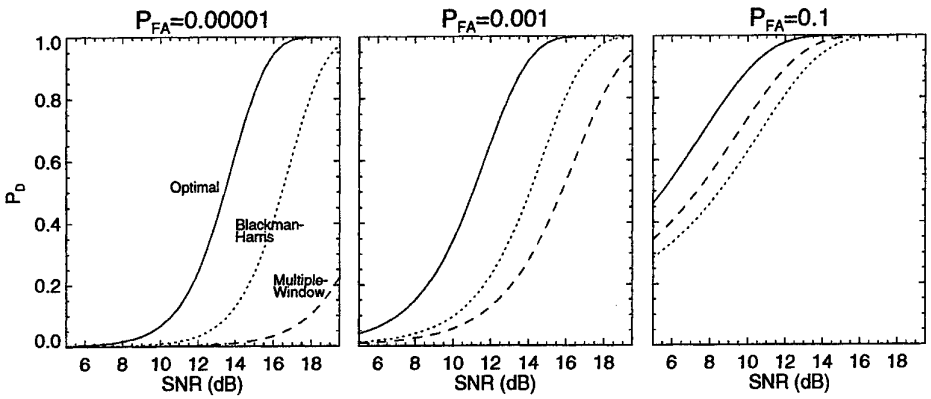


FIG. 2. Comparison of performance of the MW detector with conventional unwindowed (Optimal) and windowed (Blackman-Harris) detectors at various values of P_{FA}, in Gaussian noise. Note that for the latter two detectors the noise power spectral density $S(f_0) = \sigma^2$ is assumed known *a priori*, while for the MW detector it is unknown. In all cases the target SNR is defined as $N|\mu|^2/\sigma^2$. The MW detector uses $N = 128, W = 4/N, K = 5$.

It follows from (3) that, provided f_0 is known, the (matched filter) estimate $\hat{\mu} = y_w(f_0)/\sqrt{G}$ of μ is obtained from a single realisation of $y_w(f_0)$. Also, if $S(f_0) = \sigma^2$ is known then the decision rule for the Neyman-Pearson detector is expressed in terms of the likelihood ratio

$$l(y_w) = \frac{|y_w(f_0)|^2}{\sigma^2} = \frac{G|\hat{\mu}|^2}{\sigma^2} \underset{\mathrm{H}_1}{\overset{\mathrm{H}_0}{\lessgtr}} t \tag{4}$$

It is straightforward to show that $l(y_w)$ has a non-central chi-squared distribution with two degrees of freedom[5], $l(y_w) \sim \chi'^2(2, d)$, where d is the output target SNR, given by $d = G|\mu|^2/\sigma^2$.

In practice, however, a single realisation of $y_w(f_0)$ is not sufficient to satisfy the Neyman-Pearson criterion for detection because $S(f_0)$ is unknown. The role of conventional CFAR processing is to estimate $S(f_0)$ by sampling neighbouring ARD cells. In favourable (ie. homogeneous) noise conditions the effective number of independent samples is large, (say, tens of samples) and the noise estimate is unbiased and of very low variance. Achieved performance in this case will therefore approach that of the detector in (4), which is shown in figure 2 over a wide range of P_{FA} values, with the Blackman-Harris window and with no window ("Optimal"). As discussed in section 2, however, the performance of a conventional CFAR detector can be significantly degraded if the noise is inhomogeneous.

In contrast with the above, the orthonormal multiple windows can be used to generate K approximately independent realisations of the windowed DFT, denoted by the vector $\boldsymbol{y} = [y_0(f_0) \ \cdots \ y_{K-1}(f_0)]^T$. Again, under the assumption of slowly varying $S(f)$ in the analysis band $|f - f_0| < W$, it can be shown that

$$\boldsymbol{y} \sim \mathcal{CN}(\mu \boldsymbol{V}, S(f_0)\boldsymbol{I}) \tag{5}$$

where $\boldsymbol{V} = [\sqrt{G_0} \ \cdots \ \sqrt{G_{K-1}}]^T$ is the vector of net processing gains for the individual multiple windows. Applying least-squares regression analysis[4, 5] to (5)

one obtains the decision rule for the Neyman-Pearson detector,

$$l(\boldsymbol{y}) = \frac{(2K-2)|\hat{\mu}|^2}{2(\boldsymbol{y}-\hat{\mu}\boldsymbol{V})^H(\boldsymbol{y}-\hat{\mu}\boldsymbol{V})} \underset{\mathrm{H_1}}{\overset{\mathrm{H_0}}{\lessgtr}} t \tag{6}$$

which is distributed as non-central F,

$$l(\boldsymbol{y}) \sim F'(2, 2K-2, d_{\mathrm{MW}}). \tag{7}$$

where $d_{\mathrm{MW}} = \frac{G_{\mathrm{tot}}|\mu|^2}{S(f_0)}$ is the output target SNR, $\hat{\mu} = \boldsymbol{V}^T\boldsymbol{y}/G_{\mathrm{tot}}$ and $G_{\mathrm{tot}} = \boldsymbol{V}^T\boldsymbol{V} = \sum_{k=0}^{K-1} G_k$ is the total, net processing gain of the K multiple windows. It follows from (6) and (7) that $S(f_0)$ need not be known to achieve any desired P_{FA}, though P_{D} is of course strongly dependent on $S(f_0)$ through d_{MW}. The performance of this MW detector is included in figure 2.

To interpret figure 2, note that there are two key factors that govern MW detector performance. First is the total processing gain G_{tot} of the multiple windows, which for the present case is around 2.5 dB better than the Blackman-Harris window, independent of P_{FA}. Second, and much more significant here, is the number of degrees of freedom, $2K-2 = 8$ in (7) which, when expressed in terms of SNR "gain", is strongly dependent on P_{FA}, particularly at low to moderate numbers of degrees of freedom.

Of the three P_{FA} values in figure 2, the middle one (P_{FA}=0.001) is most typical of the Jindalee radar. For real skywave radar data, in a region of homogeneous noise, one therefore expects the performance of the MW detector to be of the order of a few dB poorer than a conventional detector using the Blackman-Harris window. To confirm this a Monte-Carlo-type analysis is carried out in the next section using real noise data from the Jindalee radar.

4. ANALYSIS USING JINDALEE NOISE DATA

To more accurately tune the MW detector for typical operating conditions, synthetic coloured noise was generated by approximating some real Jindalee data using a 4th order AR process. P_{D} was then calculated by injecting synthetic targets into realisations of the AR noise, from which the optimum values for the MW detector were estimated to be $W = 5/N$, $K = 6$. The effect of pre-whitening the radar data, as discussed in [2, 4] for example, was also investigated. For this purpose a simple high-pass FIR bandpass filter was used to suppress surface clutter and was implemented using the *filtfilt* algorithm from MATLAB, so as to avoid filter transients. It was found that pre-whitening generally improved detection performance and the optimum values $W = 4/N, K = 7$ were estimated.

The performance of the tuned MW detectors was then estimated for a single dwell of Jindalee data by injecting synthetic targets, as described in [6]. The dwell used was representative of benign operating conditions, similar to that of figure 1, except that there were no meteor returns present. The "conventional CFAR detector" used for comparison in this section is one of a number of detectors that can be invoked in the Jindalee radar.

False alarm performance was first analysed as a function of Doppler shift, by dividing Doppler space into a series of bands and summing up all detections in

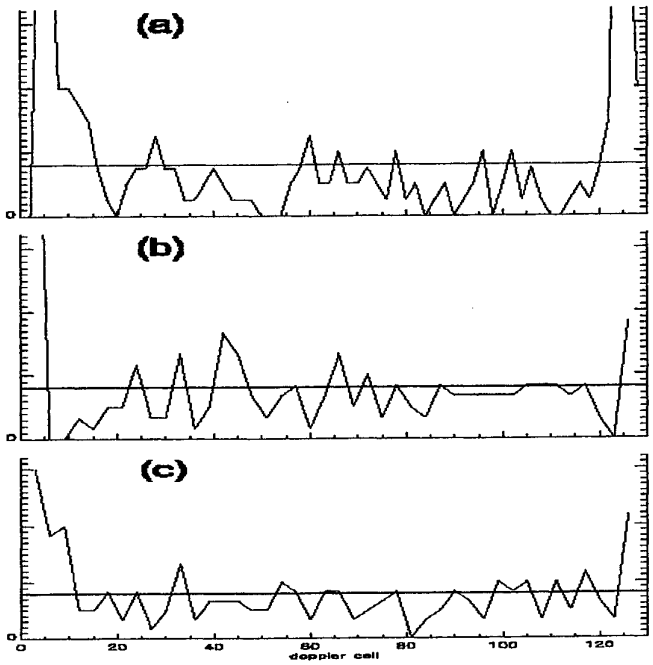


FIG. 3. Achieved P_{FA} of (a) conventional CFAR detector, (b) MW detector without pre-whitening and (c) with pre-whitening. The desired P_{FA} is indicated by the horizontal line. Also, we use $N = 128$, with $W = 5/N, K = 6$ in Figure (b) and $W = 4/N, K = 7$ in Figure (c).

the absence of injected targets. The results of figure 3 indicate that the false alarm performance of the conventional and MW detectors is generally reasonable for most Dopplers of interest, taking into account expected statistical fluctuations. The MW detector appears to produce the most uniform P_{FA}, and it is evident from figures 3(b) and (c) that pre-whitening mitigates sidelobe leakage near the surface clutter returns.

To estimate P_D two subsets of the ARD dwell data were used, consisting of 1250 and 684 cells in the high and low (near clutter) Doppler regions, respectively. The results of figure 4(a) show that, even with pre-whitening, the detection performance of the MW detector is roughly 3 dB poorer than a conventional detector at high Doppler. This is roughly consistent with the model results of section 4 and is accounted for by the larger number of degrees of freedom used by the conventional detector. In figure 4(b) it is evident that the performance gap is even greater near surface clutter. The cause of this is unclear, but is presumably related in some way to sidelobe leakage.

5. CONCLUSION

It is concluded that for skywave radar the simple MW detector analysed in this paper is, in general, unlikely to rival a detector using conventional CFAR processing. Although it may be possible to demonstrate superior performance for the MW detector in more complicated examples of Jindalee noise data, it is clear that the smaller number of degrees of freedom used by the MW detector put it at an unfair disadvantage from the outset. A fairer approach would be to generalise the MW

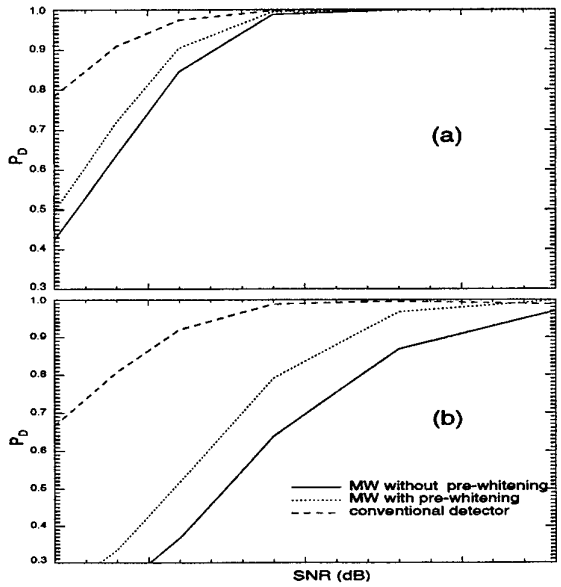


FIG. 4. Comparison of P_D for the MW detector and a conventional CFAR detector in (a) a high Doppler, clutter-free region and (b) a low Doppler, near-clutter region of ARD space. A typical value of P_{FA} was used. The SNR axes for both plots are the identical, with each minor tick mark on the corresponding to 1 dB. For the MW detector the values $W = 4/N, K = 7$ were used with pre-whitening, while $W = 5/N, K = 6$ were used without pre-whitening.

technique to multi-dimensional processes (see [7], for example) thereby allowing the MW detector to exploit extra degrees of freedom from the range and azimuth dimensions. However, in terms computational load and ease of implementation, a more practical approach may be to construct a hybrid detector in which the multiple Doppler windows are used simply to provide extra degrees of freedom to a conventional CFAR processor.

REFERENCES

1. M. D. E. Turley, "Hybrid CFAR techniques for HF radar," in *Proc. IEE Radar 97 conference*, (Edinburgh, UK), October 1997.
2. D. J. Thomson, "Spectrum estimation and harmonic analysis," *Proceedings of the IEEE*, vol. 70, pp. 1055–1096, 1982.
3. D. J. Thomson, "Quadratic-inverse spectrum estimates: applications to palaeoclimatology," *Phil. Trans. Royal Soc. Lond. A*, no. 332, pp. 539–597, 1990.
4. A. Drosopoulos and S. Haykin, "Adaptive radar parameter estimation with Thomson's multiple–window method," in *Adaptive Radar Detection and Estimation* (S. Haykin and A. Steinhardt, eds.), ch. 7, pp. 381–459, Toronto, CANADA: John Wiley and Sons, Inc, 1992.
5. M. G. Kendall and A. Stuart, *The Advanced Theory of Statistics*. High Wycombe, United Kingdom: Charles Griffin and Co. Ltd., 1977.
6. J. Praschifka, J. Lomen, and M. D. E. Turley, "Analysis of detection performance for signals in unknown noise, as applied to HF skywave radar," in *Proc. International Symposium on Signal Processing and its Applications*, (Gold Coast, Australia), 1996.
7. T. P. Bronez, "Spectral estimation of irregularly sampled multidimensional processes by generalised prolate spheroidal sequences," *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. 36, pp. 1862–1873, December 1988.

A Parametric Adaptive Matched Filter for Airborne Radar Applications

J.R. Roman[1], M. Rangaswamy[2], D.W. Davis[1] Q.Z. Zhang[3], J.H. Michels[4] and B. Himed[4]
[1]Scientific Studies Corporation, Palm Beach Gardens, FL
[2]ARCON Corporation, Waltham, MA
[3]University of Central Florida, ECE Dept., Orlando, FL
[4]U.S. Air Force Research Laboratory/SNRT, Rome, NY

Abstract

The parametric adaptive matched filter (PAMF) for space-time adaptive processing is introduced via the matched filter (MF), multichannel linear prediction, and the multichannel LDU decomposition. Two alternative implementations of the PAMF are discussed. Issues considered include sample training data size and constant false alarm rate. Probability of detection is estimated using simulated phased array radar data for airborne surveillance radar scenarios, and signal-to- interference-plus-noise ratio is estimated for airborne phased array radar measurements. For large sample sizes, the PAMF performs almost as well as the MF; performance degrades slightly for small sample sizes. In both sample size ranges the PAMF is tolerant to targets present in the training set.

I. Introduction

This paper presents a new model-based space-time adaptive processing (STAP) algorithm for airborne surveillance phased array radars operating in Gaussian interference. STAP is an area of current interest to the Air Force Research Laboratory (AFRL) for programs such as the Advanced Airborne Surveillance Program (AASP), Multi-Channel Airborne Radar Measurement (MCARM), and the Space Based Radar (SBR). The airborne/spaceborne surveillance radar application presents specific challenges and constraints, but detection performance, computational load, and secondary (training) data requirements are key issues in all cases. STAP for radar target detection was proposed first by Brennan and Reed [2]. The method of [2] consists of:

1. interference covariance matrix estimation from target-free training data
2. weight vector calculation
3. test statistic formation and threshold comparison.

The threshold exhibits a dependence on the true covariance matrix. Consequently, the constant false alarm rate (CFAR) property is lost. A modification to attain CFAR was proposed in [3]. A key result of [2] is the rule-of-thumb, referred to as "Brennan's rule", for training data support so that 3 dB normalized signal-to-interference-and-noise ratio performance is attained. Specifically, the Brennan rule states that for an array with a JN element-pulse (or spatio-temporal) product, $K = 2JN$ independent, identically-distributed, target-free training data vectors are needed to attain performance corresponding to a 3 dB level below optimum. The training data support requirement increases drastically as the problem dimensionality grows (increased J and/or N). Moreover, the training data support available in practice is limited by the temporal and spatial non-stationarity of the interference. Also, system characteristics, such as fast-scanning arrays and receiver bandwidth, impose further restrictions on the amount of training data that can be collected effectively.

Calculation of the weight vector in the conventional method requires the inverse of the $JN \times JN$ spatio-temporal interference sample covariance matrix. This operation has a computational cost on the order of $O(J^3N^3)$, which grows exponentially with increased problem dimensionality. Thus, it is imperative to reduce the training data and computational requirements of STAP algorithms for real-time applications. Parametric (or model) -based methods offer a high-performance alternative to conventional joint-domain architectures and their various approximations [4], as well as to reduced-rank techniques [5]. For radar applications, parametric-based methodologies were formulated first for single-channel systems [7]. More recently, the method has been extended to multichannel systems [8], [9], as well as to multichannel systems in non-Gaussian clutter environments [10]. The method of this paper is based on approximating the interference spectrum with an auto-regressive (AR) model of low order. The fact that a low-order AR model provides an accurate representation of simulated and measured interference for a variety of system and scenario conditions leads to reduced computational requirements. Furthermore, the modeling fidelity is attained using a small fraction of the Brennan rule training data set, thus presenting reduced secondary data requirements. In addition, the method offers significant improvement in detection performance over the conventional adaptive matched filter (AMF) [3]. Specifically, it is demonstrated herein that with a large sample support the PAMF approximates closely the detection performance of the optimal known-covariance matched filter (MF) [3]. It is demonstrated also that the PAMF provides significantly improved detection performance over the AMF using only a small fraction of the secondary data required by the AMF. Furthermore, the PAMF is tolerant to the presence of targets in the secondary data, for both small and large secondary data set sizes.

Multichannel parameter identification methods constitute an inherent part of the PAMF. The identification algorithms considered herein are the Strand-Nuttall (SN) and the least-squares.

II. PHASED ARRAY RADAR DETECTION

A side-looking linear phased array radar configuration is considered herein, with J array channels and a coherent processing interval (CPI) of N pulse repetition intervals (PRI's). A binary hypothesis test is applied to the JN-element complex baseband array measurement vector, $\mathbf{x} \in C^{JN}$. One CPI is the time elapsed during the collection of returns from K_T range bins, and the data collected in one CPI is referred to as the data cube. The data vector contains an unwanted disturbance $\mathbf{d} \in C^{JN}$ with positive definite covariance matrix Rd, and may contain an additive desired signal ae with fixed but unknown complex amplitude "a" and known signal steering vector $\mathbf{e} \in C^{JN}$. The disturbance consists of partially-correlated clutter c, directional broadband interference (jamming) **i**, and thermal white noise **w**, with covariance matrices $\mathbf{R}_c$, $\mathbf{R}_i$, and $\mathbf{R}_w$, respectively. It is assumed that the disturbance components are additive and pairwise independent, and each is a stationary, zero-mean, Gaussian-distributed process. Thus, $\mathbf{x}$ $CN(a\mathbf{e}, \mathbf{R}_d)$, that is, x satisfies the complex normal distribution with mean $a\mathbf{e}$ and covariance $\mathbf{R}_d = \mathbf{R}_c + \mathbf{R}_i + \mathbf{R}_w$. The binary detection problem is to select between hypotheses $H_0 : a = 0$ and $H_1 : a \neq 0$, given a single realization of $\mathbf{x}$. For each pulse in the CPI, the array output sequence is processed to generate a scalar detection test statistic, which is compared to a threshold. If the test statistic exceeds the threshold, H_1 is declared. Otherwise, H_0 is selected. In practice, $\mathbf{R}_d$ is unknown, and must be estimated from data considered to be "signal free". This constitutes the adaptive detection problem. In the conventional approach, $\mathbf{R}_d$ is replaced with its maximum likelihood estimate obtained from $K < K_T$ independent "secondary" or "training" data vectors $\mathbf{d}_k$, $k = 1, 2, \ldots, K$, with d_k $CN(0, Rd)$. For these conditions, the maximum likelihood estimator is the sample matrix estimator, $\hat{\mathbf{R}}_d = \frac{1}{K}\sum_{k=1}^{K} \mathbf{d}_k \mathbf{d}_k^H$. Adaptive detection in radar systems of the type considered herein is accomplished by a "moving window" processing approach, wherein the detection test is applied to each range bin in the data cube. The range bin selected for testing at a particular instant is referred to as the "primary data". The filter applied to the primary data is generated adaptively, utilizing the secondary data to design the filter and to extract information relevant to the determination of the threshold. In this paper a detection method is presented that uses prediction error filtering (PEF), in which the filter coefficients contain the disturbance correlation information in compact form. These filters use a time series representation of the data. Thus, it is convenient to introduce the sequence representation of the data as $\mathbf{x}(n)$, $n = 0, 1, \ldots N-1$, with $\mathbf{x}(n) \in C^J$, and $\mathbf{x} = [\mathbf{x}^T(0) \ldots \mathbf{x}^T(N-1)]^T$.

Defining corresponding sequences for the disturbance process and its components, the detection problem is re-stated as:

$$\begin{aligned} H_0 &: \ \mathbf{x}(n) = \mathbf{d}(n) \ \ n = 0, 1, \ldots N-1 \\ H_1 &: \ \mathbf{x}(n) = a\mathbf{e}(n) + \mathbf{d}(n) \ \ n = 0, 1, \ldots N-1. \end{aligned} \tag{1}$$

The target steering vector sequence $\mathbf{e}(n)$ is of the form $\mathbf{e}(n) = \frac{1}{\sqrt{N}} e^{j2\pi n \bar{f}_{td}} \mathbf{e}(\bar{f}_{ts})$, $n = 0, 1, \ldots N-1$ where $\bar{f}_{td}$ and $\bar{f}_{ts}$ are the target normalized Doppler and spatial frequencies, respectively, and $\mathbf{e}(n) \in C^J$ is the target spatial steering vector, defined as $e(n) = \frac{1}{\sqrt{J}}[1 \ e^{j2\pi \bar{f}_{ts}} \ldots e^{j2\pi(J-1)\bar{f}_{ts}}]^T$.

Thus, the concatenated (block) target steering vector $\mathbf{e}$ is of the form $\mathbf{e} = [\mathbf{e}^T(0) \ldots \mathbf{e}^T(N-1)]^T$. A similar definition holds for other sequences. Block vector and vector sequence representations are used interchangeably throughout this paper.

III. MATCHED FILTER CONFIGURATIONS

For known $\mathbf{R}_d$ and unknown signal amplitude, a CFAR test statistic was proposed in [3]. This test statistic takes the form

$$\Lambda_{MF} = \frac{|\mathbf{e}^H \mathbf{R}_d^{-1} \mathbf{x}|^2}{\mathbf{e}^H \mathbf{R}_d^{-1} \mathbf{e}} \tag{2}$$

where MF denotes matched filter. This test is a normalized version of that proposed in [2]. For the AMF test statistic, the unknown $\mathbf{R}_d$ is replaced by its maximum likelihood estimate $\hat{\mathbf{R}}_d$, so that

$$\Lambda_{AMF} = \frac{|\mathbf{e}^H \hat{\mathbf{R}}_d^{-1} \mathbf{x}|^2}{\mathbf{e}^H \hat{\mathbf{R}}_d^{-1} \mathbf{e}} \tag{3}$$

and is referred to as the CFAR AMF.The MF and AMF detection statistics admit various interpretations, each providing unique insight. One interpretation is derived from the matrix square-root factorization of the sample block covariance matrix and provides the reason for the matched filter name. A second interpretation is based on the relation that exists between multichannel linear prediction and the multichannel (or block) LDU decomposition, and provides the insight for the PAMF, which is the main motivation for this development. That groundwork is laid out next in the context of the MF, but the approach is valid also for the AMF. Let $\mathbf{s} = \mathbf{R}_d^{-\frac{1}{2}} \mathbf{e}$ and $\mathbf{v} = \mathbf{R}_d^{-\frac{1}{2}} \mathbf{x}$. The MF test statistic is expressed as

$$\Lambda_{MF} = \frac{|\mathbf{s}^H \mathbf{v}|^2}{\mathbf{s}^H \mathbf{s}} \tag{4}$$

Notice that under the null hypothesis $\mathbf{v}$ has identity covariance. Thus, the transformation $\mathbf{R}_d^{-\frac{1}{2}}$ is a block whitening filter for the disturbance. Under the alternative hypothesis, however, the target component in $\mathbf{x}$ is rotated and scaled by the whitening transformation. This requires that the whitening transformation be applied also to the detector steering vector, $\mathbf{e}$, so that the rotated and scaled steering vector, $\mathbf{s}$, matches the rotated and scaled target component in $\mathbf{x}$. The numerator of Λ_{MF} is a matched filter. The normalization provided by the denominator term results in CFAR performance. An alternate interpretation of Λ_{MF} is obtained by utilizing the block **LDU** decomposition of $\mathbf{R}_d$, which is of the form $\mathbf{R}_d = \mathbf{ADA}^H$. In this relation, $\mathbf{A} \in C^{JN \times JN}$ is a lower block-triangular matrix with $J \times J$ identity matrices along the main block diagonal, and $\mathbf{D} \in C^{JN \times JN}$ is a block-diagonal matrix with Hermitian matrices $\mathbf{D}_i \in C^{J \times J}$, $i = 0, 1, \ldots, N-1$ along the main block diagonal. The matrix $\mathbf{A}^{-1}$ is of the form

$$\mathbf{A}^{-1} = \begin{bmatrix} \mathbf{I}_J & [\mathbf{O}] & \ldots & [\mathbf{O}] & \\ \mathbf{A}_1^H(1) & \mathbf{I}_J & [\mathbf{O}] & \ldots & [\mathbf{O}] \\ \mathbf{A}_2^H(1) & \mathbf{A}_2^H(2) & \mathbf{I}_J & \ldots & \\ \ldots & \ldots & \ldots & \ldots & \\ \mathbf{A}_{N-1}^H(1) & \mathbf{A}_{N-1}^H(2) & \ldots & \mathbf{I}_J & \end{bmatrix} \tag{5}$$

where $\mathbf{I}_J$ is the $J \times J$ identity matrix, and the block element matrices $\mathbf{A} \in C^{J \times J}$ have no specific structure. These matrices are defined with the Hermitian operator in order to allow for consistent notation between various model identification algorithms. The rows of $\mathbf{A}^{-1}$ denote the matrix coefficients of the nth-order multichannel linear predictor for the process x(n). The corresponding matrix Dn is the covariance matrix of the nth-order prediction error vector. In terms of the block LDU decomposition and applying the square-root factorization of $\mathbf{D}$, the inverse of the block covariance matrix can be represented as $\mathbf{R}_d^{-1} = \mathbf{A}^{-H}\mathbf{D}^{-1}\mathbf{A}^{-1}$. Using this in eq (2) gives

$$\Lambda_{MF} = \frac{|\mathbf{u}^H\mathbf{D}^{-1}\epsilon|^2}{\mathbf{u}^H\mathbf{D}^{-1}\mathbf{u}} \tag{6}$$

where the block vectors $\mathbf{u} = \mathbf{A}^{-1}\mathbf{e}$ and $\epsilon = \mathbf{A}^{-1}\mathbf{x}$ have been defined implicitly. The block vectors ϵ and ν contain the multichannel element vectors $\epsilon(n)$ and $\nu(n)$, and these are given by

$$\begin{aligned} \epsilon(n) &= \textstyle\sum_{k=0}^{n} \mathbf{A}_n^H(k)\mathbf{x}(n-k) \quad n = 0, 1, \ldots, N-1 \\ \nu(n) &= \mathbf{D}_n^{-\frac{1}{2}} \textstyle\sum_{k=0}^{n} \mathbf{A}_n^H(k)\mathbf{x}(n-k) \quad n = 0, 1, \ldots, N-1 \end{aligned} \tag{7}$$

respectively, with for all n. This implies that e(n) is the output of an n^{th}-order moving-average (MA) filter in block mode, with input the block vector $\mathbf{x}$. Such a filter is denoted as MA(n). Since these filter coefficients are the linear prediction coefficients, the vectors in the sequence $\{\epsilon(n) | n = 0, 1, \ldots, N-1\}$ are uncorrelated in pairs (given the minimization criterion associated with linear prediction). This step is equivalent to temporal whitening of the input data sequence, $\{\mathbf{x}(n) | n = 0, 1, \ldots, N-1\}$. Furthermore, $\epsilon(n)$ is the n^{th} prediction error vector with covariance matrix $\mathbf{D}_n$. Thus, for all n, the covariance matrix of $\nu(n)$ is the identity matrix, $\mathbf{I}_J$. Since the transformation generates uncorrelated elements along the spatial dimension at each time instant, this step is a spatial whitening transformation (or spatial block whitening filter). Thus, the vector sequence $\{\nu(n) | n = 0, 1, \ldots, N-1\}$ is temporally and spatially uncorrelated. Similar expressions are obtained for the rotated block steering vectors $\mathbf{u}$ and $\mathbf{s}$. Thus, the MF statistic can be viewed as the magnitude squared (power) of the inner product between the two block vectors ν and $\mathbf{s}$, where ν is a concatenation of the filtered data sequence, and $\mathbf{s}$ is a concatenation of the filtered detector steering sequence. This quantity is then normalized by the inner product of the filtered detector steering sequence, as shown above. For the adaptive case, the matrix coefficients are replaced with their estimates.

IV. PARAMETRIC MATCHED FILTER

The above discussion suggests an approximation to the MF with a simplified structure. First, for both residuals, ϵ and ν in (1) and (2), respectively, retain only the vector sequence for the filter of order P where $1 \lesssim P \lesssim N-1$. Second, let the MA filtering step be a moving window rather than a block window. These two modifications imply that the temporal and spatial filters have a sequential form:

$$\nu(n) = \mathbf{D}_P^{-\frac{1}{2}} \sum_{k=0}^{P} \mathbf{A}^H(k)\mathbf{x}(n-k+P) \quad n = 0, 1, ..., N_\epsilon - 1 \tag{8}$$

where, $\mathbf{D}_P$ is the covariance matrix of the P^{th}-order predictor error (P^{th}-order filtering path), and $N_\epsilon = N - P$. The above-defined filter outputs have the same symbols as the corresponding variables in the MF, but this is adopted to limit the number of symbols introduced. The intended case should be clear from the context. In the present context, $\{\epsilon(n) | n = 0, 1, \ldots, N_\epsilon - 1\}$ is the output sequence of a MA(P) filter in moving- window mode and with input the data

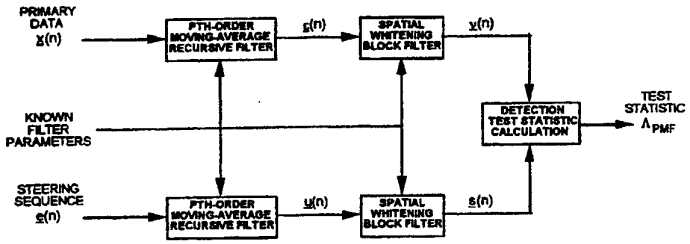


Fig. 1. Detection Architecture of PMF

sequence $\{\mathbf{x}(n)|n = 0, 1, \ldots, N-1\}$. If the matrix coefficients $\{\mathbf{A}^H(k)|k = 0, 1, \ldots, P\}$ and $\mathbf{D}_P$ are determined as the P^{th}-order linear prediction coefficients and the predictor error covariance matrix, respectively, and if a system of order P is an appropriate model for the disturbance, then the MA filter output sequence vectors are uncorrelated in pairs and $\mathbf{D}_P$ is the covariance matrix of $\epsilon(n)$ i.e., under the stated conditions, the MA(P) filter is a whitening filter. In addition, vector sequence $\{v(n)|n = 0, 1, ..., N_\epsilon - 1\}$ is also temporally uncorrelated, and each sequence element has covariance matrix $\mathbf{I}_J$. Lastly, the block vector v has J(N-P) elements. In general, one or more of the conditions for whitening is not met, and consequently the filter output sequence has residual color (non-white). Thus, $\mathbf{e}(n)$ and $v(n)$ are referred to hereafter as the temporal and spatio- temporal residual sequences, respectively. As a result of the two modifications to the MF, the steering sequence is filtered analogously. That is, the temporal and spatio-temporal steering residual sequences are given by

$$\mathbf{s}(n) = D_P^{-\frac{1}{2}} \sum_{k=0}^{P} \mathbf{A}^H(k)\mathbf{e}(n-k+P) \quad n = 1, 1, \ldots, N_u - 1 \tag{9}$$

where $N_u = N_\epsilon = N - P$, and the block vector s has $J(N-P)$ elements. Based on the above discussion, the parametric MF (PMF) detection statistic is defined as

$$\Lambda_{PMF} = \frac{|\sum_{n=0}^{N-P-1} \mathbf{s}^H(n)v(n)|^2}{\sum_{n=0}^{N-P-1} \mathbf{s}^H(n)\mathbf{s}(n)} \tag{10}$$

An architecture for the PMF is shown in Fig. 1. Since the PMF is an approximation to the MF, it is likely that it lacks CFAR performance. Although difficult to prove analytically, simulation-based analyses indicate that the PMF offers CFAR-like behavior for a variety of cases.

V. PARAMETRIC ADAPTIVE MATCHED FILTER

When the filter parameters are unknown, they must be estimated adaptively, and multichannel linear prediction can be applied to estimate parameters $\mathbf{D}_P$ and $\mathbf{A}^H(k)$ of the MA(P) filter using the secondary data. Furthermore, model types such as time series (besides the AR) and state variables can be used in the context of this method [11]. Additionally, for each model type there are alternative parameter identification algorithms, and implementation structures (such as tapped delay lines or lattice filters for time series). This range of options provides a general form to the parametric adaptive MF (PAMF) detection statistic, denoted as Λ_{PAMF}. The form of the Λ_{PAMF} is as in eq (10), but the quantities are generated using estimated parameters obtained via an appropriate model identification algorithm. Thus, the PAMF is formulated as a data-adaptive version of the PMF, further generalized to include a wide variety of whitening filter types. An architecture for the PAMF is presented in Fig. 2. As for the previous case, all variables in the PAMF equations and in Fig. 2 are distinct to those of other detection statistics, but the same symbols are used for notational simplicity. The PAMF lacks the CFAR property because the estimation error in the filter parameters is a function of the true disturbance covariance matrix. In addition, the estimation error varies as a function of the parameter estimation algorithm applied. This is true even with knowledge of the best model order. Nevertheless, the results in Section VII (as well as others not included) indicate that some PAMF configurations (distinct filter implementations) exhibit CFAR-like behavior over a wide range of parameters and conditions, analogous to the PMF.

Implementation of the PAMF as in Fig. 2 requires selection of a model type to represent the disturbance (null hypothesis condition), a parameter identification algorithm, and a whitening filter configuration. The work reported herein focuses on multichannel AR model types, and two multichannel identification algorithms; namely, the Strand-Nuttall (SN)[12], [13] and least-squares (LS) [14] algorithms. Since each algorithm has distinct characteristics and performance in the context of the PAMF, the two configurations are denoted as PAMF-SN and PAMF-LS, respectively.

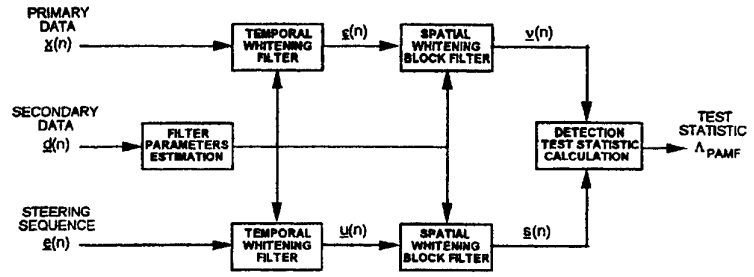


Fig. 2. Detection Architecture of PAMF

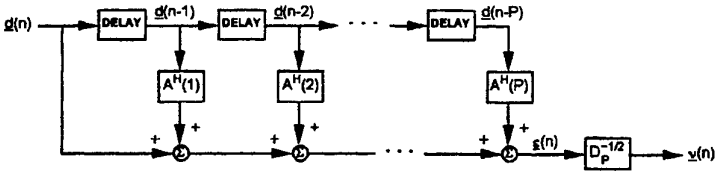


Fig. 3. Tapped Delay Line Implementation for MA (P) Filter

The inverse of an AR model is an MA model, and the MA whitening filter is implemented herein as a multichannel tapped delay line. Linear state variable models and associated identification algorithms are considered in [11], [15]. Specification of the AR(P) system, automatically specifies the inverse MA(P) system. For a given system order, the number of AR(P) (or MA(P)) complex-valued parameters is J^2P, without considering the $\frac{J(J+1)}{2}$ distinct elements of the residual covariance matrix. Fig. 3 shows a tapped delay line for an MA(P) temporal whitening filter, followed by the spatial whitening block filter generated using the residual covariance matrix. Due to lack of space, the interested reader is referred to [11] for implementation details of the model identification algorithms.

VI. PERFORMANCE ANALYSES

A. Definitions and Criteria

Two different power measures were adopted. The first is the per-pulse, per-channel, input SINR, defined as $\mathbf{SINR}_{IN} = \frac{|a|^2}{\sigma_d^2}$, where a is the target amplitude (as defined previously), and σ_d^2 denotes the variance (power) of each element of the disturbance vector at each time instant. The second measure is the output SINR for the MF, SINR=$|a|^2\mathbf{e}^H\mathbf{R}_{dd}^{-1}\mathbf{e}$. SINR is useful for comparing detection performance of STAP algorithms and test statistics (including the MF), and analytic expressions for P_D are available as functions of SINR (in the case of the MF test statistic). However, SINR requires knowledge of the true disturbance block covariance matrix. SINR_{IN} can be established in most cases, but analytic expressions are unavailable. Thus, SINR is adopted for simulated data analyses, and $\mathbf{SINR}_{IN}$ for MCARM data analyses. In addition, clutter-to-noise ratio (CNR) and jammer-to-noise ratio (JNR) are defined as per-pulse, per-channel variance ratios. Detection analyses with simulated data evaluate probability of detection (P_D) for a fixed value of probability of false alarm (P_{FA}); whereas analyses with measured data evaluate the detection test statistics. A high value of the test statistic implies a high P_D since for the methods considered herein the test statistic is the output SINR, and for the conditions considered herein P_D is a monotonically-increasing function of output SINR. In all analyses the time-averaged sample residual covariance is used.

B. Detection Performance Using Simulated Radar Data

A modified version of the airborne surveillance phased array radar clutter model in [16] was used to generate simulated radar data (this software is based on analytic models similar to those in [4]). Temporal de-correlation effects caused by internal clutter motion are included by multiplying the mth lag of the clutter model ACS by a factor of the form , where μ_t is the one-lag clutter temporal correlation coefficient. The parameter μ_t is a function of the pulse repetition

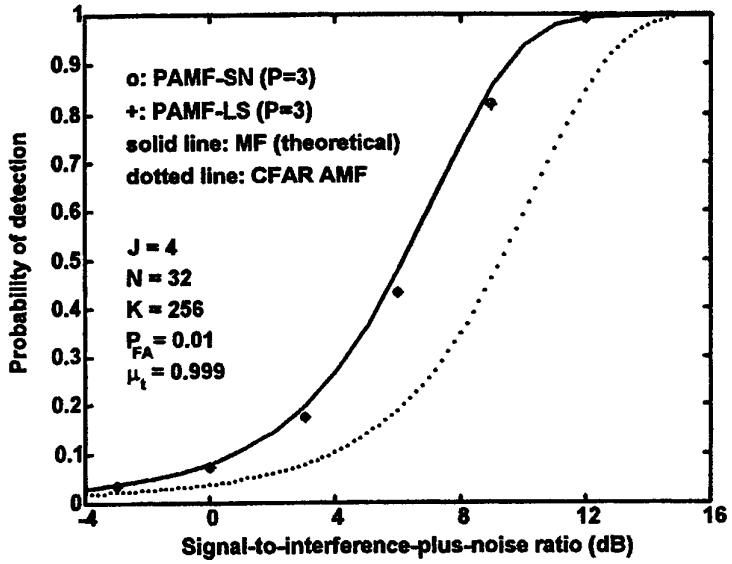


Fig. 4. Probability of Detection vs SINR (K=256, Crab=0 deg)

frequency (PRF), the radiation wavelength, and the standard deviation of the clutter velocity components. Simulation-based detection performance results are presented for two analyses with typical radar system parameters and scenario conditions. The first analysis compares the performance of three detection statistics (PAMF-SN, PAMF-LS, and CFAR AMF) via plots of P_D vs. SINR. For the first analysis, $J = 4$ array channels and $N = 32$ pulses. Thus, this is referred to as the $J << N$ Analysis. The second analysis compares the P_D obtained with the PAMF-SN and the PAMF-LS at one SINR value, SINR = 9 dB. In this second analysis, $J = 14$ and $N = 16$; this is the $J \approx N$ Analysis.

$J << N$ Analysis: All cases considered in this analysis have in common the parameters listed in Fig. 4, as well as $f_{ts} = 0.0$, $f_{td} = 0.3336$, $\sigma_w = 1$ (normalized receiver noise standard deviation), and CNR = 40 dB. Crab angle (g), secondary data size (K), and the presence of jamming are variable, as specified next. P_D as a function of SINR is presented for three scenario cases: Case 1 is for g = 0 deg and no jamming. Case 2 is for g = 20 deg and no jamming. Case 3 is for g = 0 deg and two barrage jammers; one jammer is at fjs = - 0.35 with JNR = 45 dB, and the other jammer is at $f_{js} = 0.2$ with JNR = 50 dB. For each case, two values of secondary data size are considered: $K = 2JN = 256$ and $K = 2J = 8$. Due to limitation of space only sample plots are provided in this paper. Detailed results will be presented at the conference as well as in a companion refereed journal publication. Fig. 4 presents P_D vs. SINR for the scenario and system conditions listed above. In this figure the MF curve (solid line) is the upper bound in performance. This curve was calculated using the analytic relation in [3]. For the other statistics, each P_D value estimate is determined via Monte Carlo (MC) analysis. First, a threshold is determined that provides $P_{FA} = 0.01$ using $N_{MC} = 50$ repetitions of $N_{PFA} = 2,000$ independent data realizations each. Second, this threshold is used to estimate P_D, also using $N_{MC} = 50$ repetitions of $N_{PFA} = 2,000$ independent data realizations each. The CFAR AMF curve (dotted line) is a spline interpolation to 0.2 dB spacing of simulation-based results (obtained with $K = 2JN = 256$) generated at an interval of 1.0 dB along the SINR axis. P_D results for the parametric test statistics were calculated at 3 dB SINR intervals to reduce simulation time. In all cases and for both sample size conditions, $P = 3$ provided the best performance for the PAMF test statistics, which is less than the Nuttall upper bound, $P_{SNU} = 4.24$, and the LS constraint, $P_{LSU} = 25.40$. Additional results of our investigation not reported here provide the following observations. First, the PAMF using small sample size out-performs the CFAR AMF using large sample size. Second, the PAMF-LS and the PAMF-SN perform similarly. Third, for large sample size (K = 256), the PAMF performs close to the optimal MF. And fourth, for small sample size (K = 8), the PAMF performance degrades several percentage points relative to the large sample case, but remains close to the optimal MF. These observations are valid also for a wide variety of scenario and system conditions, although the relative performance of these two PAMF versions can vary. Systems in which the number of channels is approximately the same as the number of pulses constitute one such example.

$J \approx N$ Analysis: Table 1 presents P_D results at SINR = 9 dB and K = 256 for the three cases considered. As in the first analysis, $f_{ts} = 0.0$; however, $f_{td} = 0.1624$, which is much closer to the clutter ridge than the value used for the detection plots. Besides P_D, the table includes the calculated standard deviation of the estimated P_D, denoted as

Case		MF	CFAR AMF		PAMF-SN (P=1)		PAMF-LS (P=1)	
Crab Angle(deg)	No. of Jammers	$\bar{P}_D$	$\bar{P}_D$	SD($\bar{P}_D$)	$\bar{P}_D$	SD($\bar{P}_D$)	$\bar{P}_D$	SD($\bar{P}_D$)
0	0	0.8635	0.4571	0.0397	0.4523	0.1017	0.8474	0.0109
20	0	0.8635	0.4610	0.0382	0.2160	0.1015	0.8277	0.0276
0	2	0.8635	0.4565	0.0375	0.3906	0.1018	0.8469	0.0111

TABLE I

DETECTION PERFORMANCE OF THE PAMF FOR THE THREE SIMULATION CASES AT SINR = 9DB (J = 14; N = 16; K = 256; $\bar{f}_{td} = 0.1624$)

SD[P_D]. The MF and AMF statistics are presented also. Model order is P = 1 for both parametric methods (here $P_{SNU} = 0.86$ and $P_{LSU} = 14.87$). Inspection of Table 1 indicates the lack of CFAR for the parametric detection statistics, although the variability in P_D for the PAMF-LS is only on the order of 2is as good as or better than in Figs. 4. The performance of the PAMF-SN is comparable to or less than that of the AMF, which is a noticeable degradation in relation to the results in Figs. 4. This is foretold by the Nuttall upper bound.

C. Detection Performance Using Measured Radar Data

MCARM database [17] analyses were carried out using one elevation channel, four azimuth channels, and thirty-two pulses from acquisition 575, flight 5 (file rl050575). Range bins (RBs) 142 through 469, inclusive, are used (range-dependent power loss is compensated). Two distinct filter adaptation procedures were applied to study the effects of secondary data size and content. In Procedure A a fixed-window filter is designed using K = 2JN = 256 RBs, selected as RBs 142 through 269, inclusive, and RBs 341 through 468, inclusive. A single fixed-window filter is designed, and the detection test statistic is generated for RBs 270 through 340, inclusive. In Procedure B a moving-window filter is designed using K = 2J = 8 RBs. The moving window consists of eleven adjacent RBs: [four secondary bins — one guard bin — test bin — one guard bin — four secondary bins]. This 11-bin window is applied with RB 270 and RB 340 as the first and last test bins, respectively, and the window is slid from bin-to-bin between these two edge RBs. Target-present and target-absent conditions are studied by inserting artificial targets with an $\overline{SINR}_{IN}$ of -30 dB at RBs 291 and 293, and with an $\overline{SINR}_{IN}$ of -10 dB at RBs 238, 269, 373, and 400. Disturbance power ( ) is estimated as a five-bin average, centered on the RB in which the target is placed. All targets have $\bar{f}_{ts} = 0.0$ and $\bar{f}_{td} = 0.1$. Model order values $P = [2, 3, 4]$ were evaluated for each algorithm and procedure, and the model with best performance was selected. The selection criteria was a combination of highest target peak value, and lowest false-target peak values. Model orders 3 and 4 performed similarly in most cases. The moving-window CFAR AMF filter required diagonal loading of the sample block covariance (at 55 dB below the maximum diagonal element). A typical plot is shown in Fig. 5. Detailed results will be presented at the workshop as well as in a companion journal paper. Fig. 5 is for the fixed-window filter, shows that the PAMF fixed-window filter (with K = 256) bring out the targets at RBs 291 and 293 with at least 25 dB (almost 30 dB for the PAMF-LS) above the 0 dB mean value. Also, the highest background peak is at least 15 dB below the lowest target peak for the PAMF test statistics. The AMF fails to produce a peak at the target RB locations due to the four large-amplitude targets (-10 dB) with the same steering vector in the secondary data. Additional results for the case of a moving-window filter as well as fixed window-filter will be presented at the workshop.

VII. SUMMARY AND CONCLUSIONS

The parametric-based approach introduced herein combines the use of prediction error filtering methods with model identification algorithms to achieve data whitening, which is then followed by matched filtering. This method is referred to as the parametric adaptive matched filter (PAMF). The PAMF admits a variety of distinct implementations, based on the model identification algorithm used and the whitening filter architecture adopted. Herein the Strand-Nuttall (SN) and the multichannel formulation of the least-squares (LS) method, both for AR models, were considered.

Detection performance results were reported for a simulated radar data analysis involving the two PAMF algorithms and the CFAR AMF. Airborne radar measurements collected from the AFRL MCARM program were processed to generate detection test statistic vs. range bin, at specific input SINR levels. This establishes an algorithm's ability to extract the signal from the range bin under test and to reject unwanted disturbance processes.

The parametric methods out-perform the AMF, providing a significant detection performance enhancement for both simulated and measured data. Specifically, for simulated data, the PAMF methods with sample support satisfying the "Brennan rule" perform close to the known-covariance MF. For simulated data and small sample support, the performance of both PAMF statistics is still close to that of the MF curve. However, for cases wherein the number of array elements is approximately the same as the number of pulses, the PAMF-LS maintains its level of performance, but the PAMF-SN degrades significantly. Of relevance, the PAMF is tolerant to the presence of targets in the secondary

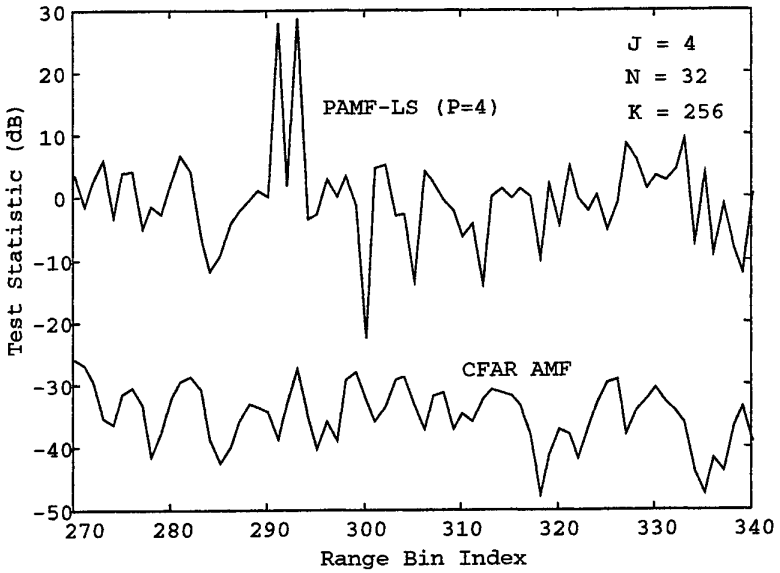


Fig. 5. PAMF-LS and AMF Test Statistics (fixed-window filter)

data, for both small and large secondary data sets. Comparative performance evaluation of the PAMF and other dimensionality-reducing methods is an on-going acivity, to be reported in the future.

A current limitation of the PAMF is the lack of CFAR property. This is due to the dependence of the estimation error in the filter parameter estimators on the "true" disturbance covariance. Furthermore, the estimation error varies from one parameter estimation algorithm to another, with the PAMF-LS providing the most CFAR-like behavior. Considerable progress towards CFAR-like behavior has been realized by utilizing the error covariance matrix estimated using the prediction error filter residuals, rather than the model-based estimate obtained from the parameter estimation algorithms. CFAR options for the PAMF is an area of current research.

Other current and future work covers the combination of the parametric-based processing method with detection rules other than the MF. This is being considered for the case of compound-Gaussian clutter disturbance statistics also.

References

[1] I. Reed, J. Mallett, and L. Brennan, "Rapid convergence rate in adaptive arrays," *IEEE Trans. on Aerospace and Electronic Systems*, vol. **AES-10**, pp. 853–863, 1974.

[2] F. Robey, D. Fuhrmann, E. Kelly, and R. Nitzberg, "A CFAR adaptive matched filter detector," *IEEE Trans. on Aerospace and Electronic Systems*, vol. **AES-28**, pp. 208–216, 1992.

[3] A. Jaffer, M. Baker, W. Balance, and J. Staub, "Adaptive space-time processing techniques for airborne radar," Tech. Rep. RL-TR-91-162, Rome Laboratory , July 1991.

[4] J. Goldstein and I. Reed, "Theory of Partially Adaptive Radar," *IEEE Trans. on Aerospace and Electronic Systems*, vol. **AES-33**, no.4, pp. 1309–1325, 1997.

[5] P. Metford, S. Haykin, and D. Taylor, "An Innovations approach to discrete-time detection theory," *IEEE Trans. on Information Theory*, vol. **IT-28**,(2), pp. 376–380, 1982.

[6] J. Michels, P. Varshney, and D.D.Weiner, "Multichannel signal detection involving temporal and cross-channel correlation," *IEEE Trans. on Aerospace and Electronic Systems*, vol. Vol. 31, No. 3, pp. 866–880, 1995.

[7] L. Timmoneri, I. Proudler, A. Farina, and J. McWhirter, "QRD-based MVDR algorithm for adaptive multipulse antenna array signal processing," *IEE Proc.F, Commun., Radar, & Signal Process.*, vol. Vol. 141, No. 2, pp. 93–102, 1994.

[8] M. Rangaswamy, J. H. Michels, and D. D. Weiner, "Multichannel detection for correlated non-Gaussian random processes based on innovations," *IEEE Trans. on Signal Processing*, vol. **SP-43**, pp. 1915–1922, 1995.

[9] J. Roman, M. Rangaswamy, D. Davis, Q. Zhang, J. Michels, and B. Himed, "Parametric-Based and Reduced-Rank STAP algorithms for Airborne Radar Applications," Tech. Rep. AFRL-SN-RS-TR-1999-TBD, U.S. Air Force Research Laboratory, 1999.

[10] O. Strand, "Multichannel complex maximum entropy (auto-regressive) spectral analysis," *IEEE Trans. on Automatic Control*, vol. **AC-22**, pp. 634–640, 1977.

[11] A. Nuttall, "Multivariate Linear Predictive Spectral Analysis Employing Weighted Forward and Backward Averaging," Tech. Rep. TR-5501, Naval Underwater Systems Center, New London, CT, October 1976.

[12] S. Marple, *Digital Spectral Analysis With Applications*. New Jersey: Prentice Hall, 1987.

[13] J. Roman, D. Davis, and J. Michels, "Multichannel Parametric Models for Airborne Phased Array Clutter," in *Proceedings of the IEEE National Radar Conference*, (Syracuse, NY), 1997.

[14] J. Roman and D. Davis, "Multichannel System Identification and Detection Using Output Data Techniques," Tech. Rep. RL-TR-97-5(Vol II), U.S. Air Force Research Laboratory, 1997.

[15] Westinghouse-Electronic-Systems, "Multichannel Airborne Radar Measurement (MCARM)," Tech. Rep. RL-TR-96-49, 1, U.S. Air Force Research Laboratory, Rome Site, 1995.

TARGET TRACKING WITH DISSIMILAR SENSORS USING SET-BASED ESTIMATION

S.K. SALEEM and L. SCIACCA
Weapons Systems Division- WSI
Defence Science and Technology Organisation
PO BOX 1500
Salisbury, SA 5018, AUSTRALIA

E-mail: Khusro.Saleem@dsto.defence.gov.au Tel: +61 08 8259-6614

Abstract

A multiple sensor target tracking algorithm is presented. The algorithm combines polar coordinate data from a range-angle sensor, such as a RADAR and an angle-only sensor, such as an acoustic array, and outputs Cartesian coordinate track data. This non-linear estimation problem is solved using a set-based estimation technique, which does not rely on statistical assumptions about the sensor measurement noise. Issues including data association and data-fusion are automatically solved using this approach.

1 Introduction

Tracking targets using sampled noisy range and angle radar measurements is a difficult problem. It is further complicated when the aim is to combine this with sampled noisy angle-only measurements from a passive sensor [1]. There are several reasons for this. Firstly, target tracking is generally performed in an x-y Cartesian coordinate system whereas sensor measurements are made in a range-bearing coordinate system. This results in a non-linear estimation problem which is difficult to solve [2]. Secondly, tracking using passive sensors is also a highly non-linear estimation problem and also results in an unobservable system which makes conventional statistical estimation techniques difficult to apply [1,2]. Thirdly, conventional statistical estimation approaches involve making assumptions about the sensor measurement noise, most commonly- that it is Gaussian stationary noise. The problem is that in real sensor systems, the noise is biased, non-Gaussian and non-stationary, due to effects such as dynamic video quantization, digital sampling quantization, receiver saturation, finite numerical precision, etc. Finally, combining data from active and passive sensors is generally treated as a separate problem, commonly termed jointly as 'data-association' and 'data-fusion'. This is generally solved as a statistical hypothesis testing problem which relies on the statistical assumptions mentioned previously, and these assumptions are often flawed.

This paper presents an alternative approach which utilizes set-based estimation principles [5]. Set-based estimation assumes as little as possible about the system other than to place bounds on those quantities to be estimated or any uncertainties (measurement errors). This leads to a recursive procedure whereby the estimate at time k is the set $\Omega(k)$ which is consistent with an initial set S_0 and all the measurement sets S_i up to time k, $\Omega(k)=\bigcap_{i=0}^{k} S_i$.

A fundamental problem is to obtain feasible representations for the sets [5]. A representation must accurately and tightly bound the true parameters and any uncertainties, and be computationally tractable. Common representations include polytopes and ellipsoids.

In this paper, rectangles are the chosen representation. Although rectangles do not yield the tightest bound, they are easily represented and provide a tractable solution to this problem. Another important advantage of this approach is that it yields a robust algorithm despite the highly non-linear nature of the problem. Finally, the data-association and data-fusion problems are solved automatically using this approach. Set-based estimation, also commonly called 'bounded-error estimation', is a simple technique which has been applied, successfully, to solving target tracking problems [3,4]. In [4] an algorithm similar to the one proposed here is developed for a single sensor tracking application. This paper extends the work to the case of two dissimilar sensors.

An outline of this paper is as follows. Section 2 formulates the problem to be solved. The sensor – target geometry, and the target and measurement models are presented. Section 3 presents the set-based estimation algorithm that performs target tracking and data-association and data-fusion. Section 4 concludes the paper.

2 Problem formulation

The sensor and target geometry is illustrated in Figure 1. The set-based algorithm presented here recursively computes the rectangular set over-bounding the x and y coordinates at time index k that are consistent with all the observations up to time index k.

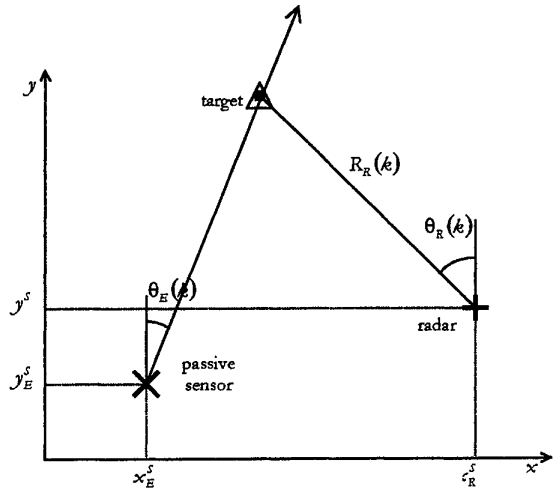


Figure 1 Sensor and target geometry

A rectangular set is aligned with the x-y coordinate system and is parameterized by upper and lower bounds in x and y; the current state estimate is an element of the current set estimate denoted by

$$\mathbf{s}(k|k) \in \Omega_s(k|k) = \left\{ x, y, v_x, v_y : \begin{array}{c} x_{\max}(k|k) \geq x \geq x_{\min}(k|k) \\ y_{\max}(k|k) \geq y \geq y_{\min}(k|k) \\ v_{x,\max}(k|k) \geq v_x \geq v_{x,\min}(k|k) \\ v_{y,\max}(k|k) \geq v_y \geq v_{y,\min}(k|k) \end{array} \right\}. \quad (1)$$

The state vector for this estimation problem can be defined by

$$\mathbf{s}(k|k) \stackrel{\Delta}{=} \left[x_{\max}(k|k), x_{\min}(k|k), y_{\max}(k|k), y_{\min}(k|k), v_{x,\max}(k|k), v_{x,\min}(k|k), v_{y,\max}(k|k), v_{y,\min}(k|k)\right]^T.$$

The state equation assumes a constant heading constant velocity target and is given by

$$\mathbf{s}(k+1|k) = \mathbf{F}\mathbf{s}(k|k) \qquad (\ )$$

$$\mathbf{F} = \begin{bmatrix} \mathbf{I} & T \cdot \mathbf{I} \\ \hline \mathbf{I} & \mathbf{I} \end{bmatrix} \qquad (\ )$$

$\mathbf{I}$ T

$$z_R^p(k) = [R_R(k), \theta_R(k)]^T + [n_R; n_{\theta_R}]^T :$$

$$z_E^p(k) = \theta_E(k) + n_{\theta_E} :$$

x

$$|n_R| \le \varepsilon_R, |n_{\theta_R}| \le \varepsilon_{\theta_R}, |n_{\theta_E}| \le \varepsilon_{\theta_E}$$

$\theta_R(k)$ $\theta_E(k)$

Such bounds are easily defined in practice and are more robust than assuming a statistical distribution. Referring to equation (1), note that the state is defined in Cartesian coordinates, therefore the measurements must be transformed from polar to Cartesian coordinates.

$$\mathbf{z}_R^c(k) \in \Omega_R(k) = \left\{ x, y : \begin{matrix} z_x^R(k) + \varepsilon_x^R(k) \ge x \ge z_x^R(k) - \varepsilon_x^R(k) \\ z_y^R(k) + \varepsilon_y^R(k) \ge y \ge z_y^R(k) - \varepsilon_y^R(k) \end{matrix} \right\} \qquad (\ )$$

U

$$z_x^R(k) = R_R(k)\sin\theta_R(k) + x_R^S$$

$$z_y^R(k) = R_R(k)\cos\theta_R(k) + y_R^S$$

$$\varepsilon_x^R(k) = (R_R(k) + \varepsilon_R)\sin(R_\theta(k) + \varepsilon_\theta) - (R_R(k) - \varepsilon_R)\sin(R_\theta(k) - \varepsilon_\theta)$$

$$\varepsilon_y^R(k) = (R_R(k) + \varepsilon_R)\cos(R_\theta(k) - \varepsilon_\theta) - (R_R(k) - \varepsilon_R)\cos(R_\theta(k) + \varepsilon_\theta)$$

x_R^S y_R^S

$$\mathbf{z}_E^c(k) \in \Omega_E(k) = \left\{ x, y : \left| y - \frac{1}{\tan\theta_E(k)}\left(x + x_E^S\right) - y_E^S \right| \le \delta \right\} \qquad (\ )$$

$$\delta = R_E^{MAX} \cos\varepsilon_{\theta_E}$$

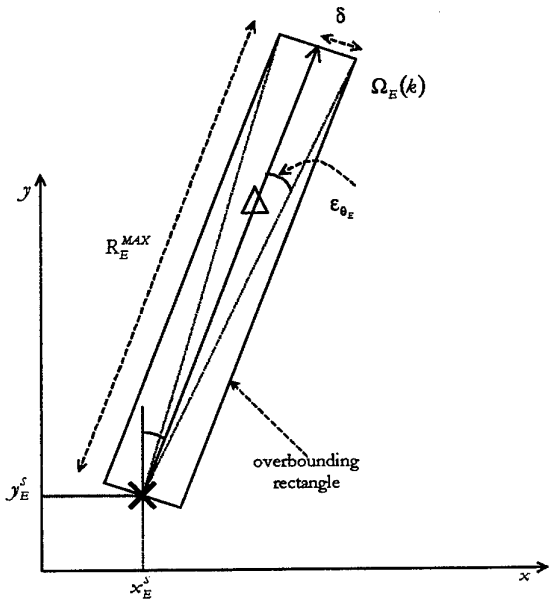


Figure 2 Passive sensor measurement transformation

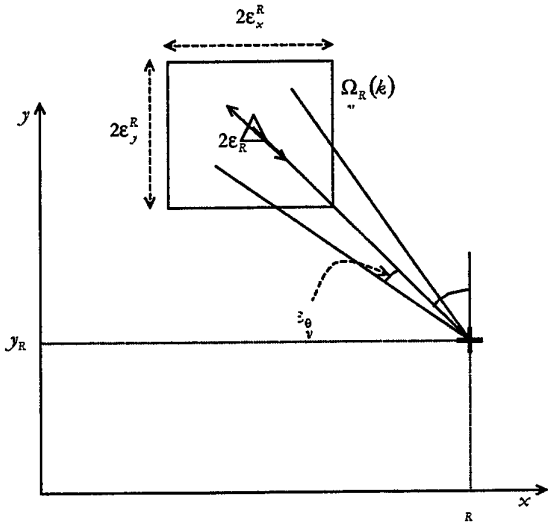


Figure 3 Radar sensor measurement transformation

3 Set-based tracking algorithm

$\Omega_s(k-1|k-1)$

1 $\Omega_s(k-1|k-1)$ k $\Omega_s(k|k-1)$

$$\Omega_s(k|k)=\Omega_s(k|k-1)\bigcap\Omega_R(k)\bigcap\Omega_E(k) \qquad 6$$

1 6

6

6

$\Omega_s(k|k)=\Omega_s(k|k-1)\bigcap\Omega_R(k)$

$$x_{\max}(k|k)=\begin{cases}z_x^R(k)+\varepsilon_x^R(k), & \text{if } z_x^R(k)+\varepsilon_x^R(k)\le x_{\max}(k|k-1)\\ x_{\max}(k|k-1), & \text{if } z_x^R(k)+\varepsilon_x^R(k)> x_{\max}(k|k-1)\end{cases} \qquad 7$$

$$x_{\min}(k|k)=\begin{cases}z_x^R(k)-\varepsilon_x^R(k), & \text{if } z_x^R(k)-\varepsilon_x^R(k)\ge x_{\min}(k|k-1)\\ x_{\min}(k|k-1), & \text{if } z_x^R(k)-\varepsilon_x^R(k)< x_{\min}(k|k-1)\end{cases}$$

$$y_{\max}(k|k)=\begin{cases}z_y^R(k)+\varepsilon_y^R(k), & \text{if } z_y^R(k)+\varepsilon_y^R(k)\le y_{\max}(k|k-1)\\ y_{\max}(k|k-1), & \text{if } z_y^R(k)+\varepsilon_y^R(k)> y_{\max}(k|k-1)\end{cases}$$

$$y_{\min}(k|k)=\begin{cases}z_y^R(k)-\varepsilon_y^R(k), & \text{if } z_y^R(k)-\varepsilon_y^R(k)\ge y_{\min}(k|k-1)\\ y_{\min}(k|k-1), & \text{if } z_y^R(k)-\varepsilon_y^R(k)< y_{\min}(k|k-1)\end{cases}$$

O

$$v_{x,\max}(k|k)=\begin{cases}\frac{1}{kT}\left[x_{\max}(0)-x_{\max}(k|k-1)\right], & \text{if } x_{\max}(k|k)\le x_{\max}(k|k-1)\\ v_{x,\max}(k|k-1), & \text{if } x_{\max}(k|k)= x_{\max}(k|k-1)\end{cases} \qquad 8$$

$$v_{x,\min}(k|k)=\begin{cases}\frac{1}{kT}\left[x_{\min}(k|k-1)-x_{\min}(0)\right], & \text{if } x_{\min}(k|k)\ge x_{\min}(k|k-1)\\ v_{x,\min}(k|k-1), & \text{if } x_{\min}(k|k)= x_{\min}(k|k-1)\end{cases}$$

$$v_{y,\max}(k|k)=\begin{cases}\frac{1}{kT}\left[y_{\max}(0)-y_{\max}(k|k-1)\right], & \text{if } y_{\max}(k|k)\le y_{\max}(k|k-1)\\ v_{y,\max}(k|k-1), & \text{if } y_{\max}(k|k)= y_{\max}(k|k-1)\end{cases}$$

$$v_{y,\min}(k|k)=\begin{cases}\frac{1}{kT}\left[y_{\min}(k|k-1)-y_{\min}(0)\right], & \text{if } y_{\min}(k|k)\geq y_{\min}(k|k-1)\\ v_{y,\min}(k|k-1), & \text{if } y_{\min}(k|k)=y_{\min}(k|k-1)\end{cases}$$

N

$$\Omega_s(k|k)=\Omega_s(k|k-1)\bigcap\Omega_E(k)$$

1 $\Omega_s(k|k-1)$

$\Omega_E(k)$

1

$\Omega_s(k|k)$

1

$$\mathbf{s}(k|k-1)\in\Omega_s(k|k-1)=\left\{x,y,v_x,v_y:\begin{array}{r}17\geq x\geq 11\\ 17\geq y\geq 10\\ -1\geq v_x\geq 1\\ -1\geq v_y\geq 1\end{array}\right\}$$

$$\mathbf{z}_E^c(k)\in\Omega_E(k)=\{x,y:|y-x+3|\leq 4\}.$$

Step 1 yields the following boundaries:

$$(x<25,y\leq 17),(x\geq 17,y\leq 17),(x\leq 18,y\geq 10),(x\geq 10,y\geq 10)$$
$$(x\leq 17,y>9),(x\leq 17,y\leq 17),(x\geq 11,y\geq 3),(x\geq 11,y\leq 11)$$

Step 2 yields the following consistent boundaries:

$$(x\geq 17,y\leq 17),(x\leq 17,y\leq 17),(x\geq 11,y\leq 11)$$

Step 3 yields the following final boundaries

$$\mathbf{s}(k) \in \Omega_s(k|k) = \left\{ x, y, v_x, v_y : \begin{matrix} 17 \geq x \geq 11 \\ 17 \geq y \geq 10 \\ -1 \geq v_x \geq 1 \\ -1 \geq v_y \geq 1 \end{matrix} \right\}.$$

Note in this case the boundaries have not changed. This is due to the fact that rectangles due not tightly over-bound the intersection of two sets. Step 4 has been omitted for simplicity. Finally, the set-based tracking algorithm is easily initialized by assigning an arbitrarily large set to $\Omega_s(0)$.

4 Conclusions

Set based estimation techniques provide the basis for robust and simple tracking where one is required to work with non-linear coordinate systems or where assumptions of Gaussian and stationary noise are inappropriate. The set based tracker also has the benefit of simple implementation in fixed point arithmetic microprocessors. It is well known that in practical tracking systems, ad hoc techniques are usually employed to improve Kalman filter stability including fixing Kalman gains after a few track updates or reverting to alpha-beta trackers with fixed gains. Set based tracking has no such problems as there is no matrix inversion. The techniques also lends itself well to manoeuvre detection.

The technique described in this paper is being investigated as the basis for a distributed sensor tracking system including integrated data association and sensor management where there are communication bandwidth constraints.

Simulation results will be presented at the workshop.

5 References

[1] Y.Bar-Shalom and T.E.Fortmann, *Tracking and Data Association*, Academic Press, N.Y., 1988

[2] Y.Bar-Shalom and X-R.Li, *Multitarget-Multisensor Tracking: Principles and Techniques*, Lecture Notes, 1995

[3] V.Broman and M.J.Shensa, "Polytopes, A Novel Approach to Tracking", Proc 25th IEEE Conf Decision and Control, Dec 1986, pp 1749-1752

[4] R.J.Evans and F.Barker, "The OCCAM Filter- A New Technique for Track-While-Scan", Proc IEEE Radar Conf, 1980, pp 342-347

[5] F.C.Schweppe, "Recursive State Estimation: Unknown but Bounded Errors and System Inputs", IEEE Trans Automatic Control, Vol 13, No 1, Feb 1968, pp 22-28

Source Localization with Distributed Electromagnetic Component Sensor Array Processing

Chong-Meng Samson See *and Arye Nehorai †

Abstract

We propose an approach to achieve high-performance localization of multiple sources using a small aperture array of spatially-distributed electric and magnetic component sensors. The approach is based on exploiting of all available electromagnetic information along with the time delay information. Using simulated data, we demonstrate that this approach outperforms both a single vector-sensor and scalar-sensor arrays in accuracy of direction-of-arrival (DOA) estimation.

1 Introduction

The problem of estimating electromagnetic wave parameters using sensor arrays has attracted significant attention over recent years and lead to the development a number of high resolution algorithms, such as MUSIC, ESPRIT and WSF. These algorithms have focused on direction-of-arrival estimation in such areas as wireless communications and radar.

Most existing array processing methods rely on the spatial diversity of the sensor array to estimate the DOA. A drawback of this approach is that the performance accuracy becomes highly dependent on the size of the array's *electrical* aperture. In many applications, the array is expected to operate over a wide frequency range. To avoid ambiguities in the array manifold, the physical size of such broadband array is constrained by the highest operating frequency and the number of sensors. Poorer performance at lower frequencies will result due to their larger wavelengths, especially when small number of receiver channels is available. The costly approach to alleviate this problem is to aim for larger "unambiguious" array geometry by increasing the number of receiver channels. Another way to overcome this problem is to use multiple sets of sensor arrays where each set is optimized to operate over a smaller bandwidth. This may not be feasible in mobile- or fast-deployment sensor array applications. Hence, there is a need to develop DOA estimation methods that use a small-aperture array that achieve good performance over a wide operating frequency.

The DOA estimator's performance can be improved by using polarization-sensitive sensor array to exploit the polarization diversity of the signals by estimating their signal polarization parameters along with their DOA [2] [3] [4]. In a recent development, Nehorai and Paldi [1] introduced the concept of vector-sensor array processing where the complete electromagnetic information of the signal is measured and processed. They apply the Poynting relationship between the electric and magnetic measurements to enable estimation of the DOA of multiple signal sources using a *single* vector-sensor. Direction-finding with a vector sensor (SuperCART antenna array) was demonstrated in [7]. SInce it does not rely on spatial diversity, a DOA estimator using a single vector sensor should exhibit consistent performance over its operating frequency band and should easily work with wide-band signals[1]. When operating as an array of vector sensors, the electromagnetic and time-delay measurements can be simultaneously used to estimate the DOAs. This allows the use of smaller aperture-arrays while maintaining good performance over a wide frequency bandwidth. However, employing an array of vector sensors may be expensive because large number of receivers is necessary. For example, a 3-vector sensors array will require an 18-channel receiver.

*C.M.S. See is with DSO National Laboratories, 20 Science Park Drive, Singapore 118230. Tel: 065-7727295. Fax: 065-7766476. Email:schongme@dso.org.sg

†A. Nehorai is with the University of Illinois, Chicago, USA. Tel: (312)996-2778. Fax:(312)413-0024. Email: nehorai@eecs.uic.edu. The work of A. Nehorai was supported by the Air Force of Scientific Research under Grants F49620-97-1-0481 and F49620-99-1-0067, the National Science Foundation under Grant MIP-9615590 and the office of Naval Research under Grant N0014-98-1-0067.

This paper proposes a simple and effective alternative for achieving good DOA estimation performance with small aperture-arrays. The approach uses an array of spatially distributed scalar magnetic and electric sensors. We shall call the proposed array as distributed electromagnetic component array (DEMCA). It is assumed that the array of scalar magnetic and electric sensors should, in aggregate, measure at least the all the 3D electric and magnetic components of the electromagnetic wave. The proposed DEMCA will afford the following three advantages: Firstly, the full electric and magnetic field components measure by the magnetic and electric sensors; thereby effectuating derivation of the sources' directional information. Secondly, their spatial distribution will allow extraction of additional sources' directional information by way of the differential-delay measurements. Finally, DEMCA's structure will significantly economize the number of receivers needed to simultaneously utilize the time-delay and *complete* electromagnetic information for DOA estimation.

2 Measurement Model

Adopting the conventions in [1], the measurement model of the vector sensor is given by

$$\begin{bmatrix} \mathbf{y}_E(t) \\ \mathbf{y}_H(t) \end{bmatrix} = \begin{bmatrix} \mathbf{I}_3 \\ (\mathbf{u}\times) \end{bmatrix} \mathbf{VQw}s(t) + \begin{bmatrix} \mathbf{e}_E(t) \\ \mathbf{e}_H(t) \end{bmatrix}, \tag{1}$$

where

$$(\mathbf{u}\times) = \begin{bmatrix} 0 & -u_z & u_y \\ u_z & 0 & -u_z \\ -u_y & u_x & 0 \end{bmatrix}, \tag{2}$$

$\mathbf{u}$ is the unit direction vector from sensor to source and u_x, u_y and u_z are the x, y and z components. The matrices $\mathbf{V}$, $\mathbf{Q}$ and vector $\mathbf{w}$ are given by

$$\mathbf{V} = \begin{bmatrix} -\sin\theta_1 & -cos\theta_1 \sin\theta_2 \\ \cos\theta_1 & -\sin\theta_1 \sin\theta_2 \\ 0 & \cos\theta_2 \end{bmatrix}, \tag{3}$$

$$\mathbf{Q} = \begin{bmatrix} \cos\theta_3 & \sin\theta_3 \\ -\sin\theta_3 & \cos\theta_3 \end{bmatrix} \tag{4}$$

and

$$\mathbf{w} = \begin{bmatrix} \cos\theta_4 \\ j\sin\theta_4 \end{bmatrix} \tag{5}$$

where θ_1, θ_2, θ_3 and θ_4 are the azimuth, elevation, ellipse's orientation and ecentricity angle.

Extending from (1) and assuming that the signal sources are *narrowband*, we can write the measurement model of the distributed component sensor array in a multiple source environment as [6]

$$\underbrace{\begin{bmatrix} \mathbf{y}_E(t) \\ \mathbf{y}_H(t) \end{bmatrix}}_{\mathbf{y}(t)} = \sum_{k=1}^{d} \mathbf{a}(\boldsymbol{\theta}^{(k)}) s_k(t) + \underbrace{\begin{bmatrix} \mathbf{e}_E(t) \\ \mathbf{e}_H(t) \end{bmatrix}}_{\mathbf{n}(t)}, \tag{6}$$

$$\mathbf{a}(\boldsymbol{\theta}^{(k)}) = \boldsymbol{\Gamma}(\theta_1^k, \theta_2^k)\boldsymbol{\Omega} \begin{bmatrix} \mathbf{I}_3 \\ (\mathbf{u}_k\times) \end{bmatrix} \mathbf{V}_k\mathbf{Q}_k\mathbf{w}_k$$

where $\boldsymbol{\theta}^{(k)} = [\theta_1^{(k)}, \theta_2^{(k)}, \theta_3^{(k)}, \theta_4^{(k)}]$ denotes the directional and polarization parameters of the k^{th} source signal. $\boldsymbol{\Gamma}(\theta_1,\theta_2)$ is a diagonal matrix whose n^{th}diagonal entry is given by $[\boldsymbol{\Gamma}(\theta_1, \theta_2)]_{nn} = a_n(\theta_1, \theta_2)e^{j\omega_c\tau_n}$, where τ_n is the differential delay of the signal source between the n^{th} component and the phase center and $a_n(\theta_1, \theta_2)$ is the response of the n^{th} component sensor; ω_c is the carrier frequency and $\boldsymbol{\Omega}$ is a selection matrix elements of 1 and 0. For example, when orthogonal triads of magnetic and electric sensors are used, $\boldsymbol{\Omega} = \mathbf{I}_6$. If an additional x electric-component sensor is used, the selection matrix becomes

$$\boldsymbol{\Omega} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ & & & \mathbf{I}_6 & & \end{bmatrix}. \tag{7}$$

From (6), observe that the electromagnetic sources directional information are all embedded in

$$\mathbf{\Gamma}(\theta_1^k, \theta_2^k)\mathbf{\Omega}\begin{bmatrix} \mathbf{I}_3 \\ (\mathbf{u}_k\times) \end{bmatrix}\mathbf{V}_k.$$

This allows the differential delay measurements resulting from diverse placement of the component sensors and electromagnetic field measurements to be jointly exploited in estimating the source parameters. Given both the complete electromagnetic and spatial information, good parameter estimation with a smaller aperture array can be expected over a wide frequency range. It suffices to point out that the distributed-component sensors array model in (6) generalizes the vector-sensor array [6].

We can express (6) compactly in matrix form as

$$\mathbf{y}(t) = \mathbf{A}\mathbf{s}(t) + \mathbf{n}(t) \tag{8}$$

where

$$\mathbf{A} = \left[\mathbf{a}(\boldsymbol{\theta}^{(1)})\cdots\mathbf{a}(\boldsymbol{\theta}^{(d)})\right] \tag{9}$$

and $\mathbf{s}(t) = [s_1(t)\ldots s_d(t)]^T$.

3 Cramer Rao Bound

We use the Cramer-Rao bound (CRB) to examine the performance gain achievable by our approach. Using the notations, statistical assumptions and results in [1] [6], the CRB is given by

$$\begin{aligned} \mathbf{C}_{crb}(\mathbf{\Theta}) &= \frac{\sigma^2}{N}\mathrm{Re}\{\mathbf{J}^{-1}\}, \\ \mathbf{J} &= \mathrm{btr}((\mathbf{1}\boxtimes\mathbf{U})\boxdot(\mathbf{D}^H\mathbf{\Pi}_c\mathbf{D})^{bT}) \end{aligned} \tag{10}$$

where

$$\begin{aligned} \mathbf{P} &= \mathrm{E}\left(\mathbf{s}(t)\mathbf{s}^H(t)\right), \\ \mathbf{U} &= \mathbf{P}(\mathbf{A}^H\mathbf{A}\mathbf{P} + \sigma^2\mathbf{I})^{-1}\mathbf{A}^H\mathbf{A}\mathbf{P}, \\ \mathbf{\Pi} &= \mathbf{I} - \mathbf{A}(\mathbf{A}^H\mathbf{A})^{-1}\mathbf{A}^H, \\ \mathbf{D} &= \left[\mathbf{D}_1^{(1)}\cdots\mathbf{D}_4^{(1)}\cdots\mathbf{D}_1^{(d)}\cdots\mathbf{D}_4^{(d)}\right], \\ \mathbf{D}_l^{(k)} &= \frac{\partial\mathbf{a}(\boldsymbol{\theta}^{(k)})}{\partial\theta_l^{(k)}}, \\ \mathbf{\Theta} &= \left[\boldsymbol{\theta}^{(1)T}\cdots\boldsymbol{\theta}^{(d)T}\right]^T \end{aligned}$$

and where σ^2 is the noise power and N is the number of independent snapshots. In order to circumvent the intrinsic singularities due to the reference coordinate system, the mean square angular error (MSAE) was proposed in [1] and is given by

$$MSAE_{cr} \triangleq N[\cos^2\theta_2 C_{crb}(\theta_1) + C_{crb}(\theta_2)]. \tag{11}$$

4 Numerical Example

By using a numerical example, we shall demonstrate the greater efficacy of the distributed electromagnetic component sensor array (DEMCA) processing when it is compared with scalar-array processing that relies on an *electric-only,* diversely polarized and co-polarized antenna array. Since the motivation of this development is the design of a small-aperture sensor array, we shall make the comparison based on the principle of "equal aperture, equal number of channels". We assume a six-channel receiver and use a six element uniform circular array in this analysis. This will allow the comparison between the performance of a vector-sensor as well as a six element diversely and co-polarized array with the proposed DEMCA. The diversely-polarized array used in this study is

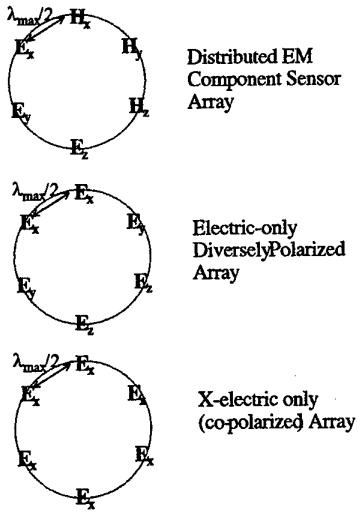


Figure 1: Array Geometry of Distributed EM Component Sensor Array, x-electric (co-polarized) array and electric-only diversely polarized array. $E_x(H_x)$, E_y (H_y) and $E_z(H_z)$ are the electric (magnetic) component sensors.

an array of x, y and z-electric component sensors. The difference between the diversely-polarized and the proposed sensor array is that the former uses only electric component sensors while the latter uses both electric and magnetic component sensors to form a six-element sensor array with a six-channel receiver. The three sensor arrays are depicted in Figure 1. Note that the inter-element spacing is fixed at $\lambda_{\max} = \frac{c}{2f_{\max}}$, where c is the speed of light and $f_{\max}$ is the maximum operating frequency.

An example of the DOA estimation performance as a function of frequency is shown in Figure 2. We considered two uncorrelated sources with $\boldsymbol{\theta}^{(1)}= [1°, 10°, 45°, 0°]^T$ and $\boldsymbol{\theta}^{(2)}= [5°, 9°, -45°, -5°]^T$. The signal-to-noise ratio is fixed at 10dB. Therein the inter-element spacing of the uniform circular array is fixed at $\frac{c}{2f_{\max}}$. Observe from the figure that the distributed EM sensor array has consistent performance over a wide operating bandwidth. In addition, it achieved four orders of magnitude of gain in accuracy of DOA estimation over the *x electric* array and one order of magnitude over the *electric-only,* diversely-polarized array at $\frac{f}{f_{\max}} = 0.3$. This result clearly demonstrates the gain obtainable from the full exploitation of the spatial and electromagnetic information afforded by DEMCA.

Figure 3 plots the DOA estimation performance as a function of the azimuthal angle of separation between uncorrelated two sources having 10dB SNR. The normalized operating frequency is fixed at $\frac{f}{f_{\max}} = 0.3$. The graph shows that the proposed DEMCA demonstrates significant performance gain especially for closely-spaced sources. This feature is particularly useful in applications with short integration time or at low signal-to-noise ratio.

5 Concluding Remarks

We have presented a new approach for the localization of electromagnetic sources through the joint exploitation of spatial diversity and electromagnetic information using spatially-distributed electric and magnetic componet sensors. Performance analysis via numerical examples illustrated the potential gain of the proposed approach over the scalar and diversely polarized array. The analysis indicated that the distributed component EM sensor array should allow the use of small array apertures while maintaining desired resolution and performance accuracy over a wide operating bandwidth.

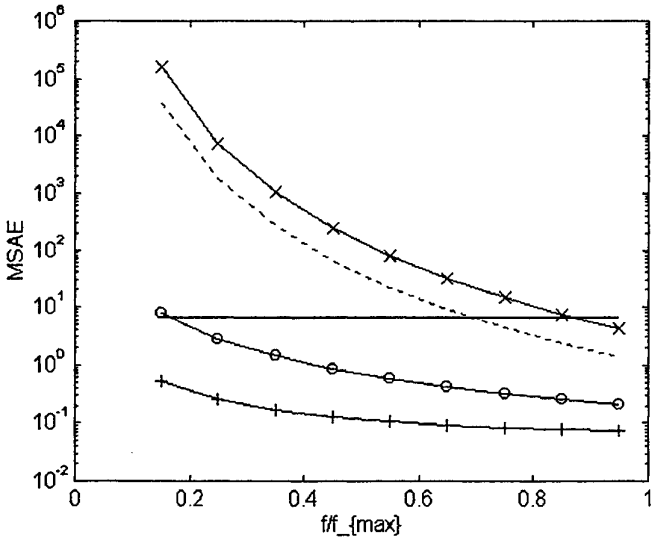


Figure 2: CRB vs Frequency. +: DEMCA. o : Diversely polarized dipole array. -: Vector sensor. ... : Scalar array of omni-directional sensors. × : Scalar array of x-electric sensor (dipole).

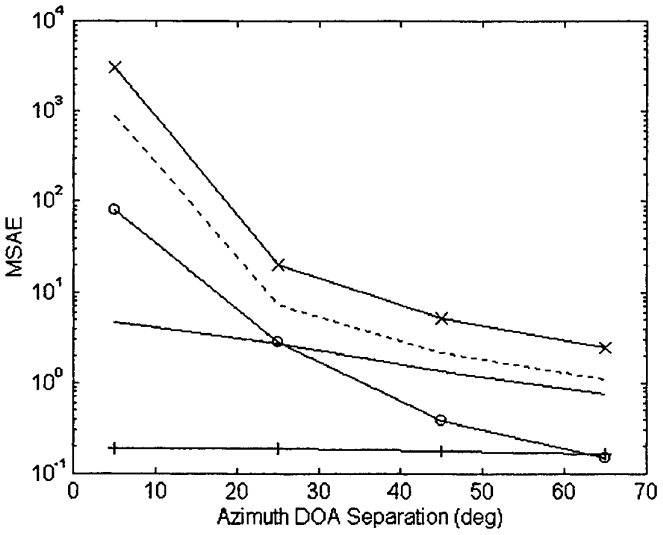


Figure 3: CRB vs Angular Separation

Acknowledgement

The authors would like to thank Mr Bob Washington, University of Illinois, Chicago, whose comments were helpful in improving the presentation of this paper.

References

[1] A. Nehorai and E. Paldi, "Vector-Sensor Array Processing for Electromagnetic Source Localization," IEEE Trans. on Signal Processing, Vol.42, No.2, pp.376-398, Feb. 1994. (a shorter version of this paper has also appeared in Proc 25th Asilomar Conf. Signals Syst. Comput., Nov. 1991, pp. 566-572)

[2] E.R. Ferrara Jr. and T.M. Parks, "Direction finding with an array of antenna having diverse polarization", IEEE Trans. Antennas Propagat., vol. AP-31, pp.231-236, Mar. 1983.

[3] R. Schmidt, "A Signal Subspace Approach to Multiple Emitter Location and Spectral Estimation, " P.h.D. dissertation, Stanford University, C.A., Nov. 1981.

[4] I. Ziskind and M. Wax, "Maximum Likelihood Localization of Diversely Polarized Sources by Simulated Annealing," IEEE Trans. Antennas and Propagation, Vol. 38, pp. 1111-1114, July 1990.

[5] J. Li, "Direction and Polarization Estimation Using Arrays with Small Loops and Short Dipoles," IEEE Trans. Antennas and Propagation, Vol. 41, No. 3, pp. 379-387, March 1993.

[6] C.M.S. See and A. Nehorai, "Distributed Electromagnetic Component Sensor Array", 7th Annual Adaptive Sensor Array Processing Workshop, March 1999. Full version in preparation.

[7] G.F. Hatke, "Performance Analysis of the SuperCART Antenna Array", Project Report AST-22, Lincoln Laboratory, Massachusetts Institute of Technology, 23 March 1992.

Nonlinear Equalization Using SVMs

Daniel J. Sebald and James A. Bucklew

Department of Electrical and Computer Engineering, University of Wisconsin-Madison
1415 Engineering Drive; Madison, WI 53706-3840

E-mail: sebald@cae.wisc.edu; bucklew@engr.wisc.edu

Support vector machines are investigated as a method for performing nonlinear equalization in communication systems. The support vector machine has the advantage that a smaller number of parameters for the model can be identified in a manner that does not require the extent of prior information or heuristic assumptions that some previous techniques require. An enhanced method of using a bank of support vector machines allows utilization of previously detected symbols to increase performance. Simulation results are compared against results from other researchers using techniques such as neural networks and decision feedback. We find that the support vector machine generalizes well given a set of training data, with the tradeoff typically being decreased performance from the optimum Bayesian solution.

Key Words: nonlinear equalization; support vector machine; SVM; ISI; decision feedback

0. INTRODUCTION

A support vector machine (SVM) [2] embodies the concepts of generalized learning theory developed by Vapnik [8] and others. A SVM uses training data as an integral element of the function estimation model as opposed to simply using training data to estimate parameters of an apriori model using maximum likelihood techniques. In this sense, it is more general in terms of noise and correlation properties than methods such as radial basis function (RBF) networks [4]. Furthermore, the optimization, or learning, method of SVMs is more manageable and generalizes better than techniques such as Volterra filters [1, 5] and neural networks [3].

These are strong motivations for investigating the use of SVMs for nonlinear equalization, or more appropriately *detection*. In the case of equalization, it is desirable that a modem require a small set of training data to characterize the transmission channel. Also, the model should be efficient for real-time applications. SVMs train with relatively small amounts of data, and once training of the SVM has completed, the detection stage is efficient, comparable to Volterra filters and neural networks.

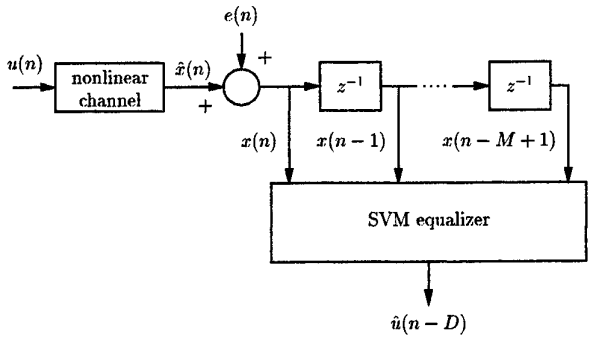


FIG. 1. Model of the nonlinear transmission system, originating from Chen, et al. [3].

Standard SVM results are compared against that for a neural network presented in [3]. Then results for a SVM-bank (SVMB) structure are compared against that for a decision-feedback equalizer presented in [6]. The latter system has the advantage that it uses detected symbols to modify decision boundaries. The result is increased performance over the standard SVM in typical communication channels having appropriate signal-to-noise ratios (SNR).

1. DETECTION VIEWED AS PATTERN RECOGNITION

As pointed out in [3], equalization may be viewed as a classification problem. In such a scenario, the output of a communications channel can be grouped as a vector and used as the input to a classification machine whose output should match as best as possible some delayed version of the original signal entering the channel. Figure 1 shows a discrete-time pulse amplitude modulation model of a communications channel. The transmitted data sequence, $u(n)$, is an independent, equiprobable binary sequence taking values $\{-1, +1\}$. The output of the channel, $x(n) \in \mathbb{R}$, is the sum of a deterministic, nonlinear function of $u(n)$, $\hat{x}(n)$, and an additive noise, $e(n)$. The goal of the equalizer is then to mimic the desired output $u(n - D)$. Call the equalizer detection output $\hat{u}(n - D)$.

The deterministic portion of the channel model consists of a linear, finite impulse response (FIR) filter followed by a polynomial nonlinearity. Let

$$\tilde{x}(n) = \sum_{k=0}^{N-1} h_k u(n-k), \quad \text{and} \quad \hat{x}(n) = \sum_{p=1}^{P} c_p \tilde{x}^p(n),$$

where $\{h_k\}$ are the FIR filter coefficients and $\{c_p\}$ are the polynomial coefficients. By grouping the output of the channel into vectors

$$\mathbf{x}(n) = \left[x(n) \;\; x(n-1) \;\; \ldots \;\; x(n-M+1)\right]^T$$

and taking the desired classification to be a training sequence input to the channel delayed by D samples, i.e., $y_n = u(n - D)$, a SVM can be trained to solve the detection problem. Varying the delay D results in different performance of the equalizer because the correlation between $u(n - D)$ and $\mathbf{x}(n)$ changes with D.

2. SVM

A SVM is a method for separating clouds of data in the feature space, i.e., the space generated from nonlinear mappings of the pattern space data $\mathbf{x}(n)$, using an *optimal* hyperplane. Given an input vector $\mathbf{x}$, an SVM classifies according to

$$\hat{y} = \text{sign}\{f(\mathbf{x})\}$$

where $\hat{y}$ is the estimate to the classification and

$$f(\mathbf{x}) = \sum_{i \in S} \alpha_i y_i \Phi(\mathbf{x}_i) \cdot \Phi(\mathbf{x}) + b = \sum_{i \in S} \alpha_i y_i K(\mathbf{x}_i, \mathbf{x}) + b. \tag{1}$$

Here $\{\alpha_i\}$ are Lagrange multipliers, S is the set of indices for which $\mathbf{x}_i$ is a *support vector*, i.e., a vector for which $\alpha_i \neq 0$ after optimization, and $K(\cdot,\cdot) : \mathbb{R}^M \times \mathbb{R}^M \mapsto \mathbb{R}$ is a kernel satisfying the conditions of Mercer's theorem [7, 8]. We see in (1) that after training, only a subset of the training data enters the model (i.e., data reduction) and operations are only performed on data in the pattern space, not in the feature space (i.e., more manageable than previously studied nonlinear techniques).

According to Vapnik [8], for training data which is non-separable the dual optimization problem is to maximize

$$W(\alpha) = \sum_{i=1}^{L} \alpha_i - \frac{1}{2} \sum_{i,j=1}^{L} \alpha_i \alpha_j y_i y_j K(\mathbf{x}_i, \mathbf{x}_j)$$

under constraints

$$\sum_{i=1}^{L} \alpha_i y_i = 0, \quad \text{and} \quad 0 \leq \alpha_i \leq C, \quad i = 1, \ldots, L.$$

where increasing C penalizes errors more heavily [2]. This is a quadratic programming problem that may be solved with traditional optimization techniques.

Preliminary studies on character recognition problems [8] suggest that the type of SVM kernel is inconsequential so long as the capacity is appropriate for the the amount of training data and complexity of the classification boundary. In the simulations here the polynomial kernel $K(\mathbf{x}, \mathbf{z}) = (\mathbf{x} \cdot \mathbf{z} + 1)^d$ is used exclusively. The polynomial order d is a parameter which controls the capacity of the SVM. The greater d, the more complex classification boundary the SVM can create.

3. SIMULATIONS AND RESULTS

The constellation points are the noise-free channel outputs resulting from the various inputs and are classified according to the value of D. Let the noise-free channel outputs be grouped as a vector $\hat{\mathbf{x}}(n) = \left[\hat{x}(n) \;\; \hat{x}(n-1) \;\; \ldots \;\; \hat{x}(n-M+1)\right]^T$. Then the constellation sets and classification regions can be expressed as $C_{s_i,D} = \left\{\hat{\mathbf{x}}(n) | u(n-D) = s_i\right\}$ and $R_{s_i,D} = \left\{\mathbf{x}(n) | \hat{u}(n-D) = s_i\right\}$, $i = 1, 2$. The optimum classifier assumed here is the Bayesian maximum likelihood detector under conditions of equiprobable probabilities and zero/one cost [7].

TABLE 1
Simulation statistics for a standard SVM detector.

D	$P_{e,\text{train}}$	$P_{e,\text{test}}$	$P_{e,\text{unproc}}$	$\sigma_{P_{e,\text{train}}}$	$\sigma_{P_{e,\text{test}}}$	$\sigma_{P_{e,\text{unproc}}}$	$\bar{N}_{\text{sv}}$	$\bar{N}_{\text{msv}}$
0	0.163	0.172	0.602	0.0166	0.00722	0.00825	205	10.0
1	0.0458	0.0589	0.905	0.00727	0.00473	0.00408	66.5	9.40
2	0.0366	0.0421	0.503	0.00800	0.00575	0.00506	51.1	9.30

The first simulation is with the nonlinear channel $\hat{x}(n) = \tilde{x}(n) - 0.9\,\tilde{x}^3(n)$, $\tilde{x}(n) = u(n) + 0.5\,u(n-1)$, additive white Gaussian noise of power $\sigma_e^2 = 0.2$, and SVM parameters $M = 2$, $C = 5$ and $d = 3$. Results are an average of ten trials with 500 samples in the training set and 5000 samples in the test set. Statistics for the simulations are given in Table 1. Most encouraging is that the standard deviations for probability of error for the SVM are approximately one eighth of the probability of error, and the error probabilities for training and test data are approximately the same. This confirms that the SVM is very good at generalizing.

We now test the behavior of the SVM on a channel having zero-mean, colored noise with $E[e(n)e(n-1)] = \rho$. Figure 2 shows an example SVM classifier for channel $\hat{x}(n) = \tilde{x}(n) + 0.1\,\tilde{x}^2(n) + 0.05\,\tilde{x}^3(n)$, $\tilde{x}(n) = 0.5\,u(n) + u(n-1)$, $\sigma_e^2 = 0.2$, $\rho = 0.48$, $M = 2$, $D = 0$, $d = 3$, and $C = 5$. The number of trials, training samples and test samples were the same as in the previous simulation. Region $R_{+1,D}$ is shaded while $R_{-1,D}$ is left unshaded. Included on the pattern space is the signal constellation where $C_{+1,D}$ is marked by a large $\bullet$ and $C_{-1,D}$ is marked by a large $\times$. Training data is also displayed, using a small $\circ$ to indicate $u(n-D) = +1$ and a small $\times$ to indicate $u(n-D) = -1$. The optimum Bayesian solution is given in [3]. The SVM chooses a decision boundary similar to the optimum and is logical in terms of the training data. The optimum $R_{-1,0}$ for this example includes a disconnected region, but the SVM cannot match the polygon nature of the optimum.

A third example compares the SVM bit error rate (BER) against the optimum BER as a function of SNR for the channel $\hat{x}(n) = \tilde{x}(n) + 0.2\,\tilde{x}^2(n)$, $\tilde{x}(n) = 0.3482\,u(n) + 0.8704\,u(n-1) + 0.3482\,u(n-2)$, $\rho = 0.48$, $M = 3$, $D = 1$, $d = 3$, and $C = 0.1$. The number of training samples was again 500, while the number of trials and test samples were varied to compensate for greater relative variance for low BER estimates. The BER results, given in Fig. 3, show that the SVM requires approximately 2.0–2.5 dB more SNR to match the Bayesian performance. Above 15 dB SNR, the SVM result is essentially the same as the neural network solution given in [3].

The last simulation considers channels having severe ISI for which the standard SVM detector does not perform well. It has been shown for linear channels [6] that, depending upon the nature of the ISI, significantly better performance can often be achieved by incorporating previously detected symbols into the detector. This concept is exemplified in the decision feedback equalizer (DFE). The decision feedback idea can be incorporated into an SVM by simply lengthening its input vector by appending previous SVM outputs. That is, let

$$\mathbf{x}(n) = \left[\, x(n) \;\; x(n-1) \;\; \ldots \;\; x(n-M+1) \;\; \hat{u}(n-D-1) \;\; \ldots \;\; \hat{u}(n-D-M') \,\right]^T$$

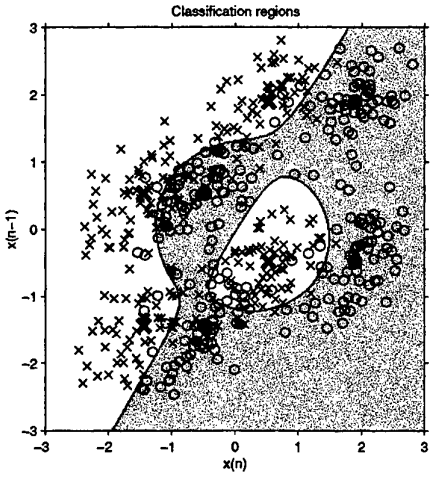


FIG. 2. Example of typical classification region of a standard SVM detector.

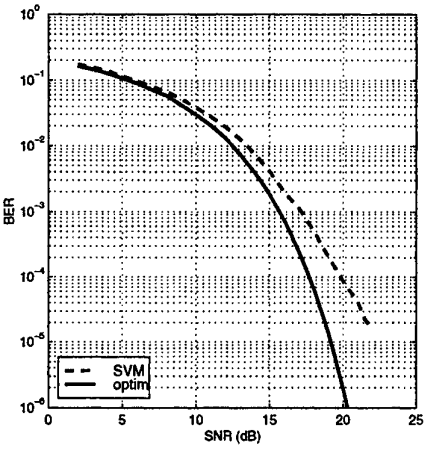


FIG. 3. Average BER for the standard SVM detector versus the Bayesian detector.

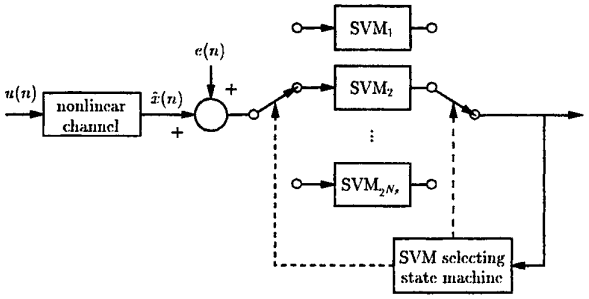


FIG. 4. A bank of SVMs with selection based upon the state of previous decisions. Each SVM constructs a different decision boundary based upon the constellation for a given state.

where M' is the number of previously detected symbols fed back. We call this approach the decision feedback SVM (DFSVM). A DFSVM with correct decisions fed back is called the perfect DFSVM (PDFSVM).

Chen, et al. [4] described an important property of the pattern space related to previously detected symbols and proposed a novel method of utilizing this property in their system. The SVM can be altered to utilize the same ideas by building a bank of SVMs controlled by a state machine having previously detected symbols as an input. This approach is illustrated in Fig. 4. Individual SVMs are optimized using subsets of training data conditioned upon previous symbols.

The channel model for this last example is that studied by Proakis in [6] for DFE. That channel is linear, i.e., $\hat{x}(n) = \tilde{x}(n)$, with output $\tilde{x}(n) = 0.407\,u(n) + 0.815\,u(n-1) + 0.407\,u(n-2)$. The parameters for the various types of SVM in the simulation were $M = 2$, $D = 1$, $d = 3$, and $C = 0.5$. For the DFSVM and PDFSVM $M' = 1$. The added noise was white. The number of training samples was 500, and results were averaged over ten trials. The number of test points was varied to account for variation in BER. The optimum non-ISI binary signaling is $P_{\mathrm{e,opt}} = \frac{1}{2}\mathrm{erfc}(\sqrt{\gamma})$, where $\mathrm{erfc}(\cdot)$ is the complimentary error function and γ is the SNR for a real channel. Figure 5 shows the performance as a function of SNR

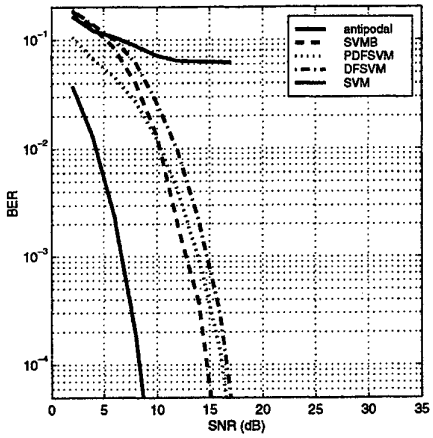


FIG. 5. Average BER for the SVM, SVMB, DFSVM, and PDFSVM detectors compared against the optimum binary, non-ISI detector.

for the various SVM approaches. The SVMB outperforms the DFSVM, requiring approximately 2.0 dB less SNR to achieve the same BER, and it performs about the same as the DFE of [6].

4. SUMMARY

Simulations have shown that the SVM provides a robust method for addressing nonlinearities in communication channels exhibiting ISI. The method performs as well as neural networks and Volterra filters, and has several advantages over these methods. The use of a bank of SVMs controlled by a state machine allows incorporation of decision direction. This significantly increases performance for certain channel scenarios. Currently, the SVM can only be used in block adaptive applications since no sample-by-sample adaptive version is known to exist.

REFERENCES

1. S. Benedetto, E. Biglieri, and V. Castellani. *Digital transmission theory.* Prentice-Hall, Englewood Cliffs, NJ, 1987.
2. Christopher J. C. Burges. A tutorial on support vector machines for pattern recognition. *Data mining and knowledge discovery,* 2(2):1–47, 1998.
3. S. Chen, G. J. Gibson, C. F. N. Cowan, and P. M. Grant. Adaptive equalization of finite nonlinear channels using multilayer perceptrons. *EURASIP Signal Processing,* 20(2):107–119, June 1990.
4. Sheng Chen, Bernard Mulgrew, and Stephen McLaughlin. Adaptive Bayesian equalizer with decision feedback. *IEEE Trans. Signal Processing,* 41(9):2918–2927, September 1993.
5. Robert D. Nowak and Barry D. Van Veen. Volterra filter equalization: a fixed point approach. *IEEE Trans. Signal Processing,* 45:377–388, February 1997.
6. John G. Proakis. *Digital communications.* McGraw-Hill, New York, 3rd edition, 1995.
7. Henry Stark and John W. Woods. *Probability, random processes, and estimation theory for engineers.* Prentice-Hall, Englewood Cliffs, NJ, 1986.
8. Vladimir N. Vapnik. *Statistical learning theory.* Wiley, New York, 1998.

On the Statistical Nature of Real Sinusoids Associated with Rotating Machinery

Peter Sherman
Iowa State University, Ames, IA 50011
Email: shermanp@iastate.edu

Abstract

This paper represents the first phase of an ongoing investigation into the nature of sinusoidal types of random processes associated with real world phenomena. While perfect sinusoids exist only in the mathematical sense, their use as an approximation model in relation to real world phenomena has, and continues to be widespread; often with much value. The motivation for this work is the belief that knowledge of their deviation from such a model can provide additional useful information. The focus of this paper is on sinusoids in relation to random processes associated with rotating machinery. The tools used include mathematical limit theorem results, standard signal processing tools including spectral estimation and Kalman filtering, and basic statistics. Some noteworthy results include the normality of amplitude and frequency, characterization of the same as stationary random processes, and potential to improve condition monitoring of rotating machinery.

1. Introduction

The concept of a sinusoid arises in the study of phenomena in practically every area of science and engineering. Examples include vibration of rotating machinery [SHW], species extinction rates [RAS], earthquake prediction [SAF], atmospheric wind profiles [WIS], sonar [JOH], heart rate variability [MMM], and music quality, to name just a very few. However, a true sinusoid only exists in the mathematical sense. Even a sine wave generator does not generate a perfect sinusoid. A perfect sinusoid is characterized by three constants, namely, amplitude, frequency and phase. Because the phase parameter reflects only the relation of the sinusoid to the time at which the sinusoid is first observed, the value of this parameter is dictated by the observer. Both amplitude and frequency, however, are subject to change over time. And so it is these parameters which will be the focus of this paper. It is worth pointing out that the frequency of a sinusoid is simply the inverse of its period, since in many applications it is period, and not frequency variability that is of interest.

2. Amplitude and frequency (or period) variability can occur slowly over time (relative to the period of the sinusoid), as well as within a single period. Here, the period refers to the time period associated with 360 degrees of angle. In many rotating machinery applications the speed of the machine will vary to some degree slowly over time. If the basic natures of the associated signals, such as vibration, sound or pressure, are not influenced by the slow-time variations, then it may be possible to recover truly periodic signals by performing a time-to-angle transformation [LS1]. Even then, however, there may well remain intracycle variability of sinusoidal data.

2. Tools Used for Analysis

The tools used for analysis of the data include (I) the minimum variance (MV) and associated autoregressive (AR) spectral families, along with their theoretical convergence properties [FFS], (ii) standard power spectral density (PSD) estimates, (iii) extended Kalman filtering , (iv) fixed order AR models and (v) histogram, scatter plot, and correlation coefficient information. The families of MV and AR spectra are used to identify nominally sinusoidal components, and to provide input to the extended Kalman filter for tracking the time-varying amplitude and frequency of a real sinusoid. The PSD is used primarily for comparative purposes because it is the tool of choice for analysis of frequency information contained in (assumedly stationary) random processes. Fixed order AR models are used in an attempt to better characterize the time-varying behavior of amplitude and frequency estimates. Finally, basic statistical tools, including histogram and moment estimates, are used to get a better idea of the distributional properties of frequency and amplitude information.

2. The Machinery Data to be Analyzed

The data chosen for analysis is from the Westland Helicopter data base [WES]. This data base includes vibration data taken from a military helicopter under well controlled test conditions, and for a variety of planted fault conditions. Our analysis will focus on the vibration associated with accelerometer number 6 for the no-fault condition, and for a pinion bearing fault near to the measurement location. The data were originally sampled at 103116.08 Hz. Because of the preponderance of energy below 20,000 Hz, this data was decimated by a factor of 5.The no-fault data was chosen because of the presence of a very strong sinusoidal component at 3150 Hz. This affords us the opportunity to study a real world sinusoid without complications associated with a significant amount of noise corruption. The pinion bearing data was chosen for two reasons. Since the strength of this strong sinusoid was notably attenuated, this afforded the opportunity to study a potentially more complex real world sinusoid. It also provided the opportunity to explore the potential for using only sparse information associated with a single sinusoid, as opposed to the totality of information contained in a PSD, to characterize the influence of a mechanical fault. It should be noted that the above frequency of interest corresponds to a number of component gear mesh frequencies, but *not* to the spiral bevel pinion gear mesh frequency, which is 1109 Hz. Moreover, the pinion bearing theoretical defect frequency is at 311 Hz.

3. Analysis

The top plots in Figure 1 show the raw and band pass filtered time series corresponding to the no-fault and fault conditions. It is interesting to note that the influence of the fault is to enhance the modulation, while decreasing the peak level of the sinusoid. Neither of these influences is to be expected, given the nature of the fault. In fact, the expected influence would be an increase of energy at the bearing defect frequency; which in this case is 311 Hz. But the PSD plots in Figure 1 reveal a decrease of energy at that frequency, and an increase at a slightly lower frequency. Other than this very low frequency region, the only significant difference due to the fault is the peak values at approximately 3150 Hz. The plots of the filtered data in this region show that there

appear to be two sinusoidal components in this region, regardless of the fault. The fault results in a reduction of the stronger peak on the order of 15 dB.

Recall, however, that our goal is not to mathematically decompose a real sinusoid into true sinusoids, but rather to capture its time-varying amplitude and frequency characteristics. To this end, we proceed now to investigate the convergence properties of the MV(n) and AR(n) spectra as n goes to infinity. In [FFS] it was shown that the MV(n) spectra converge monotonically to the line spectrum associated with sinusoids. It was also shown that the corresponding AR(n) spectra converge to infinity at the same frequencies, while converging to the continuous PSD at all other frequencies. While not shown here due to space limitations, the MV(n) and AR(n) estimated spectra for the no-fault and faulted raw data clearly suggest the presence of a true sinusoid at approximately 3150 Hz. The same cannot be said of the spectra for the fault data. Even though the MV(n) spectra do not exhibit the asymptotic 3 dB drop between orders, as predicted in [SHL], they also do not suggest convergence. The Corresponding AR(n) spectra exhibit the same multiple peak structure as that of the PSD in Figure 1. So it is possible that it is this increased modulation effect, relative to the no-fault data, that is responsible for the non-convergence behavior of these spectra.

Since the spectra use lagged-product correlation estimates, it is also possible that the lack of exhibited convergence could be due, in part, to statistical variability associated with these estimates. While there has been some progress in obtaining the statistics of the AR(n) [LS2] and MV(n) [LIS] correlation-based estimates for mixed spectrum random processes, our analysis here will not consider this statistical influence due to the risks of distracting the reader from more fundamental issues, and of a very lengthy and involved analysis. Recall, that the goal of this paper is to explore both the nature of real sinusoids and the tools we have chosen to use for that purpose. We believe that the MV(n) and AR(n) spectral families have significant potential for that purpose. But how they are used is equally important. For example, it is commonly held that such spectra, for high enough orders, can capture all the important spectral information in a given bandwidth without the need for filtering. However, when attempting to take advantage of the convergence, as opposed to modeling capabilities of the MV(n) and corresponding AR(n) spectra, there is a definite advantage in filtering. In particular, for the no-fault data the MV(n) spectra were noted to have completely converged, as the n=40, 80 and 160 spectra are identical. This strongly suggests the existence of a true sinusoid. However, the corresponding AR(n) spectra did not exhibit the corresponding +3 dB asymptotic increase per order doubling. Since the sinusoid at approximately 3150 Hz is not a true sinusoid, this contradictory behavior is not totally unexpected; especially in view of the PSD information in Figure 1. There, it is observed that while there appear to be two closely spaced sinusoids for both the no-fault and fault data, the no-fault data has one very dominant peak. The fact that the MV spectrum has lower resolving ability than the corresponding AR spectra would explain this contradictory behavior. Specifically, at lower orders both spectra would exhibit behavior consistent with a single sinusoid, while at higher orders the AR spectra would actually reduce in magnitude with the emerging presence of two peaks. This behavior of these spectra for the fault data is more obvious. Since both the PSD in Figure 1 and the AR spectra suggest two sinusoids of relatively equal power, it would appear to support our speculation regarding the reason for the

contradictory behavior of the MV and AR spectra. It is our belief that the convergence properties of the families of MV and AR spectra offer significant potential for characterization of random processes involving real sinusoids. However, from these results, as well as a similar contradictory behavior noted in their use in the analysis of diesel vibration [SHW], it is apparent that more research is needed to better understand their joint behavior for real sinusoids.

In order to attempt to capture the time-varying amplitude and frequency behavior of the real sinusoid at approximately 3150 Hz, as opposed to modeling it as two true sinusoids, we used an extended Kalman filter (EKF), similar to that used in [SHW]. Specifically, the band pass filtered data was assumed to consist of a single sinusoid plus white noise. The sinusoid amplitude and frequency were modeled as uncorrelated random walks. The choice of this model is based, to a large extent, on our ignorance of their true behavior. As will be seen, however, it provides a sufficient characterization to allow more realistic models to be studied. A variety of model covariance values, as well as band pass filters were investigated. The results were extremely robust with respect to all of these. Furthermore, for both the no-fault and fault data the EKF model captured 99.999% of the total energy in the data.

The time-varying amplitudes and frequencies for the no-fault and fault data are illustrated in Figure 2. There are clear differences in these time series for the no-fault versus the fault data. For example, the heavier frequency modulation behavior associated with the fault data results in temporal regions wherein the filtered data is close to zero. These regions cause the EKF to lose tracking ability, and yield frequency estimates which are well outside of the actual range of activity; hence the very large excursions in estimated frequency. These excursions are periodic, with a period corresponding to the modulation period of approximately 0.05 sec. The frequency and amplitude histograms for the no-fault and fault data are shown in Figure 3. The frequency histogram for the fault data was truncated in order to alleviate the frequency estimates related to poor EKF tracking. Both frequency and amplitude information is dramatically influenced by the presence of the fault. Recall that the frequency range of interest here has no known relation to the characteristic frequency region associated with such a fault. Nonetheless, such strong differences suggest that there may well be other frequency regions of equal, if not greater ability to capture the presence of a fault.

Scatterplots of frequency versus amplitude for the no-fault and fault data are shown in Figure 4. Again, there is a distinct difference between the no-fault and fault conditions. The no-fault condition reveals a mild negative correlation (-0.41) between amplitude and frequency. For the fault condition the correlation is almost zero (0.07). The very narrow range of frequencies for the real sinusoid shown in Figure 4 results in the very peaked nature of the fault histogram. Even though the overall correlation between frequency and amplitude information is only modest for the no-fault data, and is essentially zero for the fault data, the coherence analysis suggested that there are indeed correlations in specific frequency ranges.

To further explore the time series structure of the frequency and amplitude estimates provided by the EKF then MV(n) and AR(n) tools used to analyze the measurement data were applied. One notable result indicated in the MV spectra is the presence of a strong sinusoidal component (0 dB) in the no-fault frequency time series at twice the frequency being studied. This component, shown in figure 5, is essentially absent (- 90 dB) for the fault condition. In this same time series the AR spectra indicate a very clear difference in the continuous spectral structure between the no-fault and fault conditions. In particular, the no-fault frequency time series exhibits an AR(2) shape, with a spectral peak around 2000 Hz. For the fault condition it becomes an AR(1) shape, and increases by 20 dB at lower frequencies. A closer look at the low frequency behavior of these spectra of required that the frequency and amplitude time series be decimated in order to take advantage of the convergence properties of the MV and AR spectra (c.f. [SHL] for a detailed discussion). These data were decimated by a factor of 10. Application of the MV and AR tools to this data revealed that for the no-fault condition the 20 Hz modulation behavior discussed above is revealed as strong sinusoidal components at this same frequency in both the frequency and amplitude MV (and AR) spectra. There is no evidence whatsoever of such a periodicity in the frequency time series for the fault data.; even though the amplitude time series for both the no-fault and fault conditions retains a 20 Hz periodicity.

5. Summary and Conclusions

The purpose of this effort was to investigate the potential of a combination of signal processing and basic statistical tools for characterization of real sinusoids. This was done in the context of sinusoids associated with no-fault and fault vibration data from a military helicopter power system. The fault addressed was a pinion bearing fault having a characteristic fault frequency of 311 Hz. Rather than investigating this frequency region it was decided to investigate the region around 3150 Hz. In that region it was observed that not only were strong sinusoids present, but that the bearing fault had a significant influence of the data structure. For both the no-fault and fault conditions it was noted that two sinusoidal components spaced 20 Hz apart were present, but that the fault resulted in a significant attenuation of one of the two. The 3150 Hz frequency happens to be a gear mesh frequency. Because it is also the 10^{th} harmonic of the fault frequency it is possible that it is the 10^{th} harmonic which is responsible for the change in the spectral structure in this region. However, since it is a relatively high harmonic, and since the change was so significant, one might posit that there are other influences of the fault on the power system vibration characteristics. While there is no direct support here for such speculation, the investigation resulted in a number of very significant findings. First, it was found that a model for a single real sinusoid, having time-varying amplitude and frequency, was able to completely characterize the band pass data including the two theoretical sinusoidal components. Analysis of the amplitude and frequency time series provided some novel insight into the real sinusoid that goes well beyond a simplistic two-sinusoid model suggested by traditional spectral analysis. For example, it was noted that even though the amplitude and frequency of the real sinusoid associated with the fault data was close to zero, there was a strong correlation in specific frequency regions. It was also found that the frequency and amplitude data both had a strong periodic component at

approximately two times the frequency of interest, regardless of machine condition. While the same was true for the amplitude time series in the range corresponding to the fault frequency, for the frequency time series the presence of the fault all but eliminated any periodic behavior. Finally, both histogram and scatter plot information revealed such distinct differences between the no-fault and fault conditions that discernment of these two conditions based on this information would be trivial. It is not the intent to suggest that information related to a single real sinusoid should be used instead of classic fault frequency information. However, our analysis suggests that a fault may influence data in a far greater way than has been considered to date.

Acknowledgements This work was supported, in part, by the U.S. Air Force Office of Scientific Research, grant # F49620-98-1-0252.

References

[FFS] Foias, C., Frazho, A.E. and Sherman, P.J. "A geometric approach to the maximum likelihood spectral estimator for sinusoids in noise," IEEE Transactions on Information Theory, IT-34(5), 1066-1070, September 1988.

[JOH] Johnson, D. "The application of spectral estimation methods to bearing estimation problems", *Proceedings of the IEEE,* **70**(9), 1018-1028, September 1982.

[LS1] Lau, S. S. and Sherman, P. J. "The influence of period variation on time/frequency analysis of the Westland helicopter data", ", *Proc. 9th IEEE SP Workshop on Statistical Signal & Array Processing*, 180-183, Portland, OR, September 14-18, 1998.

[LS2] Lau, S. S. and Sherman, P. J. "Asymptotic statistical properties of the AR model for mixed spectrum estimation", *Proc. IEEE ICASSP '98,* 2289-2292, Seattle, WA, May 12-15, 1998.

[LIS] Liu, Xiao Hu and Sherman, P. J. "Asymptotic statistical properties of the Capon MV spectral estimator for mixed spectrum processes", *Proc. 9th IEEE SP Workshop on Statistical Signal & Array Processing*, 328-331, Portland, OR, September 14-18, 1998.

[MMM] Myers, G. A., *et al.* "Power spectral analysis of heart rate variability in sudden cardia death: comparison of other methods", *IEEE Transactions on Biomedical Engineering*, **BME-33**(12),1149-1156, December 1986.

[RAS] Raup, D. M. and Sepkoski, J. J. "periodicity of extinctions in the geologic past", *Proceedings of the National Academy of Science* USA, **81**, 801-805, 1984.

[SAF] Safek, E. "Analytical approach to calculation of response spectra from seismological models of ground motion", *Earthquake Engineering & Structural Dynamics*, **16,** 121-134, 1988.

[SHL] Sherman, P.J. and Lou, K.N. "On the family of ML spectral estimates for mixed spectrum identification," IEEE Transactions on Acoustics, Speech and Signal Processing, ASSP-39(3), 644-655, March 1991.

[SHW] Sherman, P.J. and White, L.B. "Periodic spectral analysis of diesel vibration data", *J. Acoustic Soc. Of America*, **98**(6), 3285-3301, Dec. 1995.

[WES] See the website: http://wisdom.arl.psu.edu/Westland/data/

[WIS] Wikle, C. and Sherman, P. J. "Using the family of MV spectra to determine periodicities in atmospheric data," Journal of Climate, 8(10), 2352-2363, October 1995.

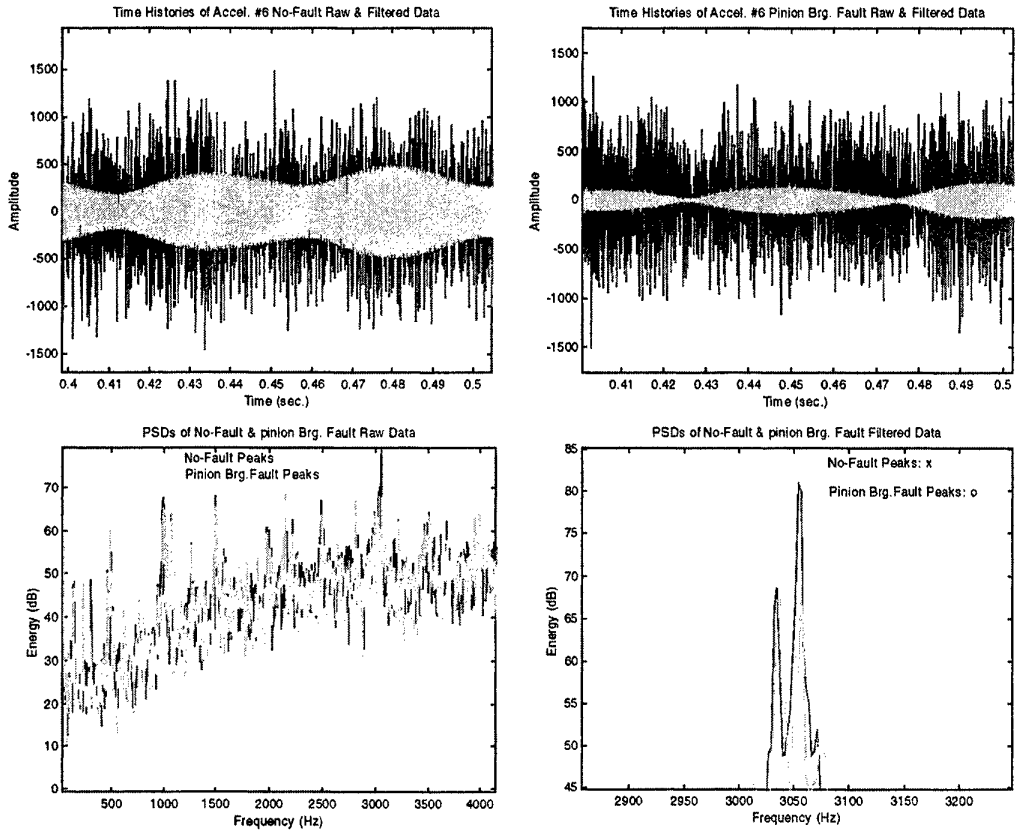


Figure 1. Plots of raw and filtered no-fault (left) and fault (right) data (Top); PSD estimates of the above data (Bottom).

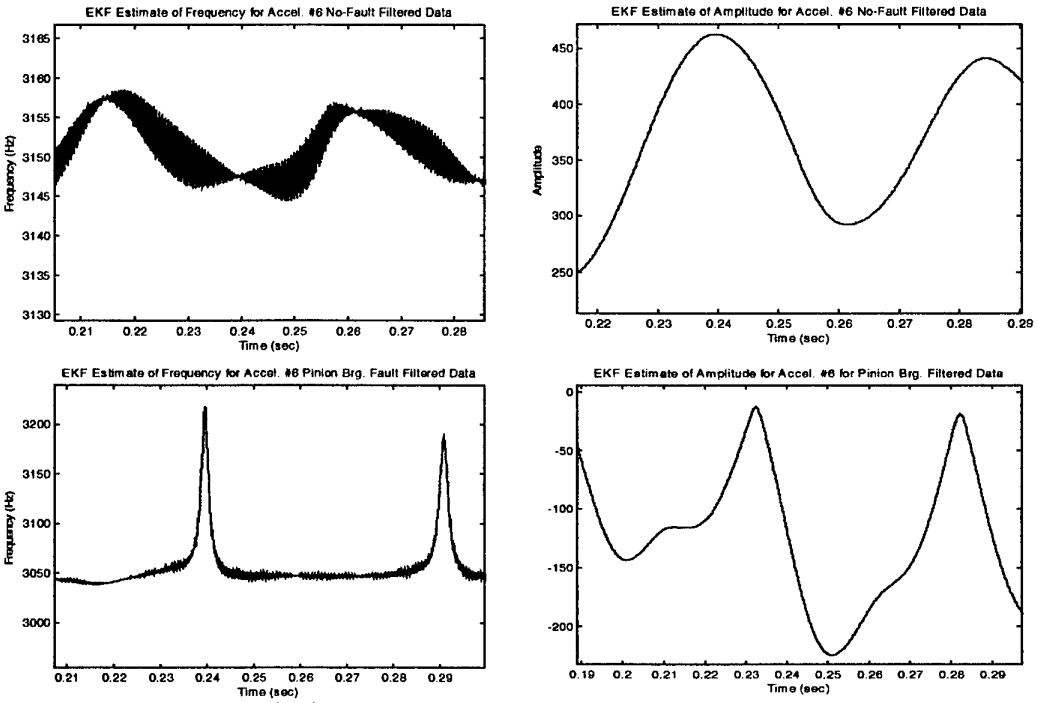


Figure 2. EKF frequency (left) and amplitude (right) estimates for filtered no-fault (top) and pinion bearing fault (bottom) filtered data.

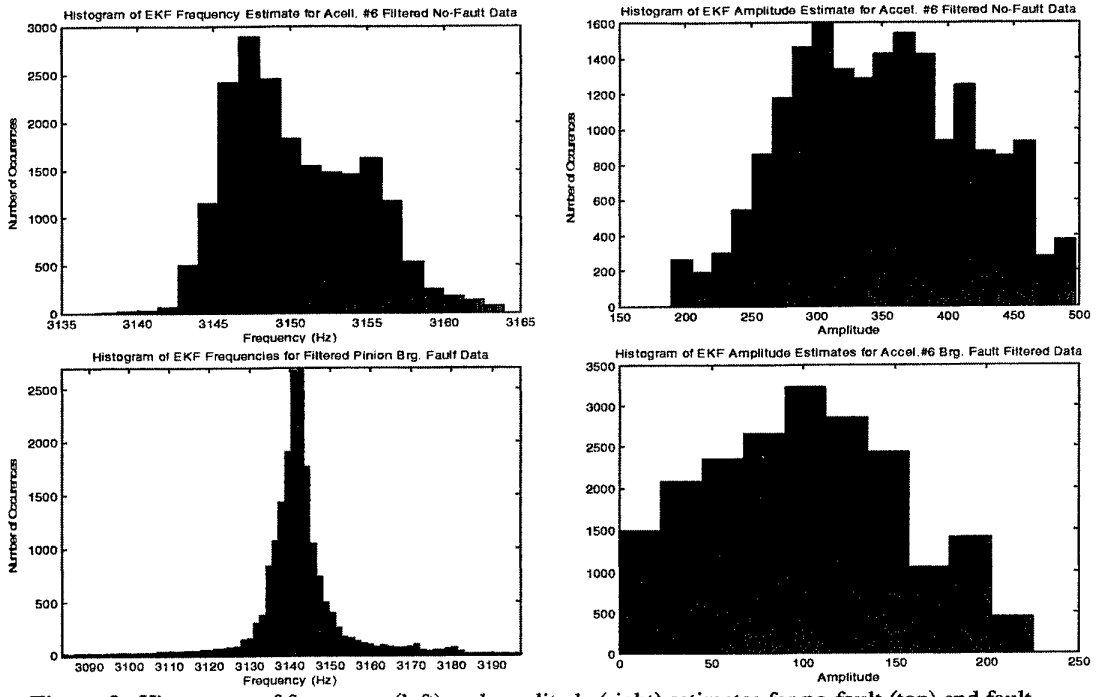


Figure 3. Histograms of frequency (left) and amplitude (right) estimates for no-fault (top) and fault (bottom) data

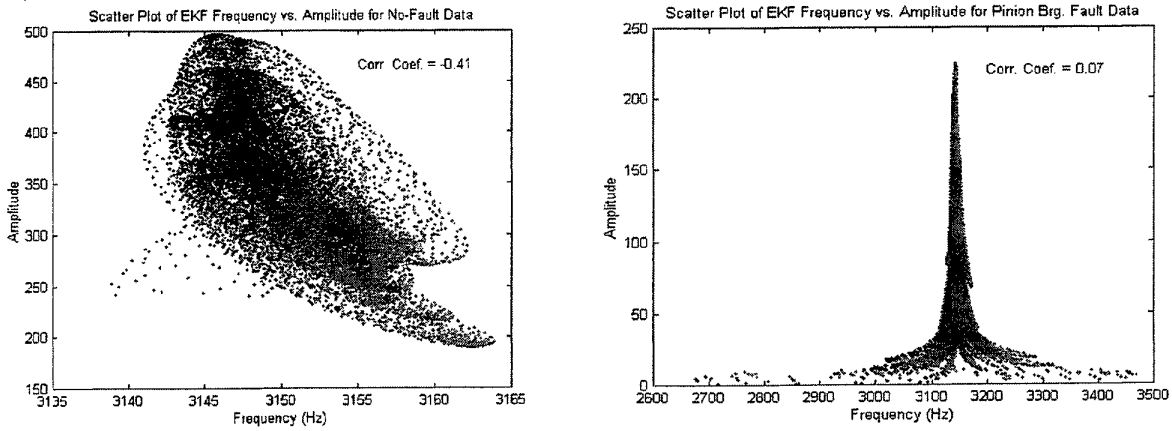


Figure 4. Amplitude versus frequency scatterplots for the filtered no-fault (left) and fault (right) data.

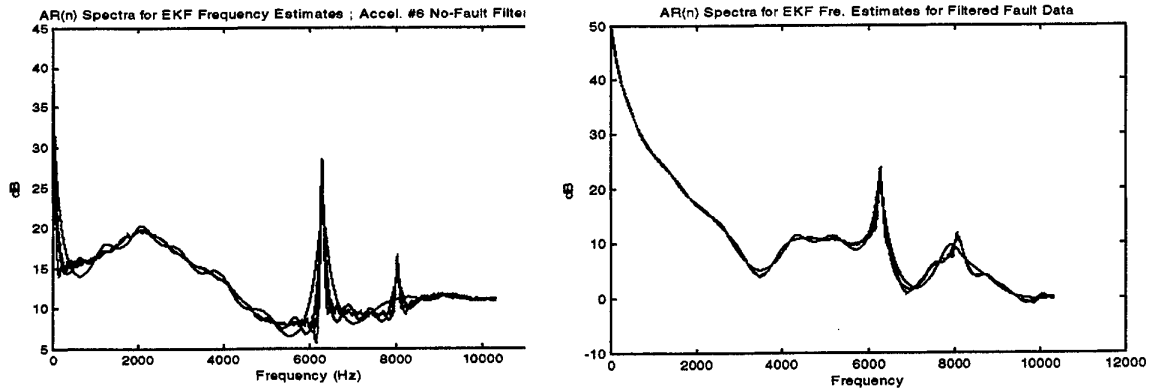


Figure 5. AR(n) spectra for estimated time-varying frequency fo no-fault (left) & fault (right) data.

Approaches to Measuring Network Change

P.J. Shoubridge[†], M. Kraetzl[†], H. Bunke[*] and W.D. Wallis[◇]

[†] *Defence Science Technology Organisation, Australia,* [*] *University of Bern, Switzerland,*
[◇] *Southern Illinois University, United States*

E-mail: peter.shoubridge@dsto.defence.gov.au

It is difficult to track change in telecommunication networks when the logical structure is dynamic. Measures of network difference are required to indicate significant changes of concern. This paper discusses possible approaches towards developing suitable measures of change.

Key Words: network management, change detection, graph matching

1. INTRODUCTION

There are various problems requiring the ability to measure network change. In the management and control of dynamic telecommunication systems, the early detection of significant network events and abnormal trends is an important network performance monitoring capability providing advance warning of possible fault conditions [19], or at least assisting with the identification of causes and locations of known problems. Such capabilities are increasingly necessary as telecommunication networks become more complex and dynamic, for example large heterogeneous enterprise networks [3].

Network performance monitoring is typically undertaken using statistical techniques to analyse variations in traffic distribution [9, 11] or changes in topology [20]. Network visualisation techniques are also used to monitor changes in telecommunication networks [1]. A useful complement to these approaches is a measure of change in the network, capturing both topology and traffic flow, to highlight when and where in the network significant events may be occurring [18]. Other network management tools can then be focussed on likely problem regions of the network for further analysis. This paper provides an overview of approaches being investigated to identify a suitable measure of network change for such applications.

A measure of network change can be determined by representing a given network, observed at time t, by a directed graph (digraph). Edge direction can be used to indicate the direction of traffic flow between two adjacent nodes in the network with an edge label representing traffic flow. A second graph can be used to represent the same network at a later time $t + \Delta t$, where Δt is some arbitrary time interval. This second graph can be compared with the original graph using a measure of network

difference between the two graphs to indicate the degree of change occurring in the network over the time interval Δt. By continuing network observations over subsequent time intervals, the graph difference measures provide a trend of the network's dynamic behaviour as it evolves over time.

The problem then becomes one of finding good graph distance measures that are sensitive to significant change events but insensitive to typical variations in network topology or flow. Following from this, is the requirement to detect significant events given a suitable distance measure. The detection problem is not being considered in this paper. In addition to a graph distance measure, it is also necessary to readily identify where in the network significant change events have occurred. This requires the association of location with measured change.

2. GRAPH MATCHING

The communications network being considered is represented as a graph. A graph consists of a set of vertices representing network nodes, and a set of edges which are ordered pairs of vertices; a pair of vertices denote the endpoints of an edge. Two vertices are adjacent if they are the endpoints of some edge.

DEFINITION 2.1. A *directed and labelled graph* G is a 6-tuple $G = (V, E, L_V, L_E, \mu, \nu)$ where:

- V is a finite set of primitive objects called *vertices*;
- L_V is a finite set of labels (for the vertices);
- $\mu : V \Rightarrow L_V$ is a function assigning labels to vertices;
- $E \subseteq V \times V$ is the set of *edges*;
- L_E is a finite set of labels (for the edges);
- $\nu : E \Rightarrow L_E$ is a function assigning labels to edges.

Edges are directed with the vertex pair $(i, j) \in E$ denoting traffic flow from node i to node j in the network. Vertex labels μ are used to uniquely identify network nodes while edge labels ν are assigned traffic flow parameters measured over some time interval Δt. The number of vertices in G is denoted $|V|$ and likewise the number of edges $|E|$.

Two graphs can be considered the same if a graph isomorphism exists between them [2]. Graph isomorphisms can be detected by mapping the vertices of one graph G_1 onto the vertices of a second graph G_2. A valid vertex mapping is found if the edge structure of G_1 is preserved in G_2 by the mapping. If all the vertices of G_1 can successfully be mapped to all the vertices of G_2, a graph isomorphism has been found [7].

A technique known as error correcting graph matching (ecgm) can be used to measure the distance or dissimilarity between two graphs. Unlike strict graph isomorphism detection, this approach enables inexact matching. Error correcting graph matching evaluates the minimum number of edit operations required to modify an input graph such that it becomes a graph isomorphism of some reference graph. This can include the possible insertion and deletion of edges and vertices, as well as possible label substitutions [15]. Generally, error correcting graph matching algorithms assign costs to each of the edit operations and use efficient tree search

techniques to identify the sequence of edit operations resulting in the lowest total edit cost [5, 13]. The resultant lowest total edit cost is a measure of the distance between the two graphs.

Consider first changes in network structure or topology only. In general graph matching problems there exist more than one possible sequence of edit operations due to the occurrence of multiple possible vertex mappings. The ecgm algorithms search for the edit sequence that results in the minimum edit cost. However, with performance monitoring of a communications network, vertex label substitution is not a possible edit operation because vertex labels reference unique physical or logical nodes within a network. As a result, the combinatorial search reduces to the simple identification of elements (vertices and edges) inserted or deleted from one graph $\overline{G}_1$ to produce the other graph $\overline{G}_2$.

If the cost associated with the insertion and deletion of individual elements is 1, the edit sequence cost becomes the difference between the total number of elements in both graphs, and all graph elements in common:

DEFINITION 2.2. Let the graph $\overline{G}_1 = (V_1, E_1, \mu_1, \nu_1)$ represent the communication network operating at time t_1, and let $\overline{G}_2 = (V_2, E_2, \mu_2, \nu_2)$ describe the same network at time t_2, where $t_2 = t_1 + \Delta t$. The *network edit distance* $d(\overline{G}_1, \overline{G}_2)$ is:

$$d(\overline{G}_1, \overline{G}_2) = |V_1| + |V_2| - 2|V_1 \cap V_2| + |E_1| + |E_2| - 2|E_1 \cap E_2| \qquad (1)$$

Clearly the edit distance, as a measure of topology change, increases with increasing degree of change experienced by the network over Δt. Edit distance $d(\overline{G}_1, \overline{G}_2)$ is bounded below by $d(\overline{G}_1, \overline{G}_2) = 0$ when $\overline{G}_2$ and $\overline{G}_1$ are isomorphic (i.e. there is no change), and above by $d(\overline{G}_1, \overline{G}_2) = |V_1| + |V_2| + |E_1| + |E_2|$ when $\overline{G}_1 \cap \overline{G}_2 = \emptyset$, the case where the networks are completely different.

The expression for graph edit distance provides a measure of difference between two graphs in terms of topology. While traffic flow represented as edge labels ν can also be incorporated into an ecgm algorithm, it is difficult to design a suitable cost function that satisfies both topology and traffic variations. For example, what is the relationship between traffic fluctuations of an order of magnitude compared to the insertion of a new vertex? Alternatively, could an integer representation of traffic flow be successfully mapped to a topology problem where the number of multi-edges indicate the volume of traffic flow? Such issues are currently being investigated.

3. GRAPH STRUCTURES

Distance measures between graphs have also been proposed using graph matching techniques that focus on underlying structures or properties within the graphs. For example, finding the maximum common subgraph of two graphs is an indication of the commonality between graphs [10, 12]. These techniques tend to be based on subgraph isomorphism algorithms [14, 21].

An induced subgraph, where common vertices must have all incident edges in both graphs, can be defined as follows:

DEFINITION 3.1. The *induced subgraph* H of a digraph $G = (V, E, \mu, \nu)$ generated by the subset W of V is the digraph $H = (W, E_W, \mu_W, \nu_W)$ where:

- $\mu_W : W \Rightarrow L_V$ is the restriction of μ to W;
- $E_W = E \cap W \times W$;
- $\nu_W : E_W \Rightarrow L_E$ is the restriction of ν to E_W.

A distance metric has been defined based on the determination of the maximal common subgraph of two graphs [6]:

$$d(G_1, G_2) = 1 - \frac{|\text{mcis}(G_1, G_2)|}{\max\{|G_1|, |G_2|\}} \tag{2}$$

where $mcis(G_1, G_2)$ denotes the maximal common induced subgraph of G_1 and G_2, and $|G|$ denotes the number of vertices in the graph G. There is no reason why the number of edges could not be used as $|G|$. The resulting distance measure is still a metric and may prove useful.

It is interesting to note that this metric is related to error correcting graph matching [4].

The most general form of the distance metric (2) is given by:

$$d(G_1, G_2) = 1 - \frac{m(G_1, G_2)}{M(G_1, G_2)} \tag{3}$$

where $m(G_1, G_2)$ is a measure of similarity between G_1 and G_2, and $M(G_1, G_2)$ is a measure of the size of the problem.

The size of the problem may also be defined as the number of vertices in the union of the two graphs. This resulting distance measure:

$$d(G_1, G_2) = 1 - \frac{|\text{mcis}(G_1, G_2)|}{|G_1 \cup G_2|} \tag{4}$$

can also be shown to be a metric.

In the telecommunications application, it seems natural to view the size of the problem as the union of the two graphs. Using the union rather than the larger of the two graphs distinguishes variations in the size of the smaller of the two graphs. If only the size of the larger graph is used to represent problem size, the distance between graphs will remain unchanged even if the smaller graph changes its size, assuming that the size of the maximum common subgraph remains constant. This latter metric may provide a more accurate measure of the relative graph difference.

There are other approaches that may prove useful in the application of network change measurement. Instead of measuring change by concentrating on network elements, as in the cases for ecgm and mcis, better indicators of significant change could perhaps be obtained by examining higher level structures within the network. One such structure being investigated is that based on vertex neighbourhoods.

An input graph G_1 is used to generate a neighbourhood graph N_1 that contains an undirected edge between two vertices if the same vertices in G_1 share one or

more common adjacent vertices, i.e. are connected to common neighbours. Similarly a graph N_2 can be generated from $\bar{G}_2$. The distance between $\bar{G}_1$ and $\bar{G}_2$ is now derived from a graph matching distance measure taken between the two neighbourhood graphs N_1 and N_2. If for example edges are deleted in N_1 or N_2, they have far greater significance because two underlying vertices in $\bar{G}_1$ or $\bar{G}_2$ are now no longer connected via common neighbours. It is unclear at this time what advantages such an approach might have, but it is expected to improve indications of significant change in the underlying network.

In addition to these techniques, the analysis of graph adjacency matrices has been shown to yield useful distance metrics between graphs. The adjacency matrix of a graph $\bar{G}$ is defined as:

DEFINITION 3.2. The *adjacency matrix* $A = [a_{\bar{i}\bar{j}}]$ of graph $\bar{G}$ is a $|V| \times |V|$ matrix where

$$a_{\bar{i}\bar{j}} = \begin{cases} \nu & \text{if edge } (i,j) \in E \\ 0 & \text{otherwise} \end{cases}$$

If the two graphs to be compared are uniquely labelled, it is possible to measure the difference between the graphs using a Hamming distance measure [16]. This metric defines the distance between the two graphs as the number of elements in which their respective adjacency matrices differ. This approach is somewhat similar to the basic ecgm distance measure shown in equation (1). Analysis of graph adjacency matrices using spectral graph theory [8] have also been shown to provide useful graph distance measures [17].

4. LOCATION OF CHANGE

Consider the scenario where a graph distance measure has indicated that two digraphs $\bar{G}_1$ and $\bar{G}_2$ have significantly different network topologies over a single time interval Δt. Of particular interest is the distribution of this change during the transition from $\bar{G}_1$ to $\bar{G}_2$. The following technique ranks all vertices within the two digraphs $(\bar{G}_1 \cup \bar{G}_2)$ in increasing order of the number of network topology change events experienced by individual vertices.

Differences between digraphs $\bar{G}_1$ and $\bar{G}_2$ are represented by a change matrix $\bar{C}$ that indicates where edges have been deleted from $\bar{G}_1$ or inserted into $\bar{G}_2$. The matrix $\bar{C}$ has a row and column for every vertex contained within the two graphs. The existence of a directed edge deleted from $\bar{G}_1$ (or inserted into $\bar{G}_2$) is represented in the matrix $\bar{C} = [c_{\bar{i}\bar{j}}]$ by the corresponding row column entry $c_{\bar{i}\bar{j}} = 1$. Indices i and j denote the respective source and destination vertices of the deleted or inserted directed edge. Any edges (i,j) that remain incident to the same pair of vertices in both $\bar{G}_1$ and $\bar{G}_2$ result in the corresponding entry $c_{\bar{i}\bar{j}} = 0$, indicating that no change has occurred. For all other entries $c_{\bar{i}\bar{j}} = 0$.

If a permutation of the matrix $\bar{C}$ is found such that the row sums and column sums of $\bar{C}$ are arranged in ascending order, entries in $\bar{C}$ where $c_{\bar{i}\bar{j}} = 1$ will tend to occupy a lower right block of $\bar{C}$. The corresponding row and column vertices will be ranked in ascending order of the number of incident edge deletion and/or insertion operations, with the last row and column vertex pair denoting the vertices that experienced the greatest change in the transition from graph $\bar{G}_1$ to $\bar{G}_2$.

5. CONCLUSION

Several techniques have been presented that could successfully measure the degree of change occurring within a communications network. Preliminary simulation investigations into the use of ecgm and mcis distance metrics indicate that effective change measurement is feasible. It is expected that applying ecgm or mcis distance metrics to graphs that represent higher level structures within the underlying network (e.g. common neighbours) should in general produce improved results. However, this is still to be confirmed.

REFERENCES

1. Becker, R.A., Eick, S.G. and Wilks, A.R. (1995) "Visualizing network data," IEEE Trans. Visualization & Computer Graphics 1(1), 16–28.
2. Bondy, J.A. and Murty, U.S.R. (1976). *Graph Theory with Applications* United Kingdom: The MacMillan Press.
3. Boutaba, R., Guemhioui, K.E. and Dini, P. (1997) "An outlook on intranet management," IEEE Communications Magazine, October, 92–99.
4. Bunke, H. (1997). "On a relation between graph edit distance and maximal common subgraph," Pattern Recognition Letters 18, 689–694.
5. Bunke, H., Messmer, B.T. (1997). "Recent advances in graph matching," International J. Pattern Recognition and Artificial Intelligence 11(1), 169–203.
6. Bunke, H., Shearer, K. (1998). "A graph distance metric based on the maximal common subgraph," Pattern Recognition Letters 19, 255–259.
7. Corneil, D.G. and Gotlieb, C.C. (1970). "An efficient algorithm for graph isomorphism," Journal of the ACM 17, 51–64.
8. Chung, F.R.K. (1997). *Spectral Graph Theory* Rhode Island: American Mathematical Society.
9. Higginbottom, G.N. (1998). *Performance Evaluation of Communication Networks* Massachusetts: Artech House.
10. Horaud, R., Skordas, T. (1989). "Stereo correspondence through feature grouping and maximal cliques," IEEE Trans. Pattern Anal. Machine Intell. 11(11), 1168–1180.
11. Jerkins, J.L. and Wang, J.L. (1998). "A close look at traffic measurements from packet networks," IEEE GLOBECOM, Sydney, Australia.
12. Levinson, R. (1992). "Pattern associativity and the retrieval of semantic networks," Comput. Math. Appl. 23, 573–600.
13. Messmer, B.T. and Bunke, H. (1998). "A new algorithm for error-tolerant subgraph isomorphism detection," IEEE Trans. Pattern Anal. & Machine Intell. 20, 493–504.
14. Myaeng, S.H., Lopez-Lopez, A. (1992). "Conceptual graph matching: a flexible algorithm and experiments," Journal of Experimental and Theoretical Artificial Intelligence 4, 107–126.
15. Sanfeliu, A. and Fu, K.S. (1983). "A distance measure between attributed relational graphs for pattern recognition," IEEE Trans. Systems, Man and Cyber. 13, 353–362.
16. Sanil, A., Banks, D., Carley, K., (1995). "Models for evolving fixed node networks: model fitting and model testing," Social Networks 17, 65–81.
17. Sarkar, S., Boyer, K.L., (1998). "Quantitative measures of change based on feature organization: eigenvalues and eigenvectors," Computer Vision and Image Understanding 71(1), 110–136.
18. Shoubridge, P.J., Kraetzl, M. and Ray, D. (1999). "Detection of abnormal change in dynamic networks," IEEE Information, Decision and Control, IDC'99, conference, Adelaide, Australia, 557–562.
19. Thottan, M. and Ji, C. (1998). "Proactive anomaly detection using distributed intelligent agents," IEEE Network, September/October, 21–27.
20. White, C.C., Sykes E.A. and Morrow J.A. (1995) "An analytical approach to the dynamic topology problem," Telecommunication Systems 3, 397–413.
21. Ullman, J.R. (1976). "An algorithm for subgraph isomorphism," Journal of the ACM 23(1), 31–42.

A Bayesian Risk Approach to Multi-Mode Detection

Dana Sinno, Douglas Cochran, and Darryl R. Morrell
Arizona State University, Tempe, AZ 85287-7206, U.S.A.

This paper considers a situation in which a detection system is configurable in such a way as to provide multiple modes of operation that differ in their detection performance and geographical coverage. A technique for optimal mode selection based upon minimizing Bayesian risk is formulated and demonstrated for the case of a two-mode system.

0. PROLOGUE

In the first Joint Australia/U.S. Workshop on Defence Signal Processing, held in Adelaide in 1997, two of the authors (Sinno and Cochran) presented a paper involving estimation using a configurable sensor system [6]. During the Workshop, Dr. Paul Miller of the Australian Defence Science and Technology Organisation told us about a real-world scenario in which searches for ground vehicles are carried out over vast uninhabited areas by helicopters outfitted with dual-mode radar systems. We understood the operating modes of the radar system to be such that they could be loosely described as "broad search" and "focused" modes and that the strategy for switching between modes during a search was left to the helicopter crew.

We subsequently proposed a mathematical model of this kind of scenario and a Bayesian approach to choosing mode switching strategies [1]. This formulation made use of a payoff function consisting of two terms, one of which captures the performance of the sensing strategy in detecting the presence of a target in the search area and the other measuring its effectiveness at localizing the target.

When we were invited to participate in the second Australia/U.S. Workshop, revisiting this problem and exploring an alternative approach that addresses a shortcoming of our technique in [1] seemed a natural way to connect our contribution with the previous Workshop and the many fine technical interactions it seeded. The approach in this paper is based upon Bayesian risk analysis and it eliminates concerns regarding correlation between the terms of the payoff function arising in our previous treatment of the problem.

1. INTRODUCTION

This paper considers a situation in which a detection system is configurable in such a way as to provide multiple modes of operation that differ in their detection performance and geographical coverage. The development that follows focuses on the case of a detector with two operating modes: a "broad search" mode that provides wide coverage and a "focused"

mode that provides better detection performance but covers less area. The system is invoked in a sequence of tests to detect and localize a target within a framework that is formulated precisely in the following section.

As noted in the Prologue, this problem was addressed in an earlier paper [1] using a cost functional approach in conjunction with a Bayesian method for incorporating the results of earlier tests in deciding which mode to use in each test. The approach presented here uses a more classical approach, based on minimization of Bayesian risk, that allows more precise designation of the priority of correct detection relative to that of correct localization.

2. PROBLEM SETUP

The situation described above is modeled as follows. The entire region of interest C is partitioned into N disjoint cells $C_1, ..., C_N$. Operating in the broad search mode (*Mode A*), the detector tests for the presence of a signal source in C. In the focused mode (*Mode B*), however, the test may be applied to exactly one cell C_n.

To account for difference in detector performance in the two operating modes, detector performance is modeled as arising from the problem of detecting a known signal in white gaussian noise of known variance. This model provides a well understood solution (i.e., the matched filter) in each test, admits several straightforward generalizations, and allows detection performance in *Mode B* to be distinguished from that in *Mode A* by simply raising the signal-to-noise ratio (SNR). More specifically, in each mode of operation the detector encounters a problem of the form

$$\begin{aligned} H_0 &: \mathbf{X} = \mathbf{N} \\ H_1 &: \mathbf{X} = S + \mathbf{N} \end{aligned} \tag{1}$$

where S is a known signal M-vector with energy $||S||^2 = 1$ and $\mathbf{N}$ is a zero-mean white gaussian M-vector having known variance σ^2; i.e., $\mathbf{N} \sim \mathcal{N}[\mathbf{0}, \sigma^2\mathbf{I}]$ where $\mathbf{I}$ is the $n \times n$ identity matrix. Since $||S||$ is fixed, the SNR (and hence the performance of the detector) in each mode can be adjusted by varying σ^2.

Assuming at most one signal source is present, denote by H_1 and H_0 the events that the signal source is, respectively, present in and absent from C. Let $h_0 = H_0$ and, for $n = 1, ..., N$, denote by h_n the event that the signal source is present in cell C_n. With these definitions, $H_1 = \cup_{n=1}^{N} h_n$. Regardless of whether it is operating in *Mode A* or *Mode B*, the system yields a decision $\rightarrow h_n$ with $0 \leq n \leq N$.

Recall that the optimal solution, in terms of minimal probability of error, to a detector problem of the form (1) is a test on the inner product $S^T\mathbf{X}$ where the detection threshold is a function of the *a priori* probability that a signal is present [2, 5]. The probabilities of detection and false alarm for each test are given by error functions of the detection thresholds. In particular, the tests applied in both operating modes will be of this form, but their detection thresholds and probabilities of detection and false alarm will all be different (even when *Mode B* is applied to different cells) because of their dependence on $\Pr(h_n)$, $n = 0, ..., N$.

3. A BAYESIAN RISK FORMULATION

Using the notation of [5], define a random "state of nature" parameter θ by $\theta = n$ if h_n is true, $n = 0, 1, ..., N$. A prior distribution for θ is assumed and a test (i.e., a *Mode A* test or a *Mode B* test on a particular cell C_n) is chosen and performed yielding a binary

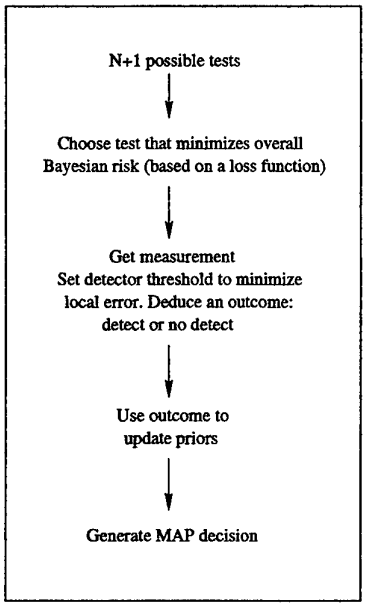


FIG. 1. Detection/localization algorithm.

outcome $\mathbf{o}$. If $\Pr(H_0|\mathbf{o}) > \Pr(H_1|\mathbf{o})$, the system decides for H_0. Otherwise, the system decides in favor of the hypothesis h_n having the largest posterior probability $\Pr(h_n|\mathbf{o})$; i.e., in this case the system decision rule ϕ takes the value n if h_n has the largest posterior probability. As shown in [1], these posterior probabilities are straightforward to compute using the detection and false alarm probabilities of the chosen test, which follow from the (prior) distribution of $\boldsymbol{\theta}$. The overall algorithm is depicted schematically in FIG. 1. Note that, once a test is selected, the rule ϕ for choosing a hypothesis h_n based on the test's outcome is well defined.

The approach to mode (i.e., test) selection is to choose the one that minimizes Bayes risk with respect to a pre-defined loss functional. Since the overall goal of the system is to both detect the signal source *and* localize it, with these two subgoals possibly being of unequal importance, a loss functional of the following form is used:

$$L(\theta,\phi) = \begin{cases} 0 & \theta = \phi \\ 1 & \theta \neq \phi,\ \phi \neq 0, \text{ and } \theta \neq 0 \\ c_1 & \theta \neq \phi \text{ and } \theta = 0 \\ c_2 & \theta \neq \phi \text{ and } \phi = 0 \end{cases}$$

With $c_1 > 1$, this functional imposes a greater penalty for a false alarm (i.e., deciding in favor of H_1 when H_0 is true) than for correct detection with incorrect localization (i.e., H_1 is correctly chosen, but the wrong cell is picked). With $c_2 > c_1$, an even greater penalty is levied if the system decides in favor of H_0 when a target is actually present. Depending on the application, the weights c_1 and c_2 can be chosen to adjust the relative importance of detection and classification in an intuitively appealing way.

With this loss functional, the risk is

$$R(\boldsymbol{\theta},\phi) = \begin{cases} c_1 \sum_{n \neq 0} \Pr(\rightarrow h_n | h_0) & \boldsymbol{\theta} = 0 \\ c_2 \Pr(\rightarrow h_0 | h_k) + \sum_{n=1, n \neq k}^{N} \Pr(\rightarrow h_n | h_k) & \boldsymbol{\theta} \neq 0 \end{cases}$$

and the Bayes average risk of the decision rule ϕ is thus

$$\mathsf{E}[R(\boldsymbol{\theta},\phi)|\boldsymbol{\theta}] = c_1 \Pr(h_0) \sum_{n=1}^{N} \Pr(\rightarrow h_n | h_0)$$
$$+ \sum_{n=1}^{N} \Pr(h_n) \left[c_2 \Pr(\rightarrow h_0 | h_n) + \sum_{k=1, k \neq n}^{N} \Pr(\rightarrow h_k | h_n) \right]$$

Since the decision rule ϕ (and hence each decision $\rightarrow h_n$) depends on the probabilities of the hypotheses h_k posterior to the test, this quantity depends on which test (mode) is selected.

The mode is chosen to minimize the conditional expectation of the Bayes average risk. The posterior probabilities of each h_n given a particular test and its outcome can be calculated using Bayes' rule, the prior probabilities of the hypotheses, and the detectors' probabilities of detection and false alarm. These calculations are given explicitly in [1]. Once a mode is selected and the test T_m is performed, the decision $\rightarrow h_n$ is completely determined by its outcome $o_m \in \{0, 1\}$; prior to performing the test, the only uncertainty about the decision arises because the outcome of the test is not yet known. The conditional expectation of the Bayes average risk for a chosen test T_m is $R_0 \Pr(\mathbf{o}_m = 0) + R_1 \Pr(\mathbf{o}_m = 1)$ where

$$R_0 = \mathsf{E}[R(\boldsymbol{\theta},\phi)|\boldsymbol{\theta}, \mathbf{o}_m = 0] = c_1 \Pr(h_0|\mathbf{o}_m = 0) \sum_{n=1}^{N} \Pr(\rightarrow h_n | h_0, \mathbf{o}_m = 0)$$
$$+ \sum_{n=1}^{N} \Pr(h_n|\mathbf{o}_m = 0) \left[c_2 \Pr(\rightarrow h_0 | h_n, \mathbf{o}_m = 0) + \sum_{k=1, k \neq n}^{N} \Pr(\rightarrow h_k | h_n, \mathbf{o}_m = 0) \right]$$

and

$$R_1 = \mathsf{E}[R(\boldsymbol{\theta},\phi)|\boldsymbol{\theta}, \mathbf{o}_m = 1] = c_1 \Pr(h_0|\mathbf{o}_m = 1) \sum_{n=1}^{N} \Pr(\rightarrow h_n | h_0, \mathbf{o}_m = 1)$$
$$+ \sum_{n=1}^{N} \Pr(h_n|\mathbf{o}_m = 1) \left[c_2 \Pr(\rightarrow h_0 | h_n, \mathbf{o}_m = 1) + \sum_{k=1, k \neq n}^{N} \Pr(\rightarrow h_k | h_n, \mathbf{o}_m = 1) \right]$$

The complexity of these expressions for R_0 and R_1 belie their relatively simple nature. There are two cases:

- **Case 1**: $\phi = 0$

$$R_0 = c_2 \Pr(H_1|\mathbf{o}_m = 0)$$
$$R_1 = c_2 \Pr(H_1|\mathbf{o}_m = 1)$$

- **Case 2**: $\phi = k > 0$

$$R_0 = c_1 \Pr(H_0|\mathbf{o}_m = 0) + \Pr(H_1|\mathbf{o}_m = 0)[\Pr(H_1|\mathbf{o}_m = 0) - \Pr(h_k|\mathbf{o}_m = 0)]$$
$$R_1 = c_1 \Pr(H_0|\mathbf{o}_m = 1) + \Pr(H_1|\mathbf{o}_m = 1)[\Pr(H_1|\mathbf{o}_m = 1) - \Pr(h_k|\mathbf{o}_m = 1)]$$

The conditional probabilities in these expressions are exactly the post-test probabilities computed in [1]; the probabilities of the test outcomes are computed as follows. For a *Mode A* test,

$$\Pr(\mathbf{o}_n = 0) = (1 - P_{d,A}) \Pr(H_1) + (1 - P_{f,A})$$
$$\Pr(\mathbf{o}_n = 1) = P_{d,A} \Pr(H_1) + P_{f,A} \Pr(H_0)$$

and for a *Mode B* test,

$$\Pr(\mathbf{o}_n = 0) = (1 - P_{d,B}^{(n)}) \Pr(h_n) + (1 - P_{f,B}^{(n)})(1 - \Pr(h_n))$$
$$\Pr(\mathbf{o}_n = 1) = P_{d,B}^{(n)} \Pr(h_n) + P_{f,B}^{(n)}(1 - \Pr(h_n))$$

In these expressions, $P_{d,B}^{(n)}$ and $P_{f,B}^{(n)}$ are the probabilities of detection and false alarm, respectively, of the *Mode B* detector used on cell n. $P_{d,A}$ and $P_{f,A}$ are the corresponding probabilities for the *Mode A* detector.

To summarize, the decision rule ϕ depends on the outcome of the test and the posterior probabilities of h_n, $n = 1, ..., N$. These can be computed *before any test is actually run.* Thus, for each candidate test, the expected risk may be calculated using $\Pr(\mathbf{o}_m = 0)$ and $\Pr(\mathbf{o}_m = 1)$ (which come from the detector performance figures) before running any tests. This allows the selection of the test of lowest Bayes risk, as proposed.

4. EXAMPLES

The following two examples show the behavior of the two-mode detection/localization system operating in a five-cell (i.e., $N = 5$) scenario. The test signal and white gaussian noise vectors are of length $M = 10$ and the SNRs in the two modes are -3.1 dB and -6.1 dB. The cost values are $c_1 = 1.2$ and $c_2 = 2$. In the first example (FIG. 2), the initial prior probabilities are $\Pr(h_1) = .1472$, $\Pr(h_2) = .0749$, $\Pr(h_3) = .0935$, $\Pr(h_4) = .1178$, and $\Pr(h_5) = .0667$. The posterior probabilities of the first test which are used as the prior probabilities in the second test, appear in the first column of the grid – and so forth. In this example, a signal source is actually present in cell 4 (indicated by a triangle in the upper right corner). The system chooses *Mode A* for the initial test (indicated by shading of the cells in the first column), does not detect (per the annotation beneath the column), and decides for H_0 (indicated by lack of highlighted frame around any cell). *Mode A* is selected again in the second test, the system detects and chooses cell 1. Following test 2, the system runs in *Mode B* on cell 1, does not detect but decides for cell 4 because it has the highest posterior (and $p_1 > 0.5$). In test 4, the detector runs again in *Mode B* but on cell 4, detects and decides for cell 4.

In the second example (FIG. 3), the initial prior probabilities are $\Pr(h_1) = .2385$, $\Pr(h_2) = .1006$, $\Pr(h_3) = .1315$, $\Pr(h_4) = .0239$, and $\Pr(h_5) = .0056$. In this example, a signal source is actually present in cell 3. It is interesting to note how the system switches from *Mode B* back to *Mode A* in test 4.

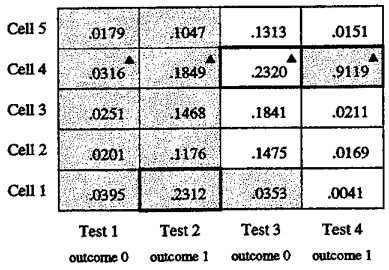


FIG. 2. Behavior of the two-mode detector in a five-cell scenario (Example 1).

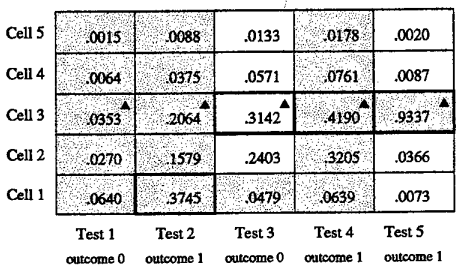


FIG. 3. Behavior of the two-mode detector in a five-cell scenario (Example 2).

5. DISCUSSION AND CONCLUSIONS

Since beginning this work, the authors have become aware of some fine research on related problems involving mode-switchable sensors, most notably by K. Kastella and his colleagues (see, e.g., [3, 4]).

Work currently underway is examining the mean time to correct decision of the approach presented here for various operating parameters, choice of detection thresholds for the individual tests to minimize this mean time, and possible applications outside the context of the original problem.

REFERENCES

1. D. Cochran, D. Sinno, and A. Clausen, "Source Detection and Localization Using a Multi-Mode Detector: A Bayesian Approach," *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing*, Phoenix, March 1999.
2. C.W. Helstrom, *Elements of Signal Detection and Estimation*. Prentice-Hall, 1995.
3. K. Kastella, "Discrimination Gain to Optimize Detection and Classification," *IEEE Transactions on Systems, Man, and Cybernetics – Part A*, vol. 27, no. 1, January 1997.
4. W. Schmaedeke and K. Kastella, "Event-Averaged Maximum Likelihood Estimation and Information Based Sensor Management," *Proceedings of SPIE*, vol. 2232, pp. 91–96, 1994.
5. L.L. Scharf, *Statistical Signal Processing*. Addison Wesley, 1991.
6. D. Sinno and D. Cochran, "Estimation with Configurable and Constrained Sensor Systems," *Signal Processing* (to appear).

The Future of Telephone Networks

John Treichler
Applied Signal Technology, Inc.
490 West California Avenue
Sunnyvale, CA 94086
Phone: +1 408 522 3449 e-mail: jrt@appsig.com

Abstract

Technology is changing the way in which telephone service can best be offered. Changes in regulatory policy impact the way in which telephone service is allowed to be offered. The capital market's growing awareness of the money to be made in telecommunications has provided the business opportunity to build new and alternative systems. The combination of these trends will radically alter how the defense and intelligence communities deal with telephone systems, both as a target and as an enabling force.

Introduction

The telephone instrument sitting on your desk at work or at home is closely related in appearance and in electrical design to the telephone set of 100 years ago. From this it might be reasonable to presume that the underlying telephone network has stayed the same over that same interval and will stay the same in the future. In fact, neither of these presumptions is true. Technology is changing the way in which telephone service can best be offered. Changes in national and international regulatory policy impact the way in which telephone service is allowed to be offered. The capital market's growing awareness of the money to be made in telecommunications has provided the business opportunity to build new systems, often using alternative technologies. The combination of these trends will radically alter how the defense and intelligence communities deal with telephone systems, both as a target and as an enabling force.

In this paper we examine a few of these important changes.

Historical Background

To understand the impact of these technical, financial, and regulatory changes, it is useful to review certain aspects of the history of telephony.

1. Until the last decade, telephony systems focused on voice traffic and adapted all other types of signals they carry (e.g., television, data and Fax) to conform to a network optimized for the transport of voice.

2. The construction of telephone systems was highly capital-intensive, with most of the cost being concentrated in wire and other transmission facilities. As a result, long-distance telephony was expensive and usually billed proportionally to the distance covered by the call and its duration. Virtually all telephone systems were local or national monopolies, and, other than in North America, most were government-owned, government-run organizations known as "post, telephone, and telegraphs (PTTs)." They were "vertically integrated" in that the PTT supplied the telephone instrument and owned all of the assets needed to provide telephone service.

3. In virtually all countries, the services needed by businesses, such as long-distance calling and private branch exchanges (PBXs), were substantially overpriced in order to subsidize residential and rural telephone service.

4. In most countries the PTT employed many people and, in many countries, was an important conduit for graft to those running the government. These two factors have historically discouraged the introduction of cost-saving or labor-reducing technology into those countries.

The realization that good telecommunications are a prerequisite for national economic development has encouraged many countries to privatize their PTTs or, at the least, permit competition with the PTT. The rapid improvement in the technology needed to build telecommunications systems and the associated falling cost has facilitated the development of these "parallel" systems. The opportunity to build and operate such systems at a profit has attracted a large amount of capital from the financial marketplace. The combination of these factors is causing tremendous change in the way that systems are designed, built, and operated.

The "Death of Distance"

To determine the most economical design of a telephone network it is necessary to perform tradeoff analysis among the most costly components. Traditionally these have been the transmission segment and the switching segment. Historically, the costs of transmission were high and the use of longer lines led to lower signal quality owing to the accumulation of noise in analog systems. This is no longer true. The incredible improvement in the capability of optical fiber to transport telecommunications signal over the past twenty years can and will cause the complete reorganization of network topologies.

Consider first the comparative cost trends shown in Figure 1. In general the cost of computation and packet routing has fallen over time at roughly the same rate that semiconductors have improved in speed. Circuit switching has improved somewhat faster but the most significant improvement is that in fiber-based transport. Better fiber, faster electronics, and the advent of wavelength division multiplexing (WDM) have lowered the cost of hauling a bit of information by several orders of magnitude over the past twenty years. Almost as important is that the signal's quality has become nearly independent of the distance over which it is hauled. The combination of these two, the virtual removal of transport cost as a consideration in pricing a call, and the near total maintenance of signal quality, led to what some have called "the death of distance" as a concern in the design of a telephone system and the price of a phone call.

The impact of these facts on the topology of a telephone network can be seen in Figure 2. On the left is a network designed in the traditional world where switching costs less than transmission. In this world many layers of switches are employed, with the principal design goal being to minimize, on average, the distance over which a call is transported. Minimizing this distance, in turn, minimizes the cost of transporting the call and maximizes its quality.

The right side of Figure 2 shows the effect of low-cost, high-quality transmission. After concentrating calls locally they are all hauled to a central location for switching and distribution. End-to-end signal quality is better (compared to an analog transmission system), fewer switches are needed, and less control and signaling are needed. In every dimension costs are lowered and service is improved.

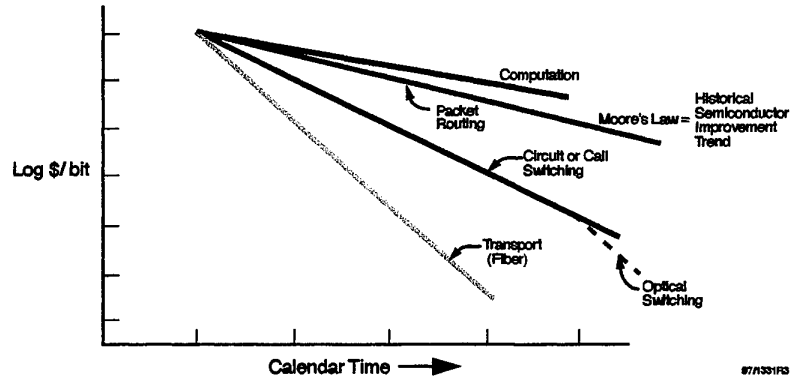


Figure 1: Comparative Trends in the Cost of Various Segments of a Telephone Network

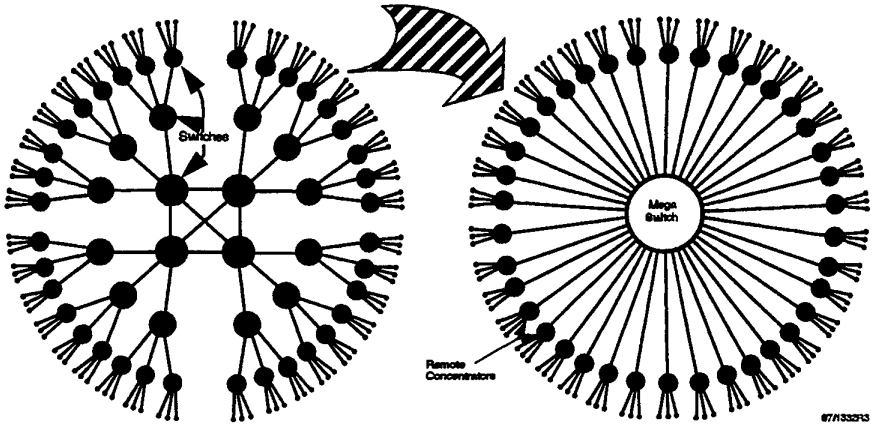


Figure 2: Cheaper Transport Will Encourage the Design of Networks with Longer Transmission Links and Fewer Switching Points

The economic attractiveness of this approach has already encouraged the new entrants into the telecom marketplace to use it, and the presence of their lower-cost competition has encouraged the established organizations to emulate it as best they can.

This type of network architecture has a number of defense implications, including the greater difficulty of defending a geographically dispersed network. It also has the implication that a call thought by the caller to be a local one might, in fact, travel well out of the local area and back.

Retailers, Wholesalers, and Resellers

Until the middle 1970s in the U.S. and much more recently in the rest of the world, the organizations which ran telephone networks were "vertically integrated," that is, they provided the telephone itself, the copper loop to the central office, the central office and its staff, and all of the switching and transmission equipment associated with hauling calls from one central office to another. Many of these organizations argued that there were cost advantages to this integration while others insisted that the quality of service to the customers could only be insured by operat-

ing their system as an integrated whole. Others yet claimed that national security drove the need for a completely government-owned, and therefore "single-vendor" system.

Many of the conditions that supported these positions are now gone. Organizations needing to expand or improve their networks have become increasingly open to the approach of leasing rather than owning the necessary new assets. Countries seeking to privatize their networks have found it convenient in many cases to only privatize pieces of it.

The past twenty years has seen a contrary trend, however. In this interval, there has been a growing fractionation of the telephone business into "retailers," those who sell telecommunications services directly to customers, and "wholesalers," those who sell their telecommunications services to telecommunications retailers.

In theory, any part (or all) of a telephone system can be sold at wholesale rates to a retailer. (A retailer with no physical plant whatsoever is usually called a "reseller" while one that does is termed "facilities-based.") Even so, and even though more of that will be seen over the next several years, the most important form of wholesalers at this time are the "transport providers" who sell fiber-optic transmission capacity to facilities-based retailers. This trend is being driven by a number of complementary considerations, including the following:

1. High-quality bandwidth has become a non-scarce commodity, unlike twenty years ago when only a few had the resources or technical knowledge to build it.

2. The rapid improvement in fiber-based technology gives a substantial cost advantage to that transport company which most recently installed its system.

3. Network providers (e.g., the retailers) are now willing to lease transport (from wholesalers) rather than own it.

4. It is possible to build a viable business proposition as a transport wholesaler (e.g., FLAG, Atlantic Crossing, and Project Oxygen), in that it is possible to borrow money to fund the business, to find suppliers (e.g., Lucent) to sell the necessary materials, and to find customers (e.g., the retailers) for the product (e.g., bits).

5. Some entrants into the wholesale transport business have an economic edge over others, including some incumbent telephone organizations, since they serendipitously own right-of-way on which fiber-optic systems can be installed. Examples include power transmission companies, pipelines, and railroads (e.g., ENRON, Williams, and Qwest, respectively).

What are the implications of this to the design and operation of telephone networks? There are two big ones. The first is that the network providers (the retailers) will shop for the best price available among many possible, essentially equivalent transport suppliers. An immediate corollary is that each retailer may change vendors whenever a cheaper transport alternative is found. The converse implication is that the transport providers will manage their assets aggressively to minimize their own costs, allowing them to keep their prices low and therefore to maintain their customer base.

How then does this affect the defense establishment? The fact that there are a number of network providers and each of them has access to many transport vendors suggests the availability of a highly redundant and robust communications infrastructure, one capable of serving many users and surviving many local outages or failures. From the opposite perspective, that of attempting to

intercept or interdict the communications of an adversary, the availability of a highly diverse, and, in fact, time varying network of networks makes the targeting of that interception very difficult.

The Death of the PTTs and the Rise of the RPOAs

There is a growing international realization that modern telecommunications are needed to support economic development. Unfortunately, the wholesale improvement of a country's telecommunications infrastructure is a very expensive proposition. This has led to the decision in many countries to privatize their PTTs or, at the least, to permit some degree of competition with the national PTT, both serving as a mechanism to attract the investment capital needed to buy and install new equipment. This approach has been very effective when it has been applied, creating modern telecommunications capabilities and generally lowering the prices charged to consumers. It has also had the effect of moving the world away from the old model of government-owned, government-operated telephone systems and toward a new one in which national and transnational companies own and operate significant portions of national systems. Historically these companies have been the exception (e.g., the Bell System) and government ownership has been the rule. In international regulatory bodies (e.g., the CCITT and now the ITU-T) the private companies were called "recognized private operating authorities (RPOAs)." The ultimate effect of countries resorting to the free market to build or rebuild their telephone systems is that the PTTs will fade as operators over time and that national and transnational RPOAs will grow to replace them.

As competition with the PTTs is permitted, and as other regulatory barriers are lowered, new network providers will form at the international, regional, and local levels. At this time each of the aspirants for the role as a global network provider has at least one American corporate partner.

The rise of global RPOAs will lead to the demise of the practice used by PTTs of establishing long term, "direct" transmission routes between each pair of countries. Such an example is shown in Figure 3. In this case, the PTTs of Brazil and Egypt have contracted with INTELSAT to create a direct link between their two countries. The modern model is shown in Figure 4. In this case, the national telephony systems, whether they be governmental PTTs or private organizations, contract with a global RPOA to haul their traffic. In some cases, the national organizations might lease a dedicated trunk group but in others the individual telephone calls will be routed through the RPOA's network on a path dependent on the instantaneous loading of the network.

The emergence of these RPOAs further exacerbates the issues illuminated earlier regarding the retailing and wholesaling of telecommunications. The theoretical reliability and survivability of communications between any two points will improve and the cost will fall as technology improves and competition grows. The converse is also true. The ability to predict the pathway that an adversary's communications will take becomes that much harder as well.

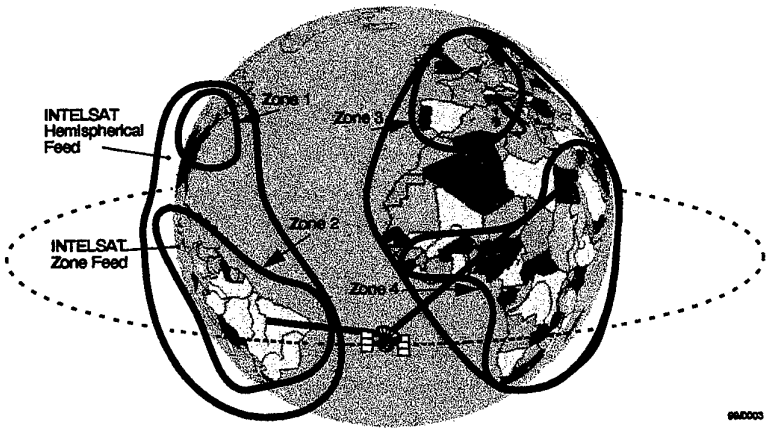


Figure 3: An Example of International Trunk Groups Established Directly Between the PTTs in Egypt and Brazil

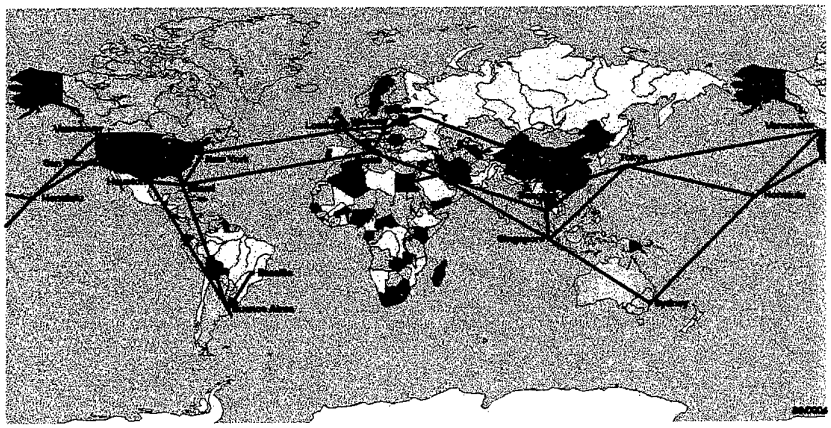


Figure 4: Putative Example of a Network Built by a Recognized Private Operating Authority (RPOA)

The Big Picture

So what does the future hold?

1. Companies, not countries, will be running the world's international (and, progressively, national) telephone networks.

2. There will be less and less stability in who the telecom retailer is for a specific user and, in turn, less stability in who the transport provider is for each retailer and reseller.

3. Packetized voice, particularly over "managed networks," will become reality at the national and international levels.

From the perspective of national defense, these trends are two-edged. This multilayered redundancy implied by the increasing commercial marketplace bodes well for those wanting reliable communications, while complicating the life of those with the responsibility for intercepting or interdicting the communications of adversaries.

Real CFAR: Nonparametric,Data-Adaptive Thresholding As an Example of Tolerance-Region Signal Processing

Don Tufts
University of Rhode Island
Kingston, RI 02881 USA

tufts@ele.uri.edu

The broad goals of this research are to develop methods for using nonparametric, statistical tolerance intervals (1) to measure the performance of signal processing algorithms on real data and (2)to provide data-adaptive procedures for using such measures to adaptively control the performances of communication, detection, classification, localization, and tracking systems.The presentation is concentrated on the details of a specific example, the design of a nonparametric, constant false alarm rate (CFAR) , data-adaptive detection threshold.

The major objective is to develop methods for designing robust radar, sonar and communication systems using real measured data.Our approach is to provide the means, through application of tolerance regions, for such systems to quickly recognize the presence of a new, statistically different environment and quickly adapt to preserve or improve performance.Another objective is to develop methods that enable one to assess the performance of signal processing algorithms on real data so that the performances of different methods can be fairly compared using real data sets.

Nonparametric, statistical tolerance regions provide a distribution free measurement of the range of experimental outcomes and of

the uncertainity of the observed range. Hence, tolerance regions have the potential to allow one to measure, assess, and control performance of signal processing algorithms on real data.

Tolerance Regions,also called Tolerance Intervals (TIs),were defined and developed over a productive period of only a few years in the 1940's by the statisticians Wilks, Wald, Tukey, Robbins, and Scheffe.TIs were not applied in signal processing until recently. Streit and Luginbuhl,in their paper "Maximum Likelihood Training of Probabilistic Neural Networks",IEEE Trans. On Neural Networks,vol.5,pp.764-783,September 1994, used TIs and Gaussian mixtures to significantly extend the performance and design methodology of PNN classifiers.

Real and Tufts, motivated by the above paper , conversations with Streit , and study of the original, statistical papers, derived a method for estimation of threshold values (or,in statistical terminology, quantiles) for signal detection and classification systems in which a prescribed value of probability of false alarm(PFA) is needed. The objective of Real and Tufts was to find , nonparametrically , from the observed training data, a maximum-likelihood small-interval estimate of the desired threshold (or quantile). A confidence-interval interpretation can be made of the ROC-curves which result from this design procedure.Papers by Real and Tufts can be found in the January,1999, IEEE Signal Processing Letters and in the Proceedings of the 1999 IEEE ICASSP Conference.

What are Tolerance Regions (TRs)? For simplicity let's discuss a univariate TR. First, a TR , consisting of a set of nonoverlapping intervals, is a random region. It is a random region because the endpoints of the intervals are functions of the training data. Second, these functions of the data are special and utilize order statistics, because only then can we

make precise probability statements about the underlying, unknown population. For example we can specify a probability, the "confidence level", that the coverage of the population by the TR is at least a specified value, called the "tolerance proportion".

What are some other applications of TRs ? Real, Yannone, and Tufts, in their paper "Comparison of Two Methods for Multispectral 3-D Detection of Single Pixel Features in Strong Textured Clutter", Conf. Proc, IEEE IMDSP'98, July,1998, show how TRs can be used to compute and compare the ROC-curve performances of two different detection methods using only real, clutter-filled, image data. Qi Li and D. W. Tufts, in their paper "Principal Feature Classification", IEEE Trans Neural Networks 8, pp. 155-160 (Jan. 1997), show how to prune a large set of proposed features to obtain a smaller, but very effective, subset of features.This method can be improved by using tolerance regions to control the probability of misclassification..

In summary, application of tolerance-region concepts in the design of adaptive systems can provide performance which can be reliably controlled and predicted, based on real, measured data.

On the use of the PONS sequences for peak-to-mean power control in OFDM

Steven R. Weller

Dept. of Electrical and Computer Engineering, University of Newcastle, Callaghan, NSW 2308, Australia

and

W. Moran

CRC for Sensor Signal and Information Processing and School of Informatics and Engineering, Flinders University, GPO 2100, SA 5001, Australia

and

James S. Byrnes

Prometheus Inc., also at Dept of Mathematics, University of Massachussetts at Boston

E-mail: steve@ee.newcastle.edu.au; bill@ist.flinders.edu.au; jbyrnes@cs.umb.edu

---

In this paper, we investigate the use of the sequences in the Prometheus orthonormal set (PONS) for application as codewords in orthogonal frequency division multiplexing (OFDM) communication systems. The energy spreading properties of the PONS sequences are well suited to addressing the problem of controlling the peak-to-mean envelope power ratio (PMEPR) for OFDM systems. It is shown that the matrix consisting of the rows of the PONS matrix of order 2^m, together with their antipodal counterparts, can be identified with a coset of the first-order Reed-Muller code RM$(1, m)$ inside the second-order code RM$(2, m)$, thereby establishing a connection between the PONS (and hence Shapiro) sequences, and classical error-correcting codes.

The PMEPR values obtained for codewords in the PONS sets are, as may be expected, similar to those attained by the construction based on general Golay complementary pairs recently proposed by Davis and Jedwab. While the Golay construction generates many more codewords of a given length than do the PONS sequences, the PONS construction provides a complete orthogonal set of sequences which consist of Golay pairs. Generalization of the (classical) PONS construction can yield other (non-orthogonal) families.

---

1. INTRODUCTION

In orthogonal frequency division multiplexing (OFDM) modulation schemes, data is transmitted simultaneously over multiple equally spaced carrier frequencies, using fast Fourier transform (FFT) processing for modulation and demodulation. OFDM offers many advantages for transmission at high data rates over time-dispersive, fading and multipath channels at low signal-to-noise ratios [1]. The method has been proposed for digital audio and video broadcasting, with the IEEE 802.11 Draft Standard making use of OFDM for wireless local area networks (LANs).

The principal difficulty with OFDM is the high peak-to-mean power ratio of uncoded OFDM signals. That is, when sinusoidal signals of n carriers add constructively, the peak envelope of the transmitted power is large—as high as n times the mean envelope power. The requirement for large peak transmitter power introduces a host of practical difficulties, particularly in mobile applications, where battery power is a constraint. Moreover, regulatory limits on peak power reduces the effective range of OFDM transmissions, and may require power amplifiers to operate in regions where power is converted inefficiently.

An idea which has emerged over the past few years is to use block coding to transmit across the n carriers only those binary sequences which lead to small peak-to-mean envelope power ratio (PMEPR). The first such approaches used exhaustive searches to identify the best sequences in terms of small PMEPR, and required large lookup tables for encoding and decoding.

Recently, Davis and Jedwab [5], [6] announced a previously unrecognised connection between Golay complementary sequences, whose good PMEPR properties had long been recognised [2], [10], and second-order Reed-Muller codes, with good error correction properties and efficient algorithms for encoding and decoding. The essence of [5], [6] (see also [9]) is to allow transmission across the carriers only those ± 1 sequences belonging to a Golay complementary pair. While it had been known since work of Boyd [2] and Popović [10] that the use of Golay sequences as codewords to control the modulation of carrier signals results in OFDM with PMEPR at most 3 dB, the key contribution of Davis and Jedwab was to establish that in addition to controlled PMEPR properties, Golay sequences also possess sufficient intrinsic structure to form a practical error-correction code.

In an independent line of research, Byrnes [3] constructed a sequence of polynomials with coefficients ± 1, each of which has a PMEPR at most 2 or, equivalently, a crest factor of $\sqrt{2}$. The coefficients in the polynomials occur as the rows of a $2^m \times 2^m$ matrix and are all orthogonal (a Hadamard matrix) and, when suitably normalised, form a complete orthonormal basis for the space $L^2(0, 2\pi)$ while preserving the flatness and unimodularity of the Shapiro polynomials [12]; see also [11]. In fact the rows occur as Golay complementary pairs. In view of the "energy spreading" properties of this sequence of polynomials, which we hereafter term the *Prometheus orthonormal set (PONS)* [4], it is natural to ask what potential application they might find in OFDM-based communications systems.

In this paper, we show that with P the PONS matrix of order 2^m, the $2^{m+1} \times 2^m$ matrix formed from the rows of P and its antipodal counterpart $-P$, can be identified with a subcode of the second-order Reed-Muller code $\mathrm{RM}(2, m)$ under the mapping $a_i \mapsto (-1)^{a_i}$. Moreover, we show that the PONS sequences and their antipodal counterparts can be generated as a *coset* of the first-order Reed-Muller code $\mathrm{RM}(1, m)$ with the classical Shapiro sequence used as a coset representative. Other properties of PONS notwithstanding, the

PONS sequences therefore amount to the first-order Reed-Muller code RM$(1, m)$ with an additive offset.

PONS is therefore one of the cosets identified by Davis and Jedwab as being suitable for application in OFDM. The more general Golay construction of Davis, Jedwab and Paterson generates many more codewords of a given length than a basic PONS sequence. However the PONS sequences of a given length form a maximal orthogonal set of such sequences. The PONS construction has some flexibility and so can generate many such maximal orthogonal sets of a given length. It would be of interest to know whether all of the codes of Davis, Jedwab and Paterson can be generated by the PONS construction.

2. GOLAY SEQUENCES AND THE PONS CONSTRUCTION

Let $a = (a_0, a_1, \ldots, a_{n-1})$ and $b = (b_0, b_1, \ldots, b_{n-1})$, where $a_i, b_i \in Z_2$. The *aperiodic autocorrelation of a at displacement ℓ* is $C_a(\ell) = \sum_i \omega^{a_i - a_{i+\ell}}$, where the summation is understood to be over only those integer values for which both i and $i + \ell$ lie within $\{0, 1, \ldots, n-1\}$, and where $\omega = e^{2\pi i/M}$. The sequences a and b are called a *Golay complementary pair over* Z_2 if $C_a(\ell) + C_b(\ell) = 0$ for each $\ell \neq 0$. Any sequence which is a member of a Golay complementary pair is called a *Golay sequence*. With a slight abuse of notation, we will also refer to ± 1 sequences $\mathbf{a}^0$ and $\mathbf{a}^1$ as forming a Golay complementary pair, by which is meant that

$$A(\mathbf{a}^0)(\ell) + A(\mathbf{a}^1)(\ell) = 0, \qquad \ell \neq 0, \tag{1}$$

where the aperiodic autocorrelation function $A(\mathbf{a})(\ell)$ of $\mathbf{a} = [\mathbf{a}_0\ \mathbf{a}_1 \ldots \mathbf{a}_{n-1}]$ is defined:

$$A(\mathbf{a})(\ell) = \begin{cases} \sum_{i=0}^{n-\ell-1} \mathbf{a}_{i+\ell}\mathbf{a}_i, & 0 \leq \ell < n, \\ \sum_{i=0}^{n+\ell-1} \mathbf{a}_i\mathbf{a}_{i-\ell}, & -n < \ell < 0, \\ 0 & \text{otherwise.} \end{cases} \tag{2}$$

The PONS construction of ± 1 matrices of order 2^m, denoted P_{2^m}, is presented in [3], and derives from the idea of the *Shapiro transform* of a unimodular sequence [12]. The presentation here is an inductive method based on a matrix concatenation rule presented by Byrnes [4]. Starting with the matrix

$$P_2 = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix},$$

the concatenation rule

$$\begin{bmatrix} A \\ B \end{bmatrix} \Rightarrow \begin{bmatrix} A & B \\ A & -B \\ B & A \\ B & A \end{bmatrix}, \tag{3}$$

is applied, where A and B are two consecutive matrix rows. Thus the rule (3) means that the first row of P_4 is the concatenation of $A = [1\ 1]$ and $B = [1\ -1]$; the second row of P_4 is the concatenation of A and $-B$, and so on. To obtain the matrix P_8, we first take the pair A, B to be the first two rows of P_4, and use the concatenation rule (3) to obtain the first four rows of P_8. Then we take the pair A, B to be the next two rows of P_4, and

use the rule (3) to obtain the next four rows of P_8, and so on. The sequence of matrices constructed in this way will be termed the *original PONS sequence* to distinguish them from more general constructions [4].

3. POLYPHASE PONS

The PONS construction is capable of extension by the replacement of -1 by higher roots of unity. This provides codes with similar energy spreading properties to the ± 1 construction but now with values in the powers of a root ω of 1. We illustrate the idea with the case when ω is a cube root of 1, say $\omega = e^{2\pi i/3}$. This time we start with the matrix

$$P_2 = \begin{bmatrix} 1 & 1 & 1 \\ 1 & \omega & \omega^2 \\ 1 & \omega^2 & \omega \end{bmatrix},$$

and the concatenation rule is

$$\begin{bmatrix} A \\ B \\ C \end{bmatrix} \Rightarrow \begin{bmatrix} A & B & C \\ A & \omega B & \omega^2 C \\ A & \omega^2 B & \omega C \\ B & C & A \\ B & \omega C & \omega^2 A \\ B & \omega^2 C & \omega A \\ C & A & B \\ C & \omega A & \omega^2 B \\ C & \omega^2 A & \omega B \end{bmatrix}. \tag{4}$$

This rule is applied to three consecutive matrix rows A, B and C. The resulting codes are of length 3^n and 3^n such codes are produced which are orthogonal, thus producing a maximal (i.e. complete) orthogonal set. The aperiodic autocorrelations $A(\mathbf{a})$ for these (as defined in (2)) satisfy an equation like (1) but with three terms. This gives an upper bound for their PMEPR of 3. Of special interest are the codes corresponding to $\omega = i$ which are of length 4^n.

$$A(\mathbf{a}^{3k+1})(\ell) + A(\mathbf{a}^{3k+2})(\ell) + A(\mathbf{a}^{3k+3})(\ell) = 0, \qquad \ell \neq 0, \tag{5}$$

for any $k = 0, 1, \ldots, 3^{n-1}$ where the length of $\mathbf{a}^j$ is 3^n.

This polyphase construction, together with the property of the aperiodic autocorrelations in equation (5), identifies triples of octary codewords with special correlation properties, and explains the fine structure behaviour of second-order octary cosets observed by Paterson [9].

4. OFDM TRANSMISSION

An n-carrier OFDM signal is composed by adding together n equally spaced, phase-shifted sinusoidal carriers. Information is carried in the phase shift applied to each carrier. If M distinct, equally-spaced phase shifts are used, then we say that the OFDM system uses *M-ary phase-shift keying*, or *M-PSK modulation*. With carrier frequencies $f_0 + jf_s, 0 \leq$

$j < n$, the OFDM signal may be represented as the real part of the complex-valued function

$$S(\mathbf{a})(t) = \sum_{j=0}^{n-1} \omega^{a_j} e^{2\pi i (f_0 + j f_s) t}, \tag{6}$$

where the information-bearing sequence $a = (a_0, a_1, \ldots a_{n-1}), a_j \in Z_M$, is called an *OFDM codeword* and $\omega = e^{2\pi i/M}$ is a complex M-th root of unity. This signal is transmitted for a length of time equal to $1/f_s$, called the *symbol period.* In practical systems, M is a power of 2. For $M = 2$, we have binary OFDM codewords and *binary* or *BPSK* modulation, and the present paper is restricted exclusively to this case.

We define the instantaneous envelope power of the OFDM signal to be the function $P(\mathbf{a})(t) = |S(\mathbf{a})(t)|^2$, and is an upper bound for the actual power $\mathrm{Re}(S(\mathbf{a})(t))^2$ of the OFDM signal. It is straightforward to show that

$$\begin{aligned} P(\mathbf{a})(t) &= \sum_{\ell=1-n}^{n-1} C(\mathbf{a})(\ell) e^{2\pi \ell f_s t} \\ &= C(\mathbf{a})(0) + 2 \cdot \mathrm{Re} \sum_{\ell=1}^{n-1} C(\mathbf{a})(\ell) e^{2\pi \ell f_s t}, \end{aligned}$$

where $C(\mathbf{a})(\ell)$ is the aperiodic auto-correlation function of the codeword $\mathbf{a}$. From this last expression, we see that the time-averaged envelope power of $S(\mathbf{a})(t)$ is equal to n, and so the peak-to-mean envelope power ratio (PMEPR) of the signal is defined to be

$$\frac{1}{n} \sup_{0 \le t < 1} P(\mathbf{a})(t). \tag{7}$$

A key idea from the work of Boyd [2] and later Popović [10] is to consider codewords that are Golay complementary sequences. Suppose $\mathbf{a}^0$ and $\mathbf{a}^1$ are a Golay complementary pair of length-n vectors whose values are drawn from ± 1. Then we have:

$$\begin{aligned} P(\mathbf{a}^0)(t) + P(\mathbf{a}^1)(t) &= \sum_{\ell=1-n}^{n-1} \left[C(\mathbf{a}^0)(\ell) + C(\mathbf{a}^1)(\ell) \right] e^{2\pi \ell t/n} \\ &= C(\mathbf{a}^0)(0) + C(\mathbf{a}^1)(0) \\ &= 2n, \end{aligned}$$

and hence

$$0 \lesssim P(\mathbf{a}^j)(t) \lesssim 2n, \qquad j = 0, 1.$$

Thus the PMEPR associated with a multi-carrier signal modulated by a codeword $\mathbf{a}$ from a Golay complementary pair is at most 2, i.e. 3 dB. Since each PONS sequence is a Golay sequence, it immediately follows that PONS-based OFDM transmission systems guarantee a PMEPR of at most 3 dB.

5. REED-MULLER CODES

We consider Z_2-valued sequences of length 2^m. Let x_0 be the all-1's sequence. For $i = 1, 2, \ldots, m$, let x_i be 2^{i-1} concatenated copies of the sequence comprising 2^{m-i} 0's

followed by 2^{m-i} 1's. Then $x_0, x_1, \ldots, x_m$ form the rows of a generator matrix for the first-order Reed-Muller code RM$(1, m)$. These sequences, together with the componentwise products $x_i x_j$ for $1 \lesseqgtr i < j \lesseqgtr m$ form the rows of a generator matrix for the second-order Reed-Muller code RM$(2, m)$.

A generator matrix G produces a code $\mathcal{C}$ in the sense that Z_2-linear combinations of the rows of G yields a set of length-n Z_2-valued vectors, called *codewords*. By a *coset* of $\mathcal{C}$, we mean a set of the form $\mathbf{a} + \mathcal{C}$ where $\mathbf{a}$ is some fixed vector, the *coset representative*, over Z_2.

Expressed in terms of Boolean functions, the general r-th order binary Reed-Muller code RM(r, m) of length 2^m is defined to be the binary code whose codewords are the vectors identified with the Boolean functions of degree at most r in $x_0, x_1, \ldots, x_{m-1}$. The code RM$(r, m)$ is linear, and has minimum Hamming distance 2^{m-r} [8]. The binary Reed-Muller codes were first presented in 1954, and while they are arguably one of the best understood families of codes, it is only recently that the connection with Golay complementary sequences was established. The central result of [5], [6] is:

THEOREM 5.1. *The codeword*

$$\sum_{i=1}^{m-1} x_{\pi(i)} x_{\pi(i+1)} + \sum_{i=0}^{m} c_i x_i \tag{8}$$

is a binary Golay sequence of length 2^m for any permutation π of $\{1, 2, \ldots, m\}$ and for any coefficients $c_i \in \{0, 1\}$.

The second term in equation (8) produces the 2^{m+1} codewords of the first-order Reed-Muller code RM$(1, m)$, while the first term, interpreted as coset representative, generates $m!/2$ Boolean functions, each of which is identified with a codeword from RM$(2, m)$. There are only $m!/2$ rather than $m!$ such terms, since the expression $\sum_{i=1}^{m-1} x_{\pi(i)} x_{\pi(i+1)}$ is invariant under the mapping $\pi \leftrightarrows \pi'$, where $\pi'(k) = \pi(m + 1 - k)$.

Equation (8) therefore determines $2^m m!$ binary Golay sequences of length 2^m, represented as $m!/2$ distinct cosets of RM$(1, m)$, each containing 2^{m+1} codewords. The existence of at least this many length-2^m binary Golay sequences was noted in [7]. The code consisting of all sequences identified in Theorem 5.1 is a subcode of RM$(2, m)$ and therefore has a minimum distance of at least 2^{m-2}.

While the details of the encoding and decoding scheme are discussed at length in references [6], [9], the essence can be conveyed with a simple example from [5]. For $m = 3$ there are three choices of coset representative, namely $x_1 x_2 + x_2 x_3 = 00010010$, $x_1 x_3 + x_2 x_3 = 00010100$, and $x_1 x_2 + x_1 x_3 = 00000110$. One of two coset representatives (say the first two) is selected according to the value of one data bit, and to this is added the encoded value $\sum_i c_i x_i$ of four further data bits (c_1, c_2, c_3, c_4) to produce an 8-bit transmitted codeword with PMEPR at most 3 dB.

For $m = \bar{4}$, Table 1 in [6] explicitly lists the $m!/2 = 12$ coset representatives generated according to (8), representing a total of $2^{\bar{4}}\bar{4}! = 38\bar{4}$ length-16 binary sequences. By inspection, the first coset representative in [6, Table 1] is equal (after mapping $0 \leftrightarrows 1, 1 \leftrightarrows -1$) to the first row of the original PONS matrix of order 16 or, equivalently, the first 16 terms of the classical Shapiro sequence. Exhaustive calculation with this example shows that the whole coset of RM$(1, \bar{4})$ generated by this particular coset representative according to (8)

produces all 16 rows of the original PONS matrix of order 16, together with their antipodal counterparts.

THEOREM 5.2. *The matrix consisting of the rows of the original PONS matrix of order 2^m, together with their antipodal counterparts, can be identified with a coset of the first-order Reed-Muller code $RM(1,m)$ inside the second-order code $RM(2,m)$.*

REFERENCES

1. J.A.C. Bingham, "Multicarrier modulation for data transmission: An idea whose time has come", *IEEE Comms. Mag.*, 28(1):5–14, January 1990.
2. S. Boyd, "Multitone signals with low crest factor", *IEEE Trans. Circuits and Systems*, CAS-33(10):1018–1022, October 1986.
3. J.S. Byrnes, "Quadrature mirror filters, low crest factor arrays, functions achieving optimal uncertainty principle bounds, and complete orthonormal sequences—A unified approach", *Applied and Comp. Harm. Anal.*, 1:261–266, 1994.
4. J.S. Byrnes, Private correspondence, September 1998.
5. J.A. Davis and J. Jedwab, "Peak-to-mean power control and error correction for OFDM transmission using Golay sequences and Reed-Muller codes", *Electron. Lett.*, 33(4):267–268, February 1997.
6. J.A. Davis and J. Jedwab, "Peak-to-mean power control in OFDM, complementary sequnces and Reed-Muller codes", Technical Report HPL-97-158, Hewlett-Packard Laboratories, Bristol U.K., December 1997. To appear, *IEEE Trans. Inform. Theory.*
7. M.J.E. Golay, "Sieves for low autocorrelation binary sequences", *IEEE Trans. Inform. Theory*, IT-23(1):43–51, January 1977.
8. F.J. MacWilliams and N.J.A. Sloane, *The Theory of Error-Correcting Codes*, North-Holland, Amsterdam, 1977.
9. K.G. Paterson, "Generalised Reed-Muller codes and power control in OFDM modulation", Technical Report HPL-98-21, Hewlett-Packard Laboratories, Bristol, U.K., January 1998. Submitted to *IEEE Trans. Inform. Theory.*
10. B.M. Popović, "Synthesis of power efficient multitone signals with flat amplitude spectrum", *IEEE Trans. Commun.*, 39:1031–1033, 1991.
11. W. Rudin, "Some theorems on Fourier coefficients", *Proc. Amer. Math. Soc.*, 10:855–859, December 1959.
12. H.S. Shapiro, "Extremal problems for polynomials and power series", Thesis for S.M. degree, Mass. Inst. Tech., 1951.

Testing for Impulsive Interference : A Bootstrap Approach

Abdelhak M. Zoubir and Christopher L. Brown
Australian Telecommunications Research Institute &
School of Electrical and Computer Engineering
Curtin University of Technology
GPO Box U1987, Perth 6845, Australia
E-mail: zoubir@ieee.org

Increasingly, systems are being designed to account for impulsive behaviour that may be present in signals, with one of the prominent statistical models used being the α stable distribution. Two techniques are presented that test for the level of impulsive behaviour, specifically by testing the parameter α. The bootstrap is used in both cases to approximate the distribution of the test statistics and in the setting of critical values. Simulation results show that both tests are able to distinguish between non-impulsive (Gaussian) and impulsive (non-Gaussian) α stable distributions.

Key Words: impulsive interference, alpha-stable distribution, parametric bootstrap, characteristic function, goodness-of-fit tests

1. INTRODUCTION

Statistical models that incorporate impulsive behaviour have found use in the analysis of atmospheric communication channels, underwater acoustic signals, radar systems, economic time series and biomedical signals [7]. The alpha stable (αS) distribution has been prominently used in many of these cases. This may be due to physical reasons – interference from spatially Poisson distributed scatterers is αS distributed – or merely due to its general form – it is a broad family of distributions.

The αS distribution is a four parameter distribution defined in terms of its characteristic function

$$\phi(t) = \exp\{j\delta t - \mid ct \mid^{\alpha} [1 - j\beta \mathrm{sgn}(t)\omega(t,\alpha)]\}$$

where

$$\omega(t,\alpha) = \begin{cases} \tan(\alpha\pi/2)\,, & \alpha \neq 1 \\ -(2/\pi)\log|t| & \alpha = 1 \end{cases}$$

and $\mathrm{sgn}(t)$ is the signum function [3]. Herein, α, $0 < \alpha \leq 2$, is the characteristic exponent, β, $-1 \leq \beta \leq 1$, the skewness parameter, c, $0 < c$, the scale parameter and δ, $-\infty < \delta < \infty$ the location parameter.

The characteristic exponent, α, is a measure of how impulsive the distribution is. The smaller α, the more impulsive is the αS distributed process, that is, the more outliers occur in an observed series. When $\alpha = 2$, its maximum value, the αS distribution is equivalent to the Gaussian distribution.

The degree of impulsiveness is an important feature in many applications. For example, in a radar application a detector designed for Gaussian interference may be used if $\alpha = 2$, rather than a more complicated detector for signals in αS interference.

To test impulsive behaviour we suggest a test for α. The procedures presented are applicable in principle to all four parameters, however we will focus on testing α only, and, more specifically, on the special case of testing for the Gaussian distribution ($\alpha = 2$) against non-Gaussian, stable distributions.

Throughout this paper we have used the parameter estimation procedure suggested by Koutrouvelis in [6]. By manipulation of the characteristic function of the αS distribution it was shown that estimators of the parameters could be found through a regression technique. The estimators are consistent and asymptotically unbiased. Tabulated values of the MSE (and a number of other statistical quantities) of the estimators were presented for some parameter settings. These quantities were derived through Monte Carlo simulations in [6]. Expressions for the asymptotic properties of the estimators are unavailable.

In the following section we discuss the hypothesis testing problem for impulsive behaviour using bootstrap methods. Following this, a test procedure is derived in section 3 that tests for the value of α. An alternative approach, using the characteristic function is presented in section 4. In section 5, simulation results are presented and conclusions are drawn in section 6.

2. HYPOTHESIS TESTING WITH THE BOOTSTRAP

The bootstrap is a simple automatic procedure that can take the place of analytic analysis. It can be used to estimate the sample distribution of statistics when standard methods cannot be applied. Observations are randomly resampled and the statistics re-computed – mimicking the process of repeating the experiment. When this is done a large number of times, the distribution of the re-computed values approximates the distribution of the statistic. The principle of hypothesis testing using the bootstrap is discussed in [4, 9].

Consider the αS family of distributions, indexed by the parameter vector $\mathbf{p} = [\alpha \ \beta \ c \ \delta]^T \in \mathcal{P}$, $F_{\{\mathbf{p}:\mathbf{p}\in\mathcal{P}\}}$. Further, define two disjoint subsets of this family $\mathcal{F}_0 = \{F_{\mathcal{P}_0}\}$ and $\mathcal{F}_1 = \{F_{\mathcal{P}_1}\}$. These two sets are indexed by parameter sets $\mathcal{P}_0$ and $\mathcal{P}_1$ which span the parameter space $\mathcal{P}$, that is, $\mathcal{P}_0 \cup \mathcal{P}_1 = \mathcal{P}$ and $\mathcal{P}_0 \cap \mathcal{P}_1 = \varnothing$

Let the observations $\mathbf{X} = X_1, X_2, \ldots, X_n$ be independent and identically distributed having distribution $F_{\mathbf{p}}$. We wish to test the hypothesis

$$\mathsf{H}\text{: } F_{\mathbf{p}} \in \mathcal{F}_0 \quad \text{versus} \quad \mathsf{K}\text{: } F_{\mathbf{p}} \in \mathcal{F}_1$$

or alternatively, defined in terms of the parameter space

$$\mathsf{H}\text{: } \mathbf{p} \in \mathcal{P}_0 \quad \text{versus} \quad \mathsf{K}\text{: } \mathbf{p} \in \mathcal{P}_1$$

If we can assume $\mathbf{X}$ is symmetrically distributed about the origin, that is, $\beta = 0$ and $\delta = 0$, and we assume unit scale, $c = 1$, the hypothesis becomes one of testing the parameter α

$$\mathsf{H}\text{: } \alpha = \alpha_0 \quad \text{versus} \quad \mathsf{K}\text{: } \alpha \neq \alpha_0$$

To test the hypothesis, we define a test statistic $T(\mathbf{X}) = T$ with distribution $K_{T,\mathbf{p}}(x) = P_\alpha\{T \leq x\}$ where $P_\mathbf{p}$ is the probability corresponding to $F_\mathbf{p}$. We can now find the bootstrap estimator of the critical value q, $\hat{q} = K^{-1}_{T,\alpha_0}(1-\gamma)$, where γ is the desired level of significance. The distribution of the test statistics of independent bootstrap resamples, T^*, approximates the distribution of T. Consequently, $\hat{q}$ can be approximated by an order statistic of T^*.

Under some assumptions on the distribution of T under H and K, it has been proven that the bootstrap test $T(\mathbf{X}) \underset{\mathsf{H}}{\overset{\mathsf{K}}{\gtrless}} \hat{q}$ is asymptotically correct and consistently uniform in $\mathbf{p}$ [8].

3. TESTING THE α PARAMETER

The test for α suggests the use of the test statistic

$$T_\alpha = \frac{\hat{\alpha} - \alpha_0}{\hat{\sigma}_{\hat{\alpha}}}$$

where $\hat{\alpha}$ is an estimate of α derived from the observations and $\hat{\sigma}_{\hat{\alpha}}$ is an estimate of its standard deviation. This quantity is an approximate pivot, meaning its distribution is *approximately* independent of any unknowns.

In Table 1, we present the proposed bootstrap based procedure for testing the α parameter. We also show the technique in block diagram form in Fig 1. Thick arrows indicate the bootstrap replications generated through the αS random number generator (RNG). Estimates of $\hat{\sigma}_{\hat{\alpha}}$, the variance of the estimate $\hat{\alpha}$ and $\hat{\sigma}^*_{\hat{\alpha}}$, the variance of the bootstrapped estimates $\hat{\alpha}^*$ are obtained using nested bootstrap stages. Details can be found in [11].

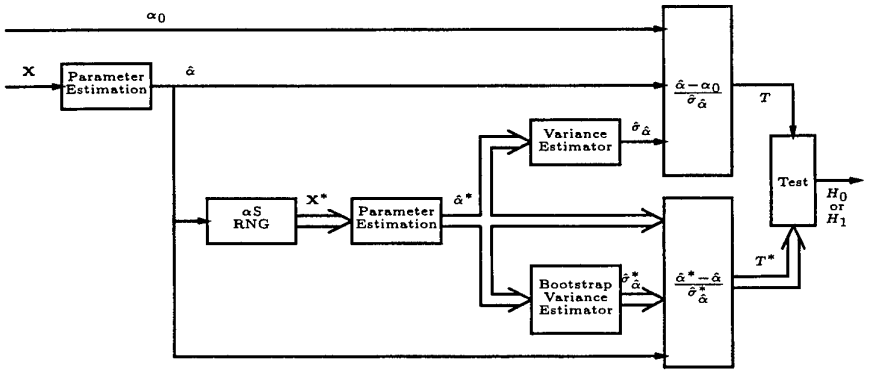


FIG. 1. Block diagram of the bootstrap test for α.

Step 1. *Parameter estimation.* Find the parameter estimate $\hat{\alpha}$ from the sample $\mathbf{X}$.

Step 2. *Parametric resampling.* Using a pseudo-random number generator, generate a random sample $\mathbf{X}^*$ of the same size as $\mathbf{X}$, from an αS distribution with parameter $\hat{\alpha}$.

Step 3. *Calculation of the bootstrap statistic.* From $\mathbf{X}^*$, calculate T^*

Step 4. *Repetition.* Repeat steps 2 and 3 many times to obtain a total of B bootstrap statistics $T_1^*, T_2^*, \ldots, T_B^*$ and the test statistic T.

Step 5. *Ranking.* Rank the collection $T_1^*, T_2^*, \ldots, T_B^*$ into increasing order to obtain $T_{(1)}^* \leq T_{(2)}^* \leq \cdots \leq T_{(B)}^*$.

Step 6. *Test.* A bootstrap test has then the following form: reject H if $T \leq T_{(C)}^*$, where the choice of C determines the level of significance of the test and is given by $C = \lceil \gamma(B+1) \rceil$, where γ is the nominal level of significance [4].

TABLE 1
The bootstrap principle for testing the hypothesis $\mathsf{H} : \alpha = \alpha_0$ against $\mathsf{K} : \alpha < \alpha_0$.

Importantly, the distribution of T_α is approximated by the distribution of

$$T^* = \frac{\hat{\alpha}^* - \hat{\alpha}}{\hat{\sigma}_{\hat{\alpha}}^*}$$

and not by the distribution of $\frac{\hat{\alpha}^* - \hat{\alpha}}{\hat{\sigma}_{\hat{\alpha}}}$ or of $\hat{\alpha}^* - \hat{\alpha}$. This has been shown to keep the actual level of significance closer to the nominal level [4, 11]. The bootstrap distribution of $T^* = \frac{\hat{\alpha}^* - \hat{\alpha}}{\hat{\sigma}_{\hat{\alpha}}^*}$ approximates the distribution of $T_\alpha = \frac{\hat{\alpha} - \alpha_0}{\hat{\sigma}_{\hat{\alpha}}}$ better under H than the distributions of $\hat{\alpha}^* - \hat{\alpha}$ approximates the distribution of $\hat{\alpha} - \alpha_0$ in terms of finding the scaling or dispersion of $\hat{T}$.

4. CHARACTERISTIC FUNCTION BASED TEST

Differences between characteristic functions have been used extensively to test for changes in distributions. This is especially so for Gaussianity testing, where the technique originated [2, 5]. However, it was frequently noted that one of the major advantages of characteristic function (cf) based goodness-of-fit testing was that it could be adapted to test for almost any distribution, as long as the cf is specified. This lead to its recent application to testing for the αS distribution, against all other distributions, in [1].

Here, a parametric form of the cf based tests is formed to accommodate the additional knowledge / assumption that we are operating within a defined family of cfs. A *parametric* estimate of the cf of the generating process is compared to the cf of the distribution under H, $\phi(t, \mathbf{p}_0)$, rather than a nonparametric estimate, such as the empirical characteristic function. This parametric estimate is found by using the estimated parameter values and the known form of the cf of αS distributions and is denoted $\phi(t, \hat{\mathbf{p}})$.

In [10] it was found that the peak absolute difference between two cfs provided a good measure of the distance between the two distributions. Drawing on this, we

define our test statistic to be

$$T_\phi = \max_t |\phi(t, \hat{\mathbf{p}}) - \phi(t, \mathbf{p}_0)|$$

The parameter vector, $\mathbf{p} = [\alpha\ \beta\ c\ \delta]^T$, has all four parameters of the distribution. This highlights an advantage of this test statistic over the statistic defined in section 3, namely, it incorporates all parameters into the test. This allows its use in a broader range of problems.

The distribution of T_ϕ is complicated and unknown, we again draw on the parametric bootstrap to determine critical values. We will approximate this distribution by

$$T^* = \max_t |\phi(t, \hat{\mathbf{p}}^*) - \phi(t, \hat{\mathbf{p}})| .$$

The procedure is presented in Fig 2.

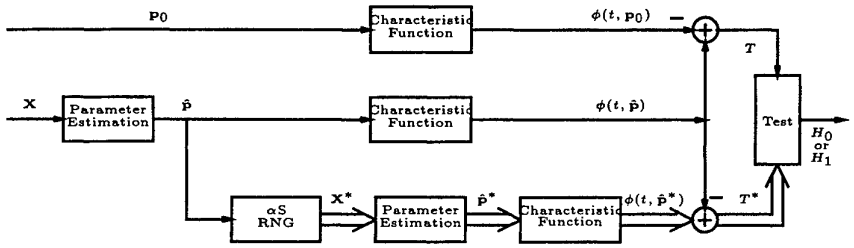


FIG. 2. Block diagram of the characteristic function based test.

5. SIMULATION RESULTS AND DISCUSSION

A simulation study was undertaken to determine the performance of the two tests. Here we consider $\alpha_0 = 2$, that is, we are testing the hypothesis that the observations are Gaussian distributed against non-Gaussian αS distributed. This is probably the most important case to consider as it tests if the observations have bounded or infinite variance.

Rejection rates for a number of values of α are presented in Tables 2 and 3 for observation sample sizes of 200 and 400 respectively. The nominal significance level was set at 10%, the number of bootstrap replications was 300 and 25 replications were used for the bootstrap variance estimator in the evaluation of the test statistics.

TABLE 2

Rejection rates (in %) for sequence lengths of 200 based on 300 replications.

α	1.7	1.8	1.9	1.95	2
T_α	97.0	88.0	56.3	27.0	7.7
T_ϕ	93.3	73.3	37.7	11.0	2.3

Inspection of the results show that directly testing α through the T_α statistic yielded higher rejection rates than the cf based technique, T_ϕ. However, it should be remembered that the cf based test is simpler to adapt to the case were more

TABLE 3
Rejection rates (in %) for sequence lengths of 400 based on 300 replications.

α	1.7	1.8	1.9	1.95	2
T_α	100	99.3	76.3	50.3	6.3
T_ϕ	100	95.3	61.3	29.0	0.3

than one parameter is to be tested. The performance of the cf based test may be improved through the use of a pivotal statistic based on $\phi(t, \hat{\mathbf{p}})$.

As expected, rejection rates decrease as α approaches $\alpha_0 = 2$ and when fewer observations are available. The achieved level is well below the nominal level, especially for the un-standardised T_ϕ.

6. CONCLUSIONS

Two tests have been presented for testing the parameter values of an αS distribution. The bootstrap procedures implemented have been shown to allow the appropriate setting of critical values for the test that have maintained the nominal level of significance. Although testing the α parameter directly yielded a more powerful test, it is to be noted that the characteristic function based procedure has a high degree of flexibility. Simulation results show that both reject the non-Gaussian alternatives tested with rates varying depending on the degree of impulsive behaviour and the number of observations available.

REFERENCES

1. C. L. Brown and S. Saliu. Testing of alpha-stable distributions with the characteristic function. In *Proceedings of the IEEE Signal Processing Workshop on Higher-Order Statistics, HOS'99*, pages 224–227, Ceasarea, Israel, June 1999.
2. T. W. Epps. Testing that a stationary time series is Gaussian. *The Annals of Statistics*, 15:1683–1698, 1987.
3. P. Hall. A comedy of errors : The canonical form for a stable characteristic function. *Bulletin of the London Mathematical Society*, (13):23–7, May 1980.
4. P. Hall. *The Bootstrap and Edgeworth Expansion*. Springer-Verlag, New York, 1992.
5. I. A. Koutrouvelis. A goodness-of-fit test of simple hypotheses based on the empirical characteristic function. *Biometrika*, 67(1):238–40, 1980.
6. I. A. Koutrouvelis. Regression-type estimation of parameters of stable laws. *Journal of the American Statistical Association*, 75(372):918–928, December 1980.
7. C. L. Nikias and M. Shao. *Signal Processing with Alpha-Stable Distributions and Applications*. John Wiley & Sons, 1995.
8. J. Shao and D. Tu. *The Jackknife and Bootstrap*. Springer, 1995.
9. A. M. Zoubir. Signal detection using the bootstrap. *Signal Processing*, (in press).
10. A. M. Zoubir and M. J. Arnold. Testing Gaussianity with the characteristic function. *Signal Processing*, 53:245–255, September 1996.
11. A. M. Zoubir and B. Boashash. The bootstrap and its application in signal processing. *IEEE Signal Processing Magazine*, 15(1):56–76, January 1998.

INDEX

DASP99 Sessions

Monday August 23

9:00am-Lecture Session M1 *(Chair-Bill Moran, Flinders University of South Australia)*

9:00-9:40 Doug Gray, CSSIP, SA, "Applications of Probabilistic Least Squares Tracking to Pulse Train Deinterleaving"

9:40-10:20 John Benedetto, "Periodicity Detection and Multidimensional Irregular Sampling Algorithms"

10:40am-Lecture Session M2 *(Chair-Rick Johnson, Cornell University)*

10:40–11:20 Moeness Amin, Villanova University, Philadelphia PA, "Direction Finding Based on Spatial Time-Frequency Distribution Matrices"

11:20–12:00 Ken Harrison, Murdoch University, Perth WA, "Quantum Algorithms"

1:00pm Poster Session *(Chair – Jon Sjogren, Air Force Office of Scientific Research)*

D. Cochran	A Bayesian Risk Approach To Multi-Mode Detection
J. Praschifka	Investigation Of Target Detection In HF Skywave Radar Using Thomson?S Multiple Window Method
K. Kastella, S. Musick	Bias Estimation In An Association-Free Nonlinear Filter
C. Hallam	A Multicriteria Shortest Path Algorithm For Plotting A Safe Path Through A Field Of Sensors
M. Zoltowski	Smoothing Of Space-Time Power Minimisation Based Preprocessor For GPS
P. Sherman	On The Statistical Nature Of Real Sinusoids Associated With Rotating Machinery
L. White	Convergence Analysis Of An HMM Based Multiuser Detector For CDMA Channels With ISI

8:00pm-Lecture Session M3 *(Chair-Alan Lindsey, Air Force Research Laboratory / IFGC)*

8:00-8:40 Stella Batalama SUNY Buffalo, "Robust Adaptive Recovery of Spread-Spectrum Signals with Short Data Records"

8:40-9:20 John Treichler, AST, Sunnyvale CA, "The Future of Telephone Networks"

Tuesday August 24

8:00am - Lecture Session T1 *(Chair-Lang White, University of Adelaide)*

8:00-8:40 Yuri Abramovich, CSSIP, SA, "Nonuniform Linear Antenna Array Design and Signal Processing for DOA Estimation of Gaussian Sources"

8:40-9:20 J. McClellan, "Quadtree Focusing for UWB SAR"

9:40am-Lecture Session T2 (*Chair-Alan Lindsey, Air Force Research Laboratory / IFGC*)

9:40-10:20 Vikram Krishnamurthy, University of Melbourne, "Finite Dimensional Algorithm for Optimal Scheduling of Hidden Markov Models Sensors"

10:20-11:00 Dale A. Lambert, DSTO, Salisbury SA, "Advisers With Attitude For Situational Awareness"

11:00-11:40 A. Nehorai, "Source Localization with Distributed Electromagnetic Component Sensor Array Processing" (Presenter-A. Nehorai)

1:00pm – 7:00 Small Sphere Impulsive Targeting – Group Experiments

8:00pm Poster Sessions *(Chair-Gordon Frazer, DSTO)*

L Sciacca	Target Tracking With Dissimilar Sensors Using Set-Based Estimation
R. Mersereau, M. Smith	A Hidden Markov Model Classifier For Anti-Tank Guided Missiles
N. Redding, J. Schroeder	Comparison Of Selected Features For Target Detection In Synthetic Aperture Radar Imagery
S. Elton	IDEA-A Signals Analysis Package For ELINT Processing And Extraction
B. Bullard	Rotary-Wing Target Radar Signal Processing And Signature Simulator
M. Rangaswamy, J. Michels	A Parametric Adaptive Matched Filter For Airborne Radar Applications
R. Bonneau	A Model For Shape And Texture Content-Based Image Compression
D. Dudgeon	"ATR Performance Modeling And Estimation"

Wednesday August 25

8:00am-Lecture Session W1 (*Chair-Mark Smith, Georgia Tech*)

8:00-8:40 Don Tufts, University of Rhode Island, "Real CFAR: Nonparametric, Data-Adaptive Thresholding As an Example of Tolerance-Region Signal Processing"

8:45-9:25 V. Velten, "Automatic Target Recognition (ATR) Via Invariance Theory For IR And RF Sensors"

9:30-10:10 Gordon Frazer, "Wigner-Ville Analysis Of HF Radar Measurements Of Surrogate Theatre Ballistic Missile"

10:25am - Poster Session *(Chair – John Treichler, Applied Signal Technology, Inc.)*

A. Lindsey	Performance Of Simplex Signaling In Circular Trellis-Coded Modulation
J. Hefferan	Object Enhancement In Time-Frequency Scans Of Communications Environments
B. Moran	On The Use Of The PONS Sequences For Peak-To-Mean Power Control In OFDM
J. S. Goldstein	Colored Noise Matched Filtering With Unknown Covariance
M. Greening	Array Element Localisation Using Simulated Annealing
O. P. Kenny	A Speech Segmentation Algorithm With Application To Speaker Identification
A. Zoubir	Testing For Impulsive Interference : A Bootstrap Approach
J. Michels, B. Himed	Space Time Adaptive Processing (STAP) In Compound-Gaussian Airborne Radar Clutter

1:00pm-Lecture Session W2 (*Chair-Jim Schroeder, CSSIP, SA*)

1:00-1:40 J. Bucklew, "Nonlinear Equalization Using SVMs"

1:40-2:20 Linda Davis, University of Adelaide, SA, "Turbo Channel Estimation and Equalization for Mobile Data Communications"

2:20-3:00 Peter Shoubridge, DSTO, "Approaches to Measuring Network Change"

Thursday August 26 *(Chair, Mark Smith, Georgia Institute of Technology)*

"Signal Processing in Defense Systems: The Future"

8:00-8:30 Dennis Healy, DARPA
8:30-9:00 Richard Linderman, AFRL
9:00-9:30 Marian Viola, DSTO

Panel Session, Q&A, Focus Groups

Abramovich, Yuri	T1	
Batalama, Stella	M3	
Benedetto, J.	M1	
Bonneau, Capt. Robert	TP-7	
Bucklew, Jim	W2	
Bullard, B.	TP-5	
Cochran, D.	MP-1	
Davis, Linda	W2	
Dudgeon, Dan	TP-8	
Frazer, Gordon & Spouse	W1	
Gaudreault, Michele	None	
Goldstein, Scott & Spouse	WP-4	
Gray, Doug & Spouse	M1	
Greening, Mike	WP-5	
Hallam, ??	MP-4	no abstract
Harrison, Ken & Ward, Jo	M2	no abstract
Healy, Dennis	TH1	
Hefferan, John	WP-2	
Himed, Braham	WP-8 (w/ Michels)	
Johnson, Rick	None	
Kastella, Keith	MP-3 (w/ Musick)	
Kenny, Owen	WP-6	
Kitchen, John	None	
Krishnamurthy, Vikram	T2	
Lambert, Dale	T2	
Lang, White	MP-7	
Leonard, Jim	None	
Lever, Ken & Spouse	None	
Linderman, Richard	TH1	
Lindsey, Alan	WP-1	
McClellan, Jim	T1	
Mersereau, Russ	TP-2 (w/ M. Smith)	
Michels, Jim, Spouse & Dtr	WP-8 (w/ Himed), TP-6 (w/Rangaswamy)	
Moeness, Amin	M2	
Moran, Bill	WP-3	
Munson, Dave	None	
Musick, Stan	MP-3 (w/ Kastella)	
Nehorai, A.	T2	
Newsam, Garry	None	
Praschifka, Justin	MP-2	
Rangaswamy, Muralidhar	TP-6 (w/ Michels)	
Redding, Nick	TP-3 (w/ Schroeder)	
Roberts, Bill	None	
Scharf, Louis	None	
Schroeder, Jim	TP-3 (w/ Redding)	
Sciacca, Len	TP-1	
Sherman, Peter	MP-6	
Shoubridge, Peter	W2	
Sjogren, Jon	None	
Smith, Mark	TP-2 (w/ Mersereau)	
Suddarth, Steve	None	
Treichler, J,	M3	
Tufts, Don	W1	
Velton, Vince	W1	
Viola, Marian	TH1	
Waller, Jeremy	None	
Zoltowoski, Mike	MP-5	
Zoubir, Abdelhak	WP-7	

Legend:

M1	9:00am	Monday Lecture Session #1
M2	10:40am	Monday Lecture Session #2
M3	8:00pm	Monday Lecture Session #3
MP	1:00pm	Monday Poster Session
T1	8:00am	Tuesday Lecture Session #1
T2	9:40am	Tuesday Lecture Session #2
TP	8:00pm	Tuesday Poster Session
W1	8:00am	Wednesday Lecture Session #1
WP	10:25am	Wednesday Poster Session
W2	1:00pm	Wednesday Lecture Session #2
TH1	8:00am	Thursday Lecture Session #1

NOTES:
missing TP-4

Schedule

Sunday, August 22

7:00 pm	10:00	Welcome Reception Beverages, finger foods.

Monday, August 23

7:30 am	9:00	Breakfast / Registration
9:00	10:20	Lecture Session M1 (2)
10:20	10:40	Break
10:40	12:00	Lecture Session M2 (2)
12:00	1:00	Lunch / Posters
1:00	3:00	Monday Posters
3:00	6:30	Free Time
6:30	8:00	Dinner
8:00	9:20	Lecture Session M3 (2)

Tuesday, August 24

7:00 am	8:00	Breakfast
8:00	9:25	Lecture Session T1 (2)
9:25	9:40	Break
9:40	11:50	Lecture Session T2 (3)
11:50	1:00	Lunch
1:00	7:00	Free Time / DASP99 Small Sphere Targeting Experiments
7:00	8:00	Dinner
8:00	10:00	Tuesday Posters

Wednesday, August 25

7:00 am	8:00	Breakfast
8:00	10:10	Lecture Session W1 (3)
10:10	10:25	Break
10:25	12:00	Wednesday Posters
12:00	1:00	Lunch
1:00	3:10	Lecture Session W2 (3)
3:10	6:30	Free Time
6:30	8:00	Dinner / Posters

Thursday, August 26

7:00 am	8:00	Breakfast
8:00	9:30	"Signal Processing in Military Systems: The Future" *Linderman, Viola, Healy*
9:30	9:40	Break
9:40	11:30	Panel Session, Q&A, Focus Groups
11:30	12:30	Lunch
1:00		Australian Guests Depart for Chicago
3:00	4:30	Wendella Boat Tour
5:00	7:00	Navy Pier: Ferris Wheel, Shops
7:30	10:00	Chicago-Style Dinner at Pizzeria DUE
10:00		Depart for LaSalle