

UNCLASSIFIED

AD NUMBER
ADB226867
NEW LIMITATION CHANGE
TO Approved for public release, distribution unlimited
FROM Distribution authorized to U.S. Gov't. agencies only; Specific Authority; 28 Jul 97. Other requests shall be referred to RL/OCSM, 26 Electronic Parkway, Rome, NY 13441-4514.
AUTHORITY
AFRL/IFOIP ltr, 15 Jun 2004

THIS PAGE IS UNCLASSIFIED

RL-TR-97-5, Vol I (of two)
Final Technical Report
May 1997



MULTICHANNEL SYSTEM IDENTIFICATION AND DETECTION USING OUTPUT DATA TECHNIQUES

Scientific Studies Corp.

Jaime R. Roman and Dennis W. Davis

28 JUL 1997

*DISTRIBUTION AUTHORIZED TO U.S. GOVERNMENT AGENCIES ONLY; SPECIFIC
AUTHORITY: ~~DDP-FAR SUPPLEMENT 252.227-7013, ALT II~~. OTHER REQUESTS FOR THIS
DOCUMENT SHALL BE REFERRED TO RL(OCSM), ROME, NY. 13441-4514*

GOVERNMENT PURPOSE LICENSE RIGHTS (SBIR PROGRAM)

Contract No: F30602-93-C-0193

Contractor: Scientific Studies Corp.

For a period of four(4) years after delivery and acceptance of the last deliverable item under the above contract, this technical data shall be subject to the restrictions contained in the definition of "Limited Rights" in DFARS clause at 252.227-7013. After the four-year period, the data shall be subject to the restrictions contained in the definition of "Government Purpose License Rights" in DFARS clause at 252.227-7013. The Government assumes no liability for unauthorized use or disclosure by others. This legend, together with the indications of the portions of the data which are subject to such limitations, shall be included on any reproduction hereof which contains any portions subject to such limitations and shall be honored only as long as the data continues to meet the definition on Government purpose license rights.

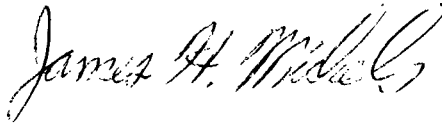
19970722 093

DTIC QUALITY INSPECTED 4

Rome Laboratory
Air Force Materiel Command
Rome, New York

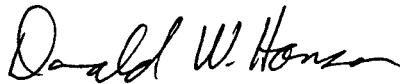
RL-TR-97-5, Vol I has been reviewed and is approved for publication.

APPROVED:



JAMES H. MICHELS
Project Engineer

FOR THE COMMANDER:



DONALD W. HANSON, Director
Surveillance & Photonics Directorate

DESTRUCTION NOTICE - For classified documents, follow the procedures in DOD 5200.22M. Industrial Security Manual or DOD 5200.1-R, Information Security Program Regulation. For unclassified limited documents, destroy by any method that will prevent disclosure of contents or reconstruction of the document.

If your address has changed or if you wish to be removed from the Rome Laboratory mailing list, or if the addressee is no longer employed by your organization, please notify Rome Laboratory/OCSM, 26 Electronic Pky, Rome, NY 13441-4514. This will assist us in maintaining a current mailing list.

Do not return copies of this report unless contractual obligations or notices on a specific document require that it be returned.

REPORT DOCUMENTATION PAGE			Form Approved OMB No. 0704-0188	
Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188), Washington, DC 20503.				
1. AGENCY USE ONLY (Leave blank)		2. REPORT DATE May 97		3. REPORT TYPE AND DATES COVERED FINAL Jul 93 - Oct 96
4. TITLE AND SUBTITLE MULTICHANNEL SYSTEM IDENTIFICATION AND DETECTION USING OUTPUT DATA TECHNIQUES, VOL I (OF TWO)			5. FUNDING NUMBERS C - F30602-93-C-0193 PE - 65502F PR - 3005 TA - RC WU - 92	
6. AUTHOR(S) Jaime R. Roman Dennis W. Davis				
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Scientific Studies Corp. 2250 Quail Ridge Palm Beach Gardens, FL 33418			8. PERFORMING ORGANIZATION REPORT NUMBER N/A	
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES) Rome Laboratory/OCSM 26 Electronic Parkway Rome NY 13441-4514			10. SPONSORING/MONITORING AGENCY REPORT NUMBER RL-TR-97-5, Vol I	
11. SUPPLEMENTARY NOTES Rome Laboratory Project Engineer: James H. Michels, OCSM, 315-330-4432				
12a. DISTRIBUTION AVAILABILITY STATEMENT USGO Agencies Only; Specific Authority; DOD FAR Sup 252-227-7013, Alt II. Other Requests: RL/OCSM, 26 Electronic Parkway, Rome, NY 13441-4514.			12b. DISTRIBUTION CODE 128 JUL 1997	
13. ABSTRACT (Maximum 200 words) The multichannel innovations-based detection algorithm (MIBDA) methodology formulated in Phase I was developed in full and evaluated in the context of two applications: (a) airborne surveillance phased array radar systems, and (b) computer-based electrocardiogram (ECG) diagnosis. A software simulation was developed to generate simulated multichannel phased array radar data. This software was exercised to represent ground clutter as the output of a state variable model (SVM), and to assess the performance of such models in the MIBDA. The MIBDA was modified and extended to the automated diagnosis of ECG traces. MIBDA performance for ECG diagnosis was assessed using the Common Standards for Quantitative Electrocardiography (CSE) database. Analysis and simulation results indicate that the MIBDA offers a valid alternative to conventional systems in both of the applications considered. This suggests also that the MIBDA can be configured to apply to a wide range of problem areas. Volume I presents the MIBDA methodology, SVM identification algorithms, MIBDA design and performance evaluation issues, and simulation results for both applications. Volume II presents the analytic and software model for airborne surveillance phased array radar systems.				
14. SUBJECT TERMS Multichannel Detection, Airborne Surveillance, State-Space Model, Innovations Processes, System Identification, ECG Diagnostics			15. NUMBER OF PAGES 238	
			16. PRICE CODE	
17. SECURITY CLASSIFICATION OF REPORT UNCLASSIFIED	18. SECURITY CLASSIFICATION OF THIS PAGE UNCLASSIFIED	19. SECURITY CLASSIFICATION OF ABSTRACT UNCLASSIFIED	20. LIMITATION OF ABSTRACT SAR	

TABLE OF CONTENTS

LIST OF FIGURES	v
LIST OF TABLES	ix
1.0 INTRODUCTION	1
1.1 Notation	6
1.2 Report Overview	7
2.0 STATE SPACE MODEL-BASED MULTICHANNEL DETECTION	9
2.1 Multichannel Detection	10
2.2 State Space Model	14
2.3 Backward State Space Model.....	21
2.4 Stochastic Block Correlation Matrices.....	23
2.5 Innovations Representation	25
3.0 MULTICHANNEL SYSTEM IDENTIFICATION	32
3.1 Output Data-Based Algorithm	32
3.2 Model Order Determination	47
4.0 INNOVATIONS SEQUENCE GENERATION	51
5.0 LIKELIHOOD RATIO DETECTION	57
6.0 AIRBORNE SURVEILLANCE PHASED ARRAY RADAR APPLICATION	60
6.1 Conventional Space-Time Processing	62
6.2 Model-Based Space-Time Processing	66
6.3 Space-Time Process Modeling and Filtering Analyses	67
7.0 ECG DIAGNOSTICS APPLICATION	86
7.1 Multichannel Electrocardiography	86
7.2 CSE Database	93

7.3	Modeling and Discrimination Using CSE Data	94
7.4	ECG Diagnosis Methodology	106
7.4.1	ECG trace pre-processing	109
7.4.2	ECG residual statistics and decision criteria	112
7.4.3	ECG diagnosis methodology validation procedure	116
7.4.4	ECG diagnosis methodology validation results	125
8.0	CONCLUSIONS AND RECOMMENDATIONS	134
APPENDIX A. PARTIAL QUOTIENT SINGULAR VALUE DECOMPOSITION		138
A.1	Quotient Singular Value Decomposition	139
A.2	Partial QSVD Algorithm	142
APPENDIX B. COMBINED SYSTEM MATRIX ESTIMATION FORMULA		145
APPENDIX C. SPATIAL FILTERING AND THE LDU DECOMPOSITION		149
APPENDIX D. HYPOTHESIS FILTER DESIGN		154
D.1	Zero-Mean Test	156
D.2	Power Test	160
D.3	Auto-Correlation Sequence Whiteness Test	165
D.3.1	Circular ACS estimate for small values of N	170
D.3.2	Circular ACS estimate for large values of N	171
APPENDIX E. AUTO-CORRELATION SEQUENCE ESTIMATORS		175
E.1	Circular ACS Estimator	175
E.2	Biased ACS Estimator	181
E.3	Unbiased ACS Estimator	183
APPENDIX F. RANDOM VARIABLE TRANSFORMATIONS		186
F.1	Product of Two Independent, Gaussian-Distributed Random Variables	187

F.2	Product of Two Independent, Special K-Distributed Random Variables	191
F.3	Difference of Two Independent, Special K-Distributed Random Variables	193
F.4	Sum of N Independent, Laplace-Distributed Random Variables	195
F.5	ACS Estimator Lags	197
APPENDIX G. MULTIPLE HYPOTHESES TESTING FOR ECG DIAGNOSIS		201
REFERENCES		206

LIST OF FIGURES

1-1	Radar array with J subarrays or individual elements	3
2-1	Innovations-based multichannel detector with on-line parameter identification	13
4-1	Two-function hypothesis filter	51
4-2	Whitening filter block diagram	52
6-1	Multichannel signal in a coherent surveillance radar array system	61
6-2	Joint-domain configuration block diagram	64
6-3	Space-time configuration block diagram	65
6-4	Time-space configuration block diagram	65
6-5	Multichannel model-based detection configuration with off- line parameter identification for space-time processing	66
6-6	Logarithm of the normalized channel output power spectrum (Blackman-Tukey spectrum estimate; true ACS case)	75
6-7	Logarithm of the innovations representation model power spectrum (true ACS case)	75
6-8	Canonical correlations of the channel output process (true ACS case)	76
6-9	Normalized mutual information for model order selection (true ACS case)	76
6-10	Map of the multivariable poles and zeros of the tenth-order state-space model (true ACS case)	77

6-11	Top view of the logarithm of the innovations representation model power spectrum (true ACS case)	77
6-12	Spatial-frequency axis projection of the logarithm of the state-space model power spectrum (true ACS case)	78
6-13	Doppler-frequency axis projection of the logarithm of the state-space model power spectrum (true ACS case)	78
6-14	Logarithm of the whitening filter power spectrum (true ACS case)	79
6-15	Top view of the logarithm of the whitening filter power spectrum (true ACS case)	79
6-16	Spatial-frequency axis projection of the logarithm of the whitening filter power spectrum (true ACS case)	80
6-17	Doppler-frequency axis projection of the logarithm of the whitening filter power spectrum (true ACS case)	80
6-18	Logarithm of the channel output power spectrum (modified, averaged periodogram; biased, time-averaged ACS case)	81
6-19	Logarithm of the innovations representation model power spectrum (biased, time-averaged ACS case)	81
6-20	Canonical correlations of the channel output process (biased, time-averaged ACS case)	82
6-21	Normalized mutual information for model order selection (biased, time-averaged ACS case)	82
6-22	Map of the multivariable poles and zeros of the tenth-order state-space model (biased, time-averaged ACS case)	83
6-23	Logarithm of the residual process power spectrum (periodogram; biased, time-averaged ACS case)	83

6-24	Real part of the normalized channel 0 true and model ACS (biased, time-averaged ACS case)	84
6-25	Imaginary part of the normalized channel 0 true and model ACS (biased, time-averaged ACS case)	84
6-26	Logarithm of the channel 4 residual power spectrum (averaged periodogram; biased, time-averaged ACS case with 0 dB SNR target)	85
7-1	Cardiac conduction system of the human heart	88
7-2	Base-to-apex lead placement configuration	90
7-3	Single-cycle ECG trace for the base-to-apex lead placement configuration	90
7-4	Selected cardiac cycles of the ECG traces used in the modeling and discrimination analysis	95
7-5	True and model spectra of the lead V5 normal ECG (single- channel model)	100
7-6	True and model spectra of the lead V5 abnormal ECG (single- channel model)	100
7-7	Covariance sequence of residual for normal whitening filter applied to normal ECG signal (single-channel model)	101
7-8	Covariance sequence of residual for normal whitening filter applied to abnormal ECG signal (single-channel model)	101
7-9	Covariance sequence of residual for abnormal whitening filter applied to abnormal ECG signal (single-channel model)	102
7-10	Covariance sequence of residual for abnormal whitening filter applied to normal ECG signal (single-channel model)	102

7-11	True and model spectra of the lead V5 normal ECG (multichannel model)	103
7-12	True and model spectra of the lead V5 abnormal ECG (multichannel model)	103
7-13	Covariance sequence of residual for normal whitening filter applied to normal ECG signal (multichannel model)	104
7-14	Covariance sequence of residual for normal whitening filter applied to abnormal ECG signal (multichannel model)	104
7-15	Covariance sequence of residual for abnormal whitening filter applied to abnormal ECG signal (multichannel model)	105
7-16	Model-based, multi-lead ECG diagnosis architecture	107
7-17	ECG trace pre-processing block diagram	110
7-18	ECG two-level discrimination tree for the three-condition ECG trace diagnostic architecture	119
7-19	Generic confusion matrix for a two-level, six-hypotheses discrimination tree	121
G-1	Multiple hypotheses test block diagram (minimum probability of error criterion with equal prior probabilities)	205

LIST OF TABLES

3-1	QSVD matrix factors for the three factorizations	42
6-1	Data vector and correlation matrix definition for the three conventional space-time processing configurations	62
6-2	Scenario, system, and simulation parameters for baseline simulation analyses	69
7-1	ECG trace pre-processing procedure for methodology validation analyses	111
7-2	ECG diagnosis methodology validation approach summary	117
7-3	Conditions for simulation-based methodology validation	127
7-4	Statistical measures for the design set time- and ensemble-averaged residual ACS	132
7-5	Statistical measures for the design set time-averaged residual ACS	132
7-6	Statistical measures for the testing set time-averaged residual ACS	133
7-7	ECG diagnosis methodology evaluation confusion matrix	133
F-1	Values of scaled SL PDF parameters N and L for cases of interest	200

1.0 INTRODUCTION

In multichannel identification problems the outputs of multiple channels (or sensors) are available, and it is desired to identify the parameters of an analytical model to represent the phenomena being observed via the channel outputs. Similarly, in multichannel detection problems the outputs of multiple channels are available, and it is desired to determine the presence (or absence) of a desired signal component in the channel data. In the combined problem of multichannel identification and detection a model is estimated for the phenomena being observed via the channel outputs, and the identified model is used to facilitate the detection of a desired signal in the channel output data. Multichannel identification and detection is thus referred to also as model-based multichannel detection. In all of these problems the channel data is available simultaneously over many channels of the same type, or over many distinct channels (each channel corresponding to a different sensor type).

This document is Volume I of a two-volume Final Technical Report which summarizes the work carried out in Phase II of this program. Specifically, this volume addresses the development of state space algorithms and methodologies for model-based multichannel detection in the context of airborne surveillance phased array radar systems and electrocardiogram (ECG) diagnostics applications. Volume II (Román and Davis, 1996) presents an analytic and software model for the multichannel output waveform in an airborne surveillance phased array radar system. In such systems the channels correspond to separate antenna apertures (or elements of a single aperture array). The desired signal may or may not be present in the channel output data at any given time. The data in each channel generally includes jamming noise (spatially-localized broadband interference), receiver noise, and "clutter" (narrowband interference). In general, signal-to-

clutter ratio (SCR) and signal-to-interference ratio (SIR) values are low. And signal-to-noise ratio (SNR) values are often low also. Model-based detection methods must discriminate between the condition of target embedded in clutter and noise, and the condition of clutter and noise only.

An ECG is a recording of electrical activity of the heart as manifested on the surface of the body. A standard digital ECG recorder detects this electrical activity at multiple discrete locations on the body surface, and converts the sensed signals into channels referred to as leads. The ECG thus provides information about the condition of the heart for a large number of abnormalities. In this application the objective is to discriminate between normal and abnormal ECGs, and to classify the abnormalities into the various detectable conditions.

In both of these applications the data is collected over time at a discrete number of locations. Thus, both applications can be categorized as space/time processing problems. Emphasis was placed on surveillance radar array systems since that is the application of main interest at Rome Laboratory (RL). The ECG diagnostics problem is of interest to RL (and to the U. S. Air Force) as demonstration of dual use for the technology developed in this Small Business Innovation Research (SBIR) program.

Figure 1-1 presents a multichannel system block diagram for a surveillance radar array consisting of multiple subarrays or array elements. The output of each subarray (or each individual array element) is a complex-valued, scalar, digital sequence, denoted as $\{x_i(n)\}$. The collection of the J scalar sequences is arranged into a J -dimensional vector, $\{\underline{x}(n)\}$, which is input to a processor (not shown in the figure). A digital, multi-lead ECG recorder has an analogous set of elements: a detector, a receiver, an analog-to-digital converter, and a pre-processor. The ECG lead information

is then fed to a processor also. For both applications, a multichannel version of such a processor was the focus of Phase II reported herein.

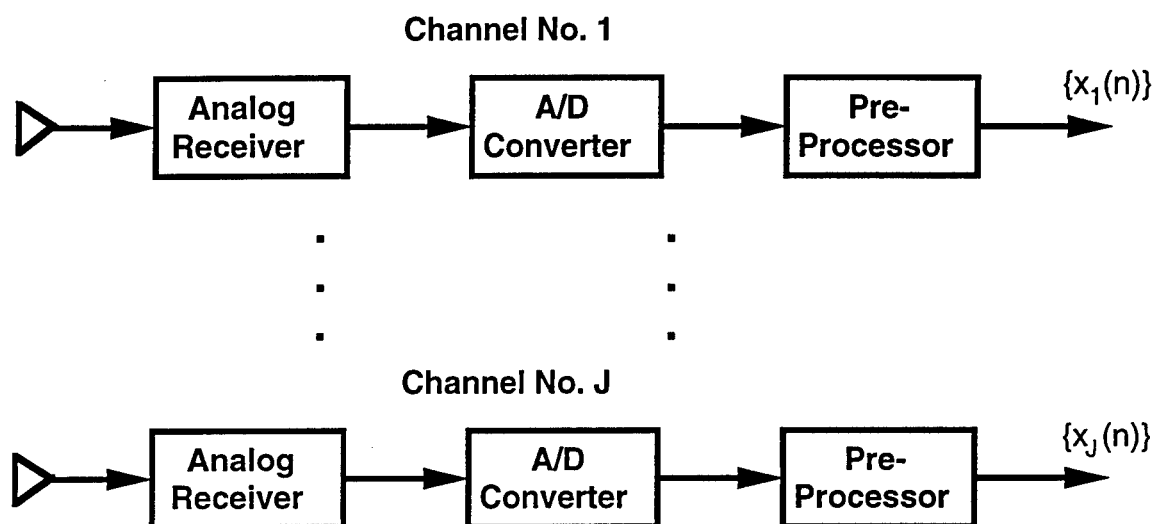


Figure 1-1. Radar array with J subarrays or individual elements.

In Phase I the multivariate (multiple input, multiple output) state space model class was adopted to represent the multichannel radar data, and new system identification techniques were applied to estimate the model parameters. Phase II continued the work along the same lines based on the success obtained in Phase I. The modeling of the complex-valued pre-processed radar signals for multichannel detection using the state space model class is one of the contributions of this work. State space models have been used in the context of target tracking (where the detected radar signal is processed further to estimate a trajectory) and for the determination of weights in antenna array sidelobe canceling and related problems, but not for multichannel detection. Model-based detection has been carried out using the more-restricted time series models (Michels, 1991; Metford and Haykin, 1985), which are included within the class of state space models and can be represented as such.

The methodology developed in Phase I was based on the recently-published algorithm developed by Van Overschee and De Moor (1993), which has several unique features. Foremost among these, the algorithm operates on output data directly to generate estimates of the parameters of a state space model (without computing output correlation matrices). This feature of the algorithm results in reduced dynamic range requirements in comparison with state space algorithms that operate on correlation matrices. The algorithm belongs to the class referred to as subspace methods because the fundamental operation of the algorithm is to decompose the vector space spanned by the channel output data into signal and noise subspaces. Implementation of this fundamental operation is carried out using the QR decomposition and the quotient singular value decomposition (QSVD) for matrix pairs. The QSVD, in turn, is based on the singular value decomposition (SVD). Efficient and stable software routines are available for the QR decomposition and the SVD (Dongarra et al., 1979).

Two other state space model identification algorithms were considered also in Phase II. Namely, the canonical correlations algorithm based on the work of Akaike (1974, 1975) and Desai et al. (1984), and the unweighted principal components algorithm proposed by Arun and Kung (1990). Both of these algorithms estimate the state space model parameters using the output correlation matrix sequence.

An important distinction in the context of radar system applications is that the vector random processes which represent the channel data are complex-valued processes in most cases. Most time series techniques and models have been formulated for complex as well as real processes. The same, however, cannot be said about state-space techniques; state-space methods and results

available in the literature have been defined almost exclusively for the case of real-valued processes, including all three algorithms considered in Phase II. The Van Overschee-De Moor algorithm was extended to the case of complex-valued processes in Phase I, and the canonical correlations algorithm was extended by Scientific Studies Corporation (SSC) to handle complex-valued processes in a program that ran in parallel with Phase I (Román and Davis, 1993b). Extending the Arun-Kung algorithm to handle complex-valued processes was carried out in Phase II.

A new algorithm for implementation of the QSVD was developed in Phase II. This algorithm simplifies the bookkeeping associated with the singular value pairs, and is more accurate and efficient than the alternatives (Van Overschee and De Moor, 1993; Paige and Saunders, 1981).

A hardware-based processor development system (PDS) was configured and integrated to serve as a testbed for the design and development of detection and identification methodologies and algorithms. The PDS consists of a Sun Microsystems' SPARCstation 10 host and a SKY Computers' SKYstation II accelerator, with FORTRAN 77 and MATLAB software (MATLAB runs only on the SPARCstation).

Two software packages were generated as part of this program to validate the methodology and the algorithms, and to carry out simulation-based analyses. One software package is programmed in FORTRAN 77, and the other is programmed in MATLAB. The FORTRAN-based package is an implementation of the model-based multichannel detection methodology using the Van Overschee-De Moor state space model identification algorithm. This package is described in a Software Users' Manual generated as a separate document (Davis and Román, 1996). The MATLAB-based package is an implementation of the model-based multichannel detection methodology using each of

the three state space model identification algorithms considered in the program (Van Overschee-De Moor; canonical correlations; Arun-Kung). Also included in the MATLAB-based software package is the simulated data generation capability is described in Volume II of this Final Report.

In summary, the analytical and simulation results obtained in this program indicate that the SSC algorithm and methodology for model-based multichannel detection has the potential to result in significant advances for surveillance radar array systems and ECG diagnostics applications.

1.1 Notation

Vector variables are denoted by underscored lower-case Helvetica and Greek letters. Matrices are denoted by upper-case Helvetica and Greek letters. Some scalars (such as the order of the state variable model) are denoted also by upper-case letters. Vector spaces are denoted by upper-case Zapf Chancery letters, such as $\mathcal{V}$. Mathematical and ancillary symbols are represented with Helvetica letters in the most part, with a few exceptions where Chicago and Times are used. The expectation operator is denoted as $E[\cdot]$; superscript T and H are used to denote the matrix and vector transpose and the Hermitian transpose operators, respectively; and an asterisk (*) denotes the complex conjugate operator. I_M denotes an M-dimensional identity matrix, $O_{N,J}$ denotes an NxJ null (zero) matrix, O_M denotes an M-dimensional (square) null matrix, and $\underline{0}_M$ denotes an M-dimensional zero vector. $|A|$ denotes the determinant of matrix A; A^{-1} denotes the inverse of matrix A; $A^\dagger$ denotes the pseudoinverse of A; $\text{range}(A)$ denotes the range (column space) of A; $\text{rank}(A)$ denotes the rank of A; $A(i,j)$ and a_{ij} are both used to denote the (i,j)th element of matrix A; and $\dim(\mathcal{V})$ denotes the dimension of vector space $\mathcal{V}$. A caret (^) over a variable denotes an estimate of the variable, a bar (-) over a

variable is used to represent the mean of the variable, and $\ln(a)$ denotes the natural logarithm of a . The symbol $\perp$ denotes "is orthogonal to;" $\cap$ denotes intersection of two vector spaces; $\oplus$ denotes the direct sum of vector spaces; $\times$ denotes the Kronecker product; $\forall$ denotes "for all;" and $\in$ denotes "is an element of."

Where possible, the symbols used herein to represent variables match the symbols used by Michels (1991). This simplifies the task of relating results and techniques presented herein to prior and current work at RL. This philosophy forces the use of non-standard symbols to represent the parameters of a state variable model. Of course, notational convention should not be a major issue provided all symbols are defined appropriately. However, it is important to mention this point in order to avoid possible confusion on the part of the reader.

1.2 Report Overview

An introduction to the model-based multichannel detection problem is presented in Section 2.0. This section includes also the definition of the state space model class and several related concepts, including the backward model associated with a forward model, and the innovations representation for a random process. The Van Overschee-De Moor parameter identification algorithm is presented in Section 3.0. As mentioned earlier, this algorithm is the primary identification algorithm in the of the SSC model-based multichannel detection methodology. Filtering of the channel data to generate the innovations sequence is discussed in Section 4.0, where it is shown that the methodology can be represented as the cascade of a joint temporal/spatial linear filter and an instantaneous linear transformation (a purely spatial filter). The innovations sequence is fed to a likelihood ratio detector which generates the detection decision, as described in Section 5.0. The surveillance radar array problem is presented in Section

6.0, along with several simulation-based results. The ECG diagnostics problem is defined in Section 7.0, including ECG trace modeling and discrimination results. Section 8.0 includes the main conclusions and recommendations borne out of this Phase II. Appendix A presents the partial QSVD algorithm for matrix pairs proposed by SSC. Appendix B presents the derivation of the "combined F " formula referred to in Section 3.0. The relationship between the LDU factorization and optimal linear filtering is presented in Appendix C. A set of statistical tests for the design of the hypothesis filters in the multichannel model-based detection methodology is presented in Appendix D. And the three most common auto-correlation matrix sequence estimators (unbiased; biased; circular) are summarized in Appendix E, including their key statistical properties. Several fundamental random variable transformations are presented in Appendix F; these transformations constitute the foundation for the statistical tests in Appendix D. Finally, the formulation for testing multiple hypotheses is summarized in Appendix G. This formulation is applied to ECG diagnosis in Section 7.4.

2.0 STATE SPACE MODEL-BASED MULTICHANNEL DETECTION

The model-based approach to multichannel detection involves processing the channel data with a multiple-input, multiple-output linear filter, and determination of a detection decision utilizing the filter output. Filter parameters can be identified on-line, as the channel data is received and processed. Alternatively, the filter parameters can be identified off-line for various conditions and stored in the processor memory to be accessed in real-time as required.

There are two general classes of linear parametric models for vector random processes: time series models and state space models. Time series models include moving-average (MA) models, auto-regressive (AR) models, and auto-regressive moving-average (ARMA) models. State space models are more general than time series models; in fact, MA, AR, and ARMA models can be represented by state space models (Appendix E). In the state space literature, the determination of the model parameters based on output data (and, sometimes, input data also) is referred to as a stochastic identification or a stochastic realization problem.

Time series models have been applied to the multichannel detection problem, and the performance results obtained provide encouragement for further research (see, for example, Michels, 1991, and the references therein). Michels (1991) adopted the AR sub-class of vector time series models to represent the multichannel output process. Given the generality of state-space models and the wealth of results available in the state-space literature, the state space model class was selected in this program to represent the multichannel signals for radar systems and other applications.

In the case of time series models, two types of model parameter estimation algorithms have been established in the literature: (a) algorithms which operate on channel output correlation matrices, such as the extended Levinson algorithm (Anderson and Moore, 1979), and (b) algorithms which operate on the channel output data directly (without the need to compute channel output correlation matrices), such as the Levinson-Wiggins-Robinson algorithm (Wiggins and Robinson, 1965) and the Strand-Nuttall algorithm (Strand, 1977; Nuttall, 1976).

In the case of state-space models, most of the existing algorithms operate on channel output correlation matrices, such as the stochastic realization approach developed by Akaike (1974, 1975). This limitation is due, in a large part, to the fact that the structure of state space models is more general than the structure of time series models, and the increase in generality has presented a significant challenge to the development of algorithms that operate on channel output data directly. Recently, however, Van Overschee and De Moor (1993) have defined a state space stochastic realization algorithm which avoids the computation of channel output correlation matrices. Furthermore, this algorithm can be implemented using robust numerical techniques. The Van Overschee-De Moor algorithm was adopted as the baseline model identification algorithm in this program.

2.1 Multichannel Detection

Detection problems in the context of radar systems can be postulated as hypothesis testing problems, where a choice has to be made among two or more hypotheses. The radar target detection problems addressed in this report involve the following two hypotheses:

H_0 : Target signal is absent

H_1 : Target signal is present

H_0 is referred to as the null hypothesis, and H_1 is the alternative hypothesis. The model-based approach to the multichannel detection problem is couched on the assumption that the vector random process at the output of the channels can be represented as the output of a white noise-driven linear system under each of the two hypotheses, and that a unique parametric model corresponds to each hypothesis. Each of the two parametric models (one for each of the two hypotheses) has an inverse, or whitening filter, and the two model inverse systems are used to process the multichannel data. Furthermore, the output of the two whitening filters must be sufficiently different to allow selection of the correct hypothesis by the evaluation of measures that are sensitive to those differences.

A particular measure that has produced robust experimental results in the model-based detection context (Metford and Haykin, 1985) is the log-likelihood ratio (LLR) test. This test is the result of solving the hypothesis testing problem using the Neyman-Pearson criterion. The LLR test in the context of model-based detection is calculated using the residual sequence at the output of each of the two whitening filters, which presents practical and implementation advantages. In such a configuration, the output of the whitening filter that corresponds to the true hypothesis is white noise, and such an output is referred to as an innovations sequence.

Figure 2-1 illustrates the architecture of an on-line innovations-based multichannel detector of the type proposed by Michels (1991), which is the multichannel extension of the single-channel detector of Metford and Haykin (1985). In the case of a radar array system, each of J radar receiver channels collects the

electromagnetic energy arriving at its aperture, and processes it to generate a discrete-time random sequence, denoted as $\{x_i(n)\}$, which contains the desired information. The J random sequences $\{x_i(n)\}$ are represented in vector form as $\{\underline{x}(n)\}$. Michels (1991) has formulated the binary detection problem for multichannel systems. Specifically, the null hypothesis, H_0 , corresponds to the case of clutter and noise present in the observation process $\{\underline{x}(n)\}$, and the alternative hypothesis, H_1 , corresponds to the case of signal, clutter, and noise present in the observation process $\{\underline{x}(n)\}$. That is, the detection decision must be made between the following two models,

$$(2-1a) \quad H_0: \quad \underline{x}(n) = \underline{c}(n) + \underline{i}(n) + \underline{w}(n) \quad n \geq n_0$$

$$(2-1b) \quad H_1: \quad \underline{x}(n) = \underline{s}(n) + \underline{c}(n) + \underline{i}(n) + \underline{w}(n) \quad n \geq n_0$$

where n_0 denotes the initial observation time, $\{\underline{c}(n)\}$ denotes the clutter process, $\{\underline{i}(n)\}$ denotes all the broadband interference processes, $\{\underline{w}(n)\}$ denotes all the array channel noise processes, and $\{\underline{s}(n)\}$ denotes the desired signal (target) process. In the model-based approach pursued herein, a distinct state variable model is associated with each of the two hypotheses, and a whitening filter is designed for each model. Each filter processes the observation sequence $\{\underline{x}(n)\}$ to generate a residual vector sequence: $\{\underline{v}(n|H_0)\}$ denotes the residual sequence at the output of the null hypothesis filter, and $\{\underline{v}(n|H_1)\}$ denotes the residual sequence at the output of the alternative hypothesis filter. These residual sequences are used in a likelihood ratio test with a pre-stored threshold to carry out the detection decision. In the literature both residual sequences are referred to as innovations sequences. This is an abuse of notation because only the residual corresponding to the true hypothesis is a true innovations in the sense defined in Section 2.5. Notwithstanding, both terms are used interchangeably in this report since such usage is widespread.

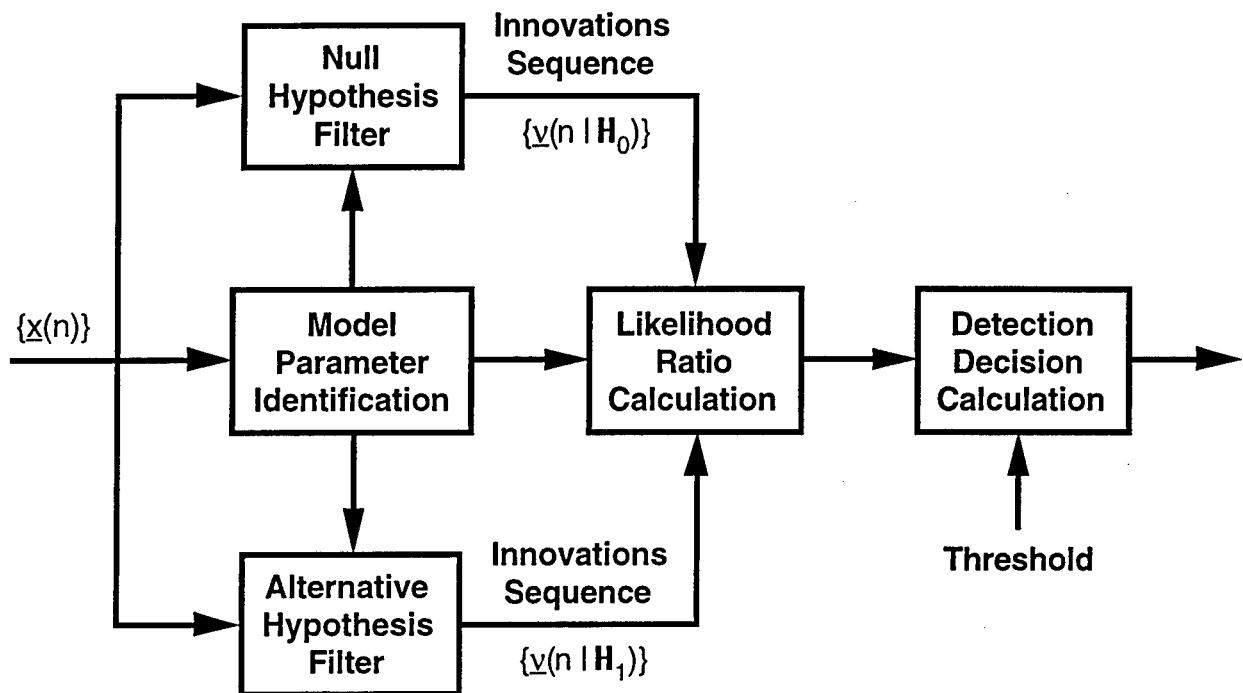


Figure 2-1. Innovations-based multichannel detector with on-line parameter identification.

As indicated in the detection configuration of Figure 2-1, the two filters can be determined in real-time by processing the observation sequence for a prescribed time interval. This approach provides the most adaptability, but may present a large computational burden for some applications. It also presents conceptual challenges, such as real-time determination of model order for each of the two filters. Alternatively, the filter design can be carried out off-line for each of the two hypotheses, and the resulting filter design implemented in the real-time configuration. The off-line approach is less robust to changes in the operational environment, but requires a simpler processor architecture, which is important in many real-time applications. Careful design of the filters off-line using adequate simulated and real data can lead to acceptable performance. Also, many pairs of fixed filters may be designed to cover distinct operational conditions. In an off-line architecture, the "Model

Parameter Identification" block in Figure 2-1 is replaced by a "Pre-Stored Filter Selection" block. The filter for the alternative hypothesis will be of higher order than the filter for the null hypothesis because the observation process for the alternative hypothesis has more information (namely, the signal component).

Michels (1991) has developed a likelihood ratio calculation and detection decision model which are compatible with the formulation adopted herein. Both of these capabilities are available at RL, and, where appropriate, the methodology presented in this report is compatible with these capabilities.

2.2 State Space Model

The class of multiple-input, multiple-output state variable models can represent effectively the channel output process for radar systems and other applications. Consider a discrete-time, stationary, complex-valued, zero-mean, Gaussian random process $\{\underline{x}(n)\}$ defined as the output of the following state space model representation for the system giving rise to the observed process:

$$(2-2a) \quad \underline{y}(n+1) = F\underline{y}(n) + G\underline{u}(n) \quad n \geq n_0$$

$$(2-2b) \quad \underline{x}(n) = H^H \underline{y}(n) + D^H \underline{w}(n) \quad n \geq n_0$$

$$(2-2c) \quad E[\underline{y}(n_0)] = \underline{0}_N$$

$$(2-2d) \quad E[\underline{y}(n_0)\underline{y}^H(n_0)] = P_0$$

Here $n=n_0$ denotes the initial time (which can be adopted as 0 since the system is stationary). Also, $\underline{y}(n)$ is the N-dimensional state of the system with $\underline{y}(n_0)$ a Gaussian random vector; $\underline{u}(n)$ is the J-dimensional, zero-mean, stationary, Gaussian, white input noise

process; and $\underline{w}(n)$ is the J-dimensional, zero-mean, stationary, Gaussian, white measurement noise process. The output (or measurement) process $\{\underline{x}(n)\}$ is also a J-dimensional vector process. Matrix F is the $N \times N$ system matrix, G is $N \times J$ input noise distribution matrix, H^H is the $J \times N$ output distribution matrix, D^H is the $J \times J$ output noise distribution matrix, and P_0 is the correlation matrix of the initial state. All these matrices are time-invariant. Matrix P_0 is Hermitian ($P_0^H = P_0$, and all its eigenvalues are real-valued) and positive definite (all its eigenvalues are positive).

System (2-2) is assumed to be asymptotically stable, which means that all the eigenvalues of matrix F are inside the unit circle. Also, system (2-2) is assumed to be reachable and observable, which implies that the dimension N of the state vector (also the order of the system) is minimal (Anderson and Moore, 1979). That is, there is no system of lesser order which has identical input/output behaviour. Lastly, system (2-2) is assumed to be minimum-phase (all its zeros are also inside the unit circle). The output distribution matrices are defined with the conjugate operator in order to have notation consistent with that of the single-output system case, where both H and D become vectors, and nominally vectors are defined as column vectors.

The input noise process correlation matrix is given as (all matrices defined hereafter have appropriate dimensions)

$$(2-3a) \quad E[\underline{u}(k)\underline{u}^H(k)] = R_{uu}(0) = Q \quad k \geq n_0$$

$$(2-3b) \quad E[\underline{u}(k)\underline{u}^H(k-n)] = R_{uu}(n) = [0] \quad k \geq n_0 \text{ and } n \neq 0$$

and the output noise process correlation matrix is given as

$$(2-4a) \quad E[\underline{w}(k)\underline{w}^H(k)] = R_{ww}(0) = C \quad k \geq n_0$$

$$(2-4b) \quad E[\underline{w}(k)\underline{w}^H(k-n)] = R_{ww}(n) = [0] \quad k \geq n_0 \text{ and } n \neq 0$$

Notice that matrices Q and C are Hermitian (that is, $Q^H = Q$, and $C^H = C$). Matrix Q is at least a positive semidefinite matrix since it is an auto-correlation matrix (all the eigenvalues of a positive semidefinite matrix are non-negative), and matrix C is assumed to be positive definite (this can be relaxed to positive semi-definite, but positive definiteness is more realistic since in the radar problem $\underline{w}(n)$ represents channel noise and other such noise processes which are independent from channel to channel).

In the most general form for this model the input and output noise processes are correlated, with a cross-correlation matrix defined as

$$(2-5a) \quad E[\underline{u}(k)\underline{w}^H(k)] = R_{uw}(0) = S \quad k \geq n_0$$

$$(2-5b) \quad E[\underline{u}(k)\underline{w}^H(k-n)] = R_{uw}(n) = [0] \quad k \geq n_0 \text{ and } n \neq 0$$

In general, matrix S is not Hermitian. Both the input and output noise processes are uncorrelated with the present and past values of the state process, and this is expressed in terms of cross-correlation matrices as

$$(2-6a) \quad E[\underline{y}(k)\underline{u}^H(k-n)] = R_{yu}(n) = [0] \quad k \geq n_0 \text{ and } n \geq 0$$

$$(2-6b) \quad E[\underline{y}(k)\underline{w}^H(k-n)] = R_{yw}(n) = [0] \quad k \geq n_0 \text{ and } n \geq 0$$

The correlation matrix of the state is defined as

$$(2-7) \quad E[\underline{y}(n)\underline{y}^H(n)] = R_{yy}(n) = P(n) \quad k \geq n_0 \text{ and } n \geq 0$$

It follows from (2-2a) and the above definitions that the state correlation matrix satisfies the following recurrence relation,

$$(2-8) \quad P(n+1) = FP(n)F^H + GQG^H \quad n \geq n_0$$

In general, matrix $P(n)$ is Hermitian and positive definite. Since system (2-2) is stationary and asymptotically stable, and since matrix Q is positive definite, then the following steady-state (large n) value exists for the recursion (2-8):

$$(2-9) \quad P(n+1) = P(n) = P \quad \text{for } n \text{ large}$$

Under steady-state conditions Equation (2-8) becomes a Lyapunov equation for the steady-state correlation matrix, P :

$$(2-10) \quad P = FPF^H + GQG^H$$

The conditions for steady-state also insure that the solution to Equation (2-10) exists, is unique (for the selected state space basis), and is positive definite (Anderson and Moore, 1979). Matrix P is unique for a given state space basis. However, if the basis of the input noise and/or the basis of the state are changed by a similarity and/or an input transformation, then a different state correlation matrix results from Equation (2-10).

The correlation matrix sequence of the output process $\{\underline{x}(n)\}$ is defined as

$$(2-11a) \quad E[\underline{x}(k)\underline{x}^H(k-n)] = R_{xx}(n) = \Lambda_n \quad \forall k \text{ and } n \geq 0$$

$$(2-11b) \quad R_{xx}(-n) = R_{xx}^H(n) \quad \forall n$$

For a system of the form (2-2), the correlation matrix $R_{xx}(n)$ can be factorized as follows,

$$(2-12a) \quad \Lambda_n = R_{xx}(n) = H^H F^{n-1} \Gamma \quad n > 0$$

$$(2-12b) \quad \Lambda_n = R_{xx}(n) = \Gamma^H [F^{n-1}]^H H = \Gamma^H [F^H]^{n-1} H \quad n < 0$$

where F^{n-1} denotes F raised to the $(n-1)$ th power, and Γ denotes the following cross-correlation matrix

$$(2-13) \quad \Gamma = E[y(n)x^H(n-1)] = R_{yx}(1) = FP(n)H + GSD \quad \forall n > 0$$

The correlation matrix sequence factorization in Equation (2-12) is the key to most correlation-based stochastic realization algorithms. The zero-lag ($n=0$) output correlation matrix is

$$(2-14) \quad R_{xx}(0) = H^H P(n)H + D^H C D = \Lambda_0$$

Matrix $R_{xx}(0)$ is Hermitian and at least positive semidefinite. In steady-state, P replaces $P(n)$ in Equations (2-13) and (2-14).

As can be inferred from the above relations, the system parameters $\{F, G, H, D, Q, C, S, P, \Gamma\}$ completely define the second-order statistics (the correlation matrix sequence $\{R_{xx}(n)\}$) of the output process, and it is said that system (2-2) realizes the output correlation matrix sequence. Conversely, the second-order statistics of the output process provide sufficient information to identify the system parameters, although not uniquely. Since the output process has zero mean and is Gaussian-distributed, the second-order statistics define the process completely.

From the system identification (stochastic realization) point of view, the problem addressed herein can be stated as follows: given the output data sequence $\{x(n)\}$ of system (2-2), estimate a set of system parameters $\{F, G, H, D, Q, C, S, P, \Gamma\}$ which generates the same output correlation matrix sequence as system (2-2).

Furthermore, the identified parameter set must correspond to a system realization of minimal order (with state vector y of minimal dimension).

It is well known (Anderson and Moore, 1979) that there can be an infinity of systems (2-2) with the same output correlation matrix sequence. The set of all systems that have the same output correlation matrix sequence is an equivalence class, and any two systems belonging to the set are said to be correlation equivalent (Candy, 1976). For example, the output correlation matrix sequence remains invariant to a similarity transformation applied to the state vector. Similarly, the output correlation matrix sequence remains invariant also to a non-singular transformation applied to the input noise and/or to the output noise. As shown by Candy (1976), the equivalence class of correlation equivalent systems is defined including other operations besides a change of basis.

Based on these comments, the solution to the system identification problem is not unique. It is also true that most of the possible system parameter solutions do not possess desirable properties. There is, however, a solution which has several features of importance. This solution is referred to as the innovations representation for system (2-2), and is discussed in Section 2.3. The identification algorithms discussed in this report generate estimates of the system parameter matrices for the innovations representation.

In general, the system matrix parameters resulting from the identification algorithm will be represented in a different basis, and should be denoted with a different symbol (say, F_1 instead of F , etc.); nevertheless, the same symbol will be used in this report in order to simplify notation.

Several definitions and notation associated with the input/output behaviour of system (2-2) are important. Consider first the L-term (finite) controllability matrix of system (2-2), C_L ; this matrix is defined as an $N \times JL$ partitioned matrix of the form

$$(2-15) \quad C_L = [G \quad FG \quad \dots \quad F^{L-1}G]$$

For a minimal-order system, matrix C_L has rank N (equal to the system order) for $L \geq N$. The controllability matrix maps the input space onto the state space. Analogously, the L-term observability matrix of system (2-2) is the following $JL \times N$ partitioned matrix,

$$(2-16) \quad O_L = \begin{bmatrix} H^H \\ H^H F \\ \vdots \\ H^H F^{L-1} \end{bmatrix}$$

and for a minimal-order system, the rank of matrix O_L is equal to N for $L \geq N$. The observability matrix maps the state space onto the output space. Classical realization theory for the deterministic case is based on the fact that a block Hankel matrix made up of the impulse response matrices of a deterministic system can be represented as the product of the observability and controllability matrices. Let $H_{L,L}$ denote a $JL \times JL$ deterministic Hankel matrix with the impulse response matrix $A(i+j-1)$ as its (i,j) th block element (a block Hankel matrix is a matrix in which the (i,j) th block element is a function of $i+j$). That is,

$$(2-17) \quad H_{L,L} = O_L C_L = \begin{bmatrix} A(1) & A(2) & \dots & A(L) \\ A(2) & A(3) & \dots & A(L+1) \\ \vdots & \vdots & \ddots & \vdots \\ A(L) & A(L+1) & \dots & A(2L-1) \end{bmatrix}$$

Equation (2-17) follows from the definition of the impulse response matrix sequence $\{A(n)\}$ for a deterministic system,

$$(2-18) \quad A(n) = H^H F^{n-1} G \quad n \geq 1$$

Matrices $\{A(n)\}$ are referred also as the Markov parameters of the deterministic system. It is well known (Kalman et al., 1969) that for $L \geq N$ the rank of the block Hankel matrix $H_{L,L}$ is equal to the system order, N . In fact, it is true also that $\text{rank}(H_{N+k,N+k}) = N$ for $k \geq 1$, and that the elements of the impulse response matrix sequence $\{A(n)\}$ satisfy a set of recursion relations of order equal to the minimal polynomial of matrix F . The block columns (and block rows) of $H_{L,L}$ satisfy the same recursion relations due to the sequential arrangement of the impulse response matrices as block elements of $H_{L,L}$.

Notice that the representation (2-18) of the impulse response matrix sequence is of the same form as the representation of the correlation matrix sequence in Equation (2-12). Thus, the matrix elements of the correlation matrix sequence $\{\Lambda_n\}$ satisfy the same set of recursion relations as the matrix elements of the impulse response matrix sequence $\{A(n)\}$, and the above-discussed properties of the deterministic Hankel matrix are also properties of the stochastic Hankel matrix defined using $\{\Lambda_n\}$ (see Equation (2-22) below).

2.3 Backward State Space Model

Associated with system (2-2) is a backward time model which is defined from the system model (2-2). Backward time models play a role in the formulation of a large class of stochastic realization algorithms. The backward time model for system (2-2) is defined as a discrete-time, stationary, complex-valued, zero-

mean, Gaussian random process with a state space representation of the form (Faurre, 1976)

$$(2-19a) \quad \underline{s}(n) = F^H \underline{s}(n+1) + \underline{v}_i(n)$$

$$(2-19b) \quad \underline{x}(n) = \Gamma^H \underline{s}(n) + \underline{v}_o(n)$$

where $\underline{s}(n)$ is the N-dimensional state vector, $\underline{v}_i(n)$ is the N-dimensional input noise vector, and $\underline{v}_o(n)$ is the J-dimensional output noise vector. Both noise vectors are uncorrelated in time (white), have mean equal to zero, and are Gaussian-distributed. The backward model output distribution matrix, Γ , is the same matrix which appears in the factorization of the output correlation matrices in Equation (2-12), and is defined in Equation (2-13).

The L-term observability matrix for the backward system (2-19) is the following $JL \times N$ partitioned matrix,

$$(2-20) \quad \mathcal{D}_L = \begin{bmatrix} \Gamma^H \\ \Gamma^H F^H \\ \vdots \\ \Gamma^H (F^H)^{L-1} \end{bmatrix}$$

The backward system is completely observable also, which implies that $\text{rank}(\mathcal{D}_L) = N$. Also of interest is the Hermitian of $\mathcal{D}_L$ with the block columns in reversed order. That is,

$$(2-21) \quad \mathcal{B}_L = \overleftrightarrow{\mathcal{D}_L}^H = [F^{L-1} \Gamma \quad \dots \quad F \Gamma \quad \Gamma]$$

where the dual-point arrow over matrix $\mathcal{D}_L^H$ indicates reversal in the order of the block columns. Notice that matrix $\mathcal{B}_L$ is like a

controllability matrix for the matrix pair (F, Γ) in reverse block column order. Thus, matrix $\mathcal{B}_L$ is referred to herein as the L-term reversed dual controllability matrix.

2.4 Stochastic Block Correlation Matrices

In the context of stochastic realization theory, the significance of the backward model follows from Equation (2-20) and the Hankel matrix of output correlation matrices, as shown next. Define a stochastic Hankel matrix $\mathcal{H}_{L,L}$ as the following $JL \times JL$ block matrix,

$$(2-22) \quad \mathcal{H}_{L,L} = \begin{bmatrix} \Lambda_1 & \Lambda_2 & \cdots & \Lambda_L \\ \Lambda_2 & \Lambda_3 & \cdots & \Lambda_{L+1} \\ \vdots & \vdots & \ddots & \vdots \\ \Lambda_L & \Lambda_{L+1} & \cdots & \Lambda_{2L-1} \end{bmatrix}$$

where the block elements $\{\Lambda_i\}$ are the elements of the output correlation matrix sequence, Equation (2-12). It follows from Equations (2-12), (2-16), and (2-22) that

$$(2-23) \quad \mathcal{H}_{L,L} = O_L \mathcal{D}_L^H$$

This equation is fundamental to stochastic realization algorithms, and allows the application of classical deterministic realization algorithms to the stochastic realization problem formulated with output correlation matrices. It also provides insight into the stochastic realization algorithm presented in Section 3.0, even though the algorithm does not require computation of the output correlation matrix sequence.

Other important matrices in stochastic realization theory include the $JL \times JL$ "future" and "past" block correlation matrices.

These matrices are the correlation matrices of future and past output block vectors defined as

$$(2-24) \quad \underline{x}_p = \underline{x}(n;n+L-1) = \begin{bmatrix} \underline{x}(n) \\ \underline{x}(n+1) \\ \vdots \\ \underline{x}(n+L-1) \end{bmatrix}$$

$$(2-25) \quad \underline{x}_f = \underline{x}(n+L;n+2L-1) = \begin{bmatrix} \underline{x}(n+L) \\ \underline{x}(n+L+1) \\ \vdots \\ \underline{x}(n+2L-1) \end{bmatrix}$$

With respect to the time instant $n+L$, vector $\underline{x}_p$ represents the past of the process $\{\underline{x}(n)\}$, and vector $\underline{x}_f$ represents the future of the process $\{\underline{x}(n)\}$. Given these definitions, the future and past block correlation matrices are given by the following $JL \times JL$ matrices:

$$(2-26) \quad \mathcal{R}_{p:L,L} = E[\underline{x}_p \underline{x}_p^H] = \begin{bmatrix} \Lambda_0 & \Lambda_{-1} & \cdots & \Lambda_{1-L} \\ \Lambda_1 & \Lambda_0 & \cdots & \Lambda_{2-L} \\ \vdots & \vdots & \ddots & \vdots \\ \Lambda_{L-1} & \Lambda_{L-2} & \cdots & \Lambda_0 \end{bmatrix}$$

$$(2-27) \quad \mathcal{R}_{f:L,L} = E[\underline{x}_f \underline{x}_f^H] = \begin{bmatrix} \Lambda_0 & \Lambda_1 & \cdots & \Lambda_{L-1} \\ \Lambda_{-1} & \Lambda_0 & \cdots & \Lambda_{L-2} \\ \vdots & \vdots & \ddots & \vdots \\ \Lambda_{1-L} & \Lambda_{2-L} & \cdots & \Lambda_0 \end{bmatrix}$$

where $\mathcal{R}_{f:L,L}$ and $\mathcal{R}_{p:L,L}$ are the future and past block correlation matrices, respectively. Both of these matrices are Hermitian as

well as block Hermitian, and they exhibit a block Toeplitz structure (a block Toeplitz matrix is a matrix in which the (i,j) th block element is a function of $i-j$).

Another matrix of interest is the block cross-correlation matrix between the future and the past, which is defined as

$$(2-28) \quad \mathcal{R}_{F:L,P:L} = E[\mathbf{x}_F \mathbf{x}_P^H] = \begin{bmatrix} \Lambda_L & \Lambda_{L-1} & \cdots & \Lambda_1 \\ \Lambda_{L+1} & \Lambda_L & \cdots & \Lambda_2 \\ \vdots & \vdots & \ddots & \vdots \\ \Lambda_{2L-1} & \Lambda_{2L-2} & \cdots & \Lambda_L \end{bmatrix} = \vec{\mathcal{H}}_{L,L} = \mathbf{O}_L \mathbf{B}_L$$

Notice that the block cross-correlation matrix $\mathcal{R}_{F:L,P:L}$ is equal to the stochastic block Hankel matrix with the block columns in reverse order, as indicated in Equation (2-28).

Equations (2-26)-(2-28) are valid for all n because the process $\{\mathbf{x}(n)\}$ is stationary. Also, for $L \geq N$, equations (2-26)-(2-28) define the correlation structure of system (2-2). In fact, the stochastic realization algorithm of Akaike (1974, 1975) is based on these block correlation matrices.

2.5 Innovations Representation

The innovations representation is a very powerful concept in the theory of linear stochastic systems due to its simplicity and its characteristics. Several texts and papers discuss this concept in detail. The discussion herein is adapted mostly from Anderson and Moore (1979), which provide a lucid presentation.

The innovations representation for a system (2-2) is a discrete-time, stationary, complex-valued, system of the form

$$(2-29a) \quad \underline{\alpha}(n+1) = F\underline{\alpha}(n) + K\underline{\varepsilon}(n) \quad n \geq n_0$$

$$(2-29b) \quad \underline{\chi}(n) = H^H \underline{\alpha}(n) + \underline{\varepsilon}(n) \quad n \geq n_0$$

$$(2-29c) \quad \underline{\alpha}(n_0) = \underline{0}_N$$

$$(2-29d) \quad E[\underline{\alpha}(n_0)\underline{\alpha}^H(n_0)] = \Pi(n_0) = \Pi_0 = [0]$$

$$(2-29e) \quad E[\underline{\alpha}(n)\underline{\alpha}^H(n)] = \Pi(n) \quad n > n_0$$

$$(2-29f) \quad \Pi(n) = \Pi \quad \text{as } n \rightarrow \infty$$

$$(2-29g) \quad R_{\chi\chi}(n) = R_{xx}(n) \quad \forall n$$

here $\underline{\alpha}(n)$ is the N -dimensional state, $\underline{\chi}(n)$ is the J -dimensional output, and the input process $\{\underline{\varepsilon}(n)\}$ is the innovations process for system (2-2). That is, $\{\underline{\varepsilon}(n)\}$ is a J -dimensional, zero-mean, white Gaussian process with correlation matrix structure given as

$$(2-30a) \quad \Omega = E[\underline{\varepsilon}(k)\underline{\varepsilon}^H(k)] = R_{xx}(0) - H^H \Pi H = \Lambda_0 - H^H \Pi H \quad k \geq n_0$$

$$(2-30b) \quad E[\underline{\varepsilon}(k)\underline{\varepsilon}^H(k-n)] = [0] \quad k \geq n_0 \text{ and } n \neq 0$$

The state correlation matrix $\Pi(n)$ has a steady-state value because the system is asymptotically stable (stationary), and the steady-state value, Π , is obtained as the limiting solution to the following recursion

$$(2-31a) \quad \Pi(n+1) = F\Pi(n)F^H + [F\Pi(n)H - \Gamma] [\Lambda_0 - H^H \Pi(n)H]^{-1} [F\Pi(n)H - \Gamma]^H \quad n \geq n_0$$

$$(2-31b) \quad \Pi(n_0) = \Pi_0 = [0]$$

Matrix K in Equation (2-29a) is given as

$$(2-32a) \quad K = [\Gamma - F\Pi H] \Omega^{-1} = [\Gamma - F\Pi H] [\Lambda_0 - H^H \Pi H]^{-1}$$

$$(2-32b) \quad K = \text{GSD } \Omega^{-1} = \text{GSD } [\Lambda_0 - H^H \Pi H]^{-1}$$

where the second relation follows from the definitions of Γ in Equation (2-13) and of Ω in Equation (2-30a). In the cases where the inverse of the correlation matrix Ω does not exist, its pseudoinverse is used instead in Equations (2-31) and (2-32).

Matrices F , H , Λ_0 , and Γ are as defined for system (2-2). That is, system (2-29) is related to system (2-2). In fact, system (2-29) as defined above is the steady-state innovations representation for system (2-2). This representation has the following important features.

- (a) First and foremost, the correlation matrix sequence of $\{\chi(n)\}$ is equal to the correlation matrix sequence of $\{\underline{x}(n)\}$, as indicated in Equation (2-29g). That is, the processes $\{\chi(n)\}$ and $\{\underline{x}(n)\}$ are correlation equivalent. This means that the innovations representation is a valid solution to the system identification problem defined herein.
- (b) Of all the correlation equivalent representations for a given output correlation sequence, the innovations representation has the smallest state correlation matrix, Π (smallest is meant in the sense of positive definiteness; that is, Π_1 is smaller than Π_2 if $\Pi_2 - \Pi_1$ is a positive definite matrix). This property of the innovations model is significant because the state correlation matrix is a measure of the uncertainty in the state.

- (c) The innovations representation is directly related to the steady-state Kalman filter (in the one-step predictor formulation) for system (2-2). In fact, the steady-state Kalman filter for system (2-2) is available immediately upon definition of the steady-state innovations representation, and viceversa. Specifically, matrix K of Equations (2-29a) and (2-31) is the steady-state Kalman gain of the optimal one-step predictor for system (2-2). This is true provided that the eigenvalues of $F-KH^H$ are stable. Thus, the innovations model is defined as above for all processes of the form (2-2), but the steady-state Kalman filter is defined only if $F-KH^H$ is stable.
- (d) The process $\{\underline{g}(n)\}$ in Equations (2-29) and (2-30) is correlation equivalent to the innovations sequence of system (2-2), which is the reason for referring to system (2-29) as the "innovations representation" for system (2-2).
- (e) The innovations model (2-29) is causally invertible. This means that the present and past of the process $\{\underline{g}(n)\}$ can be constructed from the present and past values of the output process $\{\chi(n)\}$. The converse statement is true also; that is, any causally invertible model is an innovations representation for some system. Causal invertibility of system (2-29) can be demonstrated easily. From Equation (2-29b),

$$(2-33) \quad \underline{g}(n) = -H^H \underline{\alpha}(n) + \chi(n)$$

Substituting this expression for $\underline{g}(n)$ into Equation (2-29a) results in

$$(2-34) \quad \underline{\alpha}(n+1) = [F - KH^H]\underline{\alpha}(n) + K\chi(n)$$

These relations demonstrate the causal invertibility of the innovations model (the input and output variables have traded places). Causal invertibility also provides a whitening filter for the process $\{\underline{x}(n)\}$. In fact, the whitening filter for $\{\underline{x}(n)\}$ is given by Equations (2-33) and (2-34) with $\underline{x}(n)$ in place of $\chi(n)$, and initial condition as in Equation (2-29c).

- (f) Matrix $F - KH^H$ in the inverted innovations model is a stable matrix. This follows from the fact that the matrix pair (F, H) is observable, and implies that the Kalman filter for system (2-2) is stable also.
- (g) The transfer function of the innovations model (2-29) is minimum phase. This is related to the fact that the innovations model is correlation equivalent to system (2-2), and second-order moment information (the output correlation matrix sequence) does not contain any phase information.
- (h) The innovations representation for a system of the form (2-2) is unique. Given that the innovations representation has the same output covariance sequence as system (2-2), the fact that it is unique eliminates searching for other representations for system (2-2) with the properties listed herein.
- (i) The innovations model (2-29) can be computed from the output correlation matrix sequence of system (2-2). This fact simplifies the parameter identification problem because the set of matrix parameters that must be estimated from the data is reduced to just five:

$\{F, H, \Gamma, \Pi, \Lambda_0\}$ (given these parameter matrices, the innovations covariance, Ω , and the Kalman gain, K , are obtained using Equations (2-30a) and (2-32a), respectively).

All the features listed above are of relevance to the identification approach presented in Section 3.0 because the selected parameter identification algorithm generates the innovations representation for the given output correlation matrix sequence, following feature (i).

The backward model has an associated backward innovations model which is defined by F, Γ , and the backward Kalman gain. Most of the features (a)-(i) that describe the forward innovations model are valid also for the backward innovations model, with a notable exception of feature (b), which needs to be replaced by the following statement: For each valid correlation equivalent representation for a given output correlation sequence, the state correlation matrix is smaller than the inverse of the state correlation matrix for the backward innovations model. More specifically, let Π_b denote the state correlation matrix for the backward innovations model in steady-state conditions, and let Σ denote the state correlation matrix for any valid correlation equivalent representation of an output correlation sequence. Then, $\Pi_b^{-1} - \Sigma$ is a positive definite matrix. This result provides an upper bound for the state correlation matrix of a correlation equivalent representation. Combining this with the lower bound of property (b) of the forward innovations model gives

$$(2-35) \quad \Pi \leq \Sigma \leq \Pi_b^{-1}$$

As before, the inequality between two matrices is intended in the sense of positive semi-definiteness of the matrix difference.

Of particular interest is the system representation for which the forward and backward state correlation matrices are both diagonal and equal to each other. Such a system is said to be in balanced coordinates in the stochastic sense (Desai et al., 1985). Notice that all the diagonal elements of the state correlation matrix must be less than unity in a balanced coordinates representation in order for Equation (2-35) to be satisfied. Balanced coordinates allow effective model order selection and/or model order reduction (Moore, 1981).

3.0 MULTICHANNEL SYSTEM IDENTIFICATION

The innovations representation is adopted to model the channel output process, since it reduces the model identification problem to a set of five parameter matrices, $\{F, H, \Gamma, \Pi, \Lambda_0\}$ (recall that given these parameter matrices, the innovations covariance, Ω , and the Kalman gain, K , are obtained using Equations (2-30a) and (2-32a)). Identification of the innovations representation parameter matrices is carried out using the algorithm of Van Overschee and De Moor (1993), extended to the case of complex-valued data. The Van Overschee-De Moor algorithm is based on the predictor space concept of Akaike (1974; 1975), the correlation equivalence results obtained by Faurre (1976), and the balanced stochastic realization approach of Arun and Kung (1990). The Van Overschee-De Moor algorithm is discussed in detail in the Final Report for Phase I (Román and Davis, 1993a), and is summarized herein for convenience. The algorithm is based on the decomposition of the process future into two orthogonal subspaces, wherein one subspace is spanned by the process past and the second subspace is spanned by a white noise process. Two other system model identification algorithms, canonical correlations (Desai et al., 1985) and unweighted principal components (Arun and Kung, 1990), were considered in Phase II.

3.1 Output Data-Based Algorithm

In comparison with alternative stochastic realization techniques, the Van Overschee-De Moor algorithm adopted herein has several advantages for multichannel detection applications, as listed next.

- Reduced dynamic range with respect to algorithms which require generation of the output correlation matrix sequence (correlation matrices are estimated as sums of

products of the data sequence elements, which increases the dynamic range). As such, the algorithm can be viewed as a "square-root" algorithm.

- Identifies the parameters for a model in the state-space class, which is more general than the time series class.
- Belongs to a class of algorithms referred to as "subspace methods." Subspace methods involve the decomposition of the space spanned by the output process into two orthogonal subspaces: one subspace is the space spanned by the "desired component," and the other subspace is spanned by the "noise component." The MUSIC algorithm (Schmidt, 1979; 1981), for example, also belongs to the class of subspace methods.
- An approximately balanced (in the stochastic sense) state space realization is generated, thus providing a built-in and robust mechanism for model order selection.
- Identifies the innovations representation of the system, and generates the Kalman gain directly, without having to solve a nonlinear discrete matrix Riccati equation.
- Approach differs from others in that the states of a Kalman filter for the given sequence are identified first, and then the model parameters are estimated via least-squares.
- Implementation of the algorithm involves the QR decomposition and the quotient SVD (QSVD), also known as the generalized SVD, which are stable numerical methods. Furthermore, the QSVD is applied to matrices of small dimensions.

An algorithm for implementing the QSVD as required by the Van Overschee-De Moor algorithm is presented in Appendix A. The algorithm is referred to as a partial QSVD because for certain conditions one or more columns of one of the matrices in the QSVD factorization are not generated. This is not a severe restriction because such conditions do not arise in many cases, including the Van Overschee-De Moor identification algorithm. Furthermore, the missing columns can be calculated if required (as the null space of a matrix). The partial QSVD is less complex and more accurate (from a numerical point of view) than the QSVD presented by Van Overschee and De Moor (1993).

Van Overschee-De Moor Algorithm. Consider the channel output sequence $\{\underline{x}(n)\}$. For simplicity, let the initial time $n_0 = 0$. This can be done without loss of generality because the system is stationary. Now define a block Hankel matrix $X_{0,L-1}$ with output sequence vectors assigned as block elements according to the rule $X_{0,L-1}(i,j) = \underline{x}(i+j-2)$; that is,

$$(3-1) \quad X_{0,L-1} = \begin{bmatrix} \underline{x}(0) & \underline{x}(1) & \underline{x}(2) & \dots & \underline{x}(M-1) \\ \underline{x}(1) & \underline{x}(2) & \underline{x}(3) & \dots & \underline{x}(M) \\ \underline{x}(2) & \underline{x}(3) & \underline{x}(4) & \dots & \underline{x}(M+1) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \underline{x}(L-1) & \underline{x}(L) & \underline{x}(L+1) & \dots & \underline{x}(L+M-2) \end{bmatrix}$$

Here the first subscript denotes the time index of the first element of the first row, and the second subscript denotes the time index of the first element of the last row. Matrix $X_{0,L-1}$ has JL rows and M columns (recall that J is the number of channels). The block row dimension, L , must be selected so that $J(L-1) \geq N$ (recall that N is the system order), and the column dimension, M ,

must be selected so that $M \gg L$. A more practical constraint for M is $M > 2JL$. In a similar manner define another $JL \times M$ block Hankel matrix $X_{L,2L-1}$ with output sequence vectors assigned as block elements according to the rule $X_{L,2L-1}(i,j) = x(i+j-2+L)$; that is,

$$(3-2) \quad X_{L,2L-1} = \begin{bmatrix} x(L) & x(L+1) & x(L+2) & \dots & x(L+M-1) \\ x(L+1) & x(L+2) & x(L+3) & \dots & x(L+M) \\ x(L+2) & x(L+3) & x(L+4) & \dots & x(L+M+1) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ x(2L-1) & x(2L) & x(2L+1) & \dots & x(2L+M-2) \end{bmatrix}$$

Matrices $X_{0,L-1}$ and $X_{L,2L-1}$ represent the "past" and the "future", respectively, of the output process. Akaike (1974; 1975) has demonstrated that since the order of the state space model is N , the projection of the future onto the past is an N -dimensional subspace of the M -dimensional space to which the rows of $X_{L,2L-1}$ belong. Let this subspace be called the process space, and let its complement be called the noise space. The structure of the process space (and of its matrix representation) determines the characteristics of the state space model (such as model order). The Van Overschee-De Moor algorithm is based on determining the decomposition of the future space into the two orthogonal subspaces, process space and noise space. This decomposition can be carried out using the computationally efficient and numerically robust QR decomposition (Dongarra et al., 1979).

Consider now the block Hankel data matrix $X_{0,2L-1}$, which is a $2JL \times M$ block column matrix made up of a concatenation of the past and future Hankel matrices,

$$(3-3) \quad X_{0,2L-1} = \begin{bmatrix} X_{0,L-1} \\ \text{-----} \\ X_{L,2L-1} \end{bmatrix}$$

Now apply the Hermitian operator to a "normalized" form of matrix $X_{0,2L-1}$, and carry out a QR decomposition on this matrix to obtain

$$(3-4a) \quad \frac{X_{0,2L-1}^H}{\sqrt{M}} = \frac{1}{\sqrt{M}} \begin{bmatrix} X_{0,L-1}^H & X_{L,2L-1}^H \end{bmatrix} = Q \begin{bmatrix} R^H \\ \text{-----} \\ O_{(M-2JL),2JL} \end{bmatrix}$$

$$(3-4b) \quad \frac{X_{0,2L-1}^H}{\sqrt{M}} = \begin{bmatrix} Q_A & Q_B & Q_C \end{bmatrix} \begin{bmatrix} R_A^H & R_B^H \\ [0] & R_C^H \\ [0] & [0] \end{bmatrix}$$

The normalization factor $\sqrt{M}$ is required to avoid increase in dynamic range and to match the formulation of the problem based on the correlation matrix sequence. Matrix Q is an $M \times M$ unitary matrix, submatrices Q_A and Q_B are dimensioned $M \times JL$, and submatrix Q_C is dimensioned $M \times (M-2JL)$. Matrix R^H in Equation (3-9a) is a $2JL \times 2JL$ upper-triangular matrix with rank equal to the rank of matrix $X_{0,2L-1}$. All the submatrices of R are dimensioned $JL \times JL$, and submatrices R_A^H and R_C^H are also upper-triangular. Since matrix Q is unitary, the following relations are true:

$$(3-5) \quad QQ^H = Q_A Q_A^H + Q_B Q_B^H + Q_C Q_C^H = I_M$$

$$(3-6) \quad Q^H Q = \begin{bmatrix} Q_A^H Q_A & Q_A^H Q_B & Q_A^H Q_C \\ Q_B^H Q_A & Q_B^H Q_B & Q_B^H Q_C \\ Q_C^H Q_A & Q_C^H Q_B & Q_C^H Q_C \end{bmatrix} = \begin{bmatrix} I_{JL} & [0] & [0] \\ [0] & I_{JL} & [0] \\ [0] & [0] & I_{M-2JL} \end{bmatrix} = I_M$$

Consider now the conjugate transpose of Equation (3-4), after eliminating Q_C since it is multiplied by zeros; that is,

$$(3-7) \quad \frac{X_{0,2L-1}}{\sqrt{M}} = \frac{1}{\sqrt{M}} \begin{bmatrix} X_{0,L-1} \\ \hline X_{L,2L-1} \end{bmatrix} = \begin{bmatrix} R_A & [0] \\ \hline R_B & R_C \end{bmatrix} \begin{bmatrix} Q_A^H \\ \hline Q_B^H \end{bmatrix}$$

The following two equations are obtained immediately from the partitioning in Equation (3-7),

$$(3-8) \quad \frac{X_{0,L-1}}{\sqrt{M}} = R_A Q_A^H$$

$$(3-9) \quad \frac{X_{L,2L-1}}{\sqrt{M}} = R_B Q_A^H + R_C Q_B^H$$

Equation (3-8) is a QR decomposition of $X_{0,L-1}$ (recall that R_A is lower triangular), and Equation (3-9) is a subspace decomposition of $X_{L,2L-1}$. In fact, (3-9) is the desired subspace decomposition of $X_{L,2L-1}$ (Román and Davis, 1993a). The information of the projection of the future onto the past is contained in matrix R_B . Specifically, the rank of R_B is equal to the order of the state space model representation for the future-to-past interface, and the column space of R_B is equal to the column space of the observability matrix for the state space model (Van Overschee and De Moor, 1993).

Consider Equation (3-7) and carry out a further partitioning of the QR decomposition matrices as follows:

$$(3-10a) \quad \frac{X_{0,2L-1}}{\sqrt{M}} = \frac{1}{\sqrt{M}} \begin{bmatrix} X_{0,L-1} \\ \hline X_{L,2L-1} \end{bmatrix} = \begin{bmatrix} R_{11} & [0] & [0] & [0] \\ R_{21} & R_{22} & [0] & [0] \\ \hline R_{31} & R_{32} & R_{33} & [0] \\ R_{41} & R_{42} & R_{43} & R_{44} \end{bmatrix} \begin{bmatrix} Q_1^H \\ Q_2^H \\ \hline Q_3^H \\ Q_4^H \end{bmatrix}$$

$$(3-10b) \quad \begin{matrix} & J(L-1) & J & J & J(L-1) \\ J(L-1) & \begin{bmatrix} R_{11} & [0] & [0] & [0] \\ R_{21} & R_{22} & [0] & [0] \\ \hline R_{31} & R_{32} & R_{33} & [0] \\ R_{41} & R_{42} & R_{43} & R_{44} \end{bmatrix} \\ J & \\ J & \\ J(L-1) & \end{matrix}$$

$$(3-10c) \quad \begin{matrix} & M \\ J(L-1) & \begin{bmatrix} Q_1^H \\ Q_2^H \\ \hline Q_3^H \\ Q_4^H \end{bmatrix} \\ J & \\ J & \\ J(L-1) & \end{matrix}$$

From Equations (3-7) and (3-10) it follows that the $JL \times JL$ matrices R_A , R_B , and R_C are defined with the following partitions:

$$(3-11) \quad R_A = \begin{bmatrix} R_{11} & [0] \\ R_{21} & R_{22} \end{bmatrix}$$

$$(3-12) \quad R_B = \begin{bmatrix} R_{31} & R_{32} \\ R_{41} & R_{42} \end{bmatrix}$$

$$(3-13) \quad R_C = \begin{bmatrix} R_{33} & [0] \\ R_{43} & R_{44} \end{bmatrix}$$

Refer to the partitioning in Equation (3-10) and define four other partitioned matrices as

$$(3-14) \quad R_D = [R_{41} \ R_{42} \ R_{43}] \quad R_D: J(L-1) \times J(L+1)$$

$$(3-15) \quad R_E = R_{44} \quad R_E: J(L-1) \times J(L-1)$$

$$(3-16) \quad R_F = \begin{bmatrix} R_{21} \\ R_{31} \\ R_{41} \end{bmatrix} \quad R_F: J(L+1) \times J(L-1)$$

$$(3-17) \quad R_G = \begin{bmatrix} R_{22} & [0] & [0] \\ R_{32} & R_{33} & [0] \\ R_{42} & R_{43} & R_{44} \end{bmatrix} \quad R_G: J(L+1) \times J(L+1)$$

Now carry out three QSVDs on these matrix pairs as described next. The first QSVD is applied to the matrix pair (R_B, R_C) to obtain

$$(3-18) \quad R_B^H = U_L S_L Y_L^H$$

$$(3-19) \quad R_C^H = V_L T_L Y_L^H$$

The second QSVD is applied to the matrix pair (R_D, R_E) to obtain

$$(3-20) \quad R_D^H = U_{L-1} S_{L-1} Y_{L-1}^H$$

$$(3-21) \quad R_E^H = V_{L-1} T_{L-1} Y_{L-1}^H$$

And the third QSVD is applied to the matrix pair (R_F, R_G) to obtain

$$(3-22) \quad R_F^H = U_{L+1} S_{L+1} Y_{L+1}^H$$

$$(3-23) \quad R_G^H = V_{L+1} T_{L+1} Y_{L+1}^H$$

In these three QSVDs the subscripts (L-1, L, or L+1) correspond to the term index of an associated observability matrix defined as in Equation (2-16). The dimensions and key properties of the fifteen matrix factors in the three QSVDs are listed in Table 3-1 (see Appendix A for further details on the QSVD).

In Table 3-1 and elsewhere in this report, a rectangular matrix is said to be diagonal if it has non-zero elements only along its main diagonal, in agreement with common usage. As stated in Appendix A, the diagonal elements of each matrix pair $(S_{(\bullet)}, T_{(\bullet)})$ are referred to as singular value pairs of the corresponding matrix pair $(R_{(\bullet)}, R_{(\bullet)})$. Furthermore, for each zero-valued diagonal element in $S_{(\bullet)}$ there is a corresponding unity-valued diagonal element in $T_{(\bullet)}$.

The value of the diagonal elements of matrices S_{L-1} , S_L , and S_{L+1} is indicative of model order. In fact, when the data is the output of a system of order N, only the first N diagonal entries are non-zero in each of the three matrices S_{L-1} , S_L , and S_{L+1} . This is the reason for the constraint $J(L-1) \geq N$ (since the minimum dimension of all three matrices is $J(L-1)$). Thus, for an N-th order model the matrix pairs $(S_{(\bullet)}, T_{(\bullet)})$ have a natural partition along the main diagonal corresponding to the first N entries. Specifically,

$$(3-24) \quad S_{L-1} = \begin{bmatrix} S_{L-1}^{(1)} & [0] \\ [0] & S_{L-1}^{(2)} \\ [0] & [0] \end{bmatrix} = \begin{bmatrix} S_{L-1}^{(1)} & [0] \\ [0] & O_{JL-J-N} \\ [0] & [0] \end{bmatrix}$$

$$(3-25) \quad S_L = \begin{bmatrix} S_L^{(1)} & [0] \\ [0] & S_L^{(2)} \end{bmatrix} = \begin{bmatrix} S_L^{(1)} & [0] \\ [0] & O_{JL-N} \end{bmatrix}$$

$$(3-26) \quad S_{L+1} = \begin{bmatrix} S_{L+1}^{(1)} & [0] & [0] \\ [0] & S_{L+1}^{(2)} & [0] \end{bmatrix} = \begin{bmatrix} S_{L+1}^{(1)} & [0] & [0] \\ [0] & O_{JL-J-N} & [0] \end{bmatrix}$$

$$(3-27) \quad T_{L-1} = \begin{bmatrix} T_{L-1}^{(1)} & [0] \\ [0] & T_{L-1}^{(2)} \end{bmatrix} = \begin{bmatrix} T_{L-1}^{(1)} & [0] \\ [0] & I_{JL-J-N} \end{bmatrix}$$

$$(3-28) \quad T_L = \begin{bmatrix} T_L^{(1)} & [0] \\ [0] & T_L^{(2)} \end{bmatrix} = \begin{bmatrix} T_L^{(1)} & [0] \\ [0] & I_{JL-N} \end{bmatrix}$$

$$(3-29) \quad T_{L+1} = \begin{bmatrix} T_{L+1}^{(1)} & [0] \\ [0] & T_{L+1}^{(2)} \end{bmatrix} = \begin{bmatrix} T_{L+1}^{(1)} & [0] \\ [0] & I_{JL+J-N} \end{bmatrix}$$

In practical situations where only a limited amount of noisy data is available, the cut-off between non-zero and zero-valued diagonal elements of the $S_{(.)}$ matrices disappears. This is further complicated by the finite numerical precision in the processor used to implement the algorithm.

The approach selected herein to estimate model order is to examine the diagonal elements of matrix S_L only. Besides being simple to implement, this approach is theoretically sound because the correlation matrix of the innovations model state in balanced coordinates is equal to matrix S_L ; that is, $\Pi = \Pi_b = S_L$. Also, the diagonal elements of Π are the canonical correlations (Desai et al., 1985; Román and Davis, 1993b). Model order determination is discussed further in Section 3.2.

MATRIX	DIMENSIONS	PROPERTIES
U_{L-1}	$J(L+1) \times J(L+1)$	Unitary
S_{L-1}	$J(L+1) \times J(L-1)$	Rectangular; diagonal; real-valued; with diagonal elements bound by unity and zero, and arranged in order of decreasing magnitude
Y_{L-1}	$J(L-1) \times J(L-1)$	Square; non-singular
V_{L-1}	$J(L-1) \times J(L-1)$	Unitary
T_{L-1}	$J(L-1) \times J(L-1)$	Square; diagonal; real-valued; with diagonal elements bound by unity and zero, and arranged in order of increasing magnitude
U_L	$JL \times JL$	Unitary
S_L	$JL \times JL$	Square; diagonal; real-valued; with diagonal elements bound by unity and zero, and arranged in order of decreasing magnitude
Y_L	$JL \times JL$	Square; non-singular
V_L	$JL \times JL$	Unitary
T_L	$JL \times JL$	Square; diagonal; real-valued; with diagonal elements bound by unity and zero, and arranged in order of increasing magnitude
U_{L+1}	$J(L-1) \times J(L-1)$	Unitary
S_{L+1}	$J(L-1) \times J(L+1)$	Rectangular; diagonal; real-valued; with diagonal elements bound by unity and zero, and arranged in order of decreasing magnitude
Y_{L+1}	$J(L+1) \times J(L+1)$	Square; non-singular
V_{L+1}	$J(L+1) \times J(L+1)$	Unitary
T_{L+1}	$J(L+1) \times J(L+1)$	Square; diagonal; real-valued; with diagonal elements bound by unity and zero, and arranged in order of increasing magnitude

Table 3-1. QSVD matrix factors for the three factorizations.

Now define block column partitions in matrices $U_{(j)}$, $V_{(j)}$, and $Y_{(j)}$ to correspond with the partitions in Equations (3-28)-(3-31). This results in

$$(3-30) \quad U_{L-1} = \begin{bmatrix} U_{L-1}^{(1)} & U_{L-1}^{(2)} \end{bmatrix} \quad U_L = \begin{bmatrix} U_L^{(1)} & U_L^{(2)} \end{bmatrix} \quad U_{L+1} = \begin{bmatrix} U_{L+1}^{(1)} & U_{L+1}^{(2)} \end{bmatrix}$$

$$(3-31) \quad V_{L-1} = \begin{bmatrix} V_{L-1}^{(1)} & V_{L-1}^{(2)} \end{bmatrix} \quad V_L = \begin{bmatrix} V_L^{(1)} & V_L^{(2)} \end{bmatrix} \quad V_{L+1} = \begin{bmatrix} V_{L+1}^{(1)} & V_{L+1}^{(2)} \end{bmatrix}$$

$$(3-32) \quad Y_{L-1} = \begin{bmatrix} Y_{L-1}^{(1)} & Y_{L-1}^{(2)} \end{bmatrix} \quad Y_L = \begin{bmatrix} Y_L^{(1)} & Y_L^{(2)} \end{bmatrix} \quad Y_{L+1} = \begin{bmatrix} Y_{L+1}^{(1)} & Y_{L+1}^{(2)} \end{bmatrix}$$

In Equations (3-24)-(3-32), all submatrices with superscript (1) have N columns; these submatrices are used to compute the model matrix parameters.

Matrix F can be estimated using any one of three formulas. The first formula is obtained by solving a least-squares problem formulated using the state propagation equation of the forward innovations representation; thus, the resulting estimate is referred to herein as the "forward F" and is denoted with subscript f,

$$(3-33) \quad F_f = (S_L^{(1)})^{-1/2} (\underline{Y}_L^{(1)})^\dagger Y_{L-1}^{(1)} S_{L-1}^{(1)} (\underline{U}_{L-1}^{(1)})^H U_L^{(1)} (S_L^{(1)})^{-1/2}$$

where the dagger ($\dagger$) denotes the pseudo-inverse operator, the underbar denotes that matrix $\underline{Y}_L^{(1)}$ is obtained from matrix $Y_L^{(1)}$ by deleting the last block row (J single rows), and analogously for matrix $\underline{U}_{L-1}^{(1)}$. The second formula is obtained by solving a least-squares problem formulated using the state propagation equation of the backward innovations representation, and thus is referred to herein as the "backward F" and is denoted with subscript b,

$$(3-34) \quad F_b = (S_L^{(1)})^{-1/2} \begin{bmatrix} S_L^{(1)} (U_L^{(1)})^H & T_L^{(1)} (V_L^{(1)})^H \end{bmatrix} \begin{bmatrix} U_{L+1}^{(1)} & S_{L+1}^{(1)} \\ V_{L+1}^{(1)} & T_{L+1}^{(1)} \end{bmatrix} \begin{bmatrix} (Y_{L+1}^{(1)})^\dagger & Y_L^{(1)} \end{bmatrix} (S_L^{(1)})^{1/2}$$

The third formula is obtained by solving a least-squares problem formulated using a combination of the state propagation equations of both innovations representations (forward and backward), and turns out to be a weighted linear combination of F_f and F_b . The resulting system matrix estimate is referred to herein as the "combined F " and is denoted without subscript c . Specifically,

$$(3-35a) \quad F_c = [f_{ijc}]$$

$$(3-35b) \quad f_{ijc} = \frac{s_j f_{ijf} + s_i f_{ijb}}{s_i + s_j}$$

where s_i denotes the i th diagonal element of $S_L^{(1)}$, and f_{ijc} , f_{ijf} , and f_{ijb} denote the (i,j) th elements of F_c , F_f , and F_b , respectively. Notice from Equation (3-35b) that if $f_{ijf} = f_{ijb}$, then $f_{ijc} = f_{ijf} = f_{ijb}$; also, if $s_i = s_j$, then f_{ijc} is the average of f_{ijf} and f_{ijb} . Both of these observations agree with intuition. For a short-duration data sequence (small value of M), the combined F formula should provide an improved estimate. For a long-duration data sequence the forward and backward estimates should be approximately equal, and either estimate suffices. However, the forward F calculation is preferred because it is simpler and it does not involve $V_{(\bullet)}^{(1)}$ and $T_{(\bullet)}^{(1)}$ matrices. Appendix B presents the derivation of the combined F , which is a new result obtained in Phase II.

The output distribution matrix, H , can be estimated using either one of two formulas. The first formula is obtained by solving a least-squares problem formulated using the output equation for the forward innovations representation. Thus, this

formula is referred to herein as the "forward H" and is denoted with subscript f,

$$(3-36) \quad H_f^H = [R_{31} \ R_{32}] U_L^{(1)} (S_L^{(1)})^{-1/2}$$

The second expression for estimating the output distribution matrix is based on the fact that H^H occupies the first J rows of the observability matrix, $O_{(.)}$. Consequently, this estimate is referred to herein as the "observability H" and is denoted with subscript o,

$$(3-37) \quad H_o^H = [O_L]_{\text{first J rows}} = [Y_L^{(1)} (S_L^{(1)})^{1/2}]_{\text{first J rows}}$$

With respect to accuracy, it appears that either one of these two estimates of H^H is adequate. From a computational viewpoint, the forward H formula has the matrix product $U_L^{(1)} (S_L^{(1)})^{-1/2}$ in common with the forward F formula, whereas the matrix product $Y_L^{(1)} (S_L^{(1)})^{1/2}$ in the backward F formula is the same as the observability H. Thus, an efficient approach is to estimate H using the formula dictated by which formula is used to estimate F.

Matrix Γ is estimated by solving a least-squares problem formulated using the output equation for the backward innovations representation. The resulting formula is

$$(3-38) \quad \Gamma^H = [R_{21} \ R_{22}] U_L^{(1)} (S_L^{(1)})^{1/2}$$

This expression is analogous to Equation (3-36).

Notice that the $Q_{(.)}$ matrices do not appear in the formulas for the matrix parameters F, H, and Γ . The QR decomposition is fundamental to the algorithm, but only the $R_{(.)}$ matrices have to be

calculated and stored. This is a very important feature of the algorithm because one dimension of $Q_{(\cdot)}$ is large (M), and manipulation of such matrices involves sizable storage and computational requirements.

Notice also that the QSVD factors $V_{(\cdot)}$ and $T_{(\cdot)}$ appear only in the backward F formula. In fact, only the product $V_{(\cdot)}T_{(\cdot)}$ is required in the backward F computation. This fact is important because it justifies utilization of the partial QSVD algorithm described in Appendix A.

The remaining matrix parameters for the innovations representation (2-29) in balanced coordinates are estimated easily. First, the steady-state correlation matrix of the innovations representation state, Π , and the steady-state correlation matrix of the backward innovations representation state, Π_b , are estimated as,

$$(3-39) \quad \Pi = \Pi_b = S_L^{(1)}$$

Next, the zero-lag output correlation matrix is estimated directly from the output sequence as

$$(3-40) \quad \Lambda_o = \frac{1}{N_T} \sum_{k=0}^{N_T-1} x(k)x^H(k)$$

$$(3-41) \quad N_T = M + 2L - 1$$

where N_T is the number of output data vectors (duration of the output sequence) used in the algorithm. The innovations correlation matrix is estimated using Equation (2-30a), repeated here from completeness as,

$$(3-42) \quad \Omega = \Lambda_0 - H^H \Pi H$$

And the Kalman gain is estimated using Equation (2-32a), also repeated here from completeness as,

$$(3-43) \quad K = [\Gamma - F \Pi H] \Omega^{-1} = [\Gamma - F \Pi H] [\Lambda_0 - H^H \Pi H]^{-1}$$

If Ω is singular, the pseudo-inverse operator replaces the inverse operator in Equation (3-43).

3.2 Model Order Determination

Model order determination is a necessary decision for any identification algorithm in applications where the true order of the system generating the channel output data is unknown, or where the true process generating the data may not belong to the model class adopted to represent the data. In the second case the model generated by the algorithm is a "representation model," as opposed to a "physical model" (a model based on analyses of the underlying physical processes). Determination of model order is always a difficult problem, and the solution is rarely clear-cut. The Van Overschee-De Moor identification algorithm does have several features that lead to robust model order estimation. Principally, the algorithm identifies the model parameters of the innovations representation for the multichannel process in balanced coordinates. In a system representation in balanced coordinates the position of a state in the state vector is indicative of the importance of the contribution of that state to the output correlation sequence (the first state is equal in importance or more important than the second state; etc.), and the magnitude of the corresponding correlation matrix element is representative of the relative contribution of that state.

As stated in Section 3.1, the prime mechanism for model order selection in the Van Overschee-De Moor algorithm is examination of the diagonal values of matrix S_L , which is also the steady-state correlation matrix of the state of both the forward (Π) and backward (Π_b) innovations models. Matrix S_L is diagonal, and its diagonal elements are real-valued, non-negative, bounded by unity and zero, and are arranged in order of decreasing magnitude. Furthermore, these are the canonical correlations between the past and future of the multichannel output process (Akaike [1975]; Desai et al. [1985]; Román and Davis [1993b]), which implies that the state is represented in the coordinates that allow the optimum prediction of the future of the process given the past. Thus, an effective model order selection approach is to identify the negligible diagonal elements of matrix S_L , and select the model order as the number of non-negligible diagonal elements of S_L .

In most situations involving a finite amount of noisy data, all the diagonal values in matrix S_L are different from zero and/or . This is due to the fact that the subspace decomposition is imperfect with finite amounts of data because the measurement noise $\{w(n)\}$ corrupts the past output subspace, and vice versa. In such cases model order can be estimated by identifying jump discontinuities in the magnitude of the diagonal values of S_L .

In cases where the desired state space model is only a representative model (as opposed to a physical model) for the data it is unlikely that a well-defined partition between the non-negligible and the negligible values be present. Surveillance radar arrays and medical technology applications fall in this category. In such cases the balanced coordinates and canonical correlations concepts provide important insights.

Several functions of the canonical correlations (diagonal elements of S_L) have been used in Phase II to determine model

order based on the shape of the curve. Specifically, the following functions have been used in various test cases:

- (a) canonical correlations;
- (b) normalized running sum of canonical correlations;
- (c) squared canonical correlations;
- (d) normalized running sum of squared canonical correlations;
- (e) log parameters; and
- (f) normalized mutual information parameters.

These functions are defined in (Román and Davis, 1993b). The best results have been obtained in most cases using the normalized mutual information parameters. Mutual information is used often as a criterion for model order selection (Desai et al., 1985).

For a formulation based on the future and past vectors as defined in Equations (2-24) and (2-25), Gelfand and Yaglom (1959) have defined the mutual information between the past and the future as a real-valued scalar denoted as η , and computed as

$$(3-44) \quad \eta = -\frac{1}{2} \sum_{k=1}^{JL} \ln(1-\rho_k^2)$$

where ρ_k is the k th canonical correlation (k th element of S_L), and $\ln$ denotes the natural logarithm function. Given this definition, the normalized mutual information parameter for an i th order model is defined as

$$(3-44) \quad \eta_i = \frac{-\frac{1}{2} \sum_{k=1}^i \ln(1-\rho_k^2)}{\eta} = \frac{\sum_{k=1}^i \ln(1-\rho_k^2)}{\sum_{k=1}^{JL} \ln(1-\rho_k^2)}$$

The value of this parameter represents the fraction of the mutual information in the past about the future that is retained by the

state of the i th-order innovations model for the multichannel output process. Using the normalized mutual information as criterion, the model order is the index i for which η_i exceeds a pre-selected threshold τ_η , which is a decimal between 0 and 1. Good results have been obtained consistently using $\tau_\eta = 0.9$.

Other considerations for model order determination involve the three QSVD calculations and are discussed in the Final Report for Phase I (Román and Davis, 1993a). Specifically, the first step in the QSVD for a matrix pair $(R_{(.)}, R_{(.)})$ is to carry out a SVD for a matrix formed by concatenating in a two-element block column the Hermitian transpose of the two matrices, and to determine the rank of the concatenated matrix based on an examination of the singular values. Rank determination can be a difficult task, specially when dealing with noisy data and with representation models (as opposed to physical models). In the QSVD application of interest herein, over-determination of rank (selecting a value greater than the optimum value) is more desirable than under-determination of rank because the latter option places an irreversible bound on the maximum possible model order. Thus, the approach adopted in this program is to select the rank of the concatenated matrix conservatively (over-determined) in order to allow a larger range of possible values to the model order selection step using matrix S_L .

4.0 INNOVATIONS SEQUENCE GENERATION

In the approach pursued in this program, the multichannel output data sequence under each hypothesis is modeled as an innovations representation (2-29). Thus, once the innovations model parameters have been identified, a hypothesis filter can be configured to generate the innovations sequence, $\{\xi(n)\}$, given the multichannel output data sequence. One hypothesis filter is designed corresponding to each hypothesis. Each hypothesis filter implements sequentially two distinct linear operations, as indicated in this section. All filter outputs are used in the likelihood ratio calculations (Section 5.0).

The innovations sequence at the output of a whitening filter is a white process in time, but, in general, is a correlated vector at each time instant (the innovations correlation matrix, Ω , is non-diagonal). In the context of applications involving spatially-distributed sensors, the innovations at the whitening filter output is a temporally white, spatially correlated process. Spatial whitening can be achieved using an instantaneous linear transformation. A two-processor configuration to achieve full whitening is illustrated in Figure 4-1.

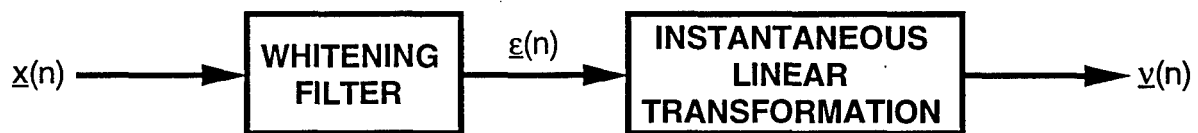


Figure 4-1. Two-function hypothesis filter.

The whitening filter for the innovations model (2-29) is a linear, discrete-time, complex-valued, time-invariant system described by the following equations:

$$(4-1a) \quad \underline{\alpha}(n+1) = [F - KH^H]\underline{\alpha}(n) + K\underline{x}(n) \quad n \geq n_0$$

$$(4-1b) \quad \underline{\varepsilon}(n) = -H^H\underline{\alpha}(n) + \underline{x}(n) \quad n \geq n_0$$

$$(4-1c) \quad \underline{\alpha}(n_0) = \underline{0}$$

where $\underline{\alpha}(n)$ is the whitening filter state vector, $\underline{\varepsilon}(n)$ is the temporal innovations associated with the observation $\underline{x}(n)$, and K is the steady-state Kalman gain matrix. The filter initial condition is set equal to zero because the innovations model initial condition is zero, Equation (2-29c). A block diagram of the whitening filter is presented in Figure 4-1, displaying the channel output vector as input, and the innovations sequence vector as output.

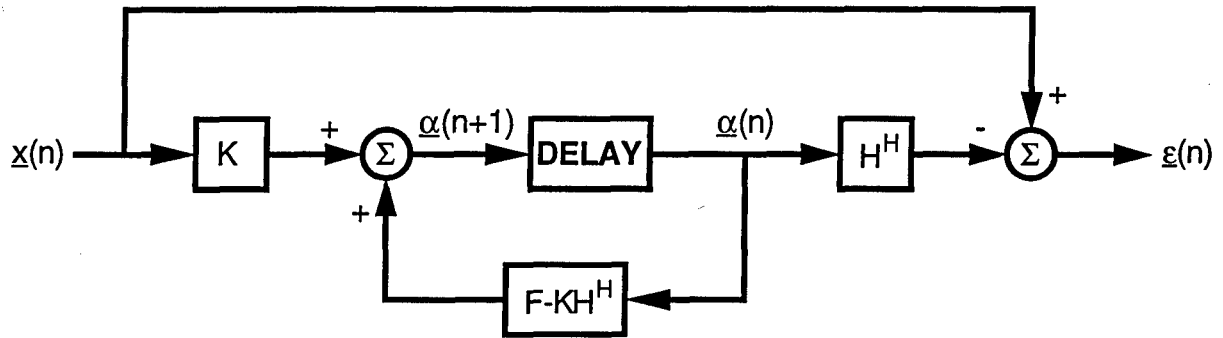


Figure 4-2. Whitening filter block diagram.

A one-step predictor formulation for the innovations model (2-29) can be defined also to generate the innovations, as described in the Phase I Final Report (Román and Davis, 1993a). Both approaches are equivalent, but the interpretation is different. The whitening filter approach is preferred herein to emphasize the fact that the desired filter output is white under matched hypothesis/filter conditions.

In the second block in Figure 4-1 a linear transformation is applied at each time instant n to the temporal innovations $\{\underline{\varepsilon}(n)\}$ in

order to generate a temporally- and spatially-whitened process $\{\underline{v}(n)\}$ as follows,

$$(4-2) \quad \underline{v}(n) = T^H \underline{\varepsilon}(n)$$

where T^H is a complex-valued, non-singular, $J \times J$ matrix. Matrix T^H is selected such that the temporal innovations covariance, Ω , is diagonalized. Let Σ denote the diagonal covariance matrix of $\underline{v}(n)$,

$$(4-3) \quad \Sigma = \begin{bmatrix} \sigma_1^2 & 0 & \cdots & 0 \\ 0 & \sigma_2^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_J^2 \end{bmatrix} = E[\underline{v}(n) \underline{v}^H(n)]$$

From Equation (4-2), Σ and Ω are related according to

$$(4-4) \quad \Sigma = T^H \Omega T$$

Diagonalization of Ω can be carried out using any one of several Hermitian matrix factorization approaches. Foremost among these are the Cholesky factorization, the LDU decomposition, and the SVD. Each of these factorizations is summarized next.

The Cholesky factorization of the temporal innovations covariance matrix Ω is defined as

$$(4-5) \quad \Omega = C C^H$$

where C is a $J \times J$ complex-valued, lower-triangular matrix with non-zero elements along the diagonal. Thus, this factorization requires that Ω be non-singular. The Cholesky instantaneous transformation matrix is denoted as T_C , and is obtained as

$$(4-6) \quad T_C^H = C^{-1}$$

And the Cholesky spatial innovations covariance matrix is the identity,

$$(4-7) \quad \Sigma_C = I_J$$

That is, the spatially-whitened innovations have unit variance. The Cholesky factorization is least desirable for spatial whitening since it requires that Ω have full rank. Michels (1991) has applied the Cholesky factorization to innovations-based multichannel detection and to correlated random process synthesis.

The LDU decomposition of the temporal innovations covariance matrix Ω is a factorization of the form

$$(4-8) \quad \Omega = LDL^H$$

where L is a $J \times J$ complex-valued, lower-triangular matrix with unity-valued elements along the main diagonal, and D is a $J \times J$ diagonal matrix with real-valued, non-negative diagonal entries. In this factorization Ω can be rank-deficient, and the rank deficiency of Ω is manifested with a corresponding number of zeros along the diagonal of D . The LDU instantaneous transformation matrix is denoted as T_P , and is obtained as

$$(4-9) \quad T_P^H = L^{-1}$$

And the LDU spatial innovations covariance matrix is equal to the diagonal matrix in the decomposition,

$$(4-10) \quad \Sigma_P = D$$

Notice that when Ω has rank-deficiency r , then exactly r spatially-whitened innovations have zero-valued variance. This condition

has to be handled appropriately when implementing the likelihood ratio detector with LDU-whitened innovations.

The LDU decomposition has an interesting interpretation in the context of spatial whitening. Therrien (1983) has shown that the LDU decomposition is related to optimal linear prediction. Specifically, the rows of matrix L^{-1} correspond to the coefficients (in reverse order) of the optimum linear prediction filters of orders 0 through $J-1$, and the diagonal elements of D are the corresponding prediction error variances. Thus, LDU-based spatial whitening is equivalent to optimal spatial filtering. This allows generation of analyses and diagnostics such as filter frequency response curves, as discussed in Appendix C.

The SVD of the temporal innovations covariance matrix Ω is defined as

$$(4-11) \quad \Omega = VSV^H$$

where V is a $J \times J$ unitary matrix, and S is a $J \times J$ diagonal matrix with real-valued, non-negative entries in the diagonal arranged in decreasing order of magnitude. In this factorization Ω can be rank-deficient also, and the rank deficiency of Ω is manifested with a corresponding number of zeros as the last diagonal entries of S . The SVD instantaneous transformation matrix is denoted as T_S , and is obtained as

$$(4-12) \quad T_S^H = V^H$$

And the SVD spatial innovations covariance matrix is equal to the diagonal matrix in the decomposition,

$$(4-13) \quad \Sigma_S = D$$

Notice that when Ω has rank-deficiency r , then exactly r spatially-whitened innovations have zero-valued variance. This condition has to be handled appropriately when implementing the likelihood ratio detector with SVD spatial whitening.

Spatial whitening of the temporal innovations as described above is useful for diagnostics and analyzing data. Additionally, LLR detection curves and related results can be generated more efficiently using appropriately-implemented spatial whitening. LDU-based spatial whitening is preferred herein due to the insight it provides as a spatial filter. Software implementation of the LDU decomposition is straightforward, specially for full rank Hermitian matrices. However, allowances need to be made in the code to handle rank-deficient matrices. SSC discovered that the LDU decomposition in the MATLAB software package generates reasonably-looking but erroneous results for rank-deficient matrices.

5.0 LIKELIHOOD RATIO DETECTION

A detection methodology for complex-valued multichannel Gaussian processes has been developed by Michels (1991) in the context of innovations-based detection. This approach has been generalized recently to include a class of non-Gaussian processes known as spherically-invariant random processes (SIRPs) and using linear estimators (Rangaswamy, Weiner, and Michels, 1993). Michels' methodology can be applied directly to the innovations sequence generated by the approach formulated herein. For brevity, only the likelihood ratio equation is presented here.

As discussed in Section 4.0, a hypothesis filter is designed for each hypothesis based on processing the multichannel data. The model order for the alternative hypothesis (H_1) whitening filter is chosen to be larger than the model order for the null hypothesis (H_0) whitening filter. Thus, for each hypothesis filter, the temporal innovations sequence is

$$(5-1) \quad \underline{x}(n|H_i) = -H^H \underline{\alpha}(n|H_i) + \underline{x}(n) \quad i = 0, 1$$

where the subscript i distinguishes between the two hypotheses. Similarly, denote the spatial innovations sequence as

$$(5-2) \quad \underline{y}(n|H_i) = T^H \underline{x}(n|H_i) \quad i = 0, 1$$

Also, the steady-state correlation matrix of the temporal innovations is denoted as $\Omega(H_i)$, and steady-state correlation matrix of the spatial innovations is denoted as $\Sigma(H_i)$.

Let $\Theta(H_0, H_1)$ denote the multichannel likelihood ratio as defined by Michels (1991) for the Gaussian signal case. Then, the log-likelihood ratio (LLR) for the temporal innovations can be expressed as,

$$(5-3) \quad \ln[\Theta(\mathbf{H}_0, \mathbf{H}_1)] = \sum_{n=n_0}^{N_T} \left[\ln \left[\frac{|\Omega(\mathbf{H}_0)|}{|\Omega(\mathbf{H}_1)|} \right] + \underline{\mathbf{x}}^H(n|\mathbf{H}_0) \Omega^{-1}(\mathbf{H}_0) \underline{\mathbf{x}}(n|\mathbf{H}_0) - \underline{\mathbf{x}}^H(n|\mathbf{H}_1) \Omega^{-1}(\mathbf{H}_1) \underline{\mathbf{x}}(n|\mathbf{H}_1) \right]$$

where $|\Omega(\mathbf{H}_i)|$ denotes the determinant of matrix $\Omega(\mathbf{H}_i)$. The LLR is compared to a threshold, $\mathcal{T}$, which is calculated adaptively to maintain a constant false alarm rate (CFAR),

$$(5-4) \quad \ln[\Theta(\mathbf{H}_0, \mathbf{H}_1)] = \begin{cases} \geq \mathcal{T} & \text{select } \mathbf{H}_1 \\ < \mathcal{T} & \text{select } \mathbf{H}_0 \end{cases}$$

A candidate CFAR approach with demonstrated good performance calculates the median of a set of the LLR values from a number of adjacent range cells (at the same azimuth) on both sides of the cell in question, and scales the calculated median value by a pre-determined constant to provide the desired false alarm rate (Metford and Haykin, 1985).

Michels (1991) has derived the LLR formula for the spatial innovations generated with LDU-based spatial whitening. Namely,

$$(5-5) \quad \ln[\Theta(\mathbf{H}_0, \mathbf{H}_1)] = \sum_{n=n_0}^{N_T} \sum_{k=1}^J \left[\ln \left[\frac{\sigma_{0k}^2}{\sigma_{1k}^2} \right] + \frac{|v_k(n|\mathbf{H}_0)|^2}{\sigma_{0k}^2} - \frac{|v_k(n|\mathbf{H}_1)|^2}{\sigma_{1k}^2} \right]$$

where $v_k(n|\mathbf{H}_i)$ denotes the k th element of $\mathbf{v}(n|\mathbf{H}_i)$, and σ_{ik}^2 denotes the k th diagonal element of $\Sigma(\mathbf{H}_i)$, as defined in Equation (4-3). Note that all the terms in Equation (5-5) are scalars. In contrast, all the terms in Equation (5-3) are functions of matrix parameters. This is significant for two reasons. First, Equation (5-5) requires less computations than Equation (5-3). Second, Equation (5-5) is applicable over a wider range of conditions. In

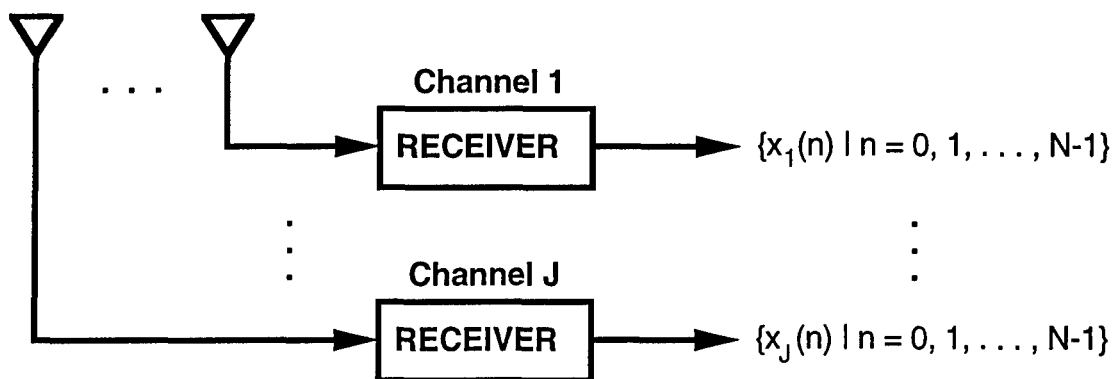
particular, Equation (5-5) can be used, with minor modifications, in the cases where either one (or both) of the temporal innovations covariances, $\Omega(\mathbf{H}_i)$, is singular. In such cases Equation (5-3) cannot be used because the determinant of a singular matrix is zero ($\ln[0]$ is undefined), and the inverse of a singular matrix is undefined. When $\Omega(\mathbf{H}_i)$ is singular, one or more of the variances σ_{ik}^2 is equal to zero. Given the linear prediction characteristic of LDU spatial filtering, $\sigma_{ik}^2 = 0$ implies that the k th variable is linearly dependent on the $k-1$ preceding variables, so that $|v_k(n|\mathbf{H}_i)| \approx 0$. In other words, LDU spatial filtering eliminates all statistically-dependent elements of the temporal innovations vector. Thus, the modification required in Equation (5-5) is to drop all the terms that involve zero-valued variances. Of course, the natural logarithm term must be expanded first ($\ln[a/b] = \ln[a] - \ln[b]$) so that only the term involving a zero-valued variance is dropped.

LLR expressions for spatial innovations generated with Cholesky- and SVD-based spatial whitening are presented by Michels (1991) and Román and Davis (1993a), respectively. Those formulas are similar to Equation (5-5) above. It is important to note that SVD-based spatial whitening can be used analogously in the cases where the temporal innovations covariance is singular.

6.0 AIRBORNE SURVEILLANCE PHASED ARRAY RADAR APPLICATION

Radar systems in general, and surveillance radar array systems in particular, constitute the primary focus of both phases of this program. In this section the space/time processing problem associated with surveillance radar array problem is formulated and several analyses are presented. The discussion herein complements the surveillance radar array data generation model and MATLAB-based software implementation presented in Volume II of this Final Report.

Consider a coherent radar system with J spatial channels (each channel is the output of either an individual array element or a sub-array composed of multiple array elements), as indicated in Figure 6-1. In a surveillance scenario (see, for example, Jaffer et al. [1991], or Rangaswamy et al. [1993]), the J -element, discrete-time, baseband, complex-valued, finite-duration, vector sequence $\{\mathbf{x}(n) | n = 0, 1, \dots, N_T - 1\}$ is the return from the radar resolution (range-azimuth) cell received at each of the J channels for the duration of the coherent processing interval (CPI), which consists of N_T data points. In the hypothesis testing formulation adopted herein, the null hypothesis (H_0) corresponds to the case of target absent, and the alternative hypothesis (H_1) corresponds to the case of target present. Under the null hypothesis, the vector sequence $\{\mathbf{x}(n)\}$ contains clutter, interference, and noise (Equation (2-1a)). Under the alternative hypothesis, $\{\mathbf{x}(n)\}$ also contains target information (Equation (2-1b)). The vector sequence is assumed to be zero-mean and Gaussian-distributed under both hypotheses. Thus, the radar return process is specified by its correlation structure; specifically, its correlation matrix sequence $\{R_{xx}(m)\}$. In turn, the structure and performance of detection algorithms are driven by this correlation structure.



$x_j(n)$: complex-valued received signal at the j th array element
corresponding to the return from the n th pulse

Receiver: demodulation, temporal sampling, etc.

Figure 6-1. Multichannel signal in a coherent surveillance radar array system.

The surveillance radar array model presented in Volume II of this Final Report includes a description of the correlation matrix sequence of each of the components present in $\{\underline{x}(n)\}$. It is convenient to state herein the key correlation features of each component, as modeled in Volume II, using the notation established in Table 6-1 below. Moving targets have both temporal and spatial correlation, and for a single target, $\text{rank}(R_S) = 1$ and $\text{rank}(R_T) = 1$. Ground clutter also has both temporal and spatial correlation, and the temporal and spatial correlations are coupled. This coupling (space-time correlation) is the reason why clutter is difficult to handle. For clutter, $\text{rank}(R_S) > 1$ and $\text{rank}(R_T) > 1$. Broadband interference only has spatial correlation, and for a single interference source, $\text{rank}(R_S) = 1$ and R_T is diagonal. Receiver noise is uncorrelated in time as well as in space. Thus, for noise both R_S and R_T are diagonal. These differences in correlation structure translate into differences in the spectral domain, and are exploited (to different degrees of success) by the various space/time processing algorithms.

SPATIO/TEMPORAL CORRELATION MATRIX

$$\underline{x} = \begin{bmatrix} x(0) \\ x(1) \\ \vdots \\ x(N-1) \end{bmatrix} \quad \mathcal{R} = E[\underline{x} \underline{x}^H] = \begin{bmatrix} R_{xx}(0) & R_{xx}(-1) & \dots & R_{xx}(-N+1) \\ R_{xx}(1) & R_{xx}(0) & \dots & R_{xx}(-N+2) \\ \vdots & \vdots & \ddots & \vdots \\ R_{xx}(N-1) & R_{xx}(N-2) & \dots & R_{xx}(0) \end{bmatrix}$$

SPATIAL CORRELATION MATRIX

$$\underline{x}(n) = \begin{bmatrix} x_1(n) \\ x_2(n) \\ \vdots \\ x_J(n) \end{bmatrix} \quad R_S = E[\underline{x}(n) \underline{x}^H(n)]$$

TEMPORAL CORRELATION MATRIX

$$\underline{x}_j = \begin{bmatrix} x_j(0) \\ x_j(1) \\ \vdots \\ x_j(N-1) \end{bmatrix} \quad R_T = E[\underline{x}_j \underline{x}_j^H]$$

Table 6-1. Data vector and correlation matrix definition for the three conventional space/time processing configurations.

6.1 Conventional Space/Time Processing

In the surveillance radar array application the objective is to detect the target while canceling the spatial interference and clutter. Conventional means to accomplish this objective determine a set of JN_T complex-valued weights that are applied to the radar return sequence $\{\underline{x}(n)\}$. These weights implement a beam pattern with nulls placed as close as possible (subject to

physical beam pattern constraints) to the direction of arrival of the incoming clutter and interference. These weights also place nulls in the temporal frequency response corresponding to the center Doppler frequency of the clutter and interference.

Wang and Cai (1994) classify the conventional space-time processing configurations for the detection of a moving target into the following three major categories:

- (a) optimum joint-domain configuration,
- (b) space-time configuration, and
- (c) time-space configuration.

The relevant data vector and covariance matrix definitions for these configurations are presented in Table 6-1. In the optimum joint-domain configuration a spatio-temporal performance criterion (signal-to-interference-plus-noise ratio) is formulated and optimized jointly (for the space and time domains). This results in a JN_T -dimensional weight vector which is applied to the JN_T -dimensional vector $\underline{x}$ formed by concatenating the N_T random vectors $\{\underline{x}(n) | n=0, 1, \dots, N_T-1\}$, as defined in Table 6-1. A block diagram for this configuration is presented in Figure 6-2 for the case of a known signal, as discussed in Wang and Cai (1994).

The other two configurations are approximations to the optimal configuration, based on formulating the problem as a cascade of two separate problems in order to reduce the computational burden. In the space-time configuration a spatial-domain (beamforming) problem is addressed first, and then a temporal-domain problem is addressed. An optimum solution (in the localized sense) is obtained for each of the two separate problems, and the solutions are applied sequentially to the data, as indicated in Figure 6-3 (also for the case of a known signal). In the time-space configuration temporal domain weighting preceeds

the beamformer. A block diagram for the time-space configuration is presented in Figure 6-4 for the known signal case. Variations of these configurations have been proposed by Jaffer et al. (1991) and Ward (1994), among others. In most of those alternatives temporal weighting is replaced with a Doppler filter bank, implemented using the DFT.

Each of the configurations listed above admits approximations defined to reduce further the computational load. This is true even for the space-time and time-space configurations, which are themselves approximations to the optimum joint-domain approach. Two important approximations to the optimum approach are the "block sliding" algorithm proposed by Jaffer et al. (1991), and the joint-domain localized generalized likelihood ratio (JDL-GLR) proposed by Wang and Cai (1994).

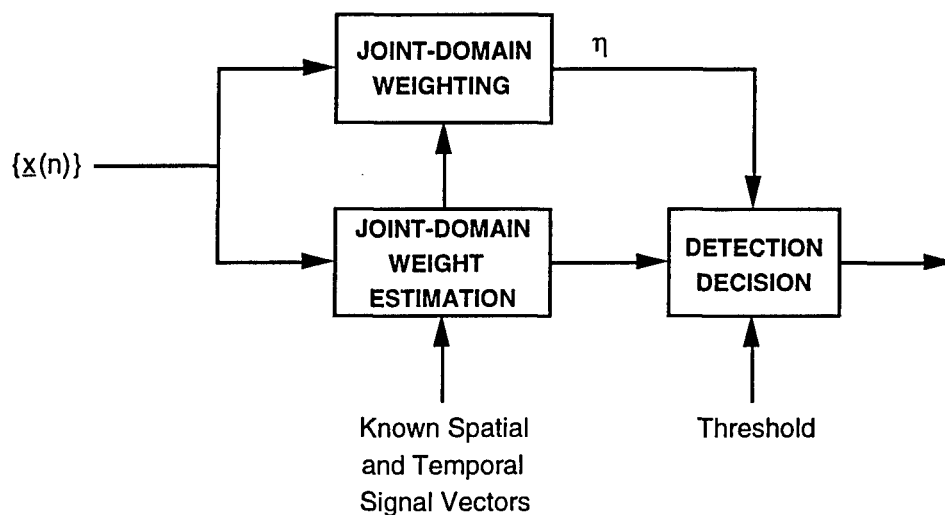


Figure 6-2. Joint-domain configuration block diagram.

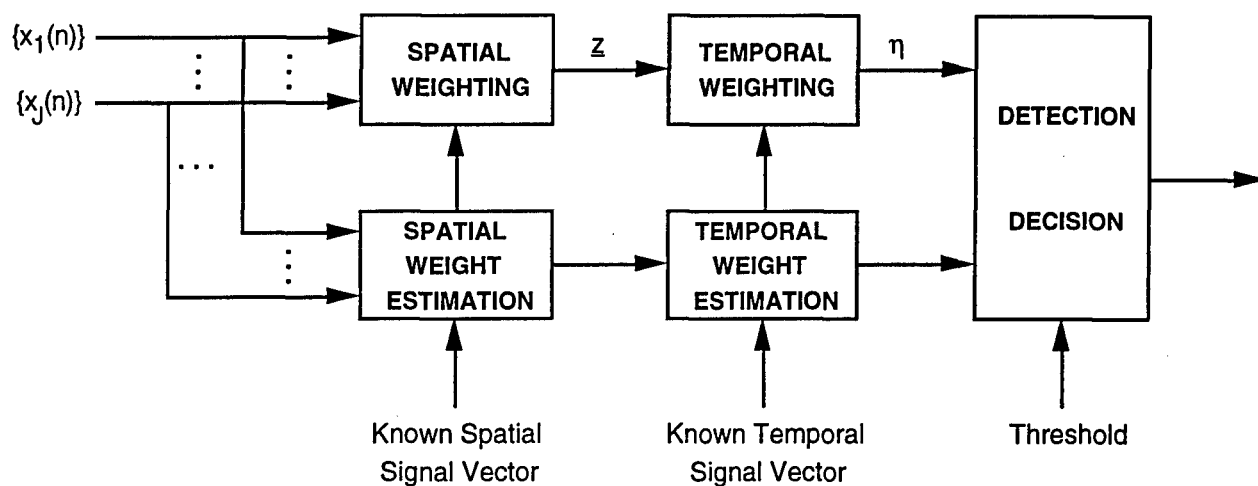


Figure 6-3. Space-time configuration block diagram.

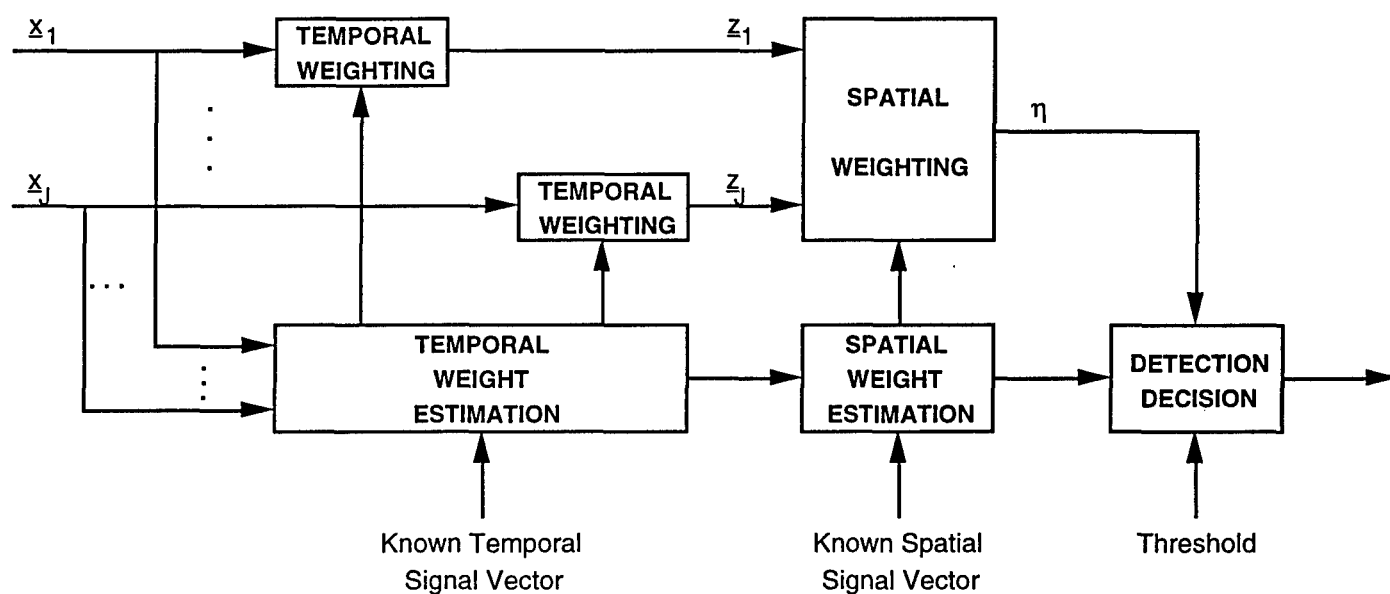


Figure 6-4. Time-space configuration block diagram.

6.2 Model-Based Space/Time Processing

The SSC model-based multichannel detection configuration developed in Phase I (Román and Davis, 1993a) applies directly to the space/time processing problem, and can be classified as a joint-domain technique. A block diagram for this configuration is presented in Figure 6-5, which is the off-line version of Figure 2-1. As described in Section 4.0, the innovations sequence, $\{\underline{v}(n|H_i)\}$, is uncorrelated in time as well as in space for the signal path corresponding to the hypothesis which is true, and is correlated in time and in space for the signal path corresponding to the hypothesis which is false. This difference is sufficient to allow making the detection decision. Of course, the sequence $\{\underline{v}(n|H_i)\}$ is a true innovations process only for the signal path corresponding to the true hypothesis.

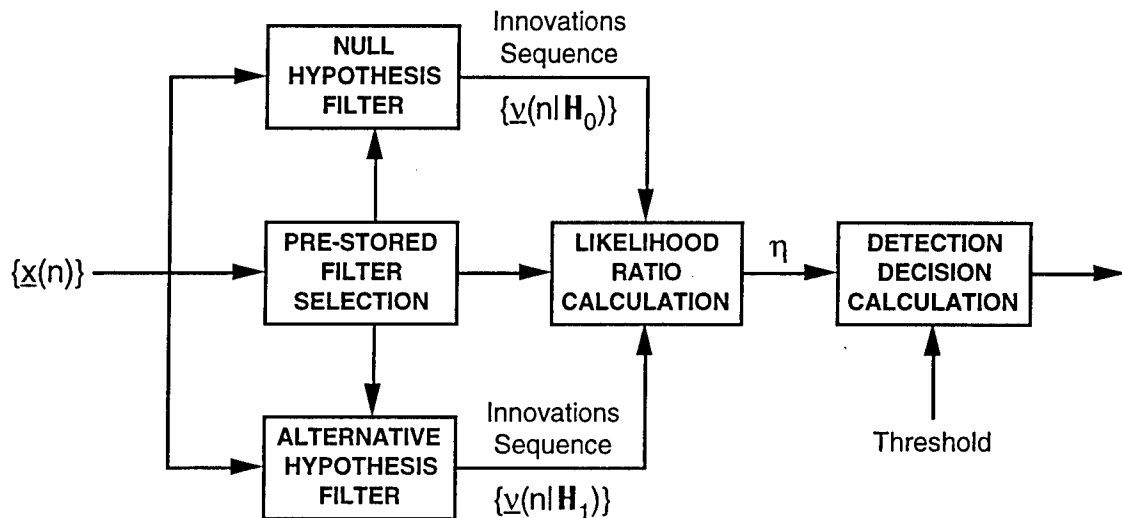


Figure 6-5. Multichannel model-based detection configuration with off-line parameter identification for space/time processing.

Each of the two filters in Figure 6-5 is a whitening filter for the respective case (null or alternative hypotheses). As

indicated in Section 4.0, the whitening takes place in two steps. In the first step a dynamic filter is used to generate the temporal innovations sequence $\{g(n|H_i)\}$ with covariance matrix $\Omega(H_i)$, which is uncorrelated in time and is less correlated in space than the radar return sequence at the filter input, $\{x(n)\}$. This reduction in spatial correlation is expected since the whitening filter in Figure 4-1 is a multichannel operator which takes into consideration both temporal and spatial correlation information. The degree of reduction in spatial correlation can be ascertained by comparing the normalized off-diagonal terms of $\Omega(H_i)$ with the normalized off-diagonal terms of R_S (the (i,j) th normalized off-diagonal element in a covariance matrix is the correlation coefficient between the i th and j th random variates). In the second step an instantaneous linear transformation is applied to whiten the temporal innovations along the spatial direction. This results in the spatially- and temporally-white innovations sequence $\{y(n|H_i)\}$ with diagonal covariance matrix $\Sigma(H_i)$.

6.3 Space/Time Process Modeling and Filtering Analyses

Sample realizations of the channel output process have been analyzed to identify state variable models for the ground clutter process and to design the two-function hypothesis filters (Figure 4-1) for the multichannel model-based detection configuration in Figure 6-5. The radar array output realizations have been generated using the space/time process simulation described in Volume II of this Final Report. Model parameters were identified using the Van Overschee-De Moor (VODM) algorithm and the canonical correlations (CC) algorithm (Román and Davis, 1993b). Hypothesis filters generated using the VODM algorithm produced less whitening of the clutter process. Thus, the results presented in this section were generated using the CC algorithm. SSC will continue to evaluate this issue, including the possibility of a software implementation error.

Table 6-2 lists the baseline parameters and conditions used in the simulations reported herein. These parameters are as described in the data generation software model of the companion Volume II (Román and Davis, 1996). Other parameters and conditions have been run, producing results similar to those presented next. Results are presented for two sets of runs; namely, the known auto-correlation sequence (ACS) case (Figures 6-6 through 6-17), and the estimated ACS case (Figures 6-18 through 6-26). In both sets of runs the system model parameters (model order; ACS lags used to identify the model; etc.) are selected to be the same.

Consider first ACS case wherein the true ACS is utilized to identify the model and whitening filter parameters. Figure 6-6 presents the channel output power spectrum estimated via the discrete Fourier transform (DFT) of the weighted two-dimensional (2-D) ACS, wherein a 60-dB sidelobe Dolph-Chebyshev lag window is applied to weight the data along each axis (the 2-D ACS for an array output sequence is defined in Volume II). This spectrum estimation approach is known as the weighted Blackman-Tukey method, independent of the lag window type that is applied. Characteristically, the spectrum in Figure 6-6 exhibits strong suppression of sidelobe leakage and reduced spectral feature resolution. Notice the clutter ridge along the cross-diagonal (clutter ridge slope $\nu=1$), with mainlobe centered at $f_{cd0} = 0.2494$ and $f_{cs0} = 0.2493$ normalized Doppler and spatial frequencies, respectively. The clutter ridge exhibits $J-1=7$ peaks, as expected. Notice also the two jammer ridges, centered at normalized spatial frequencies $f_{js} = 0.2113$ and $f_{js} = -0.3214$, and flat over the normalized Doppler frequency domain. The parameters in Table 6-2 result in a clutter-to-noise ratio (CNR) of 47.75 dB, and a jammer-to-noise ratio (JNR) of 38.0 dB.

PARAMETER TYPE	PARAMETER (UNITS)	VALUE
RADAR ARRAY SYSTEM	Number of linear array elements, J	8
	Number of points in one CPI, N	64
	Number of elevation axis elements, Je	4
	Array mainbeam azimuth angle, phi0 (deg)	30
	Peak transmitted power, Pt (kW)	200
	Pulse (uncompressed) duration, Tu (μsec)	200
	Pulse repetition frequency (Hz)	300
	Radiation frequency, fC (MHz)	450
	Receiver bandwidth, fB (MHz)	4
	Transmit pattern gain, Go (dB)	22
	Receive element gain, Ge (dB)	4
	Receive element backlobe pattern attenuation, Gb (dB)	-30
	Noise figure, Fn (dB)	3
	System losses, Ls (dB)	4
	Transmit pattern array option, patopt	UNIFORM
SURVEILLANCE SCENARIO	Platform altitude, Hp (km)	9
	Platform velocity, Vp (m/sec)	50
	Range to principal ground clutter ring, rc (km)	130
	Aircraft platform crab angle, gamma (deg)	0
TARGET	Narrowband process amplitude, a	0.99
	Target radial velocity, Vt (m/sec)	33.333
	Target azimuth angle, phit (deg)	-30
	Target elevation angle, thetat (deg)	0
	Signal-to-noise ratio, SNR (dB)	0
INTERFERENCE	Jammer azimuth angle, phii (deg)	25; -40
	Jammer elevation angle, thetai (deg)	0; 0
	Jammer power, vari	3310; 3000
GROUND CLUTTER	Number of ground patches illuminated by mainbeam, Nc	361
ARRAY NOISE	Receiver noise power per channel, varn	1
SIMULATION PARAMETERS	Number of block rows/columns in Hankel matrix, L	6
	Number of realizations used in filter design, Nrd	10
	Number of realizations used in filter evaluation, Nre	10
	Data window sidelobe level (for plots), dwindb (dB)	60

Table 6-2. Scenario, system, and simulation parameters for baseline simulation analyses.

A tenth-order innovations representation (IR) model was identified using $M=13$ lags of the true ACS ($L=6$ block rows and columns in the Hankel matrix of the CC algorithm). Figure 6-7 presents the power spectrum of the output of the tenth-order state variable model driven by a temporally- and spatially-uncorrelated sequence. The seven clutter ridge peaks are noticeable, and the spectrum has a ridge at each jammer's spatial frequency. This spectrum exhibits much higher resolution than the spectrum in Figure 6-6, as expected from an analytical model.

The canonical correlations, $\{\rho_i | i=1, \dots, J_L\}$ where $J_L=48$, for this case are presented in Figure 6-8, and the normalized mutual information parameters, $\{\eta_i | i=1, \dots, J_L\}$, obtained from the canonical correlations (Equation 3-44) are presented in Figure 6-9. Referring to Figure 6-9, the dashed (--) horizontal line represents a mutual information threshold of 0.997, and the dashed (--) vertical line represents the model order selected for that threshold. Model order ten was selected because for an equispaced element linear array model order $J-1=7$ suffices to provide whitening of channels 0 through $J-2$, but additional temporal dynamics are required to whiten channel $J-1$ also (with the jammers absent, there are seven canonical correlations with value slightly less than unity, instead of five as in Figure 6-8). A map of the IR model multivariable poles and zeros is presented in Figure 6-10, using the definition of multivariable system zeros proposed by Davison and Wang (1974; 1976). The identified tenth-order model has ten multivariable poles and ten multivariable zeros, and they reverse roles for the whitening filter (the IR model poles and zeros become the zeros and poles, respectively, of the whitening filter). This is a property of the IR and its inverse for the Davison-Wang definition of transmission zeros. It is appropriate to mention herein that the MATLAB routine *tzero* of the Signal Processing Toolbox, which calculates transmission zeros, gives incorrect results in cases where the data is complex-valued.

Routine *tzero* is an implementation of the numerical algorithm of Emami-Naeini and Van Dooren (1982) to calculate transmission zeros. However, SSC tested the MATLAB code, and discovered several inconsistencies; thus, SSC generated its own routine based on the Laub-Moore numerical algorithm for the Davison-Wang multivariable transmission zeros.

Figures 6-11 through 6-13 present three different views of the power spectrum of Figure 6-7. The top view, Figure 6-11, shows the clutter ridge with mainlobe at approximately $f_d = 0.25$ and $f_s = 0.25$ normalized Doppler and spatial frequencies, respectively, as well as the jammer ridges at approximately $f_s = 0.21$ and $f_s = -0.32$ normalized spatial frequencies. The jammer ridges are noticeable also in the projection to the spatial frequency axis, Figure 6-12. The projection to the Doppler frequency axis, Figure 6-13, complements the other figures.

Four different views of the clutter process whitening filter (both temporal and spatial whitening) power spectrum are presented in Figures 6-14 through 6-17. The 3-dimensional view in Figure 6-14 is from the same perspective as Figures 6-6 and 6-7 to allow direct comparison. The top view, Figure 6-15, shows the clutter notch at unity slope (along the cross-diagonal), a notch centered at approximately $f_s = 0.25$ to cancel the clutter mainlobe (centered at $f_{cd0} = 0.2494$ and $f_{cs0} = 0.2493$) and the jammer at $f_{is} = 0.2113$, and a notch centered at approximately $f_s = -0.32$ to cancel one of the secondary lobes and the jammer at $f_{is} = -0.3214$. Figure 6-16 presents the projection to the spatial frequency axis, wherein the notches at approximately $f_s = 0.21$ and $f_s = -0.32$ are appreciated better. Notice that the notch at $f_s = 0.21$ is deeper, as expected, since the jammer at $f_s = 0.21$ has more power (Table 6-2). The projection to the Doppler frequency axis is presented in Figure 6-17.

Consider now the case where an estimate of the true ACS is utilized to identify the IR model and whitening filter parameters. The ACS estimate is generated by averaging ten biased, time-averaged estimates of the ACS (parameter N_{rd} in Table 6-12). A modified, averaged periodogram of the channel output process is presented in Figure 6-18. The term "modified" accounts for the fact that a 60-dB Dolph-Chebyshev data window is applied along each axis to the data matrix prior to application of the 2-D DFT. Following application of the DFT, the power at each frequency is calculated to obtain the periodogram, and ten statistically-independent periodograms are averaged. As expected, this spectrum is very similar to the Blackman-Tukey spectrum in Figure 6-6, and the clutter ridge and the two jammer ridges are distinguished easily. A tenth-order IR model was identified using only $M=13$ lags of the estimated ACS, and the power spectrum of the IR model output is presented in Figure 6-19. This spectrum compares well with the spectrum in Figure 6-7, which was identified using the true ACS. Notice that the width of the jammer ridges is comparable in both spectra, although the IR model based on the true ACS has sharper features. Also, the clutter ridge mainlobe and sidelobes are defined well in Figure 6-19.

The canonical correlations, $\{\rho_i | i=1, \dots, JL\}$ where $JL=48$, for this case are presented in Figure 6-20, and the normalized mutual information parameters, $\{\eta_i | i=1, \dots, JL\}$, obtained from the canonical correlations (Equation 3-44) are presented in Figure 6-21. Notice that the first five canonical correlations are measurably less than unity (compare with Figure 6-8), and notice also that the "knee" in the curve at index value 10 is more marked than in Figure 6-8. Other independent runs based on an ACS estimated in the same manner result in very similar curves. With respect to Figure 6-21, the dashed (--) horizontal line represents a mutual information threshold of 0.96, and the dashed (--) vertical line represents the model order selected for that threshold. Model

order ten was selected to allow comparison with the previous case. A map of the IR model multivariable poles and zeros is presented in Figure 6-22. Pole locations are very similar to those in Figure 6-10, but there are marked differences in the locations of the multivariable zeros. The main difference is that five zeros are close to the origin in this case, whereas all zeros are spread out in the case based on the true ACS. This is the cause of the broader spectral features of the IR model power spectrum in Figure 6-19 in relation to the IR model power spectrum in Figure 6-7.

The whitening filter was applied to ten independent realizations (parameter N_{re} in Table 6-12) of the channel output process, and the unweighted periodograms of the ten residual sequences were averaged. The resulting spectrum is presented in Figure 6-23 in the same scale and 3-D perspective as the channel output spectrum in Figure 6-18. Notice that the residual spectrum is white, since it oscillates by just a few dB about the noise floor.

An estimate of the channel output ACS, $\{\Lambda_m | m=0, 1, \dots, M\}$, was generated using the identified system model parameters in Equations (2-12) and (2-14). Actually, ACS lags beyond M can be generated also, but M is the number of lags used to identify the model. The real and imaginary components of the true (solid line) and model (dashed line) ACS for channel 0 are presented in Figures 6-24 and 6-25, respectively. Notice that the fit is very good at all lags, specially considering that the model parameters for these results are identified using an estimated channel output ACS (as opposed to the true channel output ACS). Similar results are true for the ACS of the other channels.

The capability for moving target detection is demonstrated in Figure 6-26, wherein the unweighted periodogram of the channel 4 residual of the clutter-only (null) hypothesis filter is presented

for the case where the channel output sequence includes a target at 0 dB SNR (alternative hypothesis is true). Target parameters in Table 6-2 place the target spectral peak at the following normalized frequencies: $f_{td1} = 0.0834$ and $f_{ts1} = -0.25$. In Figure 6-26, the solid line (—) corresponds to the channel 4 residual periodogram averaged over ten residual realizations, the dashed line (--) corresponds to the theoretical (model-based) residual power for that channel, the dash-dot line (-·-) corresponds to the realized residual power averaged over ten realizations, and the dotted line (:·:) corresponds to the $\pm$ one-sigma bounds for a white process with the theoretical residual power. The theoretical residual power for channel 4, the (5,5) element of $\Omega(H_0)$, is 25.04 dB. And the realized residual power, which includes the target power, is 25.34 dB. Notice that all the spectrum points except one are within the $\pm$ one-sigma bounds, which is an acceptable condition. Notice also that the spectrum is white (approximately flat). These bounds include a factor to account for the frequency-domain averaging. The moving target is detected easily at a normalized Doppler frequency in the vicinity of $f_d = 0.083$, since it is approximately 5 dB above the noise spectrum floor.

The results presented herein indicate that the multichannel innovations-based detection configuration (Figure 6-5) using state variable model hypothesis filters is a feasible option for moving target detection in airborne surveillance scenarios using phased array radar systems. Further analyses should be carried out to establish the detection performance as a function of key parameters and in relation to other methods. Such methods include the optimum joint-domain algorithm and its approximations.

TRUE LOG POWER SPECTRUM (Weighted ACS)

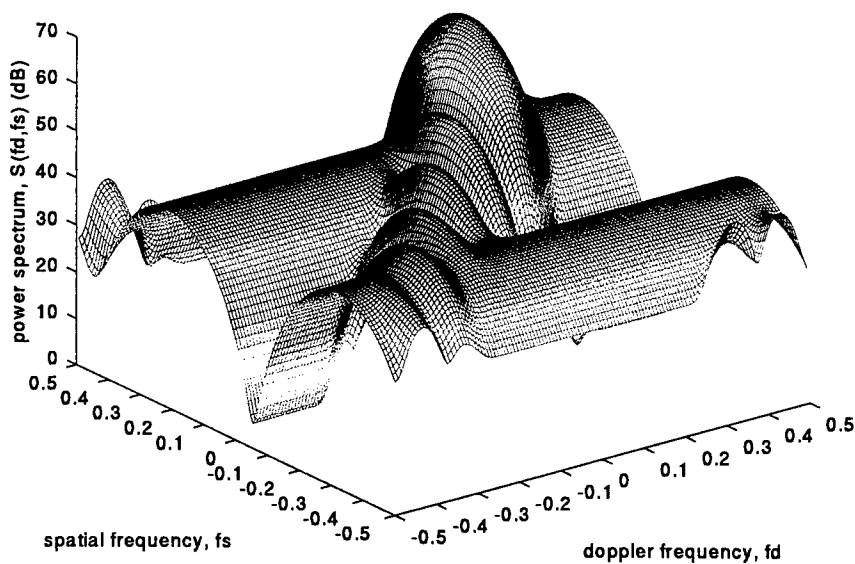


Figure 6-6. Logarithm of the normalized channel output power spectrum (Blackman-Tukey spectrum estimate; true ACS case).

MODEL OUTPUT LOG POWER SPECTRUM

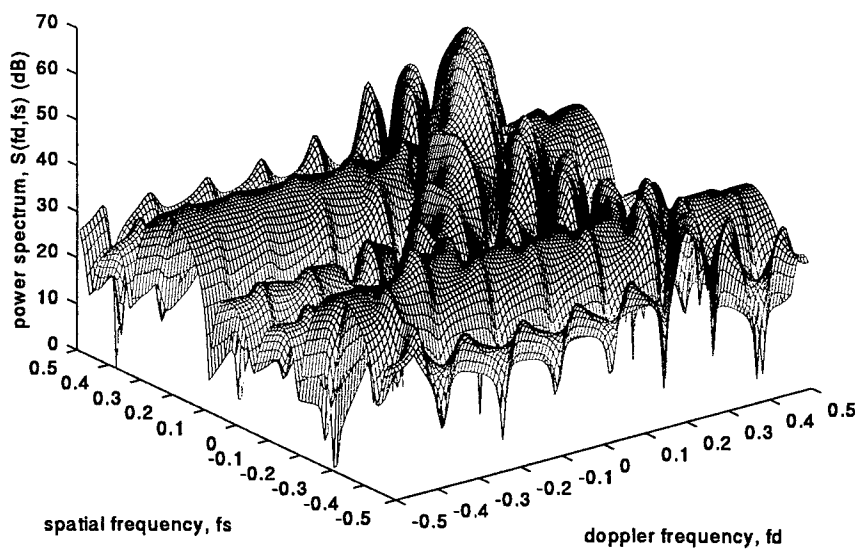


Figure 6-7. Logarithm of the innovations representation model power spectrum (true ACS case).

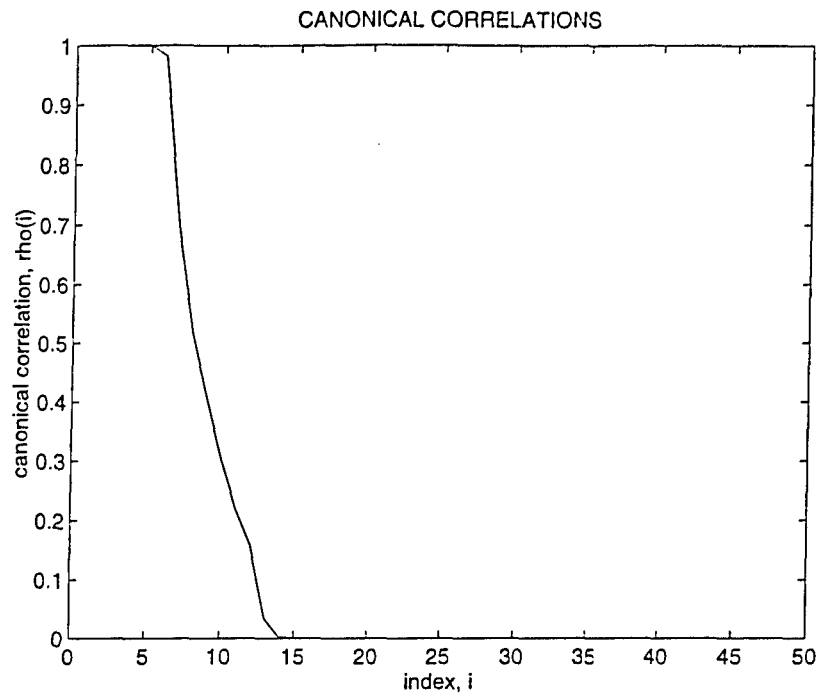


Figure 6-8. Canonical correlations of the channel output process (true ACS case).

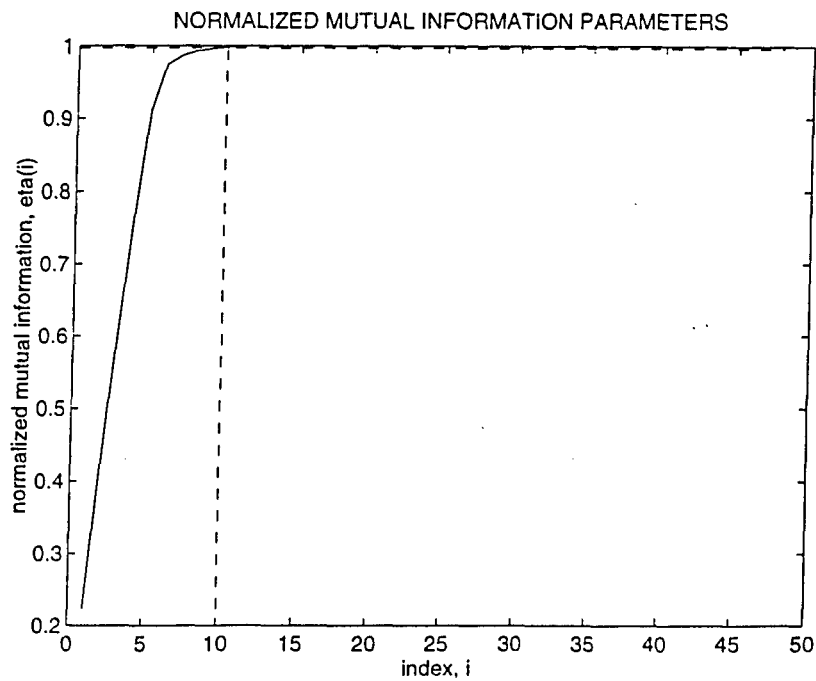


Figure 6-9. Normalized mutual information for model order selection (true ACS case).

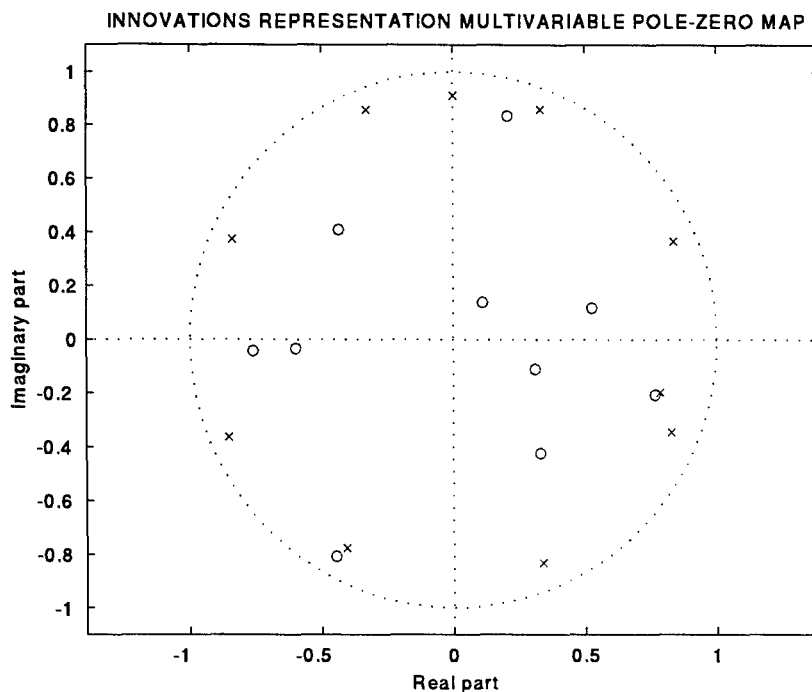


Figure 6-10. Map of the multivariable poles and zeros of the tenth-order state-space model (true ACS case).

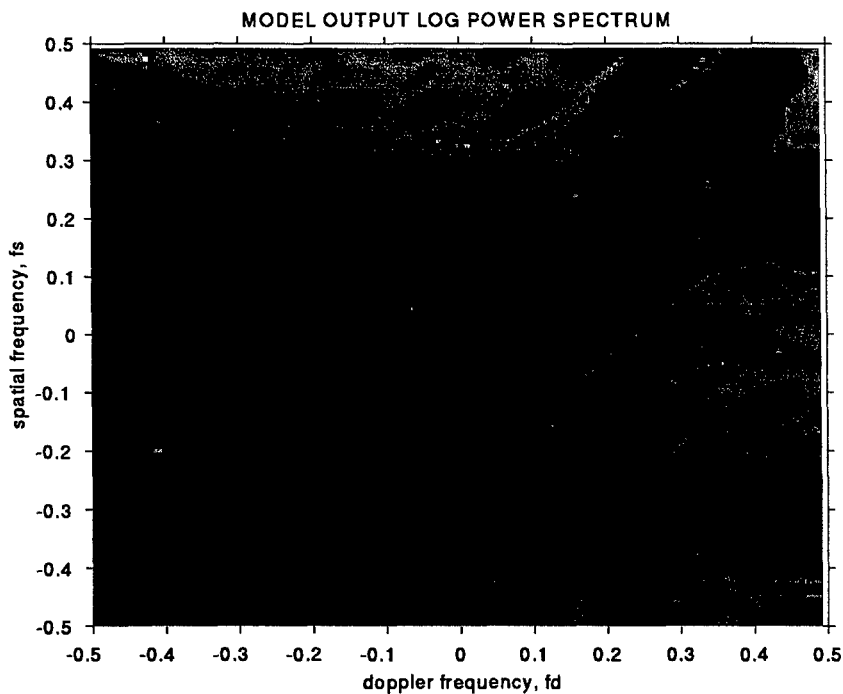


Figure 6-11. Top view of the logarithm of the innovations representation model power spectrum (true ACS case).

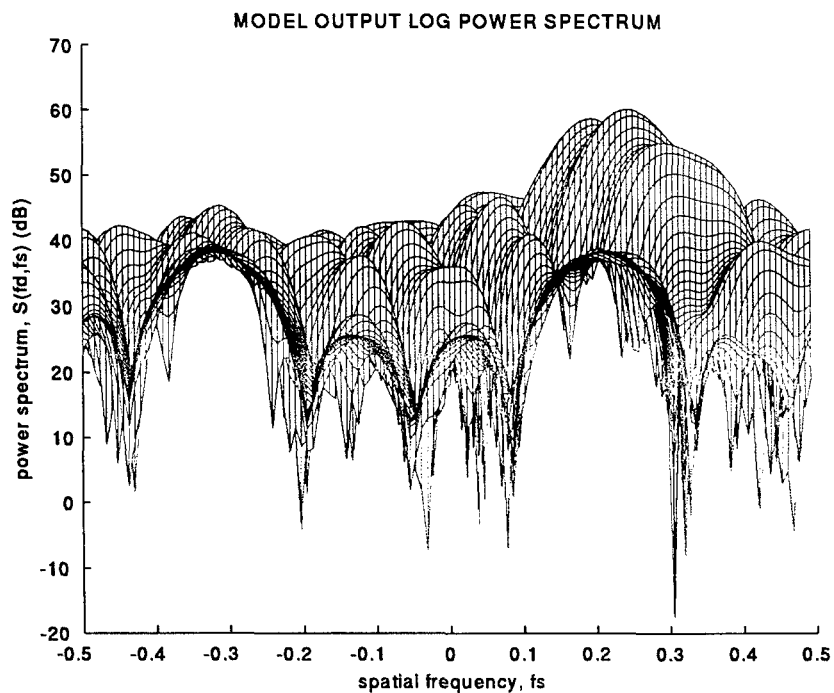


Figure 6-12. Spatial-frequency axis projection of the logarithm of the state-space model power spectrum (true ACS case).

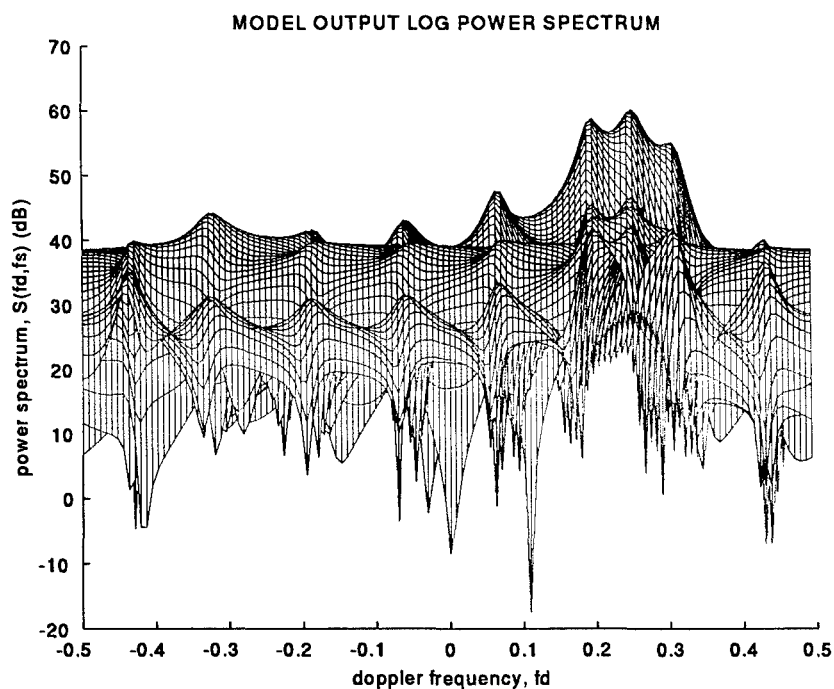


Figure 6-13. Doppler-frequency axis projection of the logarithm of the state-space model power spectrum (true ACS case).

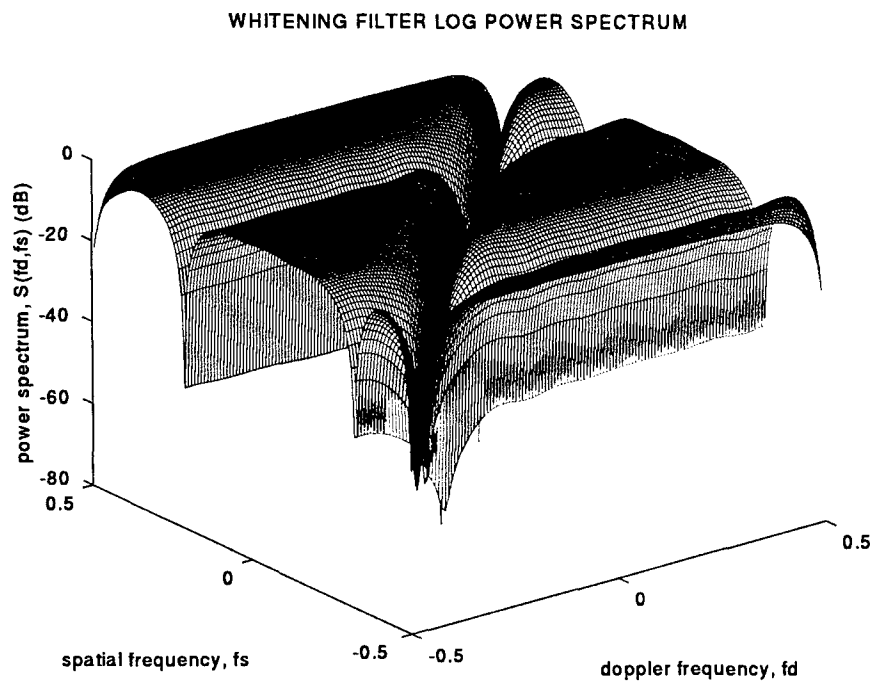


Figure 6-14. Logarithm of the whitening filter power spectrum (true ACS case).

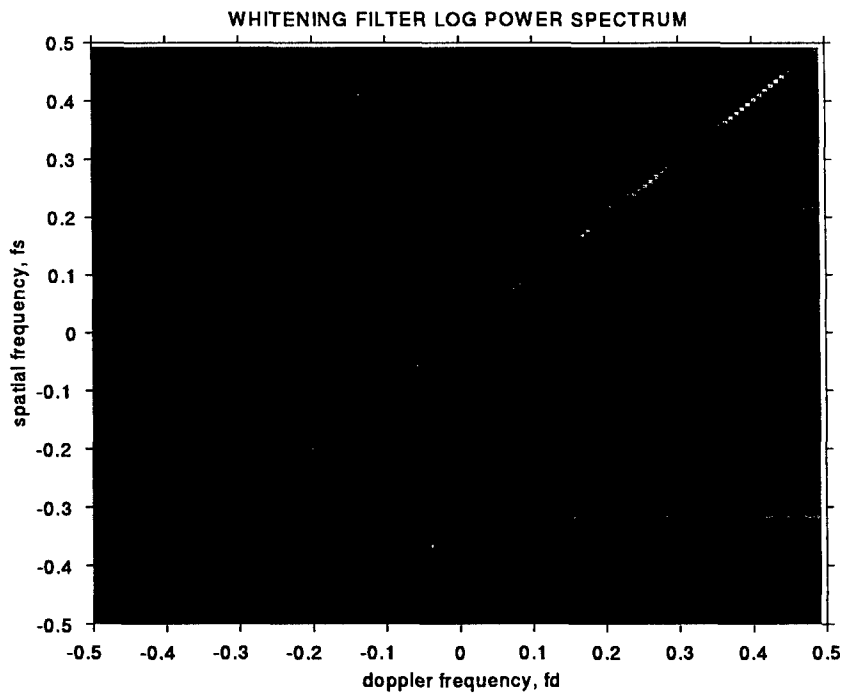


Figure 6-15. Top view of the logarithm of the whitening filter power spectrum (true ACS case).

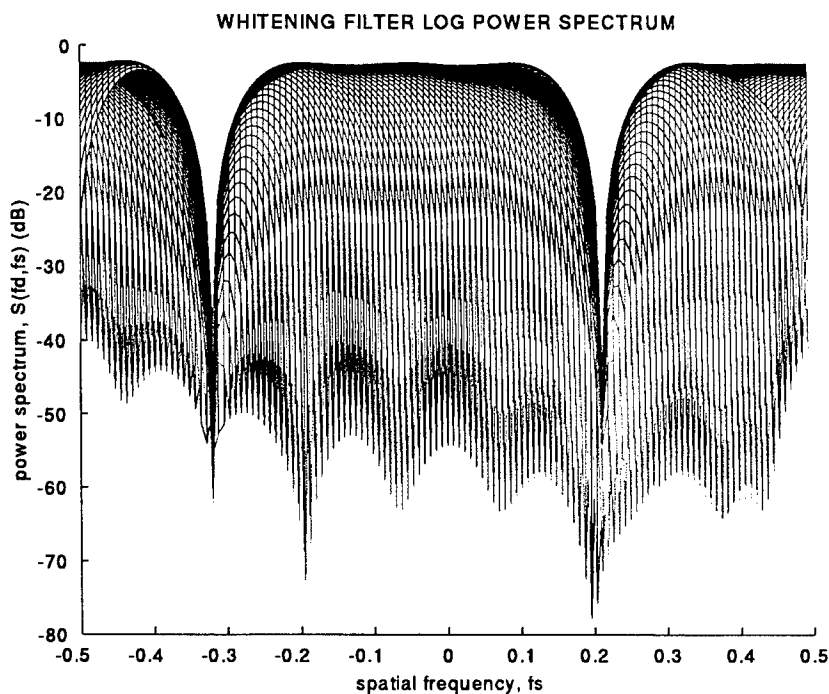


Figure 6-16. Spatial-frequency axis projection of the logarithm of the whitening filter power spectrum (true ACS case).

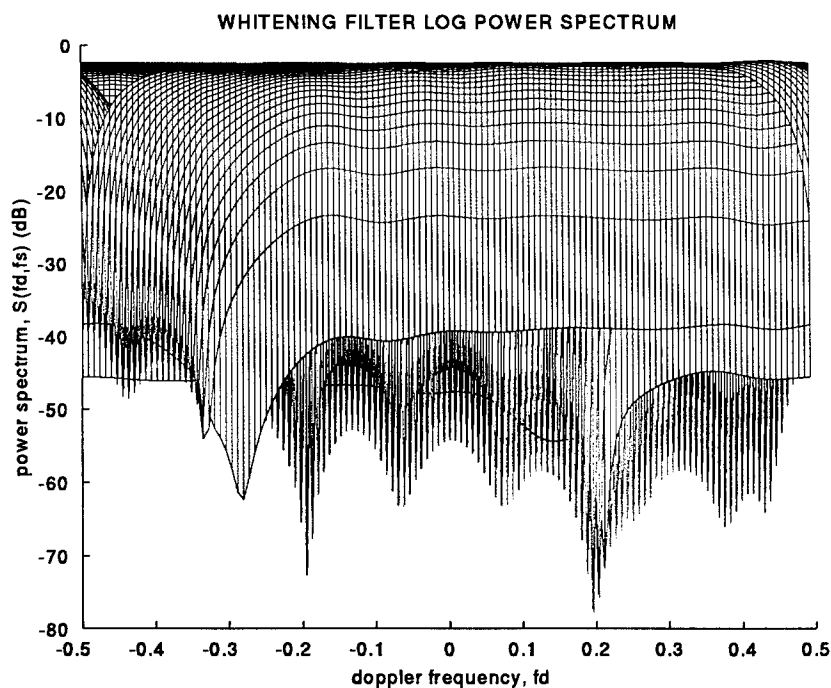


Figure 6-17. Doppler-frequency axis projection of the logarithm of the whitening filter power spectrum (true ACS case).

ARRAY OUTPUT LOG POWER SPECTRUM (Modified Averaged Periodogram)

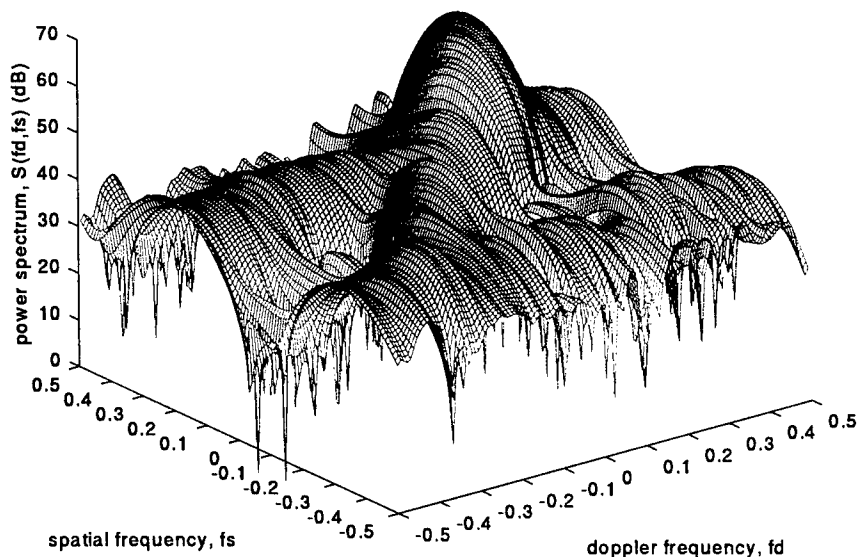


Figure 6-18. Logarithm of the channel output power spectrum (modified, averaged periodogram; biased, time-averaged ACS case).

MODEL OUTPUT LOG POWER SPECTRUM

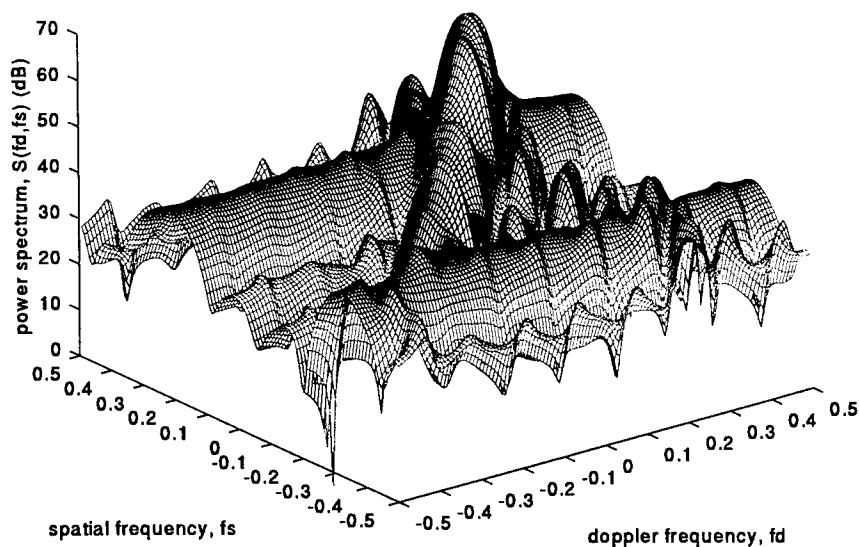


Figure 6-19. Logarithm of the innovations representation model power spectrum (biased, time-averaged ACS case).

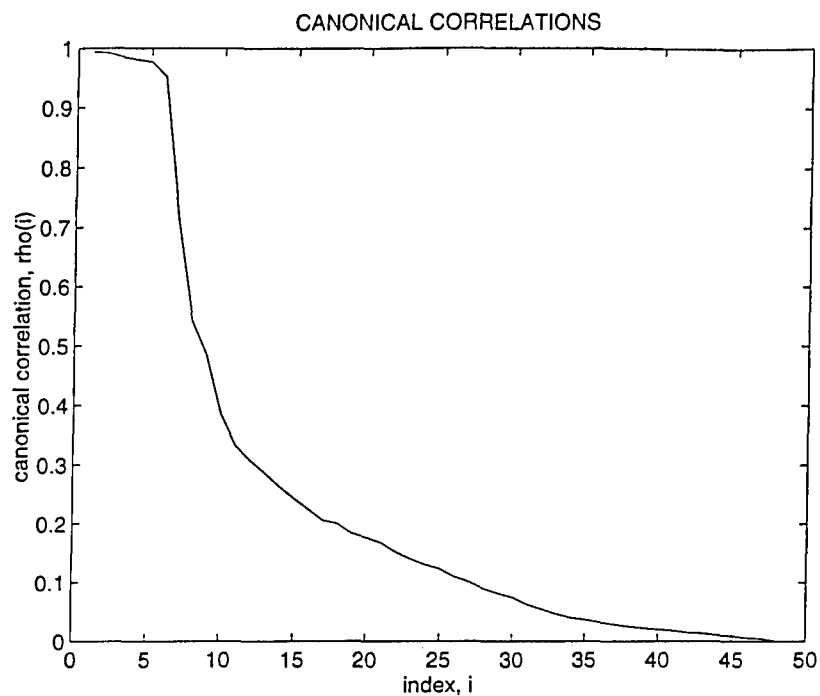


Figure 6-20. Canonical correlations of the channel output process (biased, time-averaged ACS case).

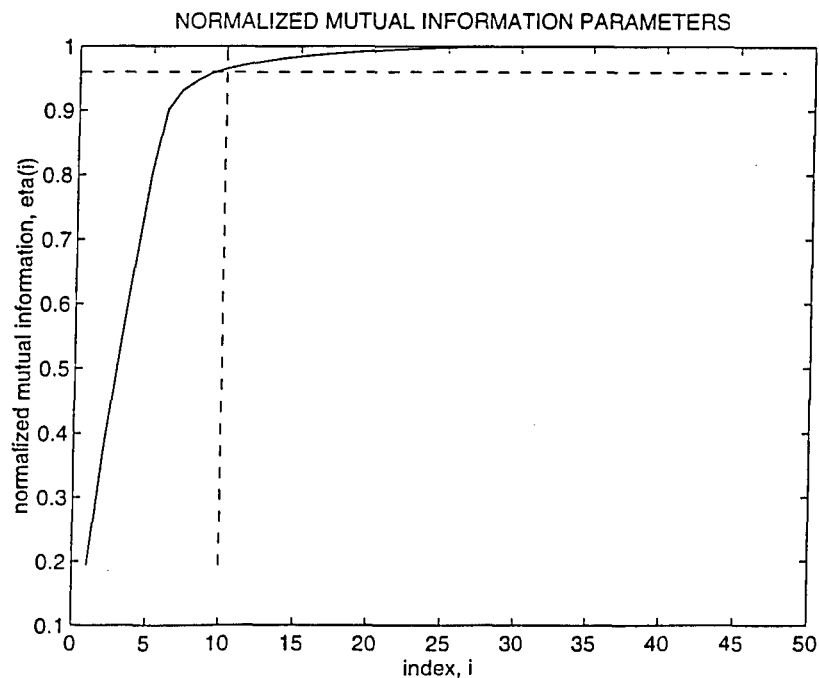


Figure 6-21. Normalized mutual information for model order selection (biased, time-averaged ACS case).

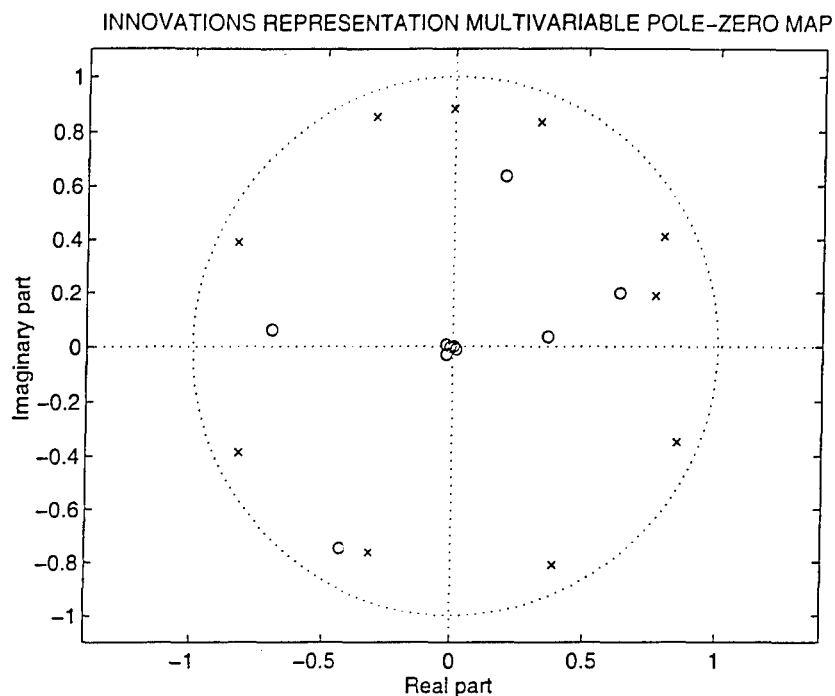


Figure 6-22. Map of the multivariable poles and zeros of the tenth-order state-space model (biased, time-averaged ACS case).

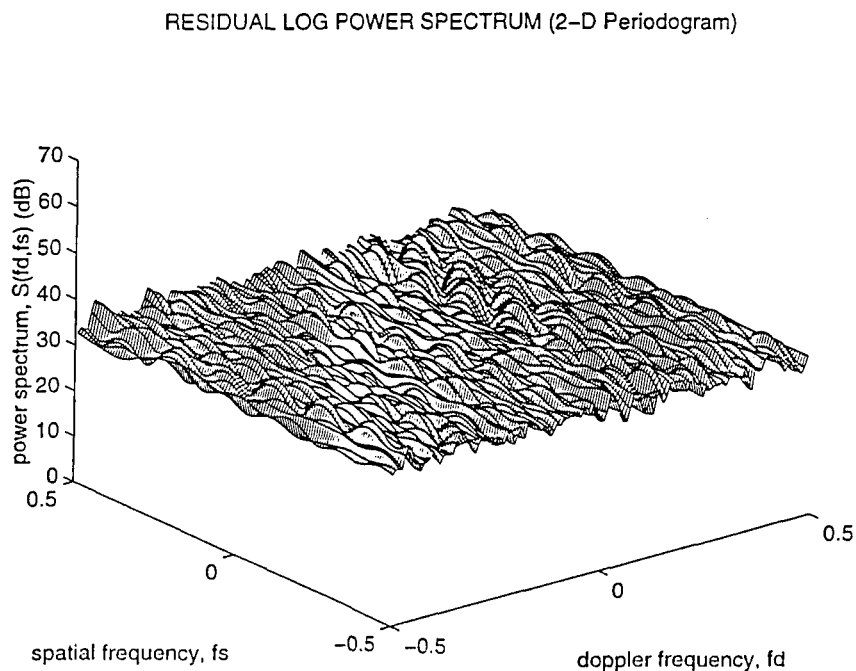


Figure 6-23. Logarithm of the residual process power spectrum (periodogram; biased, time-averaged ACS case).

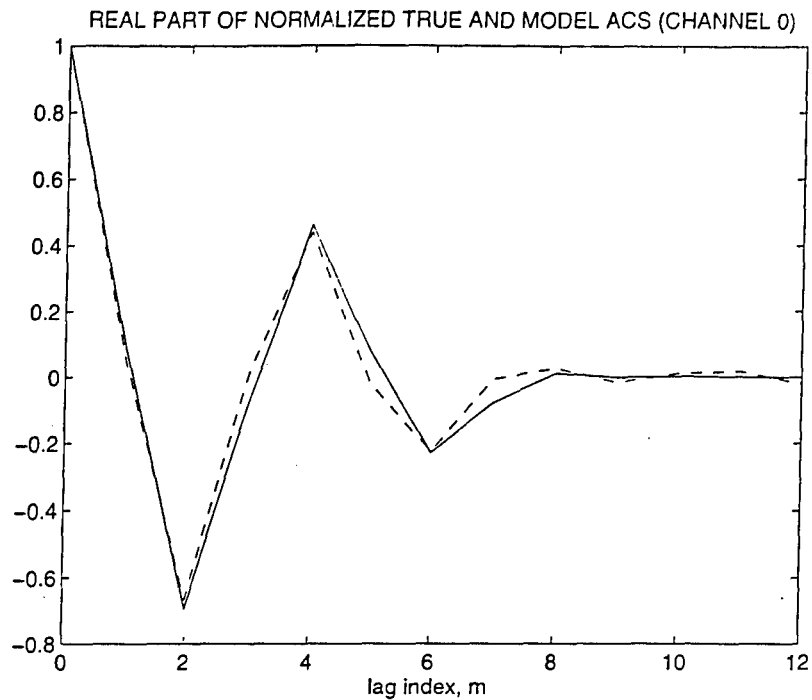


Figure 6-24. Real part of the normalized channel 0 true and model ACS (biased, time-averaged ACS case).

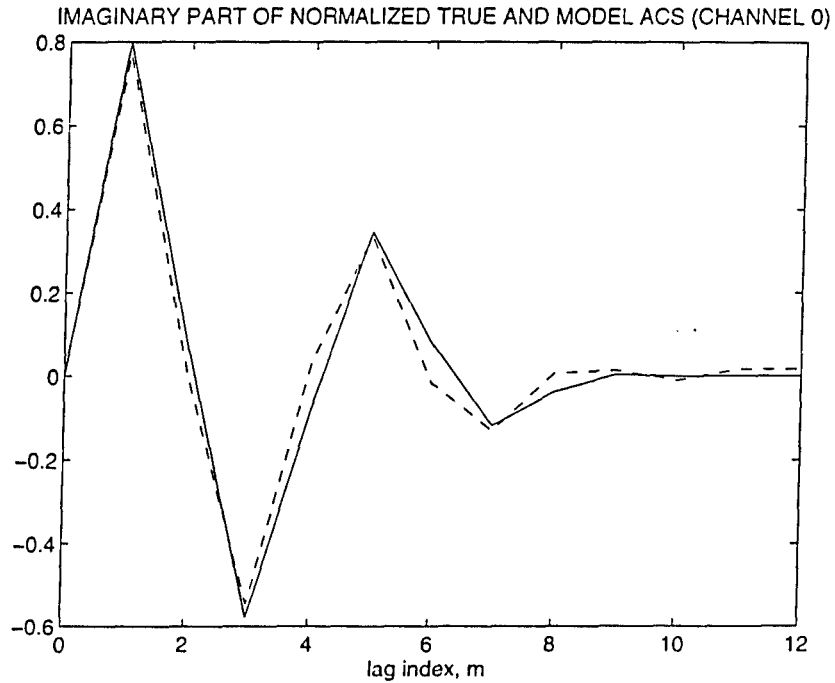


Figure 6-25. Imaginary part of the normalized channel 0 true and model ACS (biased, time-averaged ACS case).

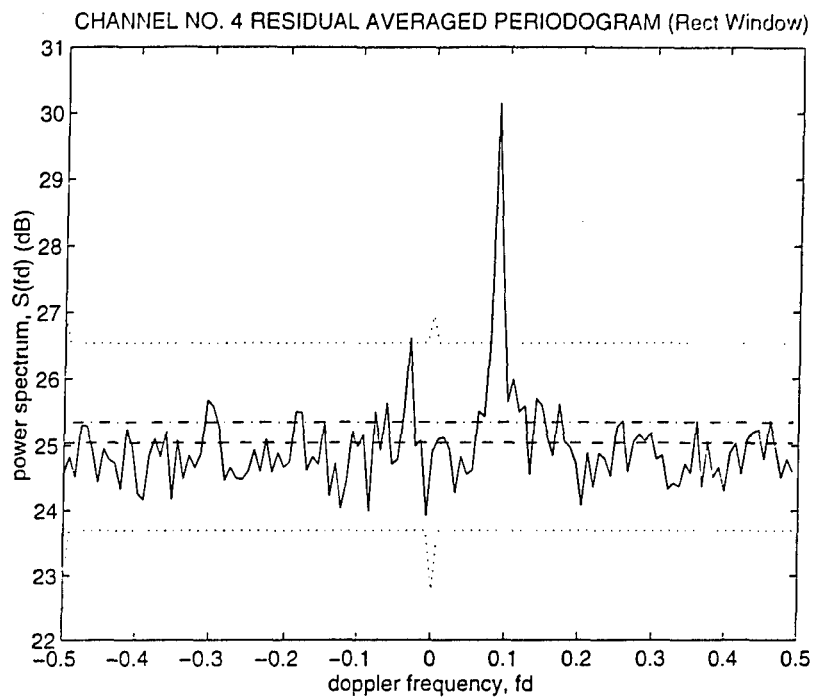


Figure 6-26. Logarithm of the channel 4 residual power spectrum (averaged periodogram; biased, time-averaged ACS case with 0 dB SNR target).

7.0 ECG DIAGNOSTICS APPLICATION

Multilead (multichannel) electrocardiography was selected in Phase II as an area for dual-use investigation since multichannel data is available inherently, and the approaches used in the industry are based on single-channel methods, as far as SSC has been able to assess. Model-based multichannel methods allow utilization of the cross-channel information in the multilead electrocardiogram (ECG) in order to enhance diagnostic capability. The multichannel identification techniques discussed in this Final Report can generate low-order models to represent effectively the cardiac abnormalities considered in this task. Modeling and diagnostic determination results are presented herein for normal ECGs and two cardiac conduction abnormalities.

Early results obtained in the first year of this Phase II program were presented at the Fourth Annual IEEE Dual Use Technologies And Applications Conference (Román and Davis, 1994). Updated and more extensive analyses were presented at the American College of Cardiology 45th Annual Scientific Session (Román et al., 1996a), and at 23rd Annual Computers in Cardiology conference (Román et al., 1996b).

7.1 Multichannel Electrocardiography

The human heart is a sophisticated pumping system that functions in a cyclical sequence of muscular contractions and relaxations of the myocardial cells in the heart muscle, as described by Wagner (1994) and Guyton (1991). These muscular actions are induced by action potentials, which are rapid changes in the electric potential of cell membranes. During an action potential cycle in a cell, the cell membrane goes from the large negative polarization state of the resting stage, through a depolarization stage to a positively-polarized state, and through

a repolarization stage back to the resting stage. In a myocardial cell the action potential cycle is activated by an external source, and a mechanical cycle of physical contraction and relaxation accompanies (with a slight delay) the electrical polarization cycle as the cell goes through an action potential cycle. Action potentials propagate from one region of a myocardial cell to the rest of the cell, and from one myocardial cell to another.

Action potentials also propagate through various groupings of specialized fibers at a rate which is several times faster than myocardial cell-to-cell propagation. These fiber groupings constitute the cardiac conduction system and are referred to as nodes, pathways, bundles, and bundle branches. Conduction system fibers lack contractile capability, but are efficient propagators of the action potential impulse. Additionally, they are capable of automatic activation of the action potential cycle, a feature referred to as self-excitatory. A sketch of the human heart and the cardiac conduction system is presented in Figure 7-1, which is adapted from Guyton (1991) and Wagner (1994). Notice in Figure 7-1 the base-to-apex reference axis which runs from the center of the base (top) of the heart to the center of the apex (bottom) of the heart, and indicates the natural orientation of the heart.

The sinus node (also referred to as sinoatrial or S-A node) of the cardiac conduction system is located at the top of the right atrium, as indicated in Figure 7-1. This node controls the rate of beat of the entire heart. Sinus node fibers are self-excitatory, which allows them to initiate the cardiac cycle. The electric action potentials that originate in the sinus node fibers pass on to the three internodal pathways that extend from the sinus node to the atrioventricular (A-V) node along the atrial walls. From the A-V node the action potential passes to the A-V bundle. In the A-V node and A-V bundle a delay is introduced in

the propagation of the action potentials. This delay allows the atria to discharge all its contents into the ventricles before the ventricles contract. The A-V bundle branches into two parts, the right bundle branch and the left bundle branch. The right bundle branch runs downward along the length of the right ventricle and divides into smaller branches which further divide into smaller and smaller branches. In this manner most parts of the right ventricle are reached directly. The left bundle branch runs downward along the length of the left ventricle and similarly divides into smaller branches to reach most parts of the left ventricle. This electrical activity is repeated every cardiac cycle.

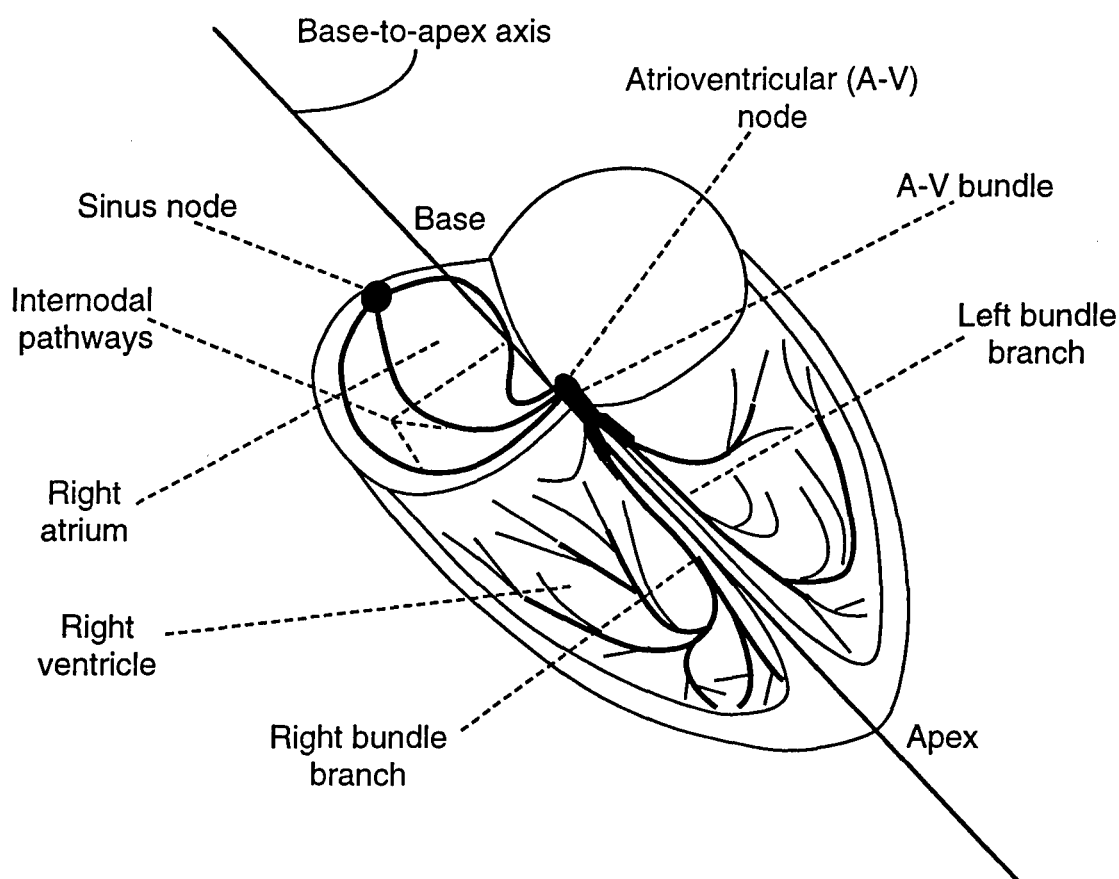


Figure 7-1. Cardiac conduction system of the human heart.

An ECG is a recording of the electrical activity of the heart taken at the surface of the body. Most modern ECG systems generate multiple, simultaneous recordings. Each recording is made with a pair of electrical leads, with a third lead serving as the ground reference. The features of an ECG waveform correspond to the activation sequence of the cardiac conduction system summarized above. Each independent lead placement configuration generates a trace with distinct characteristics. An important lead placement configuration is the base-to-apex configuration illustrated in Figure 7-2. The normal ECG waveform recorded in the base-to-apex lead configuration exhibits the form shown Figure 7-3. As indicated in Figure 7-3, the main wave features of this waveform are denoted by the letters P through U. The initial wave of the cardiac cycle, denoted as P, represents activation of the atria. Activation of the right atrium is represented by the first part of the P wave. The middle of the P wave coincides with completion of right atrial activation and initiation of left atrial activation. The final section of the P wave represents completion of left atrial activation. The A-V node is activated during the middle of the P wave, and this activation proceeds slowly toward the ventricles during the final segment of the P wave. The wave that represents electrical recovery of the atria is usually obscured by the waves representing ventricular depolarization. The next group of waves recorded is the QRS complex, which represents the activation of the ventricles. By convention, a negative wave at the onset of the QRS complex is called a Q wave. The predominant positive portion of the QRS complex is called the R wave, regardless of whether or not it is preceded by a Q wave. The negative deflection following the R wave is called an S wave. The wave in the ECG trace that represents recovery of the ventricles is called the T wave. The T wave is sometimes followed by another small positive wave called the U wave (Wagner, 1994). The source of the U wave is unknown. The various waves present in Figure 7-3 are representative of the

waves in ECG traces of other lead pairs, although other features (such as an M-shaped or split R wave) are common in other leads.

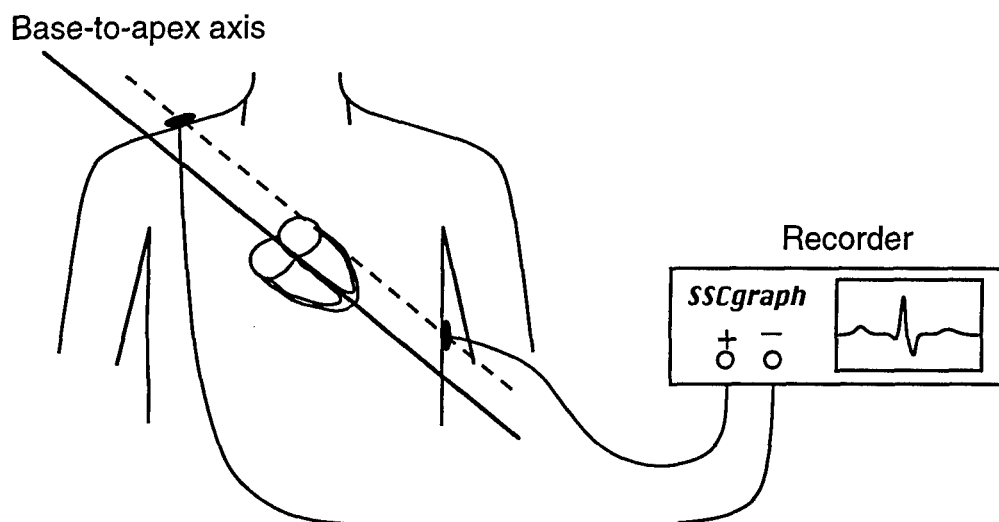


Figure 7-2. Base-to-apex lead placement configuration.

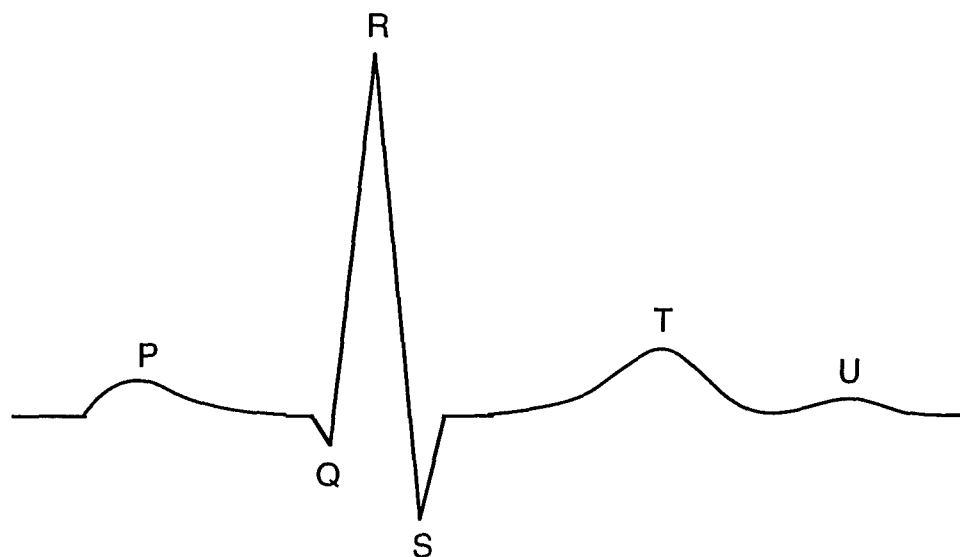


Figure 7-3. Single-cycle ECG trace for the base-to-apex lead placement configuration.

The time interval from the onset of the P wave to the onset of the QRS complex is called the PR interval, and it is a measure of the time between the onsets of activation of the atrial and ventricular myocardium. The QRS interval measures the time from beginning to end of ventricular activation. Since activation of the thicker left ventricle requires more time than the right ventricle, the terminal portion of the QRS complex represents only left ventricular activation. The ST segment is the interval between the end of ventricular activation and the beginning of ventricular recovery.

Conduction abnormalities occur when any of the components in the cardiac conduction system are damaged and fail to function as intended. Such failure can be complete or partial to various degrees. When a failure in the conduction fiber system occurs, the action potentials still propagate via the myocardial cell-to-cell mechanism. This is a slower mode, and is manifested in the ECG as an elongation of the waves in the QRS complex. Of interest herein is conduction blockage of the left or right bundle branches. These abnormalities are referred to as left bundle branch block (LBBB) and right bundle branch block (RBBB), respectively. LBBB and RBBB conditions are categorized further as either complete or incomplete, depending on the extent of the blockage.

Multichannel recording of the ECG is accomplished with a standard 12 lead configuration, as described by Wagner (1994). This set of leads consists of a six-lead frontal plane subset and a six-lead transverse plane (or precordial) lead subset. The frontal plane leads are placed on the limbs and are used to create an electrical picture of the heart at 30-degree angular intervals around the frontal plane of the heart. Actually, there is considerable redundancy in the frontal plane lead set because the traces from only two lead positions suffice to synthesize

algebraically the traces from the remaining four lead positions. The transverse plane of the heart is sensed with the precordial lead configuration, which are denoted as leads V1 through V6. Since the precordial leads provide a panoramic view of cardiac electrical activity progressing from the thinner right ventricle across the thicker left ventricle, the positive R wave normally increases in amplitude and duration from V1 to V4 or V5.

Within a single channel, distortions in the various component waveforms of the ECG and variations in the waveform interval durations can be indicative of abnormalities. Enhanced information is obtained by examining the signals from multiple leads. In fact, some abnormalities can be detected only by such means. Numerous automatic diagnostic programs exist to augment the physician's assessment (Willems et al., 1990). Most of these programs, however, are rule-based systems which operate on single-lead features in the multilead data. The underlying cross-channel information is not exploited directly in such systems.

Abnormality detection using ECGs can be formulated as a hypothesis testing problem, and the SSC model-based detection methodology can be applied. ECG traces can be modeled as a stochastic signal in additive zero-mean Gaussian noise (Zywietz, et al., 1990). Adopting a different viewpoint, ECG traces can be modeled also as a deterministic signal in additive zero-mean Gaussian noise. The SSC model-based detection methodology can be applied in either of these two contexts, and each of these two contexts involves a different modeling philosophy and identification algorithm. In a binary hypothesis formulation, the null hypothesis can be selected as the normal ECG case in additive noise, and the alternative hypothesis can be selected as the abnormal ECG case (encompassing all possible abnormalities) in additive noise. In the more general multiple hypothesis formulation, the null hypothesis can be selected as the normal ECG

case in additive noise, and one alternative hypothesis can be assigned to each abnormality selected for discrimination.

7.2 CSE Database

SSC procured the multilead database of the Common Standards for Quantitative Electrocardiography (CSE) for use in assessing the efficacy of multichannel modeling and identification techniques in ECG diagnosis. This database was developed by J. L. Willems and his associates (Willems, 1990) at the CSE Coordinating Center, Division of Medical Informatics, University of Leuven, Leuven, Belgium, over a number of years under the auspices of the Commission of the European Communities. As such it is the product of contributions from multiple European facilities. The database is available in compact disk read-only memory (CD ROM) media. This particular database was at first planned to be an annotated teaching database. However, the philosophy of the CSE coordinating center changed, and it was decided that it would become a testing database. Consequently, the diagnostic annotations are withheld from purchasers of the database. Unfortunately, the product documentation did not reflect this fact prior to our procurement of the database. SSC addressed this drawback by establishing a working relation with Dr. Victor G. Dávila-Román, a research cardiologist at Washington University School of Medicine, St. Louis, MO. Dr. Dávila-Román collaborated with SSC in defining the ECG discrimination problem reported herein, and provided diagnoses for many cases in the CSE database. In particular, Dr. Dávila-Román identified the normal, LBBB, and RBBB cases utilized in the validation procedure (Section 7.4).

In the CSE multilead database, 250 patient case records are divided into two sets of 125 ten-second digital recordings. Both normal and abnormal cases are included, and approximately 26 different abnormalities are represented. These recordings have

been taken simultaneously for the standard 12 leads and the 3 vectorcardiogram leads at 500 Hz sampling rate (2 msec sampling interval) with 16-bit resolution. This database is the premier ECG database accessible to the biomedical research community. It is regretful that the data is difficult to access and the documentation is deficient.

Jointly with the database, SSC received from the CSE Coordinating Center a copy of the CSE Database Display, Version 1.00, which is a software program developed at the Biomedical Systems Laboratory, School of Electrical Engineering, University of New South Wales (UNSW), Sydney, Australia. This biomedical research group also owns a copy of the CSE multilead database, and they developed the software to facilitate use of the database. The CSE Database Display software allows efficient access, display, and printout of the records in the CSE database. The software was provided on a 3.5" high-density IBM-compatible floppy diskette. Upon receipt of the database, SSC exercised the software with several of the ECG files, and noticed that the software was operating incorrectly in some cases. This was mentioned to Dr. Branko Celler at UNSW. Dr. Celler and his colleagues identified the problem in the software, and generated an updated version of the program. SSC received a copy of the corrected software, and has exercised it extensively. UNSW and SSC have agreed to share diagnostic information for the cases in the database.

7.3 Modeling and Discrimination Using CSE Data

Normal/abnormal ECG modeling and discrimination capability of the SSC model-based multichannel detection methodology is discussed herein. A scalar (single-input, single-output) state-space model is presented also for one of the selected leads.

These results were presented at the Fourth Annual IEEE Dual Use Technologies And Applications Conference (Román and Davis, 1994).

Leads V4, V5, and V6 of the precordial lead set were selected, resulting in a three-channel system. Waveform interval variations were excluded by using only the QRS complex segment of the ECG trace. This limits the complexity of the system model and reduces the number of computations at this early stage of the research. Two waveform records were excerpted from the CSE multilead cases: a normal QRS complex and a slightly abnormal QRS complex, corresponding to cases M01-011 and M01-081, respectively, in the CSE database filename notation. Several consecutive ECG cardiac cycles for the normal and abnormal cases are presented in Figure 7-4. Effects of spatial diversity can be investigated in the future by using non-adjacent leads such as V1, V4, and V6.

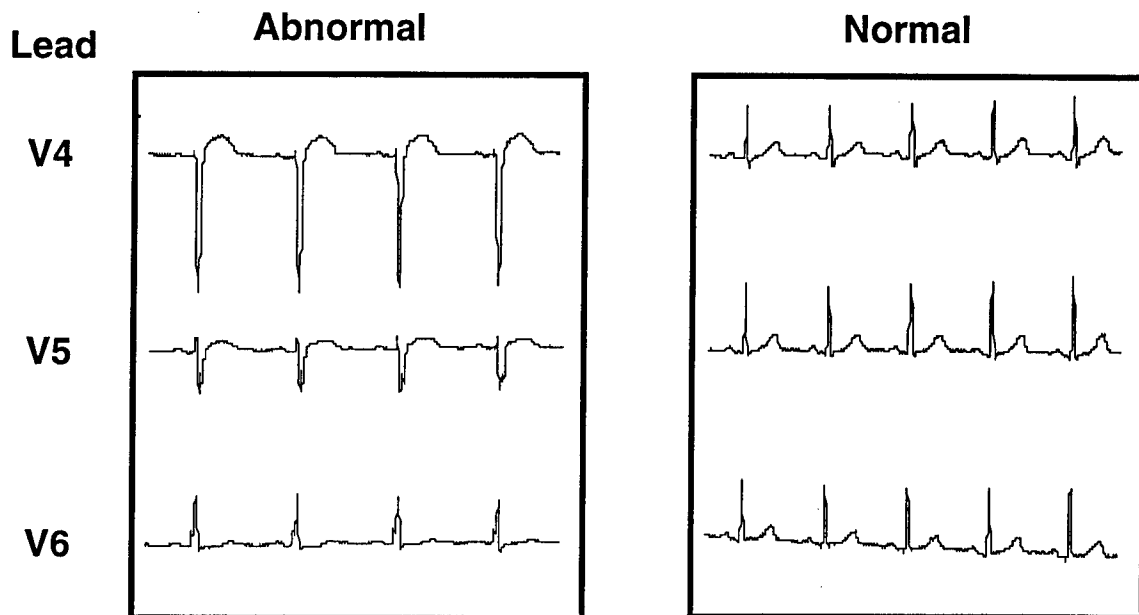


Figure 7-4. Selected cardiac cycles of the ECG traces used in the modeling and discrimination analysis.

Analysis And Simulation Procedure. The canonical correlations algorithm was selected for the analysis, using data from the selected simultaneously-recorded channels (Figure 7-4). For each channel, ten consecutive realizations of the QRS complex were selected. A QRS wave trace duration of 0.16 sec, corresponding to 80 data points, was removed from each of the ten cycles. These realizations included at least ten data points before and after the QRS complex in order to allow robustness of the covariance calculation with respect to data segmentation. Each QRS complex trace was pre-processed by removing the mean and dividing by the standard deviation of the 80-point sequence. A normalized 80-lag covariance matrix sequence was estimated for each of the ten three-channel vector data sequences, and the ten estimates were averaged to generate an averaged covariance matrix sequence. This procedure was carried out twice, first for the normal case and then for the abnormal case.

Multivariate state space models were generated for the three-channel (precordial leads V4, V5, and V6) normal and abnormal cases using the respective averaged covariance matrix sequences. Forty ($M = 2L = 40$) covariance matrix lags were used (out of the available 80 lags), and a sixth-order model was selected for both conditions (normal and abnormal). The transfer function models and temporal whitening filter residual sequences obtained are very similar for each of the three channels. Thus, results are presented herein only for channel 2 (lead V5).

A scalar state space model was generated for the scalar covariance sequence for channel 2 (lead V5). This allows direct, qualitative comparison of the single-channel results with the selected multichannel results. As in the multichannel case, forty ($M = 2L = 40$) covariance sequence lags were used for the normal QRS complex condition, and a sixth-order model was selected. For the

abnormal condition fifty ($M=2L=50$) covariance sequence lags were used, and a 14th-order model was selected.

The multichannel and single channel models for the normal and abnormal cases were used to define the temporal whitening filters. These filters were used to process 80-point QRS complex traces representative of the selected cases. Analysis of the characteristics (sample mean value; sample variance; sample covariance sequence) of the filtered residuals provides an indication of the methodology's capability for ECG diagnostics.

Simulation Results. Scalar modeling and discrimination results are presented first. A Blackman-Tukey (BT) estimate of the power spectrum computed using the 80-lag average covariance sequence was adopted as the "true" spectrum for comparison purposes. The "model" spectrum was obtained by direct evaluation of the identified model transfer function. Figure 7-5 presents the BT power spectrum and the identified single-channel model power spectrum (bold curve) for lead V5 of the normal QRS complex. Note that the relatively low order model (sixth-order) represents well the key features in the "true" spectrum. The BT power spectrum and the identified single-channel model power spectrum (bold curve) for lead V5 of the abnormal QRS complex are presented in Figure 7-6. Note that the "true" spectrum for this case has more features, which accounts for the higher model order. Comparison of the spectra in Figures 7-5 and 7-6 shows significant differences. This is expected because the respective ECG traces differ significantly (see Figure 7-4).

Figures 7-7 through 7-10 present the sample covariance sequence of the residuals obtained by filtering an 80-point QRS complex wave from the V5 lead for normal and abnormal conditions. The scalar whitening filters designed for the null and alternative hypotheses were used to generate the residuals. These figures

show that a white (uncorrelated) residual is obtained when the signal matches the whitening filter (as in Figures 7-7 and 7-9), and that a colored residual is obtained when the signal and the whitening filter are mismatched (as in Figures 7-8 and 7-10).

Multichannel modeling results are presented in Figures 7-11 and 7-12. Specifically, Figure 7-11 presents the BT power spectrum (based on the 80-lag average covariance matrix sequence) and the multichannel model power spectrum (bold curve) for lead V5 of the normal QRS complex. Note that the model does not fit the key features in the "true" spectrum as well as in the single-channel case. The BT power spectrum and the multichannel model power spectrum (bold curve) for lead V5 of the abnormal QRS complex are presented in Figure 7-12. Again, in this case the model fit to the "true" spectrum is not as good as in the single-channel case. This apparent poor spectral fit of the multichannel model is discussed in the Comments paragraphs below.

Figures 7-13 through 7-15 present the (2,2) element (which corresponds to the V5 lead) of the sample covariance matrix sequence of the vector residuals obtained by multivariate filtering an 80-point QRS complex vector sequence for the normal and abnormal conditions. The multichannel temporal whitening filters designed for the null and alternative hypotheses were used to generate the residuals. As in the scalar case, these figures show that a white (uncorrelated) residual is obtained when the signal matches the whitening filter (Figures 7-13 and 7-15), and that a colored residual is obtained when the signal and the whitening filter are mismatched (Figure 7-14).

Comments. The discrimination results presented herein indicate that both multichannel and single-channel state space models can be utilized to represent normal and abnormal ECG waveforms effectively. Additionally, the SSC model-based multichannel

detection methodology with state-space models (both multichannel and scalar) can be applied to discriminate between normal and abnormal ECGs.

Both transfer functions presented for the multichannel model do not fit the "true" spectra as well as the scalar model transfer functions. This apparent loss of performance in the multichannel case is due to the differences in the concept of transfer function (or transmission) zeros between the multichannel case and the scalar case. In the scalar case the transmission zeros of the transfer function are the roots of the polynomial in the numerator of the transfer function. However, in the multichannel case the transfer function numerator is a polynomial matrix, and the transmission zeros of the multichannel system are different from the roots of the scalar polynomials that constitute the elements of the numerator polynomial matrix. In the multichannel case the total matrix polynomial is important in determining the system response. The transfer function plots shown herein for the multichannel case are calculated using as zeros the roots of the scalar polynomials that constitute the elements of the numerator polynomial matrix, which accounts for the observed performance. Even though the plots shown do not indicate the true frequency response, they do provide an approximate indication and it is instructive to review the responses thus obtained.

Figures 7-7 and 7-9 for the scalar case exhibit residual whiteness comparable to that observed in Figures 7-13 and 7-15 for the multichannel case. However, the residual in Figure 7-8 for the scalar case where the whitening filter and the ECG wave are mismatched, shows less correlation than the residual for the same conditions shown in Figure 7-14 for the multichannel case. This is representative of the performance improvement achievable with multichannel processing over single-channel processing.

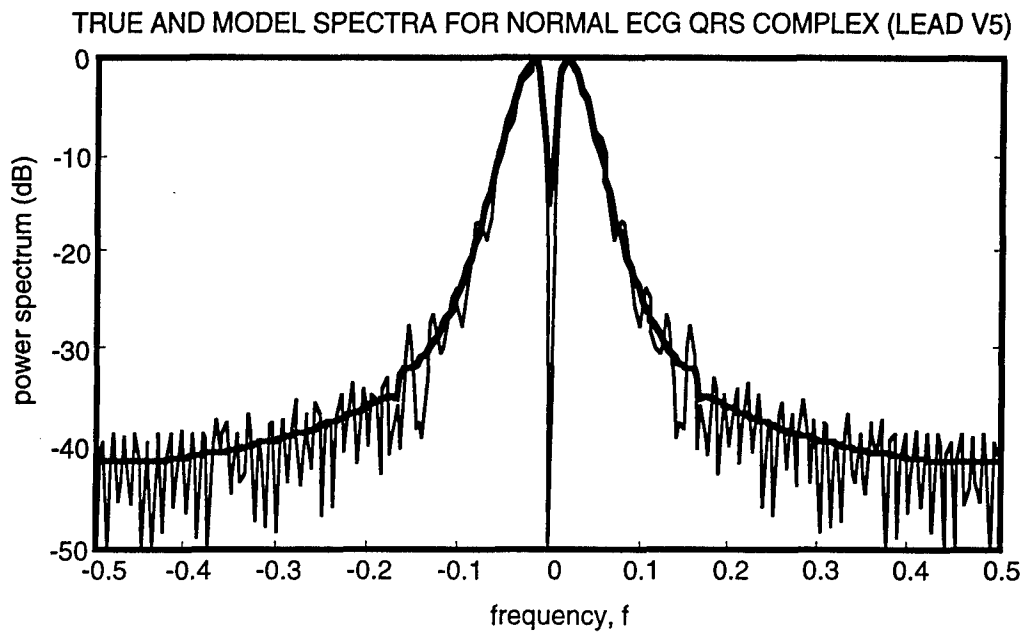


Figure 7-5. True and model spectra of the lead V5 normal ECG (single-channel model).

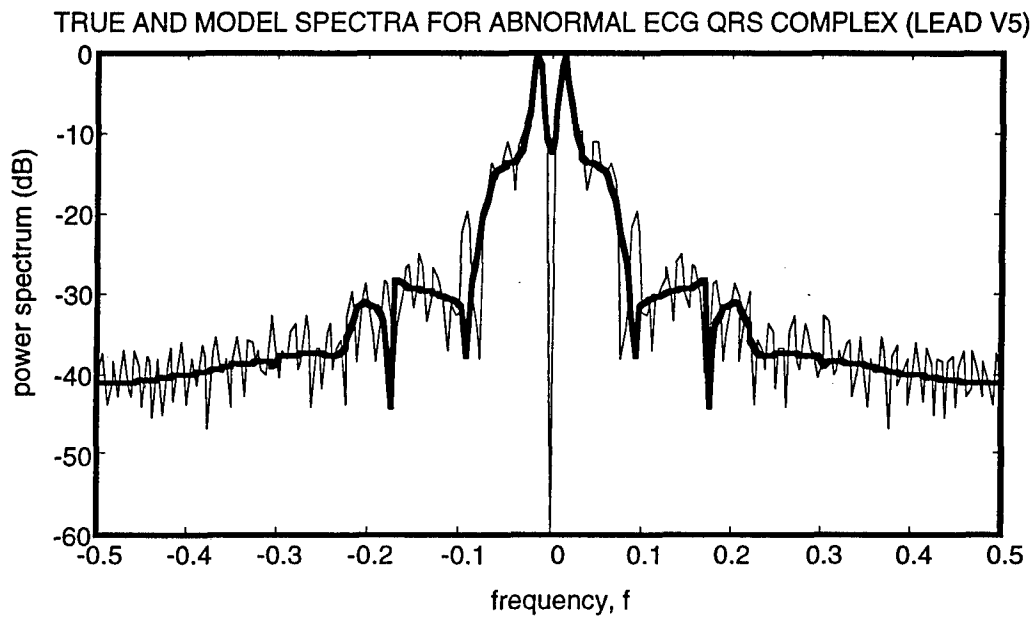


Figure 7-6. True and model spectra of the lead V5 abnormal ECG (single-channel model).

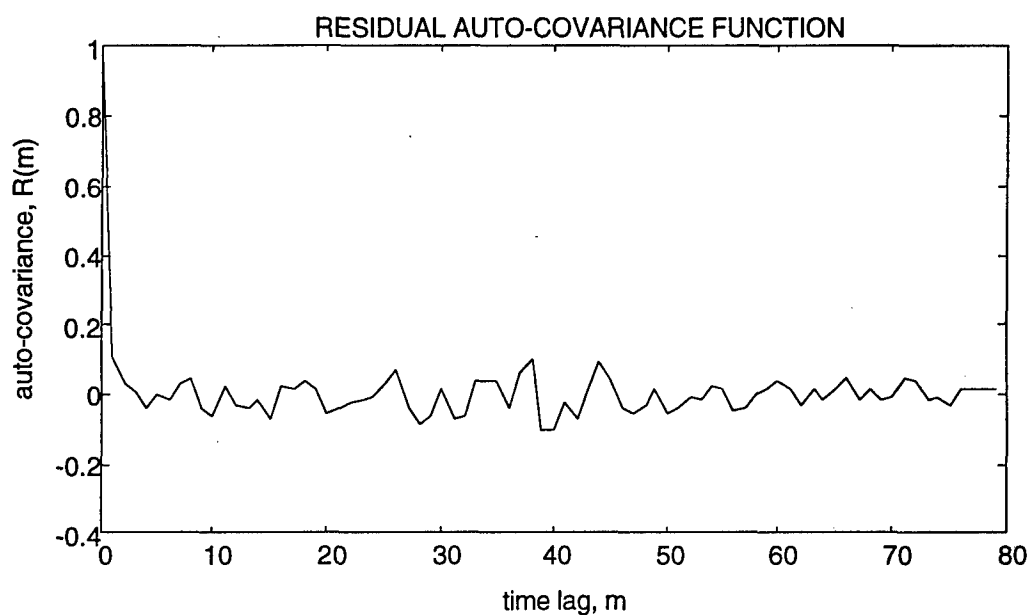


Figure 7-7. Covariance sequence of residual for normal whitening filter applied to normal ECG signal (single-channel model).

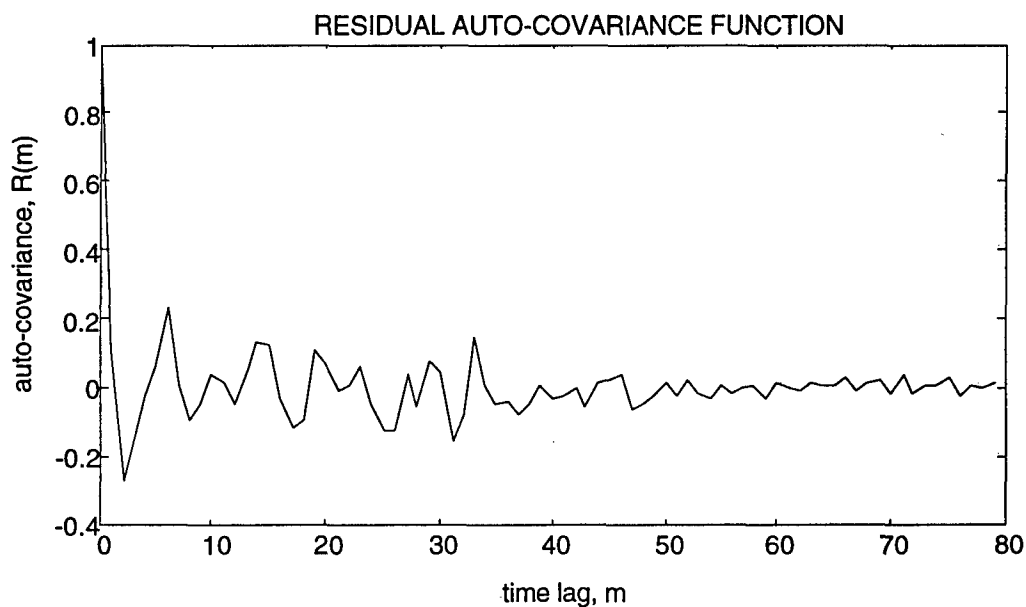


Figure 7-8. Covariance sequence of residual for normal whitening filter applied to abnormal ECG signal (single-channel model).

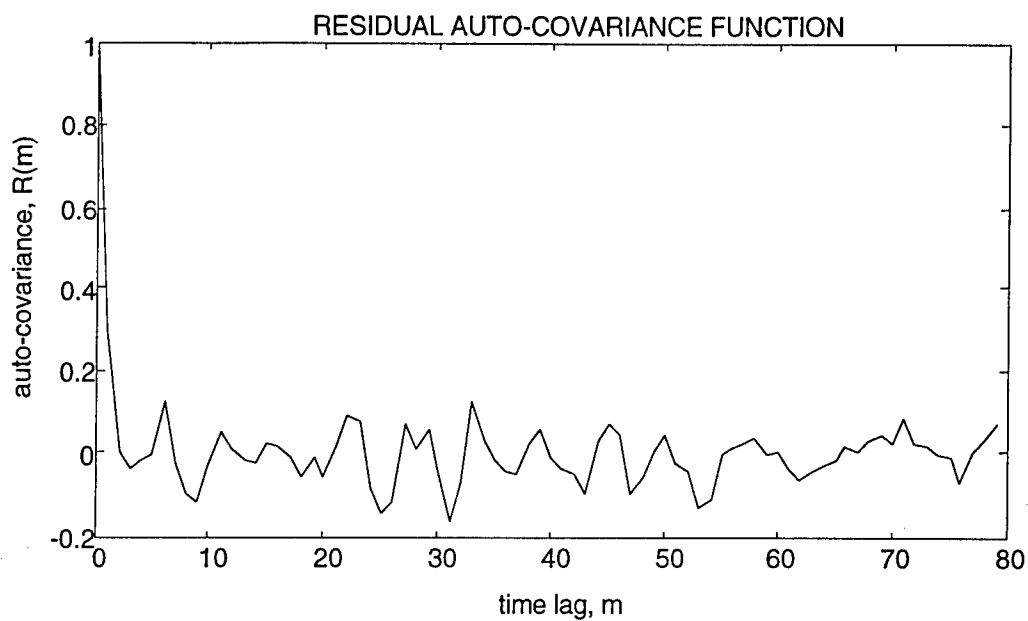


Figure 7-9. Covariance sequence of residual for abnormal whitening filter applied to abnormal ECG signal (single-channel model).

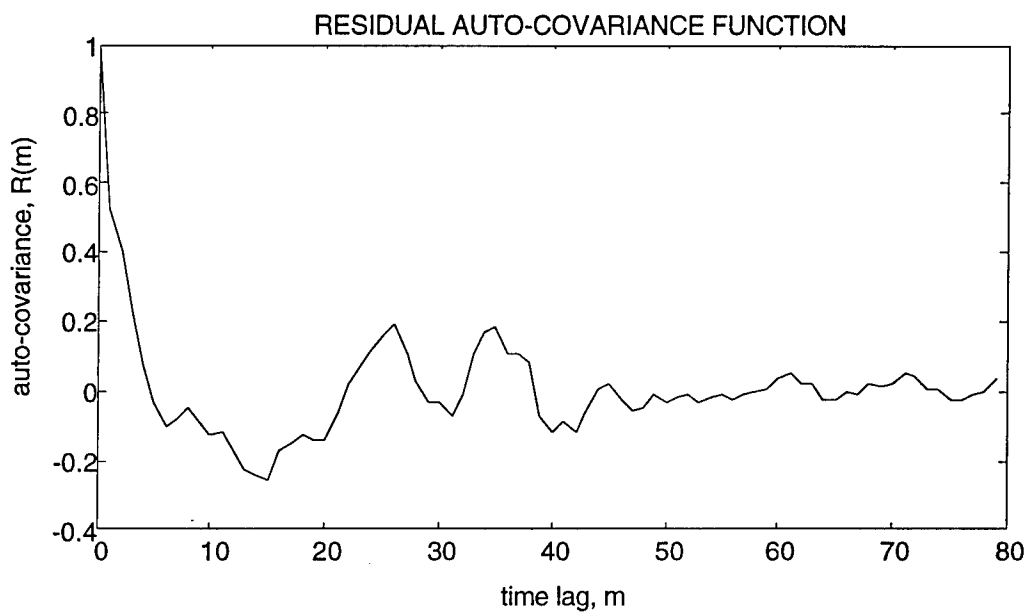


Figure 7-10. Covariance sequence of residual for abnormal whitening filter applied to normal ECG signal (single-channel model).

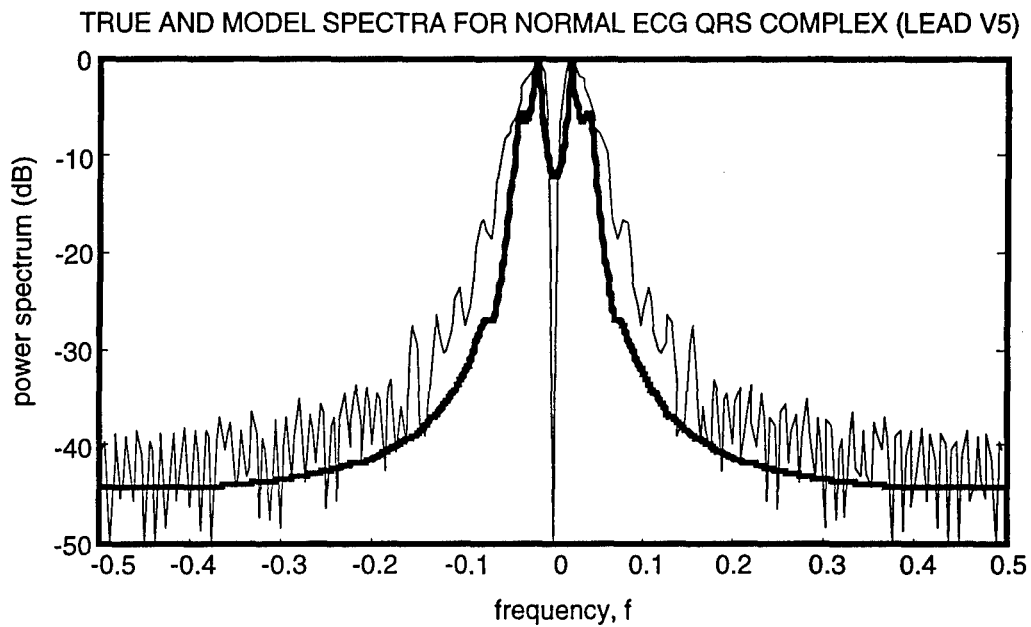


Figure 7-11. True and model spectra of the lead V5 normal ECG (multichannel model).

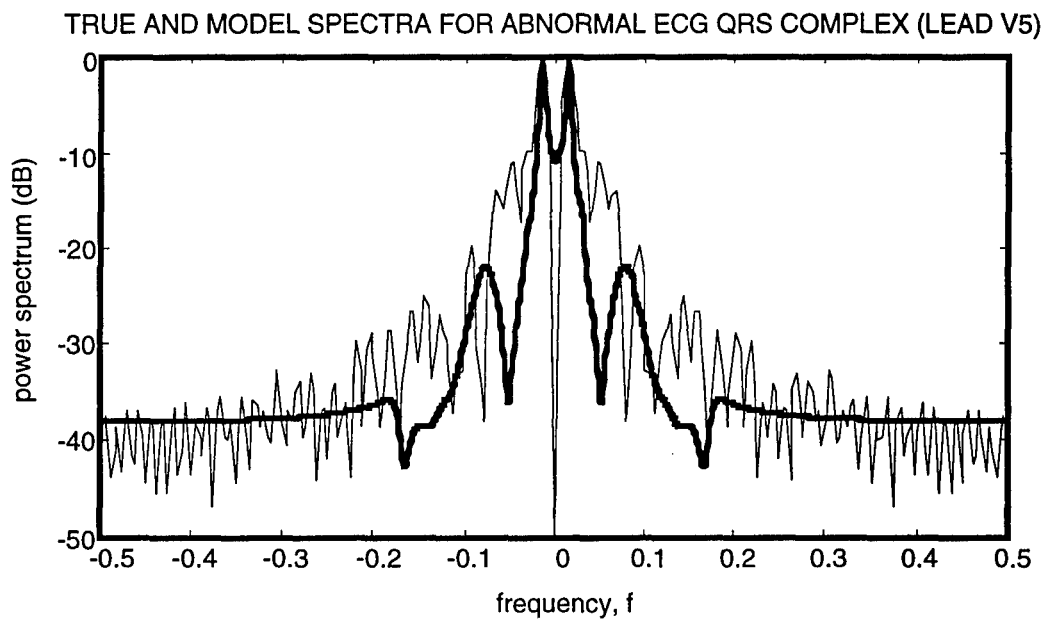


Figure 7-12. True and model spectra of the lead V5 abnormal ECG (multichannel model).

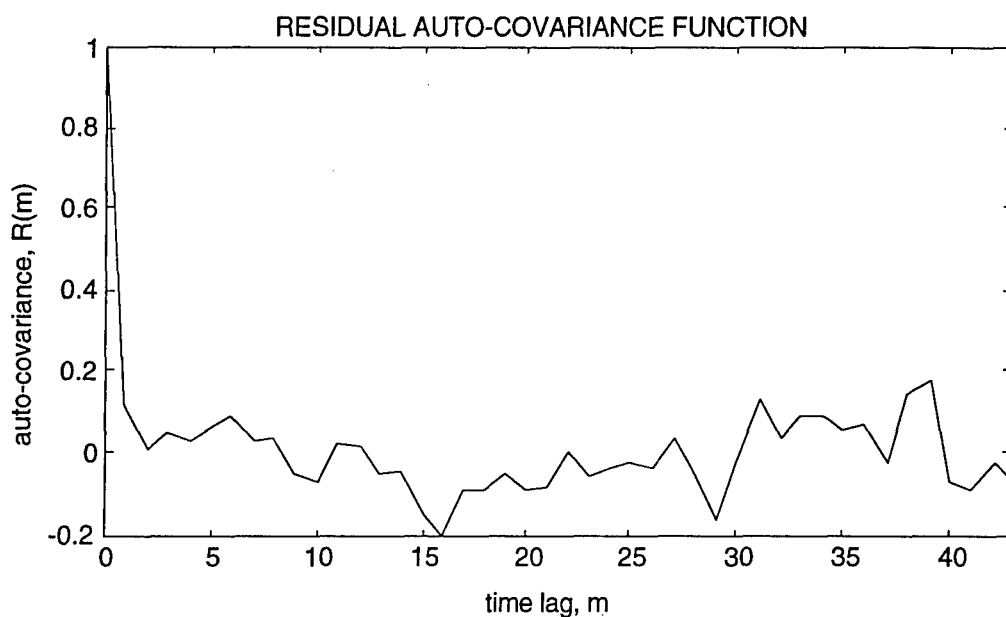


Figure 7-13. Covariance sequence of residual for normal whitening filter applied to normal ECG signal (multichannel model).

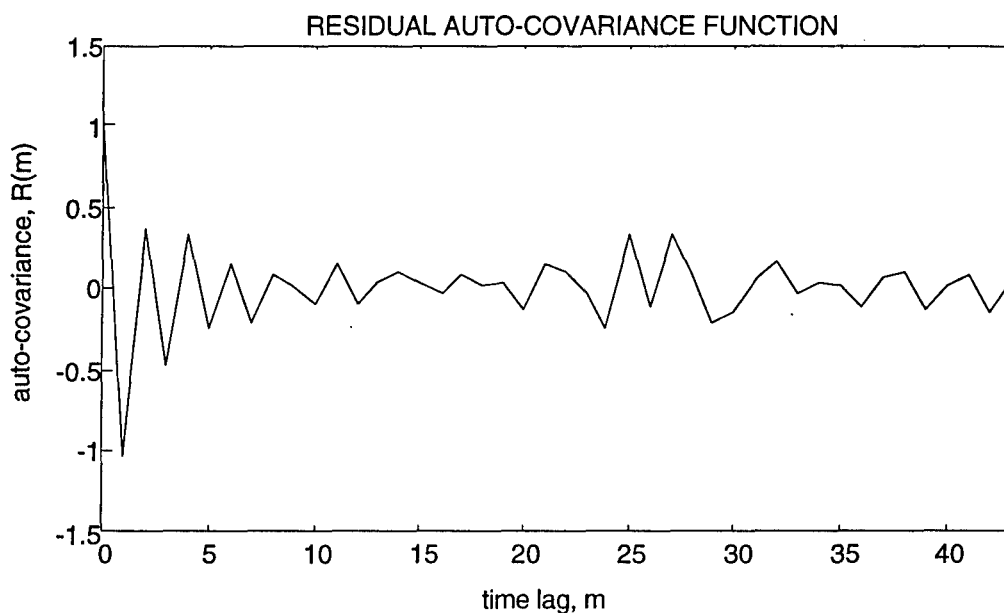


Figure 7-14. Covariance sequence of residual for normal whitening filter applied to abnormal ECG signal (multichannel model).

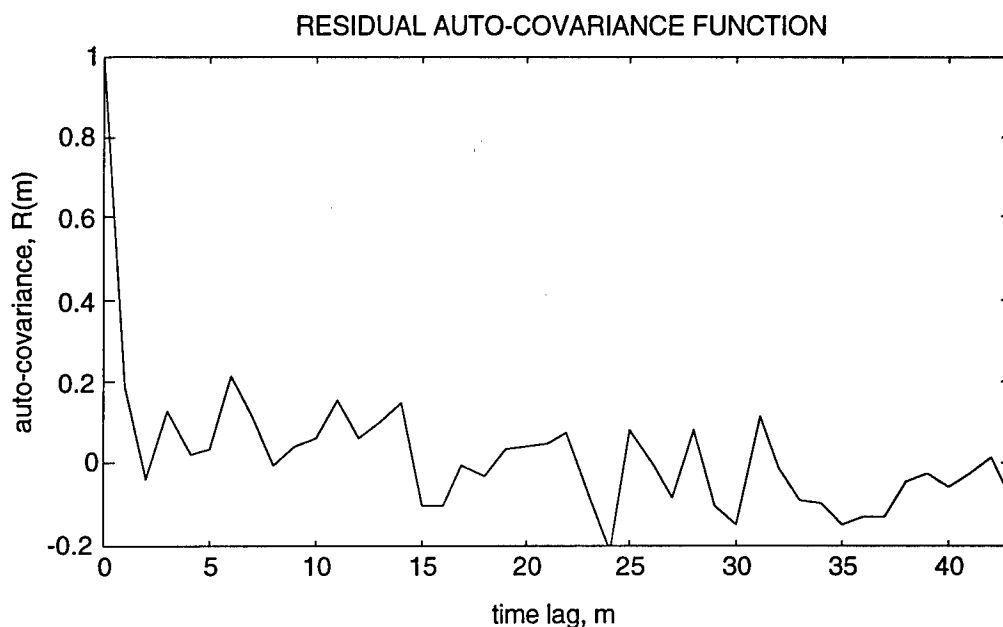


Figure 7-15. Covariance sequence of residual for abnormal whitening filter applied to abnormal ECG signal (multichannel model).

The whiteness (or lack thereof) of the residual covariance sequences in Figures 7-7 through 7-10 and 7-13 through 7-15 can be assessed using the whiteness criterion defined in Section D.3 of Appendix D. However, simulation-based analyses indicate that additional criteria are required for robust discrimination of ECG signals. These issues are addressed in Section 7.4.

Alternative Model Identification Approaches. Most ECG traces in the CSE database consist of a repeatable, deterministic component in low-level noise (high SNR), as evidenced in Figure 7-4. This suggests utilization of modeling and identification algorithms designed for deterministic signals in low-intensity noise. With this motivation, SSC investigated the applicability of the Zeiger-McEwen algorithm (Zeiger and McEwen, 1974) and the deterministic version of Kung's algorithm (Kung, 1974) to model the QRS segment of an ECG trace. Simulation-based analyses indicated that both of

these algorithms generate state-space models that represent the QRS segments accurately. However, in most cases the state-space models have several multivariable zeros outside the unit circle (non-minimum phase), which leads to unstable inverse systems. Also, the response of the inverse system exhibits behaviour that is difficult to explain only on the basis of the non-minimum-phase characteristic. Based on these observations, these alternative deterministic approaches were not pursued further.

7.4 ECG Diagnosis Methodology

As stated earlier, abnormality detection using ECGs can be formulated as a multiple hypotheses testing problem with the null hypothesis representing the normal ECG condition, and one alternative hypothesis assigned to each abnormality selected for discrimination. Thus, the SSC model-based detection methodology presented in Figure 2-1 is a likely candidate for an ECG diagnosis methodology. SSC analyzed this issue extensively, and concluded that a model-based methodology does provide a feasible approach for the diagnosis of abnormal cardiac conditions, although some modifications to the configuration utilized in the surveillance radar array application are required.

SSC has defined the model-based methodology presented in Figure 7-16 to generate multi-lead ECG diagnoses. The methodology in Figure 7-16 differs from the methodology in Figure 2-1 (as expanded upon in Sections 3 through 5) in several aspects. First, parameter identification and filter design is implemented on-line in the architecture of Figure 2-1, and off-line in the architecture of Figure 7-16. This difference is important, but less fundamental than others because off-line parameter identification is the most likely approach in many applications. Second, Figure 2-1 implements a binary hypothesis formulation, whereas Figure 7-16 implements a multiple hypotheses formulation.

This is a fundamental difference because the multiple hypotheses problem is more complicated than the binary hypothesis problem (a summary of multiple hypotheses testing is presented in Appendix G). Third, as indicated in Figure 7-16, the input ECG trace is "pre-processed," whereby artificial variations in the data are removed. Fourth, given the residual sequences, a set of residual features is calculated in Figure 7-16, instead of a likelihood ratio as in Figure 2-1. Fifth (and last), the detection decision in the architecture of Figure 2-1 is implemented as a threshold comparison, whereas the diagnosis decision in the architecture of Figure 7-16 is implemented based on several criteria. The issues involved in the third, fourth, and fifth items are expanded upon in Sections 7.4.1 and 7.4.2.

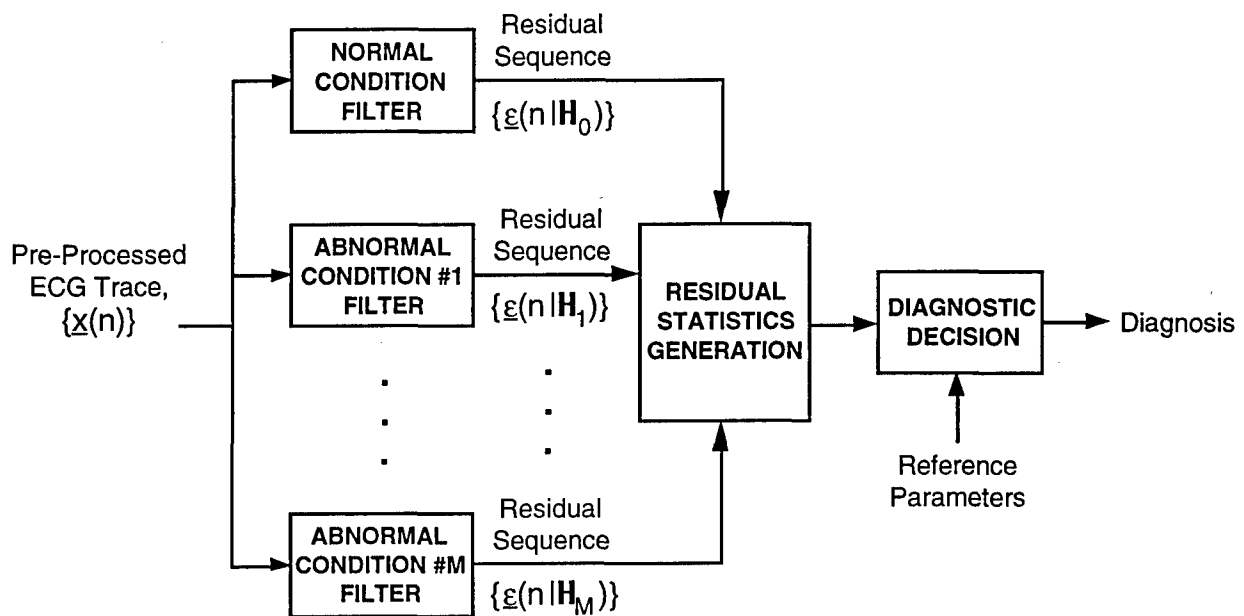


Figure 7-16. Model-based, multi-lead ECG diagnosis architecture.

Model Identification Algorithm. Although not apparent upon comparison of Figures 2-1 and 7-16, one further difference exists between the ECG diagnosis methodology and the radar detection

methodology as implemented in this report. Based on extensive modeling and condition filter design analyses, SSC discovered that the canonical correlations algorithm leads to better results in the context of ECG diagnosis than the Van Overschee-De Moor algorithm. Specifically, the Van Overschee-De Moor algorithm generates non-minimum-phase condition filters in many cases, whereas the canonical correlations algorithm generates minimum-phase condition filters in almost all ECG data cases considered in this study. This may be due to the fact that the ECG trace has a repeatable, deterministic component when viewed as a time series (see, for example, Figure 7-4). The random aspect of ECG traces is manifested predominantly over distinct realizations (different individuals; same individual on different days; differences in placement of the sensors over the body; etc.), although there are small cycle-to-cycle variations. Thus, the ECG trace can be viewed as a non-ergodic process. This is problematic to the Van Overschee-De Moor algorithm, which is applied normally to only one full cycle (or segment of a cycle) of an ECG trace. A full cycle of an ECG trace is defined herein as the epoch from the initial point of one PQRSTU segment (Figure 7-3) to the initial point of the next PQRSTU segment (thus, a cycle consists of one PQRSTU segment and a segment of noise floor). In contrast, the canonical correlations algorithm is applied to an averaged ACS estimate, where the averaging is implemented either over several cycles in one ECG trace, or for one cycle over multiple ECG traces, or over several cycles over multiple ECG traces. This allows for considerable smoothing, and allows for inclusion of ensemble statistics. For the Van Overschee-De Moor algorithm, averaging of the estimated filter parameters does not guarantee minimum-phase condition filters. Also, averaging of the ECG cycles over one trace (and/or over multiple traces) prior to processing may lead to improved results, but remains to be demonstrated. These issues will be considered in future programs.

7.4.1 ECG TRACE PRE-PROCESSING

The CSE database is a collection of true ECGs recorded with typical digital 12-lead recording equipment on actual patients. As such, the ECG traces in the CSE files exhibit most of the features that characterize actual ECGs, including random noise, 60 Hz noise (and harmonics), linear trends, cycle-to-cycle amplitude variations, repetition period variations, and iso-electric potential level variations. These features complicate the design of automatic ECG diagnostic processors and equipment, as well as the diagnosis task itself. The SSC methodology includes a pre-processing step which precedes the filtering step in Figure 7-16. Pre-processing is required to extract the PQRSTU segments (or sub-segments thereof) from the ECG trace, and to remove or modify the dominant deleterious features of the extracted segments. Pre-processing is implemented equally for each condition path, and is an integral part of condition filter design also.

The pre-processing operations discussed herein were developed for the normal, RBBB, and LBBB files in the CSE database, but it appears that the same operations are necessary also for all other files in the database. Furthermore, it is clear that some form of segmentation and signal conditioning is required for an eventual ECG diagnosis equipment based on the configuration of Figure 7-16. SSC carried out several analyses to establish the necessary pre-processing operations. The analyses focused on three types of operations: (A) segmentation of the ECG trace into cardiac cycles (or segments thereof), (B) bias offset (or removal), and (C) amplitude normalization. These three operations are presented as a generic block diagram in Figure 7-17, in the order favored by the results obtained to date. This diagram is generic because the segment duration, amplitude offset, and amplitude scale factors are unspecified. Table 7-1 lists the specific pre-processing procedure applied to the ECG traces processed to generate the

results presented in Section 7.4.4. The segmentation approach in Table 7-1 is defined for synchronously-recorded channels, as is the case in the CSE database. Variations of these three operations were considered (such as removing the sample mean from the QRS segment, and scaling the amplitude using the variance), but the procedure in Table 7-1 prepared the data best for model identification and discrimination.

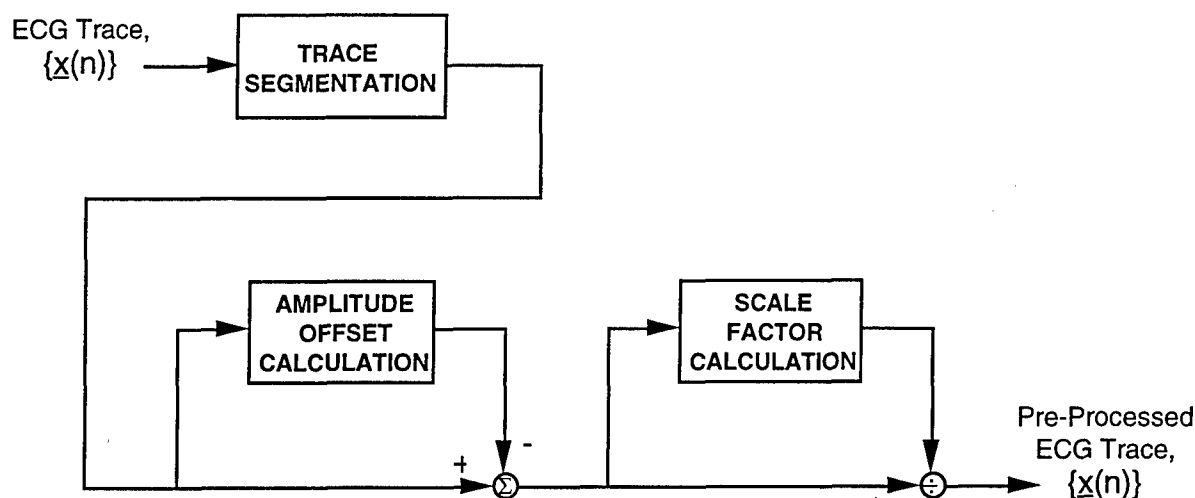


Figure 7-17. ECG trace pre-processing block diagram.

ECG TRACE PRE-PROCESSING PROCEDURE

A. TRACE SEGMENTATION

For each cardiac cycle, extract the QRS segment:

1. For each condition, select a channel to which the other channels are referenced to, and select a reference feature in the reference channel,
 - Normal: First positive-valued peak in lead I
 - RBBB: First positive-valued peak in lead V5
 - LBBB: First negative-valued peak in lead V3
2. For the reference channel, select the N-point segment starting at the Lth point preceding the reference feature (the point which defines the reference feature becomes the (L+1)th point in the segment).
3. Select the same initial point for all other channels of the same condition (all leads are synchronized).

B. AMPLITUDE OFFSET

For each QRS segment of each channel:

1. Select the first five (5) points of the segment and calculate their average; denote this quantity as b .
2. Subtract the local average, b , from each point of the QRS segment.

C. AMPLITUDE SCALING

For each QRS segment of each channel:

1. Calculate the sample root-mean-square (RMS) value for the N-point segment; denote this quantity as s .
2. Divide each point of the QRS segment by s .

Table 7-1. ECG trace pre-processing procedure for methodology validation analyses.

With respect to Table 7-1, the segmentation step yields the QRS segment of each cycle in each channel. For the analyses presented in Section 7.4.4, the length of the QRS segment is $N = 100$ points, and location offset values are: Normal, $L = 35$; RBBB, $L = 35$; and LBBB, $L = 39$. The amplitude offset step sets the isoelectric potential level to zero, and the amplitude scaling step normalizes the segment power to unity.

Many of the ECG traces in the CSE database include a linear trend feature, as exhibited by lead V6 of the normal ECG case in Figure 7-4. In many cases, however, the trend is negligible within a single cycle (or portion thereof), and removal is unnecessary. In future analyses involving the CSE files, a pre-processing step may be defined to remove linear trends and other prominent features that can impact modeling and discrimination performance.

7.4.2 ECG RESIDUAL STATISTICS AND DECISION CRITERIA

Under idealized conditions (the process modeled is a true random process with a state space representation), discrimination in a multiple hypotheses problem is accomplished using a comparative value test applied to the log-likelihood statistic, as indicated in Appendix G. But state space models for ECG traces are representation models rather than physical models; thus, it is reasonable to expect deviations from strict theoretical results. This is common to all applications involving real data, including radar systems.

Based on extensive analyses, SSC discovered an important characteristic of condition filter residual sequences. Namely, the residual sequence of the condition filter that matches the input ECG trace may be non-white, but is "more white" (less correlated) than the residual sequences of the non-matching

condition filters. Due to this feature of condition filter residuals, the log-likelihood statistic $l(\underline{g}|H_j)$ utilized in multiple hypotheses tests (see Appendix G) is inadequate for robust (>90% correct) diagnoses of ECG traces, but does provide correct diagnoses in a majority of the cases. SSC also investigated utilizing tests of whiteness to implement the diagnostic decision, since a match between the input trace and the condition filter should result in a white residual sequence. However, the threshold crossings statistic $C_r(\alpha)$ defined for the test of whiteness formulated in Section D.3 also turned out to be inadequate for robust diagnoses. However, it does provide a good indication of whiteness (or lack thereof) in most cases.

Other statistics were evaluated in search of a measure that would lead to robust diagnoses. Such a measure should be a scalar function of the residual vector sequence in order to simplify usage and software implementation. A good whiteness measure should assess relative whiteness also; that is, its value should be proportional to the degree of correlation in a sequence. Measures considered in this study include statistics of the scalar auto-correlation sequences (ACS) of the elements of the residual vector. One of the best candidate measures considered is the sum of the value of the circular estimate of the ACS for lags 1 through M_c (with M_c as defined in Appendix E). This is a reasonable candidate because at all lags (except lag $m=0$) the circular ACS of a white sequence is characterized by small, random (positive and negative) deviations from the value zero. The sum of such deviations is a small value. That is indeed the case for the ECG traces where the input data matches the condition filter type. However, that turned out to be the case also in various cases where the residual sequence is clearly non-white, but large positive and negative excursions of the estimated ACS lags almost canceled each other. These observations indicated that a measure based on the rectified ACS lag estimates is a better candidate.

Herein "rectified" denotes that the absolute value operator is applied to each element of a vector or sequence. For a white residual sequence, the sum of the rectified ACS lag estimates is a small value, although approximately twice the value for the un-rectified case. In contrast, the sum of the rectified ACS lag estimates for a non-white residual sequence is a large value in all cases. Large deviations from the zero baseline that cancel in the un-rectified case, add up in the rectified case. This measure does provide improved diagnosis capability over the un-rectified measure, but still less robust than desired.

After extensive analyses, it became clear that it is unlikely to identify a single measure that will perform with the desired degree of robustness in this application. Fortunately, it became clear also that joint consideration of three specific measures can provide the desired performance. These measures are evaluated individually, and their decisions combined using the two-out-of-three criterion. The selected measures are summarized next, including the formula used to generate the relevant statistic as well as the associated decision rule. In this context, $\mathbf{D}(\bullet)$ denotes the diagnosis decision (selection of hypothesis) based on statistic $(\bullet)$, and $\mathbf{D}$ denotes the final diagnosis decision, based on the two-out-of-three criterion applied to the three individual decisions.

Log-Likelihood (LL) Statistic:

$$(7-1) \quad \mathcal{L}(\underline{\mathbf{g}}|\mathbf{H}_i) = N \ln[\|\Omega(\mathbf{H}_i)\|] + \sum_{n=0}^{N-1} \underline{\mathbf{g}}^T(n|\mathbf{H}_i) \Omega^{-1}(\mathbf{H}_i) \underline{\mathbf{g}}(n|\mathbf{H}_i) \quad i = 0, 1, \dots, M$$

LL Decision Rule:

$$(7-2) \quad \mathbf{D}(\mathcal{L}) = \mathbf{H}_i \Leftrightarrow \mathcal{L}(\underline{\mathbf{g}}|\mathbf{H}_i) = \min_j [\mathcal{L}(\underline{\mathbf{g}}|\mathbf{H}_j)]$$

Threshold Crossings (TC) Statistic:

$$(7-3a) \quad C_r(\alpha, \underline{\epsilon} | H_i) = \sum_{k=1}^J \sum_{m=1}^{M_c} \max \{ 0, \text{sgn} [|\hat{r}_{kk}(m)| - \tau_r(\alpha)] \} \quad i = 0, 1, \dots, M$$

$$(7-3b) \quad \tau_r(\alpha) = \sqrt{2} \sigma_{\hat{r}} \text{erfinv}[1 - \alpha] = \frac{\sqrt{2} \sigma_{\epsilon}^2}{\sqrt{N}} \text{erfinv}[1 - \alpha]$$

TC Decision Rule:

$$(7-4) \quad D(C) = H_i \Leftrightarrow C_r(\alpha, \underline{\epsilon} | H_i) = \min_j [C_r(\alpha, \underline{\epsilon} | H_j)]$$

Rectified ACS Sum (RAS) Statistic:

$$(7-5) \quad S_r(\underline{\epsilon} | H_i) = \sum_{k=1}^J \sum_{m=1}^{M_c} |\hat{r}_{kk}(m)| \quad i = 0, 1, \dots, M$$

RAS Decision Rule:

$$(7-6) \quad D(S) = H_i \Leftrightarrow S_r(\underline{\epsilon} | H_i) = \min_j [S_r(\underline{\epsilon} | H_j)]$$

Two-Out-Of-Three Decision Rule:

$$(7-7) \quad D = \begin{cases} H_i & \text{if } D(\ell) = D(C) = D(S) = H_i \text{ or } D(\ell) = D(C) = H_i \\ & \text{or } D(\ell) = D(S) = H_i \text{ or } D(C) = D(S) = H_i \\ \emptyset & \text{if } D(\ell) \neq D(C) \neq D(S) \end{cases}$$

The notation used in these equations is as defined previously, and/or in Appendix D. Both the LL and RAS statistics are computed using only model parameters and the residual sequence. But

determination of the TC statistic as in Equation (7-3) requires the threshold parameter, $\tau_r(\alpha)$, as well as the residual sequence. This threshold is calculated using the PDF of the estimated ACS, which is a function of the number of points in the residual sequence, N (see Appendix D).

The combined decision rule in Equation (7-7) can be modified to provide an answer different than the null set for the cases where all three individual decision rules are distinct, $D(I) \neq D(C) \neq D(S)$, even though the input ECG trace represents one of the conditions included in the design. One approach is to select the decision associated with the most definitive statistic, meaning the statistic whose value is the farthest from the alternatives, based on a normalized distance measure. A further modification consists of comparing the most definitive statistic with an appropriately-selected threshold. Then, the decision associated with the most definitive statistic is selected if the statistic is below threshold, and the null set is selected otherwise. The second modification allows handling of ECG traces representing a condition that is not included in the design. These modifications remain to be evaluated.

7.4.3 ECG DIAGNOSIS METHODOLOGY VALIDATION PROCEDURE

A summary of the validation procedure applied to the ECG diagnosis methodology is presented in Table 7-2. This approach was applied to the processing architecture presented in Figure 7-16 for the case of three-condition diagnostic generation ($M=2$). In this architecture the condition filters are designed off-line using the canonical correlations algorithm, and applied in real time to multi-lead, sampled ECG traces, represented by the discrete-time signal vector $\{x(n)\}$. Filter residuals are processed to calculate the LL, TC, and RAS statistics, and these statistics

are used to establish a diagnostic decision as described in Section 7.4.2.

METHODOLOGY VALIDATION APPROACH

- CSE ECG database
- Segment of cardiac cycle used (QRS complex)
- Discrimination between three conditions
 - Normal condition (NC)
 - Right bundle branch block (RBBB)
 - Left bundle branch block (LBBB)
- Three independent ECG leads: I, V1, and V6
- Data files divided into design and testing sets
- Fifteen design set files
 - NC: {mo1-004, mo1-008, mo1-019, mo1-058, mo1-060}
 - RBBB: {mo1-014, mo1-033, mo1-074, mo1-076, mo1-123}
 - LBBB: {mo1-024, mo1-046, mo1-065, mo1-098, mo1-107}
- Fifteen testing set files
 - NC: {mo2-007, mo2-009, mo2-011, mo2-012, mo2-016}
 - RBBB: {mo2-015, mo2-019, mo2-033, mo2-036, mo2-046}
 - LBBB: {mo2-030, mo2-078, mo2-084, mo2-108, mo2-109}

Table 7-2. ECG diagnosis methodology validation approach summary.

With respect to the second item in Table 7-2, an ECG trace portion consisting of the QRS complex segment of the cardiac cycle is selected in order to limit the number of data points and the number of features used in off-line generation of the condition filters. The order of the resulting filters is lower than it would be if the full cardiac cycle is used, and the number of computations (off-line as well as on-line) is lower also. Methodology validation can be accomplished adequately by demonstrating discrimination between a limited number of conditions. In particular, the selected abnormality conditions, right bundle branch block (RBBB) and left bundle branch block (LBBB), present a realistic challenge (Dávila-Román, 1994). Another relevant feature of these two cardiac conduction

abnormalities is that they can be diagnosed using only three of the fifteen ECG leads (Wagner, 1994). The normal condition (NC) is selected also since any diagnosis machine has to handle normal cases. All of the CSE files used in this procedure have been inspected and classified by Dr. Dávila-Román (1994) according to standard cardiology practice (Chou, 1986), since the CSE data files are unlabeled. The CSE database should include a total of 29 RBBB cases, and 14 LBBB cases (Willems, 1994). Other CSE files will be labeled in the future, as part of further studies.

All the identified ECG traces in each of the three categories were partitioned randomly into a design set and a testing set. A reasonable, and adequate, partitioning rule is to assign approximately the same number of independent traces to each set. Fifteen data files were assigned to each set, design and testing, with five cases in each of the three condition categories. The ECG traces in the design set are used to determine the parameters of the condition filters in the processor architecture, and the ECG traces in the testing set are used to establish processor performance.

In pattern recognition terminology, the architecture in Figure 7-16 is a classifier which discriminates between $M+1$ classes (or categories) by assigning a data item (the residual sequence) to one of $M+1$ distinct classes. If all the classes can be grouped naturally into fewer categories, then the classifier carries out multi-level discrimination. Figure 7-17 presents a discrimination tree for the NC, RBBB, and LBBB conditions in the ECG trace diagnostics architecture of Figure 7-16 (with $M=2$). Discrimination trees help in the visualization of the classification objectives and formulation of the performance evaluation criteria. Notice that two levels of classification can be defined because the LBBB and RBBB classes can be combined to define the bundle branch block (BBB) class as a higher-level

category. Statistical evaluation of the performance of this classifier architecture can be carried out at each of the two discrimination levels. For evaluation at Level 1, each LBBB trace classified as an RBBB trace counts as a classification error, and viceversa. However, for evaluation at Level 0, the performance at the finer partitioning level (Level 1) is irrelevant. It is appropriate to note that each of the two classes considered herein can be partitioned further. Elements of the LBBB class can be categorized as either complete LBBB or incomplete LBBB. Analogously, the RBBB class is partitioned into complete RBBB and incomplete RBBB sub-classes.

Other abnormalities can be added to the discrimination tree in Figure 7-17 to correspond with an enhanced diagnostics problem. For example, myocardial infarction (MCI) and ventricular hypertrophy (VH) can be added at Level 0. In turn, these two classes can be partitioned further (Willems et al., 1991). The MCI class can be sub-divided into the following four Level 1 categories: anterior MCI, inferior MCI, combined MCI, and isolated apical infarction. And the VH class can be sub-divided into the left VH, right VH, and bi-VH Level 1 categories.

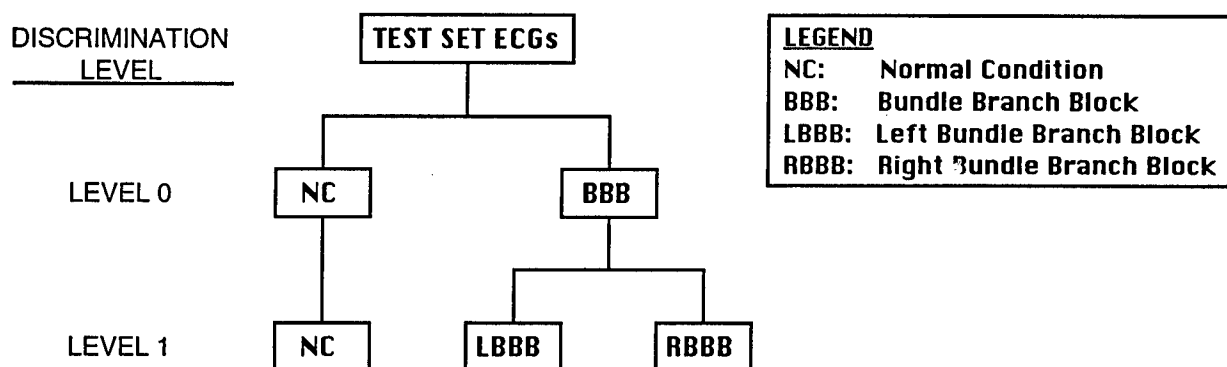


Figure 7-18. ECG two-level discrimination tree for the three-condition ECG trace diagnostic architecture.

A large database is required in order to attain statistically significant methodology validation results at the highest discrimination level (the level with the largest number of options). Also relevant is the desired resolution precision (one percent; tenth of a percent; etc.), and the desired accuracy of the evaluation result. As a first-order indication of the size of the testing set, several hundred independent ECG traces are required for each abnormality and for the normal case in order to evaluate the performance at Level 1 of the classifier in Figure 7-16 for a measurement resolution of a few percent.

The NC, RBBB, and LBBB data files identified thus far in this program are insufficient to provide conclusive evaluation results. However, prudent utilization of this data, as proposed herein, suffices to establish concept validity and to provide a first-order assessment of diagnostic accuracy. The size of the design and testing sets will increase as the truth condition of all the files in the database is established; this will be helpful in future analyses.

Classifier performance can be determined with the aid of a tool referred to as the confusion matrix, which is a scheme for tabulation of discrimination results. A generic confusion matrix is presented in Figure 7-19 for the case of three Level 0 categories (Class 1; Class 2; Other), and six Level 1 categories (Hypotheses H_0 through H_5). In this confusion matrix the Level 0 and Level 1 category "Other" allows for consideration of unknown and/or unlabeled inputs to the classifier. In the context of ECG diagnoses, the category "Other" represents ECG traces for conditions outside the set of conditions for which the diagnostic processor is designed.

The confusion matrix is completed as follows. A "zero" is placed in each empty square in the matrix at the beginning of the

performance evaluation test. Then a "one" is added to the appropriate square for each classifier decision. Upon completion of test runs, the confusion matrix contains a summary of all the decisions. Specifically, correct decisions are accounted for along the main diagonal, and incorrect decisions are accounted for in the off-diagonal elements. Level 1 performance is determined using the information in all the individual entries, whereas Level 0 performance is determined by grouping the information in the like-shaded entries.

			DECISION					
			Class 1			Class 2		Other,
			H ₀	H ₁	H ₂	H ₃	H ₄	H ₅
TRUTH	Class 1	H ₀						
		H ₁						
		H ₂						
	Class 2	H ₃						
		H ₄						
	Other, H ₅							

Figure 7-19. Generic confusion matrix for a two-level, six-hypotheses discrimination tree.

Given a completed confusion matrix, a set of statistical parameters referred to herein as performance probabilities are estimated. Performance probabilities are defined at each level, and their estimates are calculated using the relative frequency probability concept (number of outcomes divided by the number of opportunities). The most important performance probabilities are defined next, along with the formula for their estimates for two discrimination levels and M possible hypotheses, corresponding to Figure 7-18 and the results presented in Section 7.4.4 ($M = 2$).

Let Z denote the total number of ECG cases used to test the diagnostics processor, and Z_i denote the number of true cases under hypotheses H_i , for $i=0, 1, \dots, M$. Also, let D_{ij} denote the number of decisions $D = H_i$ when H_j is true, let D_C denote the total number of correct decisions, and let D_I denote the total number of incorrect decisions. Notice that D_{ij} is the entry in the j th row and i th column of the confusion matrix, and all decision variables (each D_{ij} , D_C , and D_I) are random. Given these definitions,

$$(7-8) \quad Z = \sum_{i=0}^M Z_i$$

$$(7-9) \quad Z_j = \sum_{i=0}^M D_{ij} \quad j = 0, 1, \dots, M$$

$$(7-10) \quad D_C = \sum_{i=0}^M D_{ii}$$

$$(7-11) \quad D_I = \sum_{i=0}^M \sum_{\substack{j=0 \\ j \neq i}}^M D_{ij}$$

Probability Of Correct Decision (Level 0):

$$(7-12) \quad P_{CD} = \mathcal{P}[\text{Correct Decision}]$$

$$(7-13) \quad \hat{P}_{CD} = \frac{D_C}{Z} = \frac{1}{Z} \sum_{i=0}^M D_{ii}$$

Probability Of Incorrect Decision (Level 0):

$$(7-14) \quad P_{ID} = \mathcal{P}[\text{Incorrect Decision}]$$

$$(7-15) \quad \hat{P}_{ID} = \frac{D_I}{Z} = \frac{1}{Z} \sum_{i=0}^M \sum_{\substack{j=0 \\ j \neq i}}^M D_{ij}$$

Probability Of Decision H_i When H_j Is True (Level 1):

$$(7-16) \quad P_{ij} = \mathcal{P}[\text{Decision } H_i \text{ When } H_j \text{ Is True}]$$

$$(7-17) \quad \hat{P}_{ij} = \frac{D_{ij}}{Z_j} \quad i = 0, 1, \dots, M; \quad j = 0, 1, \dots, M$$

Probability of Incorrect Decision for Hypothesis H_j (Level 1):

$$(7-18) \quad P_{IDj} = \mathcal{P}[\text{Incorrect Decision For Hypothesis } H_j]$$

$$(7-19a) \quad \hat{P}_{IDj} = \frac{\text{no. of incorrect decisions when } H_j \text{ is true}}{Z_j} \quad j = 0, 1, \dots, M$$

$$(7-19b) \quad \hat{P}_{IDj} = \frac{1}{Z_j} \sum_{\substack{i=0 \\ i \neq j}}^M D_{ij} \quad j = 0, 1, \dots, M$$

From these definitions it follows that

$$(7-20) \quad P_{CD} + P_{ID} = 1$$

$$(7-21) \quad \hat{P}_{CD} + \hat{P}_{ID} = 1$$

$$(7-22) \quad \sum_{i=0}^M P_{ij} = 1 \quad j = 0, 1, \dots, M$$

$$(7-23) \quad \sum_{i=0}^M \hat{P}_{ij} = 1 \quad j = 0, 1, \dots, M$$

Generalization of these definitions to more than two levels is trivial, but cumbersome notation-wise. Notice that the Level 1 Probability for Correct Decision for Hypothesis H_j , denoted as P_{CDj} , is the same as the Probability Of Decision H_i When H_j Is True in Equation (7-16) with $i=j$. The same follows for the estimate of the probabilities.

All the outcomes in the evaluation procedure defined herein are independent because the design set and testing set files are independent of each other and of the files in the same set, and because the processing is applied to each case file independently. Therefore, each set of random variables $\{D_{ij}|i = 0, 1, \dots, M\}$ is distributed according to the multinomial distribution characterized by the integer Z_j (number of opportunities) and the true (and unknown) performance probabilities $\{P_{ij}|i = 0, 1, \dots, M\}$. Furthermore, the random variable D_C is distributed according to the binomial distribution with parameters Z (number of opportunities) and the true performance probability P_{CD} . And similarly, D_I is distributed according to the binomial distribution with parameters Z and the true performance probability P_{ID} . Thus, the probability of D_C correct decisions in Z opportunities is given as

$$(7-24) \quad \mathcal{P}[D_C] = \binom{Z}{D_C} (P_{CD})^{D_C} (1-P_{CD})^{Z-D_C} = \frac{Z!}{(D_C)!(Z-D_C)!} (P_{CD})^{D_C} (P_{ID})^{D_I}$$

And the probability of D_I incorrect decisions in Z opportunities is given as

$$(7-25) \quad P[D_I] = \binom{Z}{D_I} (P_{ID})^{D_I} (1-P_{ID})^{Z-D_I} = \frac{Z!}{(D_I)!(Z-D_I)!} (P_{ID})^{D_I} (P_{CD})^{D_C}$$

These theoretical relations are useful even though the true performance probabilities are unknown. First of all, since D_C and D_I are binomially-distributed, the estimates in Equations (7-13) and (7-15) are unbiased, maximum-likelihood estimates of the respective true probabilities (Hastings and Peacock, 1975). Also, the binomial PDF can provide indication of the likelihood of the realized D_C (or D_I) value. As an example, let $Z=15$ and suppose that the outcome of a test is $D_C=14$. Since the mode of the PDF for D_C satisfies the inequality $(Z+1)P_{CD} - 1 \leq D_{\text{mode}} \leq (Z+1)P_{CD}$ (Hastings and Peacock, 1975), it is more likely that the unknown, true probability P_{CD} is closer to 0.9 than to 0.7. Similar considerations are valid also for the multinomially-distributed variables, but the larger number of variables that are inherently involved in the multinomial PDF complicates the issue.

7.4.4 ECG DIAGNOSIS METHODOLOGY VALIDATION RESULTS

SSC applied the approach summarized in Table 7-2 and described in Section 7.4.3 to validate the model-based ECG diagnosis methodology of Figure 7-16. The conditions for the analyses are summarized in Table 7-3, and these conditions apply for both design and testing. As noted in this table, model order 10 was selected for each of the three condition filters. This model order allows a good fit to the design set QRS segments, without modeling excessive details. The average (over the three channels of the three condition filters) residual sequence power with this model order is 25.1 dB below the input sequence power. That is, the whitening gain of the condition filters is approximately -25 dB for the design set files. Further optimization of model order may be possible, but it is more appropriate to do so with a larger design set. The condition

filters were designed off-line using the canonical correlations algorithm, which operates on the auto-correlation sequence (ACS) of the ECG trace vector. Forty-five (45) ACS lags are required by the algorithm (including the lag at $m=0$), given the selected value, 22, of the Hankel matrix block row dimension parameter. The ACS used to design the condition filters is generated as follows (Table 7-2). First, the biased, time-averaged ACS estimate (Appendix E) is generated for each of the eight QRS segments of each ECG trace in the design set. Next, this estimate is averaged over the eight QRS complexes in each of the five ECG files in the design set; this is temporal averaging of the ACS. Finally, the correlation matrices are averaged further over the five design files; this is ensemble averaging of the ACS. Both types of averaging are important because ECGs are non-ergodic. All files in the CSE database are of ten-second duration (approximately), and have at least eight good cardiac cycles (each CSE file corresponds to a different patient and/or condition; a normal cardiac cycle is of approximately one-second duration).

The design criteria described in Appendix D - mean test; power test; whiteness test (TC measure) - were applied to the residuals of the three subsets of the design set ECG files. All three design criteria were met with system order ten for each condition filter. Additionally, the residual statistical measures LL, TC, and RAS (Equations (7-1), (7-3), and (7-5)) were calculated for two different averages of the residual ACS for the design set cases. In the first average, forty circular, time-averaged residual ACS estimates are averaged. This form of averaging utilizes all the available residual estimates for a given condition, and corresponds to averaging the residual ACS over time as well as over the ensemble. In the second average, eight circular, time-averaged ACS estimates corresponding to the same ECG trace are averaged. This form of averaging involves only similar residuals.

CONDITIONS FOR SIMULATION-BASED METHODOLOGY VALIDATION

- ECG leads I, V1, and V6 labeled as channels 1 through 3, respectively.
- Condition filters designed with system order 10 and using Hankel matrix block row dimension 22.
- Eight QRS segments from each of the five independent cases used for off-line condition filter design with the canonical correlations algorithm.
- Each QRS segment pre-processed as described in Table 7-1 (bias off-set and scaling) prior to being used for condition filter design or for testing.
- Each QRS segment has $N=100$ points.
- Forty-five (45) lags of the design auto-correlation matrix sequence (for off-line filter design) obtained by averaging the biased, time-average correlation matrix sequence estimates of each of the 32 QRS segments.
- Filter residual statistics (LL, TC, and RAS) calculated for the residual ACS averaged over either thirty-two or eight circular, time-average ACS estimates.

Table 7-3. Conditions for simulation-based methodology validation.

The TC statistic requires the threshold parameter, $\tau_r(\alpha)$, which is applied to lags 1 through M_c of the diagonal elements of the averaged matrix ACS, as discussed in Appendix D. For the QRS segments considered herein $N=100$, so that all lags in a circular estimate of the ACS are approximately Gaussian-distributed. Thus, the threshold calculation is given by Equation (7-3b), which indicates that the threshold is a function of the variance of the scalar residual sequence under each hypothesis, σ_e^2 , the number of points in the residual sequence, N , and the significance level of

the whiteness test, α . A reasonable value for the level of significance is $\alpha=0.05$, and is the value selected for the analyses reported herein. The system model covariance matrix of the residual vector under each hypothesis, $\Omega(H_j)$, is known, so, the required scalar variances are known (the diagonal elements of $\Omega(H_j)$). In general, a different threshold value is required for each scalar ACS since all the diagonal elements of $\Omega(H_j)$ are different. However, for the analyses reported herein each scalar residual sequence is normalized to have unit variance prior to the generation of the circular ACS estimate. This normalization is equivalent to setting $\sigma_e^2=1$ for each scalar residual sequence; consequently, the threshold expression simplifies to (in MATLAB notation for the inverse of the error function)

$$(7-26) \quad \tau_r(\alpha) = \sqrt{\frac{2}{N}} \operatorname{erfinv}[1-\alpha]$$

Upon substitution of the known parameters α and N , the threshold is calculated as

$$(7-27) \quad \tau_r(\alpha) = \sqrt{0.02} \operatorname{erfinv}[0.95] = 0.196$$

This threshold value is used for each of the three scalar normalized ACSs in design as well as testing. For the parameters in these simulations, lags 1 through $M_c=50$ of the circular, time-average ACS estimate of a zero-mean, unit-variance, scalar white sequence should exceed the threshold of Equation (7-27) less than three times, on the average.

The RAS statistic is a function of the scalar ACS (Equation (7-5)). This statistic is computed herein also using the scalar ACS estimated for the normalized residual sequence. Given this constraint, the maximum possible value for this statistic is 50

(for $N = 100$, only lags 1 through $M_c = 50$ of the circular, time-average ACS estimate are unique).

Consider first the statistical measure results for the design set time- and ensemble-averaged residual ACS, presented in Table 7-4. For each statistical measure, the decision rule is to select the minimum (Equations (7-2), (7-4), and (7-6)). The decision for each measure is highlighted in Table 7-4; notice that the three measures produce the correct decision in all cases. Of course, the two-out-of-three decision rule also generates the correct decision in all cases. These results are expected since the measures in Table 7-4 are generated for residuals of the ECG cases used to design the condition filters. All three measures provide good discrimination margin, measured on a percentage basis. The narrowest discrimination margins are for the LL statistic for the LBBB design set cases; specifically, 6.8% and 10.0%.

Consider now the statistical measure results for the design set residual ACS with time-averaging only, which corresponds to applying the test criteria to each case individually. These results are presented in Table 7-5, with the decision for each measure highlighted. Since the measures in Table 7-5 are generated for residuals of the ECG cases used to design the condition filters, the three measures produce the correct decision in all cases, and the two-out-of-three decision rule also generates the correct decision in all cases. All three measures provide good discrimination margin, on a percentage basis. The narrowest margin is 6.4% for the LL statistic for Normal design set case m01-019.

In the testing step the testing set cases are pre-processed according to the procedure in Table 7-1, and filtered with the whitening filters designed using the design set cases. Then the statistical measures LL, TC, and RAS are calculated for the

testing set residual sequences in the same manner as discussed above for the design set. Real-time processing of ECG traces would proceed in a similar manner. This approach was carried out, and the statistical measure results for the testing set residual ACS with time-averaging only are presented in Table 7-6. Analogous to Table 7-5, the results in Table 7-6 correspond to applying the test criteria to each case individually. The rule decision for each measure is highlighted in Table 7-6, indicating that each statistical measure produces an incorrect decision in at least two cases, but two or more measures generate an incorrect decision in one case only. Therefore, the two-out-of-three decision rule produces a correct decision in fourteen cases, and an incorrect decision in one case. These observations are summarized in the confusion matrix presented in Table 7-7. Blocks that have the same shading in Table 7-7 represent the grouping of the test outcomes required for Level 0 performance evaluation.

Performance probabilities for these results are calculated using the information in Table 7-7. Level 0 probabilities are the most relevant since these results are based on a small number of test cases ($Z=15$). Specifically,

$$(7-28) \quad \hat{P}_{CD} = \frac{D_C}{Z} = \frac{14}{15} = 0.93$$

$$(7-29) \quad \hat{P}_{ID} = \frac{D_I}{Z} = \frac{1}{15} = 0.07$$

$$(7-30) \quad P_{CDNC} = \mathcal{P}[\text{Correct Decision For Normal Condition}]$$

$$(7-31) \quad \hat{P}_{CDNC} = \frac{D_{CNC}}{Z_{NC}} = \frac{4}{5} = 0.80$$

$$(7-32) \quad P_{IDNC} = \mathcal{P}[\text{Incorrect Decision For Normal Condition}]$$

$$(7-33) \quad \hat{P}_{IDNC} = \frac{D_{INC}}{Z_{NC}} = \frac{1}{5} = 0.20$$

$$(7-34) \quad P_{CDBBB} = \mathcal{P}[\text{Correct Decision For BBB Condition}]$$

$$(7-35) \quad \hat{P}_{CDBBB} = \frac{D_{CBBB}}{Z_{BBB}} = \frac{10}{10} = 1.0$$

$$(7-36) \quad P_{IDBBB} = \mathcal{P}[\text{Incorrect Decision For BBB Condition}]$$

$$(7-37) \quad \hat{P}_{IDBBB} = \frac{D_{IBBB}}{Z_{BBB}} = \frac{0}{10} = 0.0$$

P_{CD} and P_{ID} are the most fundamental probabilities and the ones estimated most accurately. Based on the inequality for the mode of the binomial distribution, the test outcome $D_C = 14$ is much more likely if the true (and unknown) probability P_{CD} is in the range $0.90 < P_{CD} < 0.95$ than if it is in the range $0.70 < P_{CD} < 0.75$. An analogous statement is true for the test outcome D_I .

The results presented in these tables are indicative of the expected performance for a larger data set assuming that the files used for design are representative of the universe of ECG traces (ensemble) for the conditions considered; this should be true for the CSE database. Nevertheless, an enhanced design for cardiac conduction abnormality diagnosis based on more independent ECG traces and further testing using more independent ECG traces is a desirable next step. A possible variation to the approach is to design each condition filter for a lead subset specific to that condition, in contrast to having the same set of leads as inputs to all filters. The work can be extended to include other cardiac abnormalities and/or to include as inputs data from other physiologic sensors operating synchronously with the 12-lead ECG.

TRUTH DESIGN CASES	CONDITION FILTER								
	NORMAL			RBBB			LBBB		
	RAS	LL	TC	RAS	LL	TC	RAS	LL	TC
NORMAL	7.73	-1,402	0	13.38	-1,065	6	18.81	960	31
RBBB	21.92	-1,109	29	8.99	-1,502	0	28.20	-750	55
LBBB	21.17	-1,467	25	41.31	-1,417	53	8.04	-1,574	1

Table 7-4. Statistical measures for the design set time- and ensemble-averaged residual ACS.

CASE (mo1)	TRUTH	CONDITION FILTER								
		NORMAL			RBBB			LBBB		
		RAS	LL	TC	RAS	LL	TC	RAS	LL	TC
_004	Normal	10.54	-1243	5	15.84	-752	17	18.39	455	24
_008	Normal	12.46	-1472	3	16.40	-1330	21	22.50	300	34
_019	Normal	13.63	-1378	3	15.82	-1290	16	17.96	-493	25
_058	Normal	10.44	-1189	3	20.18	-413	30	21.68	565	33
_060	Normal	10.38	-1402	5	12.37	-1065	7	21.87	960	33
_014	RBBB	42.15	-1353	69	17.71	-1584	16	46.07	-932	100
_033	RBBB	31.25	-898	54	15.85	-1397	8	35.33	-228	65
_074	RBBB	24.00	-1109	35	17.27	-1502	20	39.10	-750	73
_076	RBBB	20.93	-894	36	11.11	-1311	1	27.06	-609	49
_123	RBBB	20.72	-898	29	11.96	-1397	7	21.00	-228	23
_024	LBBB	52.36	-1567	105	63.39	-1595	91	31.70	-1760	66
_046	LBBB	24.54	-1496	42	48.75	-1419	76	15.02	-1614	4
_065	LBBB	56.64	-1556	105	68.18	-1575	96	18.51	-1746	18
_098	LBBB	29.80	-1467	49	50.12	-1417	79	15.02	-1574	14
_107	LBBB	30.19	253	53	35.17	-120	54	9.22	-1229	0

Table 7-5. Statistical measures for the design set time-averaged residual ACS.

		CONDITION FILTER								
		NORMAL			RBBB			LBBB		
CASE (mo2)	TRUTH	RAS	LL	TC	RAS	LL	TC	RAS	LL	TC
_007	Normal	14.65	-1331	11	18.89	-1336	20	19.74	-326	29
_009	Normal	12.04	-351	6	11.22	-366	4	18.42	2253	19
_011	Normal	15.25	-1426	11	21.11	-1433	23	21.30	-578	30
_012	Normal	19.07	-1332	18	21.29	-1226	38	27.48	537	46
_016	Normal	15.80	-1169	13	20.61	-584	30	21.28	1282	30
_015	RBBB	37.59	-1008	66	22.25	-1299	40	24.92	-740	38
_019	RBBB	27.50	-1166	41	15.86	-1391	16	23.84	-634	37
_033	RBBB	30.45	-999	26	16.43	-1077	15	22.18	-160	49
_036	RBBB	23.29	-1151	34	22.93	-1250	31	22.49	-577	32
_046	RBBB	28.29	-1372	45	21.99	-1490	34	25.23	-960	40
_030	LBBB	18.17	-1521	19	45.53	-1478	71	15.78	-1610	12
_078	LBBB	31.36	-1524	54	49.71	-1573	76	22.93	-1678	26
_084	LBBB	30.98	-1318	65	48.34	-852	82	15.62	-1116	12
_108	LBBB	39.74	-1426	73	47.64	-1578	82	32.14	-1577	57
_109	LBBB	31.43	-1574	59	52.84	-1565	79	20.74	-1696	30

Table 7-6. Statistical measures for the testing set time-averaged residual ACS.

			DECISION		
			NORMAL	BBB	
			NORMAL	RBBB	LBBB
TRUTH	NORMAL	NORMAL	4	1	0
	BBB	RBBB	0	5	0
		LBBB	0	0	5

Table 7-7. ECG diagnosis methodology evaluation confusion matrix.

8.0 CONCLUSIONS AND RECOMMENDATIONS

The work carried out in this program emphasized the development and validation of a state space methodology and algorithm for model-based multichannel detection. Emphasis was placed in the surveillance radar array application and space/time processing. Utilization of state space techniques for multichannel detection in radar systems is one novel aspect of the work reported here. The state space model class is richer than the time series model class that is used often in radar system applications. And, as demonstrated in this work, the state space model class can be used to represent effectively multichannel radar signals. Feasibility of the model-based multichannel methodology for automated ECG diagnostics was demonstrated also.

Another novel aspect of the work is the utilization of the new parameter identification algorithm developed by Van Overschee and De Moor (1993). In the process, the algorithm was extended to the case of complex-valued data (as required for radar systems), and several enhancements to the algorithm were discovered. Of particular interest is a new method to compute the QSVD that offers simpler implementation, improved performance, and less computations than the published methods. The Van Overschee-De Moor algorithm uses channel output data directly (as opposed to output correlation matrices) to estimate model parameters. This eliminates the large computational burden associated with the generation of the output correlation matrix sequence, and leads to reduced numerical precision (dynamic range) requirements. Furthermore, in a practical environment it may be possible to start processing the data as it is received. In contrast, techniques which require the computation of channel output correlation matrices have a built-in delay because the calculation of every lag requires availability of all the channel output sequence.

The Van Overschee-De Moor algorithm belongs to a class of techniques referred to as subspace methods. Subspace methods are based on decomposing the vector space spanned by the channel outputs into signal and noise subspaces. This decomposition is carried out with robust numerical techniques such as the SVD and the QR decomposition. Thus, the algorithm offers numerical and performance advantages over other techniques.

A hardware-based processor development system (PDS) was configured and integrated to serve as a testbed for the design and development of detection and identification methodologies and algorithms. The PDS consists of a Sun Microsystems' SPARCstation 10 host and a SKY Computers' SKYstation II accelerator, with FORTRAN 77 and MATLAB software (MATLAB runs only on the SPARCstation). The PDS is very effective for simulation-based analyses (single-run cases as well as Monte Carlo analyses) and for off-line processing of data collected using operational radar systems. Access to the PDS speeds up algorithm development work at both SSC (during Phase II) and RL (after delivery upon program conclusion).

Two sets of software programs were developed to validate the algorithm and methodology, and to evaluate performance. Extensive tests were carried out to validate both sets of code. One set of software programs was developed in the MATLAB simulation environment. The MATLAB-based software includes an implementation of the model-based multichannel detection methodology using each of the three state space model identification algorithms considered in the program (Van Overschee-De Moor; canonical correlations; Arun-Kung). Also included in the MATLAB-based software package is a model of airborne surveillance phased array radar scenarios that generates simulated data for evaluation of the model-based multichannel detection methodology as well as

other space/time processing algorithms. This simulated data generation capability is described in Volume II of this Final Report.

The second set of software programs constitute a FORTRAN 77 implementation of the model-based multichannel detection methodology using the Van Overschee-De Moor state space model identification algorithm. These programs can be exercised also using data generated from the MATLAB-based surveillance radar simulation. A Software Users' Manual for the FORTRAN 77 package was generated as a separate document (Davis and Román, 1996), and provides a detailed description of the FORTRAN 77 package. The FORTRAN 77 software can run on both Apple and Sun Microsystems computers.

Simulation-based analyses have demonstrated the feasibility of the SSC state space approach for modeling the multichannel clutter return in airborne surveillance phased array radar systems, and for moving target detection. The innovations-based detection methodology has demonstrated the capability to discriminate between target present and target absent hypotheses. Additionally, the methodology has been applied with equal success to a reduced-scope ECG diagnostics problem. Specifically, modeling of and discrimination between normal QRS complexes and LBBB and RBBB cardiac conduction abnormalities was established, and the automatic diagnosis of the LBBB and RBBB abnormalities was demonstrated. Both sets of results were obtained using a subset of the CSE ECG data base of real ECG traces. These results have been presented in three medical technology conferences (Román and Davis, 1994; Román et al., 1996a, 1996b).

In the process of completing the work reported here several areas have been identified for further research and development in future programs. These areas are summarized below.

Space/Time Processing

Significant progress was made in Phase II towards the development of a processor architecture capable of addressing the space/time processing problem for surveillance radar arrays. However, extensive detection performance analyses are required to establish performance over a wide variety of scenario conditions, and in relation to the optimum joint-domain method and its approximations. This is a challenging task because a standard for comparison is unavailable at the present time.

ECG Diagnostics

The results obtained in Phase II have demonstrated the feasibility of model-based multichannel methods for ECG diagnostics. However, additional detailed design and extensive testing are required in order to achieve a processor configuration capable of automated, real-time ECG diagnostics. Specifically, additional abnormalities need to be considered, and a much larger data base needs to be accessed.

Processor Development System Utilization

The PDS should be exercised further in the development of the model-based multichannel methodology and its performance evaluation in the context of surveillance radar array as well as ECG diagnostics. The PDS is essential also to the investigation of new areas such as model-based detection methodologies using two-dimensional models.

APPENDIX A. PARTIAL QUOTIENT SINGULAR VALUE DECOMPOSITION

The QSVD is a key analytical tool in the Van Overschee-De Moor identification algorithm, and as such requires an effective numerical implementation. SSC generated a MATLAB-based software subroutine implementation the QSVD algorithm proposed by Van Overschee and De Moor (1991), which, in turn, is a modification of the approach presented by Paige and Saunders (1981). The SSC QSVD subroutine was used to compute estimates of the system matrix parameters as based on the formulas presented in Section 3-1. However, in the course of running test cases it was discovered that in many cases the error in the estimate of the eigenvalues of the system matrix estimated using the combined F formula was larger than the error obtained using the forward F formula. This condition was traced to a problem inherent in the QSVD calculation associated only with the backward F formula. Thus, the QSVD algorithm proposed by Van Overschee and De Moor had to be set aside. In the process of addressing this issue SSC discovered a simpler, more robust approach to calculate the QSVD, as a modification of the algorithm recommended by Van Overschee and De Moor.

The SSC QSVD algorithm has several advantages over the Van Overschee-De Moor QSVD algorithm; specifically, it is simpler to understand and to program, and it offers improved numerical accuracy. In some cases, however, one or more columns of one of the matrix factors must be computed with another algorithm. In order to reflect that fact, the SSC QSVD algorithm is referred to herein as a partial QSVD. It turns out that the missing columns are not required in the implementation of the formulas to estimate all the system matrix parameters. Thus, the partial QSVD suffices for all the computations of interest in this report (see Section 3.1).

A.1 Quotient Singular Value Decomposition

Consider a pair of complex-valued matrices A and B for which a QSVD is desired, with A dimensioned as $m \times n$ and B dimensioned as $p \times n$. Consider also an $(m+p) \times n$ matrix C formed by concatenating A and B as

$$(A-1) \quad C = \begin{bmatrix} A \\ B \end{bmatrix}$$

In the QSVD formulation due to Paige and Saunders (1981) there are no restrictions on m , n , and p . However, let $m+p > n$ since such is the case in the context of the Van Overschee-De Moor algorithm. Let $k \leq n$ denote the rank of C ; that is, $k = \text{rank}(C)$. The SVD of matrix C is of the form

$$(A-2) \quad C = U_C S_C V_C^H = \begin{bmatrix} U_{C1} & U_{C2} \end{bmatrix} \begin{bmatrix} S_{C1} & [0]_{k,n-k} \\ [0]_{n-k,k} & [0]_{n-k,n-k} \\ [0]_{r,k} & [0]_{r,n-k} \end{bmatrix} \begin{bmatrix} V_{C1}^H \\ V_{C2}^H \end{bmatrix} = U_{C1} S_{C1} V_{C1}^H$$

here $r = m+p-n$, and matrix S_C has non-zero elements all along its main diagonal if $k = \text{rank}(C) = n$. Matrices U_C and V_C are both unitary, and the partitions of U_C , S_C , and V_C have compatible dimensions determined by the rank of C . The next step in the QSVD is to partition the $(m+p) \times k$ matrix U_{C1} as follows:

$$(A-3) \quad U_{C1} = \begin{bmatrix} U_{C11} \\ U_{C12} \end{bmatrix}$$

where U_{C11} is $m \times k$, and U_{C12} is $p \times k$. Assume now that $m < k$ and $p = k$. This occurs always in the computation of the backward F formula, and is one of the cases that will lead to ambiguous results in

most software implementations. Carrying out SVDs on matrix U_{C11} and on matrix U_{C12} results in

$$(A-4) \quad U_{C11} = US_A W_A^H = U \begin{bmatrix} S_1 & [0]_{m,k-m} \end{bmatrix} \begin{bmatrix} W_{A1}^H \\ W_{A2}^H \end{bmatrix}$$

$$(A-5) \quad U_{C12} = VT_1 W_B^H = VT_1 \begin{bmatrix} W_{B1}^H \\ W_{B2}^H \end{bmatrix}$$

where U and V and W_A and W_B are unitary matrices. S_1 and T_1 are real-valued, non-negative, diagonal matrices with the diagonal elements arranged in decreasing order of magnitude,

$$(A-6a) \quad s_i = S_1(i,i) \quad i = 1, 2, \dots, m$$

$$(A-6b) \quad 1 \geq s_1 \geq s_2 \geq \dots \geq s_m \geq 0$$

$$(A-6c) \quad t_i = T_1(i,i) \quad i = 1, 2, \dots, k$$

$$(A-6d) \quad 1 \geq t_1 \geq t_2 \geq \dots \geq t_k \geq 0$$

Consider now a re-ordering of the diagonal elements of T_1 such that

$$(A-7) \quad T_1(1,1) = t_k; \quad T_1(2,2) = t_{k-1}; \quad \dots; \quad T_1(k,k) = t_1$$

and reverse the order of the columns of V and of W_B accordingly. For the matrices resulting after these manipulations Paige and Saunders (1981) show that

$$(A-8) \quad W_A = W_B = W$$

and that the diagonal elements of S_1 and T_1 satisfy the following condition:

$$(A-9) \quad s_i^2 + t_i^2 = 1 \quad i=1,2,\dots,k$$

with $s_{m+1} = s_{m+2} = \dots = s_k = 0$ (these zero-valued elements of S_1 correspond with unity-valued elements of T_1).

Equation (A-8) is key to the Paige-Saunders QSVD formulation because it allows substitution of W in the place of W_A and W_B into Equations (A-4) and (A-5). It is then simple to show that the desired QSVD for the matrix pair (A,B) has the form

$$(A-10) \quad A = USX^H$$

$$(A-11) \quad B = VTX^H$$

with matrices S , T , and X given as

$$(A-12) \quad S = \begin{bmatrix} S_1 & [0]_{m,n-k} \end{bmatrix}$$

$$(A-13) \quad T = \begin{bmatrix} T_1 & [0]_{p,n-k} \end{bmatrix}$$

$$(A-14) \quad X = V_C \begin{bmatrix} S_{C1} W & [0]_{k,n-k} \\ [0]_{n-k,k} & I_{n-k} \end{bmatrix}$$

Notice that matrix X is nonsingular and that it does not have any particular structure. The real-valued number pairs (s_i, t_i) are the non-trivial singular value pairs of the matrix pair (A,B) . If $k < n$, there are $n-k$ trivial singular value pairs of the form $(0,0)$.

The source of the QSVD computational problem lies in the way that SVD routines compute the singular vectors for rectangular

matrices. Specifically, the singular vectors associated with zero-valued singular values are not unique, and most SVD routines select two different sets of null space singular vectors for two different matrices with the same null space. Additionally, the sign of the vectors is arbitrary, and small differences in the way that the calculations are carried out often lead to different signs for the same singular vectors.

A.2 Partial QSVD Algorithm

Recall that the relation in Equation (A-8) is key to the QSVD formulation. This relation, however, is an analytic result. Most software implementations of the QSVD will generate W_{B2} to be different from W_{A2} (see Equations (A-4) and (A-5)) when $m < k$ and $p = k$. This is due to the fact that W_{A2} spans the null space of matrix A , and representing the null space in the coordinates of any basis in that subspace leads to a correct SVD representation of A . However, for a correct QSVD representation of a matrix pair (A,B) , matrix W_{A2} must be represented in the same coordinates as matrix W_{B2} . Furthermore, the corresponding columns of W_A and of W_B must have the same sign. Most SVD numerical implementations will generate sign differences between the corresponding columns of W_A and of W_B even in the cases where $m = p = k$.

In their paper describing their QSVD formulation, Paige and Saunders (1981) mentioned that the relation in Equation (A-8) is not satisfied in some cases. However, based on their remarks it appears they did not recognize the reason for the problem, nor did they provide a procedure which avoids it. SSC has identified a procedure that avoids completely the following two issues: (a) non-uniqueness of the null-space of W_A , and (b) sign differences between the corresponding columns of W_A and W_B . The procedure generates the diagonal elements of T_1 in the correct order, Equation (A-7). Furthermore, a SVD of U_{C12} is not required, and

the calculations required in its place are numerically robust. The steps in the procedure are summarized next for the case of interest here; namely, for $m < k$ and $p = k$.

The SSC modifications to the above-defined QSVD algorithm start after Equation (A-5). In place of an SVD on matrix U_{C12} , define a new matrix D as

$$(A-15) \quad D = U_{C12} W_A$$

Notice that Equation (A-8) implies

$$(A-16) \quad D = U_{C12} W_A = U_{C12} W_B = V T_1$$

Now let $\underline{d}_i$ denote the i th column of D ,

$$(A-17) \quad D = [\underline{d}_1 \quad \underline{d}_2 \quad \dots \quad \underline{d}_k]$$

Next the $p (=k)$ diagonal elements of T_1 are obtained as,

$$(A-18) \quad t_i = |\underline{d}_i| \quad i = 1, 2, \dots, p$$

and the $p (=k)$ columns of V are obtained as,

$$(A-19a) \quad \underline{v}_i = \frac{\underline{d}_i}{|\underline{d}_i|} \quad i = 1, 2, \dots, p$$

$$(A-19b) \quad V = [\underline{v}_1 \quad \underline{v}_2 \quad \dots \quad \underline{v}_p]$$

Then matrices S , T , and X are given as in Equations (A-12)-(A-14). The essence of this modification is to avoid the SVD of an $m \times k$ matrix in Equation (A-5), avoid the re-ordering of the diagonal elements of T_1 in Equation (A-7) and the corresponding re-ordering of the columns of V , and avoid any required sign changes to the

columns of V (notice that a software or hardware implementation of the required re-orderings and sign changes is likely to involve significant calculations). These steps are replaced with the matrix product in Equation (A-15), the calculation of the norms of p vectors in Equation (A-18), and the normalization of p vectors in Equation (A-19a). Notice that since matrix W_A is unitary, the matrix product in Equation (A-15) is numerically stable. Furthermore, the calculations required in Equations (A-18) and (A-19a) are numerically stable also.

In the cases where $p > k$ and matrix V is needed directly, the above procedure generates only the last k columns of V . In such cases the remaining $p - k$ columns of V can be determined by calculating the null space of matrix D . It is important to note that such cases do not arise in the Van Overschee-De Moor identification algorithm. Indeed, only the product VT is required to implement the backward F formula, Equation (3-34).

APPENDIX B. COMBINED SYSTEM MATRIX ESTIMATION FORMULA

The combined F formula, Equation (3-35), provides an improved estimate of the system matrix in the cases where the duration of the multichannel output sequence is short (number of data vectors, N_T , is small). The formula proposed by Van Overschee and De Moor (1991) is straightforward from a conceptual viewpoint, but involves a large number of computations. In Phase II SSC formulated an approach based on Kronecker product algebra to solve the combined system matrix estimation problem which results in a simple, closed-form solution, as summarized next.

The combined F formula is the solution to a least-squares problem formulated using the state propagation equations of both the forward and backward innovations representation (Van Overschee and De Moor, 1991). Specifically, the combined least-squares problem is formulated as

$$(B-1) \quad F_c = \min_F \left\{ |Z_{L+1} - FZ_L|^2 + |W_{L+1} - F^H W_L|^2 \right\}$$

where F_c denotes the combined F , matrices Z_L and Z_{L+1} are forward Kalman state matrices, and matrices W_L and W_{L+1} are backward Kalman state matrices (Van Overschee and De Moor, 1991; Román and Davis, 1993a). After some manipulations (including minimization), the problem in Equation (B-1) can be expressed as

$$(B-2) \quad F_c Z_L Z_L^H + W_L W_L^H F_c = Z_{L+1} Z_{L+1}^H + W_L W_{L+1}^H$$

This is the combined F formula proposed by Van Overschee and De Moor (1991). Equation (B-2) is a Sylvester equation, and in the general case sophisticated techniques are required to solve it. However, it turns out that a simple, closed-form solution is possible.

Each of the individual terms in Equation (B-2) can be modified using several equivalences. Van Overschee and De Moor (1991; 1993) have shown that

$$(B-3) \quad Z_L Z_L^H = W_L W_L^H = S_L$$

$$(B-4) \quad Z_{L+1} Z_L^H = F_f S_L$$

$$(B-5) \quad W_L W_{L+1}^H = S_L F_b$$

where F_f is the forward F matrix (Equation (3-33)), F_b is the backward F matrix (Equation (3-34)), and S_L is a square diagonal matrix with non-negative diagonal elements (Equation (3-18) and Table 2-1). Substitution of Equations (B-3)-(B-5) into Equation (B-2) results in

$$(B-6) \quad F_c S_L + S_L F_c = F_f S_L + S_L F_b$$

This equation can be transformed into a simpler form via the application of Kronecker product notation. Let f_{ijc} , f_{ijf} , and f_{ijb} denote the (i,j) th element of F_c , F_f , and F_b , respectively. And now define N^2 -dimensional vectors column vectors $\underline{f}_c$, $\underline{f}_f$, and $\underline{f}_b$ by concatenating the rows of matrices F_c , F_f , and F_b , respectively. Specifically,

$$(B-7) \quad \underline{f}_c = [f_{11c} \ f_{12c} \ \cdots \ f_{1Nc} \ f_{21c} \ \cdots \ f_{2Nc} \ \cdots \ f_{N1c} \ \cdots \ f_{NNc}]^T$$

$$(B-8) \quad \underline{f}_f = [f_{11f} \ f_{12f} \ \cdots \ f_{1Nf} \ f_{21f} \ \cdots \ f_{2Nf} \ \cdots \ f_{N1f} \ \cdots \ f_{NNf}]^T$$

$$(B-9) \quad \underline{f}_b = [f_{11b} \ f_{12b} \ \cdots \ f_{1Nb} \ f_{21b} \ \cdots \ f_{2Nb} \ \cdots \ f_{N1b} \ \cdots \ f_{NNb}]^T$$

Then consider the following $N^2 \times N^2$ block diagonal matrices defined using Kronecker product notation (Pease, 1965),

$$(B-10) \quad I_N \times S_L = \begin{bmatrix} S_L & O_N & \cdots & O_N \\ O_N & S_L & \cdots & O_N \\ \vdots & \vdots & \ddots & \vdots \\ O_N & O_N & \cdots & S_L \end{bmatrix}$$

$$(B-11) \quad S_L \times I_N = \begin{bmatrix} s_1 I_N & O_N & \cdots & O_N \\ O_N & s_2 I_N & \cdots & O_N \\ \vdots & \vdots & \ddots & \vdots \\ O_N & O_N & \cdots & s_N I_N \end{bmatrix}$$

where I_N is the $N \times N$ identity matrix, O_N is an $N \times N$ matrix of zeros, $\times$ denotes the Kronecker product, and s_i denotes the i th diagonal element of the diagonal matrix S_L . Based on these definitions it is trivial to show that the following correspondences are valid:

$$(B-12) \quad F_{(\bullet)} S_L \Leftrightarrow [I_N \times S_L] f_{(\bullet)}$$

$$(B-13) \quad S_L F_{(\bullet)} \Leftrightarrow [S_L \times I_N] f_{(\bullet)}$$

where $(\bullet)$ denotes c, f, or b. Then, it follows that Equation (B-6) is equivalent to the following expression,

$$(B-14) \quad \{[I_N \times S_L] + [S_L \times I_N]\} f_c = [I_N \times S_L] f_f + [S_L \times I_N] f_b$$

Notice that the $N^2 \times N^2$ matrix $[I_N \times S_L] + [S_L \times I_N]$ is block diagonal, with i th block element $S_L + s_i I_N$. Each block element $S_L + s_i I_N$ is itself a block diagonal matrix with j th element $s_j + s_i$. Using identical index assignments for each of the two factors on the right-hand-side of Equation (B-14) leads to

$$(B-15a) \quad (s_j + s_i) f_{ijc} = s_j f_{ijc} + s_i f_{ijb} \quad i = 1, 2, \dots, N; \quad j = 1, 2, \dots, N$$

$$(B-15b) \quad f_{ijc} = \frac{s_j f_{ijf} + s_i f_{ijb}}{s_i + s_j} \quad i = 1, 2, \dots, N; \quad j = 1, 2, \dots, N$$

which is the desired solution to Equation (B-2). This formula has interesting features. Specifically, in the case where the number of multichannel output vectors is large ($f_{ijf} = f_{ijb}$), the solution is $f_{ijc} = f_{ijf} = f_{ijb}$, as expected. Also, in the special case where all the diagonal elements of S_L are equal ($s_i = s_j$ for all i and j), the solution is $f_{ijc} = (f_{ijf} + f_{ijb})/2$, the algebraic average of the forward and backward solutions, as dictated by intuition.

APPENDIX C. SPATIAL FILTERING AND THE LDU DECOMPOSITION

The LDU decomposition is a powerful analysis tool which admits efficient numerical implementation. Furthermore, this decomposition is related to optimal linear filtering. Given a square ($J \times J$), Hermitian matrix Ω , the LDU decomposition of Ω is defined as

$$(C-1) \quad \Omega = LDL^H$$

where L is a $J \times J$ complex-valued, lower-triangular matrix with unity-valued elements along the main diagonal, and D is a $J \times J$ diagonal matrix with real-valued, non-negative diagonal entries. In this factorization Ω can be rank-deficient, and the rank deficiency of Ω is manifested with a corresponding number of zeros along the diagonal of D . Matrix L , however, is full rank with unity-valued determinant (the determinant of a diagonal matrix is equal to the product of the diagonal entries). Therrien (1983) has shown that the rows of matrix L^{-1} correspond to the coefficients (in reverse order) of the optimum linear prediction filters of orders 0 through $J-1$, and the diagonal elements of D_i are the corresponding prediction error variances. In the context of interest herein, Ω is the covariance matrix of the temporal innovations vector, and L^{-1} is a linear transformation applied to the temporal innovations in order to diagonalize Ω . Thus, the LDU decomposition is equivalent to linear spatial filtering which spatially whitens the temporal innovations.

Consider the temporal innovations sequence, $\{\underline{x}(n)\}$, under either hypothesis, and apply a linear transformation T^H to obtain a temporally- and spatially-whitened process $\{\underline{y}(n)\}$ (Equation (4-2)). Specifically,

$$(C-2) \quad \underline{y}(n) = T^H \underline{x}(n) = L^{-1} \underline{x}(n)$$

It follows from Equations (C-1) and (C-2) that the covariance matrix of $\underline{y}(n)$ is diagonal,

$$(C-3) \quad \Sigma = E[\underline{y}(n)\underline{y}^H(n)] = T^H \Omega T = L^{-1} \Omega (L^{-1})^H = D$$

The equivalence between spatial whitening and the diagonalization of Ω is demonstrated next.

Let a_{ij}^* denote the i th complex-valued coefficient of a j th-order optimal linear prediction filter (the asterisk denotes complex conjugation). Recall that linear prediction filters have the structure of an auto-regressive (AR) system. In terms of these coefficients, the structure of L^{-1} is (Therrien, 1983)

$$(C-4) \quad L^{-1} = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 & 0 \\ a_{11}^* & 1 & 0 & \cdots & 0 & 0 \\ a_{22}^* & a_{12}^* & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ a_{J-2,J-2}^* & a_{J-3,J-2}^* & a_{J-4,J-2}^* & \cdots & 1 & 0 \\ a_{J-1,J-1}^* & a_{J-2,J-1}^* & a_{J-3,J-1}^* & \cdots & a_{1,J-1}^* & 1 \end{bmatrix}$$

Given this association, the j th row of Equation (C-2) is expressed as,

$$(C-5) \quad v_j = \varepsilon_j + \sum_{i=1}^{j-1} a_{ij-1}^* \varepsilon_{j-i} = \varepsilon_j - \hat{\varepsilon}_j \quad j = 1, \dots, J$$

where the caret (^) over ε_j denotes the minimum variance estimate of ε_j . From Equation (C-5), v_j can be interpreted as the error in

the linear prediction of ε_j from $\{\varepsilon_{j-1}, \dots, \varepsilon_1\}$. Also, the elements of the instantaneous (fixed n) vector $\underline{v}$ can be viewed as a finite-length sequence $\{v_1, \dots, v_J\}$. A sequence generated in such a manner is white. Thus, the linear transformation in Equation (C-2) is a spatial whitening filter.

Consider, in particular, the last (J th) row of L^{-1} . The elements in this row are the coefficients of the highest-order linear predictor that can be defined for a sequence of length J . Therefore, these elements can be viewed as the J spatial weights that remove the residual spatial correlation from the temporal innovations. With this interpretation, the frequency response of these weights provides the spatial cancelation pattern of the spatial filter. This important point is explored further below.

The transfer function of a scalar AR system is an all-pole function, and the inverse of an all-pole AR system is an all-zero moving-average (MA) system. AR and MA time series models are causal and causally-invertible; thus, their associated state space models are innovations representations in the sense of Section 2.5. In both systems the set of coefficients is the same, but the manner in which the equations are expressed differs. Equation (C-5) represents an MA system (Appendix G). And if the summation in Equation (C-5) is transferred to the other side of the equal sign, then an AR system is obtained (Appendix G),

$$(C-6) \quad \varepsilon_j = - \sum_{i=1}^{j-1} a_{jj-i}^* \varepsilon_{j-i} + v_j$$

In order to determine the transfer function of the system in Equation (C-5) using the standard notation for MA systems (Appendix G), consider the case for $j=J$ and let

$$(C-7a) \quad b_{0,J-1} = 1$$

$$(C-7b) \quad b_{i,J-1} = a_{i,J-1} \quad i = 1, \dots, J-1$$

Now Equation (C-5) for $j = J$ can be expressed as

$$(C-8) \quad v_J = \sum_{i=0}^{J-1} b_{i,J-1}^* \varepsilon_{J-i}$$

Since the variables in Equation (C-8) are functions of a discrete parameter (the integer-valued index i), the z -transform is the appropriate tool for determination of the transfer function. Application of the z -transform to Equation (C-8) results in the expression

$$(C-9) \quad N_J(z) = \sum_{i=0}^{J-1} b_{i,J-1}^* z^i E_J(z) = B_J(z) E_J(z)$$

here z denotes the transform variable, and $E_J(z)$ and $N_J(z)$ are the z -transforms of the sequences $\{\varepsilon_1, \dots, \varepsilon_J\}$ and $\{v_1, \dots, v_J\}$, respectively. Additionally, $B_J(z)$ has been defined implicitly as

$$(C-10) \quad B_J(z) = \sum_{i=0}^{J-1} b_{i,J-1}^* z^i$$

The transfer function for this system is obtained directly from Equation (C-9) as the following scalar function,

$$(C-11) \quad T_{MA}(z) = \frac{N_J(z)}{E_J(z)} = B_J(z)$$

where the subscript MA is used to denote that Equation (C-8) is a moving-average system.

Given the transfer function $T_{MA}(z)$, the frequency response is obtained by substituting $z = \exp(j\omega_s) = \exp(j2\pi f_s)$ in Equation (C-11) to obtain (the subscript s in the frequency variables is used to denote that these are spatial frequencies)

$$(C-12) \quad T_{MA}(f_s) = B_J(f_s) = \sum_{k=0}^{J-1} b_{k,J-1}^* e^{-j2\pi k f_s}$$

The summation in Equation (C-12) is the discrete Fourier transform (DFT) of the coefficients $\{b_{k,J-1}^* | k=0, \dots, J-1\}$ defined in Equation (C-7). Therefore, the DFT of these coefficients is the spatial frequency response (beam pattern) of the set of spatial weights that whiten in space the temporal innovations. Zero-padding should be used to calculate the DFT in order to obtain sufficient detail in the spatial frequency domain.

APPENDIX D. HYPOTHESIS FILTER DESIGN

The design of each hypothesis filter is an iterative process which requires determination of the goodness of each intermediate filter design. Such determination must be carried out using measures of goodness that are robust and relevant. A key aspect of the SSC model-based detection methodology is the implementation of the detection decision using the hypothesis filter residuals. Thus, the approach adopted herein to establish the goodness of a filter design is based on examination of the characteristics of the filter residuals. The discussion is presented for the case of complex-valued data since the relevant modifications for the real-valued data case are straightforward.

Some aspects of the approach described in this appendix can be used also as criteria for detection decisions in cases where the dual-hypothesis log-likelihood ratio test of Section 5.0 is inappropriate. One such case is the ECG diagnosis application discussed in Section 7.0.

Consider the output of a hypothesis filter in the design step of the dual-hypothesis case; that is, the N -point residual sequence $\{\underline{e}(n|H_i)\}$ where $i=0, 1$. When the channel output sequence condition matches the filter design condition, the residual vector sequence is characterized as follows:

- $\mathcal{N}(\underline{0}, \Omega(H_i))$: Gaussian-distributed with mean zero and covariance matrix $\Omega(H_i)$;
- circular: independent real and imaginary components with equal variance; and
- white.

Of course, the circular feature is relevant only for complex-valued residuals. This set of characteristics suggests a residual evaluation procedure that consists of:

- (a) a zero-mean test (based on Student's t distribution);

- (b) a power test (based on Snedecor's F distribution); and
- (c) a whiteness test (based on the scaled SL distribution for small N and on the Gaussian distribution for large N).

For analytical simplicity, each of these tests is applied independently. Also, all three tests are applied to each scalar element of the residual vector individually.

Given the type of tests discussed in this appendix, it is appropriate to introduce several notational simplifications. First, a scalar residual process is assumed; this avoids having to introduce a subscript to denote a representative scalar element of the residual vector. Second, the hypothesis argument H_i is dropped; this is appropriate because the testing criteria are based on the condition that the hypothesis and the data are matched. Third, the tests are presented for complex-valued data; corresponding tests for real-valued data are similar and can be obtained by inspection of the complex-valued case results. Fourth, the true mean, variance, and auto-correlation sequence (ACS) of the complex-valued, scalar, stationary, white residual process $\{\varepsilon(n) = \varepsilon_r(n) + j\varepsilon_i(n)\}$ are defined as:

$$(D-1a) \quad \mu_\varepsilon = E[\varepsilon(n)] = 0$$

$$(D-1b) \quad \mu_{\varepsilon r} = \Re\{\mu_\varepsilon\} = E[\varepsilon_r(n)] = 0$$

$$(D-1c) \quad \mu_{\varepsilon i} = \Im\{\mu_\varepsilon\} = E[\varepsilon_i(n)] = 0$$

$$(D-2a) \quad \sigma_\varepsilon^2 = E[\{\varepsilon(n) - \mu_\varepsilon\} \{\varepsilon(n) - \mu_\varepsilon\}^*] = E[|\varepsilon(n) - \mu_\varepsilon|^2] = E[|\varepsilon(n)|^2]$$

$$(D-2b) \quad \sigma_{\varepsilon r}^2 = E[\{\varepsilon_r(n) - \mu_{\varepsilon r}(n)\}^2] = E[\varepsilon_r^2(n)] = \frac{\sigma_\varepsilon^2}{2}$$

$$(D-2b) \quad \sigma_{\epsilon_i}^2 = E[\{\epsilon_i(n) - \mu_{\epsilon_i}(n)\}^2] = E[\epsilon_i^2(n)] = \frac{\sigma_{\epsilon}^2}{2}$$

$$(D-3) \quad r_{\epsilon}(m) = E[\epsilon(n) \epsilon^*(n-m)]$$

For a zero-mean process (as considered herein), $\sigma_{\epsilon}^2 = r_{\epsilon}(0)$. These true process parameters are referred to throughout this appendix. It is important to note that the true process parameters listed in Equations (D-1)-(D-3) are the system model parameters, and thus are known in the context of hypothesis filter design as well as in the context of detection decisions.

D.1 Zero-Mean Test

The sample mean for a finite-length sequence $\{\epsilon(n) | n=0, 1, \dots, N-1\}$ is the time-average estimate of the mean of the process; that is,

$$(D-4) \quad \hat{\mu}_{\epsilon} = \frac{1}{N} \sum_{n=0}^{N-1} \epsilon(n)$$

Now denote the real and imaginary parts of the sample mean as

$$(D-5a) \quad \hat{\mu}_{\epsilon r} = \Re\{\hat{\mu}_{\epsilon}\}$$

$$(D-5b) \quad \hat{\mu}_{\epsilon i} = \Im\{\hat{\mu}_{\epsilon}\}$$

respectively. The sample mean is Gaussian-distributed with mean equal to the true mean,

$$(D-6) \quad \mu_{\hat{\mu}} = E[\hat{\mu}_{\epsilon}] = \mu_{\epsilon} = 0$$

and variance given as

$$(D-7) \quad \sigma_{\hat{\mu}}^2 = E\left[(\hat{\mu}_{\epsilon} - \mu_{\epsilon})(\hat{\mu}_{\epsilon} - \mu_{\epsilon})^*\right] = \frac{\sigma_{\epsilon}^2}{N}$$

where σ_{ϵ}^2 is the variance of the residual process, as defined in Equation (D-2). The mean and variance of the real and imaginary parts of the sample mean are:

$$(D-8a) \quad \mu_{\hat{\mu}r} = E[\hat{\mu}_{er}] = \mu_{er} = 0$$

$$(D-8b) \quad \mu_{\hat{\mu}i} = E[\hat{\mu}_{ei}] = \mu_{ei} = 0$$

$$(D-9a) \quad \sigma_{\hat{\mu}r}^2 = E\left[(\hat{\mu}_{er} - \mu_{er})^2\right] = \frac{\sigma_{\epsilon}^2}{2N}$$

$$(D-9b) \quad \sigma_{\hat{\mu}i}^2 = E\left[(\hat{\mu}_{ei} - \mu_{ei})^2\right] = \frac{\sigma_{\epsilon}^2}{2N}$$

where μ_{er} and μ_{ei} denote the real and imaginary parts, respectively, of the true process mean, as defined in Equation (D-1).

It is well known that the statistical inference method of statistics can be applied to test a real-valued sample mean with a two-sided t-test (see, for example, Frieden [1983]). Since the data is complex-valued, each component (real and imaginary) of the sample mean is tested independently. The test on the mean is discussed herein for the real component only since the approach and formulas are identical for the imaginary component. In the statistical inference method as applied to the sample mean, two hypotheses are formulated: (a) a null hypothesis representing the condition that the sample mean $\hat{\mu}_{er}$ arises from a population with process mean $\mu_{er} = 0$; and (b) an alternative hypothesis representing the condition that the sample mean $\hat{\mu}_{er}$ arises from a population with process mean $\mu_{er} \neq 0$. That is,

HYPOTHESES FOR TEST ON THE MEAN:

$$(D-10a) \quad \text{NULL:} \quad \hat{\mu}_{er} = \mu_{er} = 0$$

$$(D-10b) \quad \text{ALTERNATIVE:} \quad \hat{\mu}_{er} = \mu_{er} \neq 0$$

It is important to note that in cases where the true mean value is unknown, a test value is used in place of the unknown process mean. The two-sided test on the mean for the hypotheses (D-10) is

$$(D-11) \quad |\hat{\mu}_{er}| \begin{matrix} \mu_{er} \neq 0 \\ > \\ < \\ \mu_{er} = 0 \end{matrix} \tau_{\mu}(\alpha)$$

where $\tau_{\mu}(\alpha)$ is a real-valued, positive scalar threshold for the two-sided test on the mean at α significance level (the significance level in a statistical test is the probability of false alarm in detection theory). The sample mean threshold is determined as

$$(D-12) \quad \tau_{\mu}(\alpha) = \frac{s_{er}}{\sqrt{N-1}} \tau_t(\alpha) + \mu_{er} = \frac{s_{er}}{\sqrt{N-1}} \tau_t(\alpha)$$

where $\tau_t(\alpha)$ is a real-valued, positive scalar threshold for the two-sided t-test at α significance level, and s_{er} is defined as

$$(D-13) \quad s_{er}^2 = \frac{1}{N} \sum_{n=0}^{N-1} [\Re\{\varepsilon(n)\} - \hat{\mu}_{er}]^2 = \frac{1}{N} \sum_{n=0}^{N-1} [\varepsilon_r(n) - \hat{\mu}_{er}]^2$$

Notice that s_{er}^2 is a biased estimate of the variance of the real part of the residual, $\Re\{\varepsilon(n)\}$ (the unbiased estimate of the variance has $(N-1)^{-1}$ in place of N^{-1} as the multiplicative factor).

The two-sided t-test threshold is the upper integration limit which satisfies the following integral equation,

$$(D-14) \quad 1-\alpha = 1-P_{FA} = \mathcal{P}[|t| \leq \tau_t(\alpha)] = 2 \int_0^{\tau_t(\alpha)} p_T(t, N-1) dt$$

where P_{FA} denotes the "probability of false alarm", $\mathcal{P}[\bullet]$ denotes the probability of event $[\bullet]$, variable t is t-distributed with $N-1$ degrees-of-freedom, and $p_T(t, N-1)$ is the PDF of the t-distribution with $N-1$ degrees-of-freedom. Variable t is related to the sample mean according to

$$(D-15) \quad t = \sqrt{N-1} \frac{\hat{\mu}_{er} - \mu_{er}}{s_{er}} = \sqrt{N-1} \frac{\hat{\mu}_{er}}{s_{er}}$$

and the PDF of the t-distribution with $N-1$ degrees-of-freedom is (Hastings and Peacock, 1975)

$$(D-16) \quad p_T(t, N-1) = \frac{\Gamma\left[\frac{N}{2}\right]}{\sqrt{(N-1)\pi} \Gamma\left[\frac{N-1}{2}\right]} \left(1 + \frac{t^2}{N-1}\right)^{-\frac{N}{2}} \quad -\infty \leq t \leq \infty; N > 1$$

where $\Gamma[\bullet]$ denotes the gamma function. The t distribution (as defined herein as a function of the number of data points in the sequence, N) has zero mean for all admissible values of N , and variance

$$(D-17) \quad \sigma_t^2 = \frac{N-1}{N-3} \quad N > 3$$

For $N=2$ and $N=3$ the variance is undefined, even though the PDF is defined. As N increases, the t distribution approximates the

standard Gaussian distribution, $\mathcal{N}(0,1)$; in fact, for $N > 30$ the fit is very good.

The t-test threshold, $\tau_t(\alpha)$, is calculated numerically for a specified value of α using Equation (D-14). First the integral is evaluated for an initial value of $\tau_t(\alpha)$ as its upper limit, and the result is compared to α . If the computed value is more than α , then the threshold value is increased and the integral is evaluated again using the new upper limit. If the computed value is less than α , then the threshold value is decreased and the integral is evaluated again using the new upper limit. This process is repeated until the computed integral value is within a pre-set tolerance of α (a good value for the tolerance constant is 10^{-8}). Both thresholds, $\tau_t(\alpha)$ and $\tau_\mu(\alpha)$, are expressed herein as a function of α to emphasize their dependence on the significance level parameter. The t-test threshold can be calculated also using the complement of Equation (D-14); that is, with the integral evaluated from $\tau_t(\alpha)$ to $+\infty$. However, Equation (D-14) is simpler to implement numerically.

D.2 Power Test

The three most common ACS estimators are presented in Appendix E. Each estimator has distinct statistical features and leads to different results when utilized for model identification and other problems. However, in all three cases the sample autocorrelation at lag 0 for a finite-length sequence $\{\varepsilon(n) | n=0, 1, \dots, N-1\}$ is the unbiased time-average estimate of the power of the process; that is,

$$(D-18) \quad \hat{r}_\varepsilon(0) = \frac{1}{N} \sum_{n=0}^{N-1} \varepsilon(n) \varepsilon^*(n) = \frac{1}{N} \sum_{n=0}^{N-1} |\varepsilon(n)|^2$$

Each term $|e(n)|^2$ in the summation is exponentially-distributed, and the sum of N exponentially-distributed random variables is Erlang-distributed with shape parameter N (the Erlang distribution is a special case of the gamma distribution). Thus, the sample power is real-valued, and is Erlang-distributed with mean equal to the true power,

$$(D-19) \quad \mu_{\hat{r}}(0) = E[\hat{r}_e(0)] = r_e(0) = \sigma_e^2$$

and variance given as (Michels, 1992a; 1992b)

$$(D-20) \quad \sigma_{\hat{r}}^2(0) = E\left[\{\hat{r}_e(0) - r_e(0)\}^2\right] = \frac{\sigma_e^4}{N}$$

where σ_e is the standard deviation of the residual process, as defined in Equation (D-2). A multiplication factor transforms the Erlang-distributed variable $\hat{r}_e(0)$ into the random variable χ^2 which is distributed as chi-squared with $2N$ degrees-of-freedom; specifically,

$$(D-21) \quad \chi^2 = \frac{2N}{\sigma_e^2} \hat{r}_e(0)$$

The chi-squared distribution with $2N$ degrees-of-freedom is denoted compactly as $\chi^2(2N)$. The mean of the $\chi^2(2N)$ distribution is $2N$, and the variance is $4N$.

Statistical inference can be applied to test the sample power with a threshold-based test. In the statistical inference method as applied to the sample power, two hypotheses are formulated: (1) a null hypothesis representing the condition that the sample power $\hat{r}_e(0)$ arises from a population with process power $r_e(0)$; and (b) an alternative hypothesis representing the condition that the sample

power $\hat{r}_\varepsilon(0)$ arises from a population with process power distinct from $r_\varepsilon(0)$. That is,

HYPOTHESES FOR TEST ON THE POWER:

$$(D-22a) \quad \text{NULL:} \quad \hat{r}_\varepsilon(0) = r_\varepsilon(0)$$

$$(D-22b) \quad \text{ALTERNATIVE:} \quad \hat{r}_\varepsilon(0) \neq r_\varepsilon(0)$$

The null hypothesis can be tested with a two-sided χ^2 -test. However, such a test is difficult to implement because the χ^2 PDF is asymmetric, which complicates the iterative procedure to calculate the threshold. This difficulty is avoided by utilizing a conditional application of the one-sided F-test, which is a test to determine the equality of two variances (see, for example, Frieden [1983]). The approach is as follows. First define a random variable f as:

$$(D-23) \quad f = \begin{cases} \frac{\hat{r}_\varepsilon(0)}{r_\varepsilon(0)} & \text{if Condition A: } \hat{r}_\varepsilon(0) \geq r_\varepsilon(0) \\ \frac{r_\varepsilon(0)}{\hat{r}_\varepsilon(0)} & \text{if Condition B: } \hat{r}_\varepsilon(0) < r_\varepsilon(0) \end{cases}$$

When Condition A is true, f is F-distributed with $2N$ and ∞ degrees-of-freedom, which is denoted as $F(2N, \infty)$. Whereas when Condition B is true, f is distributed as $F(\infty, 2N)$. This allows application of a one-sided test on the power ratio f for the hypotheses (D-22) of the form

$$(D-24) \quad f \begin{matrix} \hat{r}_\varepsilon(0) \neq r_\varepsilon(0) \\ > \\ < \\ \hat{r}_\varepsilon(0) = r_\varepsilon(0) \end{matrix} \tau_f(\alpha)$$

where $\tau_f(\alpha)$ is a real-valued, positive scalar threshold for the one-sided F-test at α significance level. Notice that both Conditions A and B are handled with the same test. This is due to the reciprocal symmetry of the F distribution. The sample power ratio threshold is determined as

$$(D-25) \quad \tau_f(\alpha) = \begin{cases} \frac{\tau_{\chi^2}(\alpha)}{2N} & \text{if Condition A: } \hat{r}_\epsilon(0) \geq r_\epsilon(0) \\ \frac{2N}{\tau_{\chi^2}(\alpha)} & \text{if Condition B: } \hat{r}_\epsilon(0) < r_\epsilon(0) \end{cases}$$

where $\tau_{\chi^2}(\alpha)$ is a real-valued, positive scalar threshold for the one-sided χ^2 -test at α significance level. The one-sided χ^2 -test threshold for use in Equation (D-25) is the upper integration limit which satisfies an integral equation corresponding to the condition that is true for the sample power value to be tested. Specifically,

$$(D-26a) \quad \text{Condition A: } 1 - \alpha = 1 - P_{FA} = \mathcal{P}\left[0 \leq \chi^2 \leq \tau_{\chi^2}(\alpha)\right] = \int_0^{\tau_{\chi^2}(\alpha)} p_C(c, 2N) dc$$

$$(D-26b) \quad \text{Condition B: } \alpha = P_{FA} = \mathcal{P}\left[0 \leq \chi^2 \leq \tau_{\chi^2}(\alpha)\right] = \int_0^{\tau_{\chi^2}(\alpha)} p_C(c, 2N) dc$$

where variable χ^2 is χ^2 -distributed with $2N$ degrees-of-freedom, and $p_C(c, 2N)$ is the PDF of the χ^2 -distribution with $2N$ degrees-of-freedom (the dummy variable c is used to represent the χ^2 random variable). Equation (D-21) defines the relation between the sample power and the χ^2 variable. The PDF of the χ^2 -distribution with $2N$ degrees-of-freedom is (Hastings and Peacock, 1975)

$$(D-27) \quad p_C(c, 2N) = \frac{1}{2^N \Gamma[N]} c^{N-1} e^{-c/2} \quad 0 \leq c \leq \infty; \quad N \geq 1$$

where, as before, $\Gamma[\bullet]$ denotes the gamma function. The mean of the χ^2 distribution with $2N$ degrees-of-freedom (as defined herein as a function of the number of data points in the sequence, N) is $2N$, and the variance is $4N$ for all admissible values of N . As N increases, the χ^2 distribution approximates the Gaussian distribution with mean $[4N-1]^{1/2}$ and unit variance, $\mathcal{N}([4N-1]^{1/2}, 1)$; in fact, for $N > 15$ (more than 30 degrees-of-freedom) the fit is very good.

Equation (D-25) follows from the relationship satisfied by the random variables f and χ^2 ; namely,

$$(D-28) \quad f = \begin{cases} \frac{\chi^2}{2N} & \text{if Condition A: } \hat{r}_\epsilon(0) \geq r_\epsilon(0) \\ \frac{2N}{\chi^2} & \text{if Condition B: } \hat{r}_\epsilon(0) < r_\epsilon(0) \end{cases}$$

This relation, in turn, follows from Equations (D-2a), (D-3), (D-21), and (D-23). It is important to note that Equations (D-25) and (D-28) are valid only for the cases where one of the two degrees-of-freedom parameters of the F distribution is infinite (corresponding to a known power variable). Such is the case herein with the variable $r_\epsilon(0)$ used in the power ratio f .

The χ^2 -test threshold, $\tau_{\chi^2}(\alpha)$, is calculated numerically for a specified value of α via Equation (D-26) using the approach outlined in Section D.1 for the t-test threshold calculation. As for the t-test, the formulation based on integration from zero to the threshold value (Equation (D-26)) is preferred over the

formulation where the integration limits are the threshold value and infinity.

An equivalent test can be defined on the sample variance instead of the sample power, and the approach is analogous. The power test is selected herein because generation of the sample variance requires more computations.

D.3 Auto-Correlation Sequence Whiteness Test

A whiteness test that is applied to the estimated lags can be configured for ACS estimates determined using any one of the three ACS estimators presented in Appendix E. However, the circular ACS estimator is selected herein for two main reasons. First, the circular estimator is ideally-suited for a finite-length white noise sequence. This is due to the pairwise independence of all the elements in a finite-length white noise sequence, which insures that substitution of the elements at the beginning of the sequence in place of unavailable elements at the end of the sequence is in accordance with the structure of the true ACS. Second, the PDF of the circular estimator is the same at all lags, which allows for a single threshold test to be defined for all lags. In contrast, the PDF of the biased and unbiased estimators is different for each lag because the number of data points used in the estimate at each lag decreases as the lag index increases (see Appendix E).

In general, for complex-valued data the estimated circular ACS $\{\hat{f}_e(m)\}$ is complex-valued also at lags $m \neq 0$ (with exception of lag $m = N/2$ for N even, which is real-valued for all data realizations). In the approach established herein the whiteness test is applied independently to the real and imaginary components of the ACS. An alternative approach is to define a whiteness test for the magnitude or the magnitude-squared of the ACS. Such a

test requires knowledge of the PDF of the magnitude or of the magnitude-squared of each lag of the ACS, both of which are unknown to the authors for the case where N is small-valued. For large-valued N , the PDF of the magnitude-squared of each lag of the ACS approximates the exponential PDF.

Two cases are considered herein separately as a function of data sequence length, N . In cases where N assumes a small value the exact PDF of each ACS lag is used to establish the test, whereas the Gaussian PDF approximation is used in cases where N assumes large values. Software-based analyses indicate that $N=50$ is a good value to define the boundary between the two cases. In all cases in Sections D.3.1 and D.3.2 below, $\hat{r}(m)$ is used to denote either the real or the imaginary component of the complex-valued, circular estimate of the scalar residual ACS. That is, $\hat{r}(m) = \hat{r}_{er}(m)$ or $\hat{r}(m) = \hat{r}_{ei}(m)$, where $\hat{r}_e(m) = \hat{r}_{er}(m) + j \hat{r}_{ei}(m)$. Besides simplifying the notation, this emphasizes the fact that the statistical test defined herein is the same for both ACS components (real and imaginary). The only difference between the two applications of the test is that for even-valued N , at lag $m = N/2$ the imaginary component estimate is always zero, and the variance of the real component is twice the value of the variance of the real component in other lags. This implies that at lag $m = N/2$ when N is even, a different threshold is calculated for the real component, and no threshold is required for the imaginary component.

In both cases considered herein the whiteness test is applied to the subset of the principal lags which excludes the zeroth lag, since lag $m=0$ is tested separately (Section (D.2)). Specifically, the real (or imaginary) component of the circular ACS estimate, $\{\hat{r}(m) | m = 1, 2, \dots, M_c\}$, with M_c as defined in Equation (E-7), is tested for whiteness. For a white residual, the mean and variance of the real (or imaginary) component $\hat{r}(m)$ are (Section E.1)

$$(D-29) \quad \mu_r(m) = 0 \quad 1 \leq m \leq M_c$$

$$(D-30) \quad \sigma_r^2(m) = \begin{cases} \frac{\sigma_\varepsilon^4}{2N} & \begin{array}{l} \text{real or imaginary component; } 1 \leq m \leq M_c; \text{ } N \text{ odd} \\ \text{real or imaginary component; } 1 \leq m \leq M_c - 1; \text{ } N \text{ even} \end{array} \\ \frac{\sigma_\varepsilon^4}{N} & \text{real component; } m = M_c; \text{ } N \text{ even} \\ 0 & \text{imaginary component; } m = M_c; \text{ } N \text{ even} \end{cases}$$

These two equations follow from Equations (E-19b) and (E-20), respectively, together with the circular property of the residual sequence (circularity accounts for the factor 2 in the denominator of the first condition in Equation (D-30)).

Statistical inference also allows definition of a threshold-based test for whiteness. The formulation and form of the test is common to both cases (small N ; large N), and only the PDF type and the resulting threshold value varies between the cases. Thus, the common parts of the approach are presented next. For the context herein, two hypotheses are formulated: (1) a null hypothesis representing the condition that the real (or imaginary) component of the residual ACS lag m arises from a population with true ACS lag $r(m)$; and (2) an alternative hypothesis representing the condition that the real (or imaginary) component of the residual ACS lag m arises from a population with true ACS lag distinct from $r(m)$. That is,

HYPOTHESES FOR TEST OF WHITENESS:

$$(D-31a) \quad \text{NULL:} \quad \hat{r}(m) = r(m) \quad 1 \leq m \leq M_c$$

$$(D-31b) \quad \text{ALTERNATIVE:} \quad \hat{r}(m) \neq r(m) \quad 1 \leq m \leq M_c$$

The null hypothesis can be tested with a two-sided test, and the test is applied at each lag index $m=1,2,\dots,M_c$. A two-sided test applies because the PDFs corresponding to the two cases considered herein have mean equal to zero and are symmetric with respect to the mean. The two-sided test of whiteness for the hypotheses (D-31) is implemented in two parts. In the first part, the ACS estimate at each lag is compared with the two-sided threshold for the test of whiteness at α significance level, which is denoted as $\tau_r(\alpha)$. Threshold $\tau_r(\alpha)$ is a real-valued, positive scalar. The threshold comparison is of the form

$$(D-32) \quad \begin{array}{ccc} & \hat{r}(m) \neq r(m) & \\ & > & \\ |\hat{r}(m)| & & \tau_r(\alpha) \\ & < & \\ & \hat{r}(m) = r(m) & \end{array} \quad 1 \leq m \leq M_c$$

and is applied for $m=1,2,\dots,M_c$. Let $C_r(\alpha)$ denote the number of instances that the threshold is exceeded by all the elements of the sequence $\{\hat{r}(m) | m=1,2,\dots,M_c\}$; that is,

$$(D-33) \quad C_r(\alpha) = \sum_{m=1}^{M_c} \max\{0, \text{sgn}[|\hat{r}(m)| - \tau_r(\alpha)]\}$$

where the $\max\{\bullet, \bullet\}$ operator selects the maximum of its arguments, and the $\text{sgn}[\bullet]$ operator is defined herein as

$$(D-34) \quad \text{sgn}[a] = \begin{cases} 1 & a \geq 0 \\ -1 & a < 0 \end{cases}$$

Sometimes the $\text{sgn}[a]$ operator is defined to assume the value 0 when $a=0$. Then the second, and final, part of the whiteness test is implemented as

$$\begin{array}{c}
 \{\hat{r}(m)\} \neq \{r(m)\} \\
 (D-35) \quad C_r(\alpha) \quad \begin{array}{c} > \\ < \end{array} \quad \text{round}[\alpha M_c] \\
 \{\hat{r}(m)\} = \{r(m)\}
 \end{array}$$

where $\text{round}[\bullet]$ is the round-off operator applied to the non-negative scalar αM_c . Notice that M_c is a positive-valued integer, but α is bounded by zero and unity. Equation (D-35) states that the ACS estimate is non-white with significance level α if the number of times that the estimated ACS lags exceed the value of the two-sided threshold is larger than the expected number of threshold crossings, and is white otherwise. This part of the test involves an approximation error for parameter value combinations that are affected by the quantization inherent in the count $C_r(\alpha)$ as well as in the round-off operator. For example, if $M_c = 50$ and $\alpha = 0.1$, then $\text{round}[\alpha M_c] = \text{round}[5] = 5$, whereas if $M_c = 33$ and $\alpha = 0.1$, then $\text{round}[\alpha M_c] = \text{round}[3.3] = 3$. In the cases where quantization error is present, the test is more accurate in the ensemble sense.

The two-sided whiteness threshold is the upper integration limit which satisfies the following integral equation,

$$(D-36) \quad 1 - \alpha = 1 - P_{FA} = \mathcal{P} \left[|\hat{r}(m)| \leq \tau_r(\alpha) \right] = 2 \int_0^{\tau_r(\alpha)} p_{\hat{R}}(\hat{r}) d\hat{r}$$

where P_{FA} denotes the "probability of false alarm", $\mathcal{P}[\bullet]$ denotes the probability of event $[\bullet]$, variable $\hat{r}$ is distributed according to either the scaled SL or the Gaussian distribution, and $p_{\hat{R}}(\hat{r})$ is the appropriate distribution PDF (as defined in either Section D.3.1 or D.3.2). The scaled SL distribution (where SL stands for "sum of Laplace-distributed random variables") is the distribution of the real and imaginary components of the circular ACS lags (see Appendix F). Equation (D-36) follows from the fact that both

types of PDF considered herein have mean equal to zero and are symmetric with respect to the mean.

D.3.1 CIRCULAR ACS ESTIMATE FOR SMALL VALUES OF N

Consider the cases where the duration of the residual sequence is $N < 50$. As indicated in Appendix F, both the real component and the imaginary component of each circular lag (except lag 0) follow a scaled SL distribution. Furthermore, the parameters of the scaled SL PDF are identical for both components (real and imaginary) of all lags, except for lag $m = N/2$ when N is even-valued. The relevant PDFs for the two conditions that can arise are given next (Appendix F).

PDF for real or imaginary component of ACS lag $m = 1, 2, \dots, M_c$ for N odd, or ACS lag $m = 1, 2, \dots, M_c - 1$ for N even:

$$(D-37) \quad p_{\hat{R}}(\hat{r}) = \frac{N}{2^{2N-2} (N-1)! \sigma_{\epsilon}^2} \left[\sum_{k=0}^{N-1} \frac{(2N-2-k)! 2^{2k} N^k}{k! (N-1-k)! \sigma_{\epsilon}^{2k}} |\hat{r}|^k \right] \exp\left(-\frac{2N}{\sigma_{\epsilon}^2} |\hat{r}|\right)$$

$-\infty \leq \hat{r} \leq \infty$

PDF for real component of lag $m = M_c = N/2$ for N even:

$$(D-38) \quad p_{\hat{R}}(\hat{r}) = \frac{N}{2^N ((N/2)-1)! \sigma_{\epsilon}^2} \left[\sum_{k=0}^{\frac{N}{2}-1} \frac{(N-2-k)! N^k}{k! ((N/2)-1-k)! \sigma_{\epsilon}^{2k}} |\hat{r}|^k \right] \exp\left(-\frac{N}{2\sigma_{\epsilon}^2} |\hat{r}|\right)$$

$-\infty \leq \hat{r} \leq \infty$

Both PDFs are scaled SL PDFs, but with different parameters. However, in both cases the associated parameter a of Equation (F-49) has the same value; namely,

$$(D-39) \quad a = \frac{2}{\sigma_{\epsilon}^2}$$

$$a > 0$$

The fact that the PDF is of the same type for both conditions simplifies the statistical testing approach by requiring only one threshold (two thresholds when N is even-valued).

The threshold for the scaled SL-test is calculated numerically for a specified value of α using Equation (D-36) following the approach outlined in Section D.1 for the t-test threshold calculation. As for the t-test, the formulation based on integration from zero to the threshold value is preferred over the formulation where the integration limits are the threshold value and infinity.

D.3.2 CIRCULAR ACS ESTIMATE FOR LARGE VALUES OF N

Consider now the cases where the duration of the residual sequence is $N \geq 50$. As the value of N increases the distribution of both the real and the imaginary components of the circular ACS estimate approximates the Gaussian distribution for all lags. This follows from the central limit theorem. Furthermore, simulation-based analyses indicate that the scaled SL PDF approximates the Gaussian PDF very well (even in the tails) for $N \geq 50$. Thus, it is assumed herein that both the real component and the imaginary component of each circular lag (except lag 0) follow the Gaussian distribution for $N \geq 50$. And, given Equations (D-29) and (D-30), the parameters of the Gaussian PDF are identical for both components (real and imaginary) of all lags, except for lag $m = N/2$ when N is even-valued. The relevant Gaussian PDFs for the two conditions that can arise are given next.

PDF for real or imaginary component of ACS lag $m = 1, 2, \dots, M_c$ for N odd, or ACS lag $m = 1, 2, \dots, M_c-1$ for N even:

$$(D-40) \quad p_{\hat{R}}(\hat{r}) = \frac{\sqrt{N}}{\sqrt{\pi} \sigma_{\varepsilon}^2} \exp\left(-\frac{N}{\sigma_{\varepsilon}^4} \hat{r}^2\right) \quad -\infty \leq \hat{r} \leq \infty$$

PDF for real component of lag $m = M_c = N/2$ for N even:

$$(D-41) \quad p_{\hat{R}}(\hat{r}) = \frac{\sqrt{N}}{\sqrt{2\pi} \sigma_{\varepsilon}^2} \exp\left(-\frac{N}{2\sigma_{\varepsilon}^4} \hat{r}^2\right) \quad -\infty \leq \hat{r} \leq \infty$$

Both PDFs are zero-mean Gaussian, but with different variance. The fact that the PDF is of the same type for both conditions simplifies the statistical testing approach by requiring only one threshold (two thresholds when N is even-valued).

The threshold for the Gaussian-test is calculated numerically for a specified value of α using Equation (D-36) following the approach outlined in Section D.1 for the t-test threshold calculation. As for the t-test, the formulation based on integration from zero to the threshold value is preferred over the formulation where the integration limits are the threshold value and infinity. In particular for the Gaussian case, the error function `erf` and its inverse `erfinv` in MATLAB can be used to carry out the calculations, provided an appropriate transformation is applied first. Consider first the expression involved in the threshold calculation, Equation (D-36), for the condition represented by the form of the Gaussian PDF in Equation (D-40),

$$(D-42) \quad 1 - \alpha = 2 \int_0^{\tau_r(\alpha)} p_{\hat{R}}(\hat{r}) d\hat{r} = \frac{2\sqrt{N}}{\sqrt{\pi} \sigma_{\varepsilon}^2} \int_0^{\tau_r(\alpha)} \exp\left[-\frac{N}{\sigma_{\varepsilon}^4} \hat{r}^2\right] d\hat{r}$$

and now define a dummy variable t as

$$(D-43) \quad t = \frac{\sqrt{N}}{\sigma_\varepsilon} \hat{r}$$

Substitution of this transformation into Equation (D-42) results in

$$(D-44a) \quad 1 - \alpha = \frac{2}{\sqrt{\pi}} \int_0^\tau \exp[-t^2] dt$$

$$(D-44b) \quad \tau = \frac{\sqrt{N}}{\sigma_\varepsilon} \tau_r(\alpha)$$

where the upper limit of integration, τ , is also a dummy variable. MATLAB's error function `erf` is defined as

$$(D-45) \quad \text{erf}[\tau] = \frac{2}{\sqrt{\pi}} \int_0^\tau \exp[-t^2] dt$$

Comparison of Equations (D-44) and (D-45) leads to the following relation between α and $\tau_r(\alpha)$,

$$(D-46) \quad \text{erf}\left[\frac{\sqrt{N}}{\sigma_\varepsilon} \tau_r(\alpha)\right] = 1 - \alpha$$

Equation (D-46) can be solved for $\tau_r(\alpha)$ as a function of α using the inverse error function `erfinv` in MATLAB. Doing so results in

$$(D-47) \quad \tau_r(\alpha) = \sqrt{2} \sigma_\varepsilon \text{erfinv}[1 - \alpha] = \frac{\sigma_\varepsilon^2}{\sqrt{N}} \text{erfinv}[1 - \alpha]$$

This is the desired result for the first condition, with PDF as in Equation (D-40).

Consider now the expression involved in the threshold calculation, Equation (D-36), for the condition represented by the form of the Gaussian PDF in Equation (D-41),

$$(D-48) \quad 1-\alpha = 2 \int_0^{\tau_r(\alpha)} p_{\hat{r}}(\hat{r}) d\hat{r} = \frac{\sqrt{2N}}{\sqrt{\pi} \sigma_{\epsilon}^2} \int_0^{\tau_r(\alpha)} \exp\left[-\frac{N}{2\sigma_{\epsilon}^4} \hat{r}^2\right] d\hat{r}$$

Following the same steps as for Equation (D-42) leads to the corresponding threshold for the second condition,

$$(D-49) \quad \tau_r(\alpha) = \sqrt{2} \sigma_{\hat{r}} \operatorname{erfinv}[1-\alpha] = \frac{\sqrt{2} \sigma_{\epsilon}^2}{\sqrt{N}} \operatorname{erfinv}[1-\alpha]$$

As a specific example, for a 5% significance level ($\alpha = 0.05$) and unity standard deviation ($\sigma_{\hat{r}} = 1$), $\tau_r(\alpha) = \sqrt{2} \operatorname{erfinv}[0.95] = 1.96$.

APPENDIX E. AUTO-CORRELATION SEQUENCE ESTIMATORS

Accurate estimates of the ACS for a random sequence are required hypothesis filter design (Appendix D), identification of the state space model parameters using the canonical correlations method (Section 7.3), and detection decision criteria for some applications (Section 7.3). The most common ACS estimators are of the time-average type, since they can be implemented when a single realization is available. Such a procedure implies that the random process under consideration is ergodic.

The three most common time-average ACS estimators are summarized herein; namely, the circular, biased, and unbiased. Each estimator has distinct statistical features when utilized in specific contexts. In fact, the unbiased ACS estimator leads to undesirable results in the three contexts mentioned above, and is presented herein for completeness only.

In this report, ACS estimators are required for the channel output vector sequence as well as for the residual vector sequence and/or its scalar components. However, the estimators are presented herein for the channel output vector sequence only since all other cases are handled with a simple change of notation. Thus, consider a finite-duration realization of the complex-valued channel output vector process, $\{\mathbf{x}(n) | n=0, 1, \dots, N-1\}$.

E.1 Circular ACS Estimator

For this estimator it is convenient to define an infinite-duration sequence $\{\bar{\mathbf{x}}(n) = \bar{\mathbf{x}}_r(n) + j\bar{\mathbf{x}}_i(n)\}$ as a periodic extension of the finite-duration sequence $\{\mathbf{x}(n) | n=0, 1, \dots, N-1\}$. Specifically, the desired periodic extension is of the form

$$(E-1) \quad \bar{x}(n) = \begin{cases} x(n) & 0 \leq n \leq N-1 \\ x(n \bmod [N]) & \text{elsewhere} \end{cases}$$

In Equation (F-1) the expression $n \bmod [N]$ represents "n modulo N," and is evaluated as follows. Let the integer n be represented as

$$(E-2) \quad n = n_1 + n_2 N$$

with n_1 and n_2 integers selected such that

$$(E-3) \quad 0 \leq n_1 \leq N-1$$

$$(E-4) \quad -\infty < n_2 < \infty$$

Then,

$$(E-5) \quad n \bmod [N] = n_1$$

Notice that an integer pair (n_1, n_2) always exist for any specified integer n .

Given the infinite-duration sequence $\{\bar{x}(n)\}$, the principal (unique-valued) lags of the circular ACS estimate of $\{x(n) | n = 0, 1, \dots, N-1\}$ are defined as

$$(E-6a) \quad \hat{R}_{xx}(m) = \frac{1}{N} \sum_{n=0}^{N-1} \bar{x}(n) \bar{x}^H(n-m) \quad m = 0, 1, \dots, M_c$$

$$(E-6b) \quad \hat{R}_{xx}(m) = \frac{1}{N} \sum_{n=0}^{N-1} \left(\left[\bar{x}_r(n) \bar{x}_r^T(n-m) + \bar{x}_i(n) \bar{x}_i^T(n-m) \right] \right. \\ \left. + j \left[\bar{x}_i(n) \bar{x}_r^T(n-m) - \bar{x}_r(n) \bar{x}_i^T(n-m) \right] \right) \quad m = 0, 1, \dots, M_c$$

$$(E-6c) \quad \hat{R}_{xx}(m) = \hat{R}_{xx}^H(N-m) \quad m = M_c+1, \dots, N-1$$

$$(E-7) \quad M_c = \begin{cases} \frac{N}{2} & N \text{ even} \\ \frac{N-1}{2} & N \text{ odd} \end{cases}$$

For an N -point data sequence, the largest lag that can be generated without repetition is $N-1$. Circular ACS lags M_c+1 through $N-1$ are determined from lags 0 through M_c via the conjugate symmetry relations in Equation (E-6c). Estimates for negative-valued lags are determined using the well-known conjugate relation,

$$(E-8) \quad \hat{R}_{xx}(-m) = \hat{R}_{xx}^H(m) \quad \forall m$$

In general, the principal lags of the circular ACS estimate of a complex-valued data sequence are complex-valued, except for $m=0$. A further exception occurs for lag $m=N/2$ in the cases where the number of data points is even-valued. In such cases lag $m=N/2$ of the ACS is real-valued also and can be determined as

$$(E-9a) \quad \hat{R}_{xx}(N/2) = \frac{2}{N} \Re \left[\sum_{n=0}^{\frac{N}{2}-1} \underline{x}(n) \underline{x}^H(n+N/2) \right]$$

$$(E-9b) \quad \hat{R}_{xx}(N/2) = \frac{2}{N} \sum_{n=0}^{\frac{N}{2}-1} \left[\underline{x}_r(n) \underline{x}_r^H(n+(N/2)) + \underline{x}_i(n) \underline{x}_i^H(n+(N/2)) \right]$$

In Equation (E-9) notice that $\underline{x}(\bullet)$ appears in the summation, instead of $\bar{\underline{x}}(\bullet)$; this is a result of the simplification made to Equation (E-6) for the special case $m = N/2$ with N even. Equations (E-6) and (E-9) are equivalent.

An alternative expression for the circular ACS estimator can be derived by recognizing that $\bar{\underline{x}}(n-m)$ in Equation (E-6a) is equal to $\underline{x}(n-m+N)$ for lags $m = 1, 2, \dots, N-1$ and $n < m$; that is,

$$(E-10) \quad \bar{\underline{x}}(n-m) = \underline{x}((n-m) \bmod N) = \underline{x}(n-m+N) \quad m = 1, 2, \dots, N-1; \quad n < m$$

Substitution of this equality into Equation (E-6a) results in

$$(E-11) \quad \hat{R}_{xx}(m) = \frac{1}{N} \left[\sum_{n=0}^{m-1} \underline{x}(n) \underline{x}^H(n-m+N) + \sum_{n=m}^{N-1} \underline{x}(n) \underline{x}^H(n-m) \right] \quad 1 \leq m \leq N-1$$

with lag $m=0$ determined as

$$(E-12) \quad \hat{R}_{xx}(0) = \frac{1}{N} \sum_{n=0}^{N-1} \underline{x}(n) \underline{x}^H(n)$$

Equations (E-11) and (E-12) are convenient for determination of the mean and variance of the circular ACS estimator. And Equation (E-9) is obtained also from Equation (E-11).

Using Equation (E-11) the mean matrix of the circular ACS estimator at lags $m \neq 0$ is obtained as

$$(E-13a) \quad M_C(m) = E[\hat{R}_{xx}(m)] = \frac{1}{N} \left[(N-m) R_{xx}(m) + m R_{xx}^H(N-m) \right] \quad 1 \leq m \leq N-1$$

$$(E-13b) \quad M_C(m) = \frac{N-m}{N} R_{xx}(m) + \frac{m}{N} R_{xx}^H(N-m) \quad 1 \leq m \leq N-1$$

and the mean matrix at lag $m=0$ is obtained using Equation (E-12),

$$(E-14) \quad M_C(0) = R_{xx}(0)$$

The bias error matrix at lags $m \neq 0$, denoted as $B_C(m)$, is defined as

$$(E-15) \quad B_C(m) = R_{xx}(m) - M_C(m) = \left(\frac{m}{N}\right) [R_{xx}(m) - R_{xx}^H(N-m)] \quad 1 \leq m \leq N-1$$

and for lag $m=0$,

$$(E-16) \quad B_C(0) = R_{xx}(0) - M_C(0) = [0]$$

In general, the bias error is non-zero at lags $1 \leq m \leq N-1$ for the circular ACS estimator. Notice also that the bias error at each lag is a function of the lag index as well as the lag value at two lags. The variance matrix of the circular ACS estimator at lags $m = 0, 1, \dots, N-1$ is defined as

$$(E-17a) \quad \Sigma_C^2(m) = E \left[\left\{ \hat{R}_{xx}(m) - M_C(m) \right\} \left\{ \hat{R}_{xx}(m) - M_C(m) \right\}^H \right] \quad 0 \leq m \leq N-1$$

$$(E-17b) \quad \Sigma_C^2(m) = E \left[\hat{R}_{xx}(m) \hat{R}_{xx}^H(m) \right] - M_C(m) M_C^H(m) \quad 0 \leq m \leq N-1$$

Another important estimator performance measure is the mean-square error matrix, which is defined as

$$(E-18a) \quad S_C^2(m) = E \left[\left\{ R_{xx}(m) - \hat{R}_{xx}(m) \right\} \left\{ R_{xx}(m) - \hat{R}_{xx}(m) \right\}^H \right] \quad 0 \leq m \leq N-1$$

$$(E-18b) \quad S_C^2(m) = R_{xx}(m) R_{xx}^H(m) - M_C(m) R_{xx}^H(m) - R_{xx}(m) M_C^H(m) + E \left[\hat{R}_{xx}(m) \hat{R}_{xx}^H(m) \right] \quad 0 \leq m \leq N-1$$

$$(E-18c) \quad S_C^2(m) = \Sigma_C^2(m) + B_C(m)B_C^H(m) \quad 0 \leq m \leq N-1$$

In the general case, the variance matrix and the mean-square error matrix are complicated expressions. However, for the specific case of a white scalar sequence, which is the case of interest herein, the mean, variance, bias error, and mean-square error attain a simple form, as summarized next.

White Scalar Sequence Only:

$$(E-19a) \quad \mu_C(0) = r_{xx}(0) = \sigma_x^2$$

$$(E-19b) \quad \mu_C(m) = 0 \quad 1 \leq m \leq N-1$$

$$(E-20) \quad \sigma_C^2(m) = \frac{r_{xx}^2(0)}{N} = \frac{\sigma_x^4}{N} \quad 0 \leq m \leq N-1$$

$$(E-21) \quad b_C(m) = 0 \quad 0 \leq m \leq N-1$$

$$(E-22) \quad s_C^2(m) = \sigma_C^2(m) + b_C^2(m) = \sigma_C^2(m) = \frac{r_{xx}^2(0)}{N} = \frac{\sigma_x^4}{N} \quad 0 \leq m \leq N-1$$

Thus, for an uncorrelated sequence the bias error is zero at all lags, and the mean-square error is the same at all lags.

The circular ACS estimator is ideally-suited for statistical tests of whiteness for finite-duration sequences, as described in Appendix D. However, for contexts where the ACS of a colored (non-white) process is required, such as model identification, the circular estimator can introduce significant bias and mean-square errors to the ACS. Only for the case of an uncorrelated sequence does the extended sequence exhibit the same statistical properties as the original finite-duration sequence.

E.2 Biased ACS Estimator

The biased estimator for the m th ACS lag of the finite-duration sequence $\{\underline{x}(n) | n=0, 1, \dots, N-1\}$ is defined as

$$(E-23a) \quad \hat{R}_{xx}(m) = \frac{1}{N} \sum_{n=m}^{N-1} \underline{x}(n) \underline{x}^H(n-m) \quad m = 0, 1, \dots, N-1$$

$$(E-23b) \quad \hat{R}_{xx}(m) = \frac{1}{N} \sum_{n=m}^{N-1} \left(\left[\underline{x}_r(n) \underline{x}_r^T(n-m) + \underline{x}_i(n) \underline{x}_i^T(n-m) \right] \right. \\ \left. + j \left[\underline{x}_i(n) \underline{x}_r^T(n-m) - \underline{x}_r(n) \underline{x}_i^T(n-m) \right] \right) \quad m = 0, 1, \dots, N-1$$

For an N -point data sequence, the largest lag that can be generated without repetition is $N-1$. Estimates for negative-valued lags are determined using the well-known conjugate relation,

$$(E-24) \quad \hat{R}_{xx}(-m) = \hat{R}_{xx}^H(m) \quad \forall m$$

The m th lag in Equation (E-23) has $N-m$ terms in the summation, but the normalizing factor is N for all lags. This drives the envelope of the biased ACS estimate to exhibit monotonically-decreasing behaviour as m increases. Such a feature is desirable because the envelope of the ACS of the output of a stationary system (with system matrix F stable) is monotonically decreasing. However, this feature is the reason for the "biased" qualification attached to this estimator (see Equation (E-26) below).

The mean matrix of the biased ACS estimator is determined using Equation (E-23), which leads to

$$(E-25) \quad M_B(m) = E[\hat{R}_{xx}(m)] = \frac{N-m}{N} R_{xx}(m) = \left(1 - \frac{m}{N} \right) R_{xx}(m) \quad 0 \leq m \leq N-1$$

and the bias error matrix at all principal lags is obtained as

$$(E-26) \quad B_B(m) = R_{xx}(m) - M_B(m) = \left(\frac{m}{N} \right) R_{xx}(m) \quad 0 \leq m \leq N-1$$

Notice that this estimator is unbiased only for lag $m=0$, just like the circular ACS estimator.

The variance matrix and the mean-square error matrix of the biased ACS estimator are defined as in Section E.1. As for the circular ACS estimator, in the general case the variance matrix and the mean-square error matrix are complicated expressions. However, for the specific case of a white scalar sequence, which is the case of interest herein, the mean, variance, bias error, and mean-square error attain the simple form given next.

White Scalar Sequence Only:

$$(E-27a) \quad \mu_B(0) = r_{xx}(0) = \sigma_x^2$$

$$(E-27b) \quad \mu_B(m) = 0 \quad 1 \leq m \leq N-1$$

$$(E-28) \quad \sigma_B^2(m) = \left(\frac{N-m}{N} \right) \frac{r_{xx}^2(0)}{N} = \left(\frac{N-m}{N} \right) \frac{\sigma_x^4}{N} = \left(1 - \frac{m}{N} \right) \frac{\sigma_x^4}{N} \quad 0 \leq m \leq N-1$$

$$(E-29) \quad b_B(m) = 0 \quad 0 \leq m \leq N-1$$

$$(E-30) \quad s_B^2(m) = \sigma_B^2(m) + b_B^2(m) = \sigma_B^2(m) = \left(1 - \frac{m}{N} \right) \frac{\sigma_x^4}{N} \quad 0 \leq m \leq N-1$$

Thus, for an uncorrelated sequence the bias error is zero at all lags, and the mean-square error decreases as the lag index, m , increases.

Utilization of the biased ACS estimator in model parameter estimation, adaptive filter design, spectrum estimation, and other such contexts leads to satisfactory algorithm performance in cases where the circular and the unbiased ACS estimators fail. This includes both the radar surveillance and the ECG diagnostics applications discussed in this report (Sections 6.0 and 7.0).

E.3 Unbiased ACS Estimator

The unbiased estimator for the m th ACS lag of the finite-duration sequence $\{\underline{x}(n) | n=0, 1, \dots, N-1\}$ is defined as

$$(E-31a) \quad \hat{R}_{xx}(m) = \frac{1}{N-m} \sum_{n=m}^{N-1} \underline{x}(n) \underline{x}^H(n-m) \quad m = 0, 1, \dots, N-1$$

$$(E-31b) \quad \hat{R}_{xx}(m) = \frac{1}{N-m} \sum_{n=m}^{N-1} \left(\left[\underline{x}_r(n) \underline{x}_r^T(n-m) + \underline{x}_i(n) \underline{x}_i^T(n-m) \right] + j \left[\underline{x}_i(n) \underline{x}_r^T(n-m) - \underline{x}_r(n) \underline{x}_i^T(n-m) \right] \right) \quad m = 0, 1, \dots, N-1$$

As before, the largest lag that can be generated without repetition is $N-1$. Also, estimates for negative-valued lags are determined using the well-known conjugate relation,

$$(E-32) \quad \hat{R}_{xx}(-m) = \hat{R}_{xx}^H(m) \quad \forall m$$

The m th lag in Equation (E-31) has $N-m$ terms in the summation, and the normalizing factor is $N-m$ for all lags. This forces the error bias to be equal to zero for all lags (see Equation (E-34) below), thus justifying the "unbiased" qualification to this estimator. However, the behaviour of the estimated lags at the larger index values (m close to $N-1$) has a large variance. Due to this large variance the unbiased ACS estimator is useless in most contexts.

The mean matrix of the unbiased ACS estimator is determined using Equation (E-31), which leads to

$$(E-33) \quad M_U(m) = E[\hat{R}_{xx}(m)] = R_{xx}(m) \quad 0 \leq m \leq N-1$$

and the bias error matrix at all principal lags is obtained as

$$(E-34) \quad B_U(m) = R_{xx}(m) - M_U(m) = [0] \quad 0 \leq m \leq N-1$$

As expected, this estimator has zero bias error at all lags $m=0$.

The variance matrix and the mean-square error matrix of the unbiased ACS estimator are defined as in Section E.1. As for the previous ACS estimators, in the general case the variance matrix and the mean-square error matrix are complicated expressions. However, for the specific case of a white scalar sequence, the mean, variance, bias error, and mean-square error attain the simple form given next.

White Scalar Sequence Only:

$$(E-35a) \quad \mu_U(0) = r_{xx}(0) = \sigma_x^2$$

$$(E-35b) \quad \mu_U(m) = 0 \quad 1 \leq m \leq N-1$$

$$(E-36) \quad \sigma_U^2(m) = (m+1) \frac{r_{xx}^2(0)}{N} = (m+1) \frac{\sigma_x^4}{N} \quad 0 \leq m \leq N-1$$

$$(E-37) \quad b_U(m) = 0 \quad 0 \leq m \leq N-1$$

$$(E-38) \quad s_U^2(m) = \sigma_U^2(m) + b_U^2(m) = \sigma_U^2(m) = (m-1) \frac{\sigma_x^4}{N} \quad 0 \leq m \leq N-1$$

Thus, for an uncorrelated sequence the bias error is zero at all lags, and the mean-square error increases as the lag index, m , increases.

The scalar uncorrelated case exemplifies the large mean-square error that makes the unbiased ACS estimator useless in most contexts. In fact, it is included in this appendix only for completeness, since the unbiased ACS estimator has performed unacceptably in the two applications of interest in this program, radar array surveillance and ECG diagnostics.

APPENDIX F. RANDOM VARIABLE TRANSFORMATIONS

In general, the m th lag (for $m=1,2,\dots,M_c$) of the circular estimator of the ACS of a finite-length, complex-valued, circular (with independent, identically-distributed real and imaginary components), Gaussian-distributed, scalar, zero-mean, white sequence $\{\varepsilon(n) | n=0,1,\dots,N-1\}$ is a complex-valued random variable (such a scalar sequence represents one element of the residual vector sequence). More specifically, the circular estimate of the m th lag is a sum of N terms of the form (for simplicity, the scale factor $1/N$ is omitted)

$$(F-1) \quad \varepsilon(n) \varepsilon^*(n-m) = [\varepsilon_r(n) \varepsilon_r(n-m) + \varepsilon_i(n) \varepsilon_i(n-m)] + j [\varepsilon_i(n) \varepsilon_r(n-m) - \varepsilon_r(n) \varepsilon_i(n-m)]$$

Thus, the real (imaginary) component of each lag is a random variable which is the sum of N terms, and each term is sum (difference) of the product of two independent, identically-distributed, zero-mean, Gaussian random variables. A special condition is true for lag $m = N/2$ when N is even-valued (see Appendix E). In such a case lag $m = N/2$ is always real-valued, independent of the given data, and is determined as the sum of $N/2$ terms of the form (for simplicity, a scale factor $2/N$ is omitted)

$$(F-2) \quad \Re[\varepsilon(n) \varepsilon^*(n+N/2)] = [\varepsilon_r(n) \varepsilon_r(n+N/2) + \varepsilon_i(n) \varepsilon_i(n+N/2)]$$

From Equation (F-2), the real component of lag $N/2$ is a random variable which is the sum of $N/2$ terms, where each term is sum of the product of two independent, identically-distributed, zero-mean, Gaussian random variables.

In this appendix the PDF of the m th lag of the circular ACS estimator for a scalar data sequence is derived as the PDF of the random variable resulting from the transformations on Gaussian

random variables outlined above. The approach presented herein is based on recent analyses by Rangaswamy and Michels (1996).

A different notation is adopted in this appendix for simplicity and generality. Let u_i , for $i=1,2$, denote two real-valued, independent, random variables, both distributed as $\mathcal{N}(0, \sigma_u^2)$. These random variable are transformed by a series of operations to obtain the desired PDF results.

F.1 Product of Two Independent, Gaussian-Distributed Random Variables

Define a real-valued random variable z as the product of the two Gaussian-distributed variables u_1 and u_2 ,

$$(F-3) \quad z = u_1 u_2 \quad -\infty \leq z \leq \infty$$

The mean and variance of z are

$$(F-4) \quad \mu_z = E[z] = E[u_1 u_2] = E[u_1] E[u_2] = 0$$

$$(F-5) \quad \sigma_z^2 = E[(z - \mu_z)^2] = E[z^2] = E[u_1^2 u_2^2] = E[u_1^2] E[u_2^2] = \sigma_u^4$$

The PDF of z is obtained next using the transformation of variables method (see, for example, Beckmann [1967]). In accordance with this method, define an auxiliary random variable u as

$$(F-6) \quad u = u_2 \quad -\infty \leq u \leq \infty$$

Variables z and u have a joint two-dimensional PDF, denoted as $p_{zu}(z, u)$, and the PDF of z is obtained as a marginal PDF, by integrating $p_{zu}(z, u)$ over the allowable range of values for u . That, is,

$$(F-7) \quad p_Z(z) = \int_{-\infty}^{\infty} p_{ZU}(z,u) du \quad -\infty \leq z \leq \infty$$

In turn, the joint PDF $p_{ZU}(z,u)$ is of the form

$$(F-8) \quad p_{ZU}(z,u) = p_{U_1U_2}(u_1(z,u), u_2(z,u)) \left| \frac{\partial(u_1, u_2)}{\partial(z, u)} \right| \quad -\infty \leq z \leq \infty; \quad -\infty \leq u \leq \infty$$

where the parallel vertical bars (|) are used to denote the absolute value of the parameter inside the bars. From Equations (F-3) and (F-6),

$$(F-9) \quad u_1(z,u) = u_1 = \frac{z}{u_2} = \frac{z}{u}$$

$$(F-10) \quad u_2(z,u) = u$$

and the Jacobian that appears in Equation (F-8) is determined as

$$(F-11) \quad \frac{\partial(u_1, u_2)}{\partial(z, u)} = \begin{vmatrix} \frac{\partial u_1}{\partial z} & \frac{\partial u_1}{\partial u} \\ \frac{\partial u_2}{\partial z} & \frac{\partial u_2}{\partial u} \end{vmatrix} = \begin{vmatrix} \frac{1}{u} & -\frac{z}{u^2} \\ 0 & 1 \end{vmatrix} = \frac{1}{u}$$

where the parallel vertical bars (|) denote the determinant of the matrix enclosed within the bars (parallel vertical bars are standard notation for the determinant of a matrix as well as for absolute value). Since u_1 and u_2 are Gaussian-distributed and independent, their two-dimensional PDF is

$$(F-12) \quad p_{U_1U_2}(u_1, u_2) = \frac{1}{2\pi\sigma_u^2} \exp \left[-\frac{u_1^2 + u_2^2}{2\sigma_u^2} \right] \quad -\infty \leq u_1 \leq \infty; \quad -\infty \leq u_2 \leq \infty$$

It follows from Equation (F-7) and Equations (F-8) through (F-12) that

$$(F-13a) \quad p_z(z) = \frac{1}{2\pi\sigma_u^2} \int_{-\infty}^{\infty} \exp\left[-\frac{z^2}{2\sigma_u^2 u^2}\right] \exp\left[-\frac{u^2}{2\sigma_u^2}\right] \frac{1}{|u|} du \quad -\infty \leq z \leq \infty$$

$$(F-13b) \quad p_z(z) = \frac{2}{2\pi\sigma_u^2} \int_0^{\infty} \exp\left[-\frac{z^2}{2\sigma_u^2 u^2}\right] \exp\left[-\frac{u^2}{2\sigma_u^2}\right] \frac{1}{u} du \quad -\infty \leq z \leq \infty$$

Equation (F-13b) follows from (F-13a) because the integrand in (F-13a) is an even function of u . In order to evaluate the integral in Equation (F-13b) it is convenient to introduce a transformation on the integration variable, u . Let c denote a dummy variable defined as

$$(F-14) \quad c = u^2$$

It follows from Equation (F-12) that

$$(F-15) \quad \frac{1}{u} du = \frac{1}{2c} dc$$

Substitution of these equivalences into Equation (F-13b) leads to

$$(F-16) \quad p_z(z) = \frac{1}{2\pi\sigma_u^2} \int_0^{\infty} \exp\left[-\frac{z^2}{2\sigma_u^2 c}\right] \exp\left[-\frac{c}{2\sigma_u^2}\right] \frac{1}{c} dc \quad -\infty \leq z \leq \infty$$

This final expression is evaluated by referring to integral no. 9 on page 340 of (Gradshteyn and Ryzhik, 1980), which results in

$$(F-17) \quad p_Z(z) = \frac{1}{\pi\sigma_u^2} K_0 \left[\frac{z}{\sigma_u^2} \right] = \frac{1}{\pi\sigma_z} K_0 \left[\frac{z}{\sigma_z} \right] \quad -\infty \leq z \leq \infty$$

where $K_0[\bullet]$ denotes the modified Bessel function of the second kind of order zero, and σ_z is the standard deviation of z , as defined in Equation (F-5). The PDF for the product of two identically-distributed Gaussian random variables given in Equation (F-17) is similar to the PDF of the K distribution. Thus, it is referred to herein as the special K distribution.

The characteristic function of z , denoted herein as $\phi_z(\omega)$, is defined as

$$(F-18a) \quad \phi_z(\omega) = E[e^{j\omega z}] = \int_{-\infty}^{\infty} e^{j\omega z} p_Z(z) dz \quad -\infty \leq \omega \leq \infty$$

$$(F-18b) \quad \phi_z(\omega) = \int_{-\infty}^{\infty} [\cos(\omega z) + j \sin(\omega z)] p_Z(z) dz \quad -\infty \leq \omega \leq \infty$$

$$(F-18c) \quad \phi_z(\omega) = \int_{-\infty}^{\infty} \cos(\omega z) p_Z(z) dz + j \int_{-\infty}^{\infty} \sin(\omega z) p_Z(z) dz \quad -\infty \leq \omega \leq \infty$$

$$(F-18d) \quad \phi_z(\omega) = \int_{-\infty}^{\infty} \cos(\omega z) p_Z(z) dz = 2 \int_0^{\infty} \cos(\omega z) p_Z(z) dz \quad -\infty \leq \omega \leq \infty$$

$$(F-18e) \quad \phi_z(\omega) = \frac{4}{\pi\sigma_z} \int_0^{\infty} \cos(\omega z) K_0 \left[\frac{z}{\sigma_z} \right] dz \quad -\infty \leq \omega \leq \infty$$

As indicated in Equation (F-18d), $\phi_z(\omega)$ is real-valued because the integrand of the integral in the imaginary part is an odd function of z , which integrates to zero over the real line. In contrast, the integrand of the integral in the real part is an even function of z , which implies that the integral of the real part can be evaluated as twice the integral over the positive real line. The integral in Equation (F-18e) is evaluated by referring to integral no. 6 on page 731 of (Gradshteyn and Ryzhik, 1980), which results in

$$(F-19) \quad \phi_z(\omega) = \frac{\frac{1}{\sigma_z}}{\left(\omega^2 + \frac{1}{\sigma_z^2}\right)^{\frac{1}{2}}} = \frac{1}{\sigma_z} \left(\omega^2 + \frac{1}{\sigma_z^2}\right)^{-\frac{1}{2}} \quad -\infty \leq \omega \leq \infty$$

This result is built-upon several times in the remainder of this appendix. Notice that $\phi_z(\omega)$ is an even function of ω , which implies that $\phi_z(\omega)$ is symmetric with respect to the origin of the ω axis.

F.2 Sum of Two Independent, Special K-Distributed Random Variables

Consider now two independent, zero-mean, special K-distributed random variables z_1 and z_2 with identical distributions, and define a real-valued random variable v as

$$(F-20) \quad v = z_1 + z_2 \quad -\infty \leq v \leq \infty$$

The mean and variance of v are

$$(F-21) \quad \mu_v = E[v] = E[z_1 + z_2] = E[z_1] + E[z_2] = 0$$

$$(F-22) \quad \sigma_v^2 = E[(v - \mu_v)^2] = E[v^2] = E[z_1^2 + 2z_1z_2 + z_2^2] = E[z_1^2] + E[z_2^2] = 2\sigma_z^2 = 2\sigma_u^4$$

The PDF of v is obtained next using the characteristic function of z_1 and z_2 , as determined above. For independent random variables the characteristic function of their sum is equal to the product of the two individual characteristic functions; that is,

$$(F-23) \quad \phi_v(\omega) = \phi_{z_1}(\omega) \phi_{z_2}(\omega) = [\phi_z(\omega)]^2 \quad -\infty \leq \omega \leq \infty$$

It follows that

$$(F-24) \quad \phi_v(\omega) = \frac{\frac{1}{\sigma_z^2}}{\omega^2 + \frac{1}{\sigma_z^2}} = \frac{1}{\sigma_z^2} \left(\omega^2 + \frac{1}{\sigma_z^2} \right)^{-1} = \frac{2}{\sigma_v^2} \left(\omega^2 + \frac{2}{\sigma_v^2} \right)^{-1} \quad -\infty \leq \omega \leq \infty$$

$$(F-24b) \quad \phi_v(\omega) = \frac{a^2}{\omega^2 + a^2} \quad -\infty \leq \omega \leq \infty$$

$$(F-24c) \quad a^2 = \frac{2}{\sigma_v^2} = \frac{1}{\sigma_z^2} = \frac{1}{\sigma_u^4}$$

with $a > 0$. Parameter a is introduced herein for notational simplicity. The PDF of v is the Fourier transform of the characteristic function scaled by the factor $(2\pi)^{-1}$; that is,

$$(F-25) \quad p_v(v) = \frac{1}{2\pi} \mathcal{F}[\phi_v(\omega)] = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-j\omega v} \phi_v(\omega) d\omega \quad -\infty \leq v \leq \infty$$

This integral can be evaluated to obtain the PDF of v . However, it turns out that $\phi_v(\omega)$ is the characteristic function of the zero-mean Laplace (or two-sided exponential) distribution, which is of the form (Cooper and McGillem, 1971)

$$(F-26) \quad p_v(v) = \frac{\sqrt{2}}{2\sigma_v} \exp\left(-\frac{\sqrt{2}}{\sigma_v} |v|\right) = \frac{a}{2} e^{-a|v|} \quad -\infty \leq v \leq \infty$$

where the parallel bars ($|\cdot|$) denote the absolute value. This is the PDF of the real part of the product $\varepsilon(n)\varepsilon^*(n-m)$ presented in Equation (F-1). And this is also the PDF of the (unscaled) real component of lag $m=N/2$ of the circular ACS estimator when N is even, which is presented in Equation (F-2).

F.3 Difference of Two Independent, Special K-Distributed Random Variables

As in the preceding section, consider two independent, special K-distributed random variables z_1 and z_2 with identical distributions, and define a real-valued random variable s as

$$(F-27) \quad s = z_1 - z_2 \quad -\infty \leq s \leq \infty$$

The mean and variance of s are equal to those of v ,

$$(F-28) \quad \mu_s = E[s] = E[z_1 - z_2] = E[z_1] - E[z_2] = 0$$

$$(F-29) \quad \sigma_s^2 = E[(s - \mu_v)^2] = E[s^2] = E[z_1^2 - 2z_1z_2 + z_2^2] = E[z_1^2] + E[z_2^2] = 2\sigma_z^2 = 2\sigma_u^4$$

And the PDF of s is obtained next using the characteristic function of z_1 and z_2 , as before. For two independent random variables, the characteristic function of their difference is determined as

$$(F-30) \quad \phi_s(\omega) = E[e^{j\omega s}] = E[e^{j\omega(z_1 - z_2)}] = E[e^{j\omega z_1}] E[e^{-j\omega z_2}] \quad -\infty \leq \omega \leq \infty$$

Consider the two factors on the right-most equivalence. Each of these factors is of the form

$$(F-31) \quad E[e^{j\omega z_1}] = \phi_{z_1}(\omega) = \phi_z(\omega)$$

$$(F-32) \quad E[e^{j\omega z_2}] = \{E[e^{j\omega z_2}]\}^* = \phi_{z_2}^*(\omega) = \phi_z^*(\omega) = \phi_z(\omega)$$

where the last equality is due to the fact that $\phi_z(\omega)$ is a real-valued, symmetric function. It follows that

$$(F-33) \quad \phi_s(\omega) = \phi_z(\omega) \phi_z(\omega) = [\phi_z(\omega)]^2 \quad -\infty \leq \omega \leq \infty$$

This result is identical to the characteristic function for the sum of two independent, special K-distributed random variables, $\phi_v(\omega)$ in Equation (F-23). Thus, the characteristic function of s is

$$(F-34a) \quad \phi_s(\omega) = \frac{\frac{1}{\sigma_z^2}}{\omega^2 + \frac{1}{\sigma_z^2}} = \frac{1}{\sigma_z^2} \left(\omega^2 + \frac{1}{\sigma_z^2} \right)^{-1} = \frac{2}{\sigma_s^2} \left(\omega^2 + \frac{2}{\sigma_s^2} \right)^{-1} \quad -\infty \leq \omega \leq \infty$$

$$(F-34b) \quad \phi_s(\omega) = \frac{a^2}{\omega^2 + a^2} \quad -\infty \leq \omega \leq \infty$$

$$(F-34c) \quad a^2 = \frac{2}{\sigma_s^2} = \frac{1}{\sigma_z^2} = \frac{1}{\sigma_u^4}$$

and the PDF of s is given as

$$(F-35) \quad p_s(s) = \frac{\sqrt{2}}{2\sigma_s} \exp\left(-\frac{\sqrt{2}}{\sigma_s} |s|\right) = \frac{a}{2} e^{-a|s|} \quad -\infty \leq s \leq \infty$$

As before, the scale parameter a is positive-valued, $a > 0$. Notice that v and s are both Laplace-distributed with identical distributions (both have mean zero, and $\sigma_s = \sigma_v$).

Equation (F-35) is the PDF of the imaginary part of the product $\varepsilon(n)\varepsilon^*(n-m)$ in Equation (F-1). Thus, the real and imaginary part of each term of the form $\varepsilon(n)\varepsilon^*(n-m)$ are both Laplace-distributed, with identical distribution parameters. This is an important result because it allows identical treatment for the real and imaginary components of the estimated ACS lags.

F.4 Sum of N Independent, Laplace-Distributed Random Variables

Consider a set of N independent, zero-mean, Laplace-distributed random variables $\{v_1, v_2, \dots, v_N\}$ with identical distributions. Notice that a set of Laplace-distributed variables $\{s_1, s_2, \dots, s_N\}$ representing the difference of two special K random variables can be selected instead, and the results thus obtained will be identical. Therefore, the results presented below are valid for that case also. Now define a real-valued random variable y as

$$(F-36) \quad y = v_1 + v_2 + \dots + v_N \quad -\infty \leq y \leq \infty$$

The mean and variance of y are

$$(F-37) \quad \mu_y = E[y] = E[v_1 + v_2 + \dots + v_N] = E[v_1] + E[v_2] + \dots + E[v_N] = 0$$

$$(F-38a) \quad \sigma_y^2 = E[(y - \mu_y)^2] = E[y^2] = E[v_1^2] + E[v_2^2] + \dots + E[v_N^2]$$

$$(F-38b) \quad \sigma_y^2 = N\sigma_v^2 = 2N\sigma_z^2 = 2N\sigma_u^4$$

The PDF of y is obtained next using the characteristic function of the N Laplace-distributed variables $\{v_i\}$. The characteristic function of the sum of N independent random variables is equal to the product of the individual characteristic functions. For the

case herein of identically-distributed random variables $\{v_i | i=1, \dots, N\}$, the result is

$$(F-39a) \quad \phi_y(\omega) = \prod_{i=1}^N \phi_{v_i}(\omega) = [\phi_v(\omega)]^N \quad -\infty \leq \omega \leq \infty$$

$$(F-39b) \quad \phi_y(\omega) = \left[\frac{a^2}{\omega^2 + a^2} \right]^N = \frac{a^{2N}}{(\omega^2 + a^2)^N} \quad -\infty \leq \omega \leq \infty$$

$$(F-39c) \quad a^2 = \frac{2N}{\sigma_y^2} = \frac{2}{\sigma_v^2} = \frac{1}{\sigma_z^2} = \frac{1}{\sigma_u^4}$$

The expression for parameter a is repeated herein for convenience, with an additional equivalence in terms of σ_y^2 , which follows from Equation (F-38b).

The PDF of y is the Fourier transform of the characteristic function scaled by the factor $(2\pi)^{-1}$; that is,

$$(F-40a) \quad p_Y(y) = \frac{1}{2\pi} \mathcal{F}[\phi_y(\omega)] = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-j\omega y} \phi_y(\omega) d\omega \quad -\infty \leq y \leq \infty$$

$$(F-40b) \quad p_Y(y) = \frac{1}{2\pi} \int_{-\infty}^{\infty} [\cos(\omega y) - j \sin(\omega y)] \phi_y(\omega) d\omega \quad -\infty \leq y \leq \infty$$

$$(F-40c) \quad p_Y(y) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \cos(\omega y) \phi_y(\omega) d\omega - j \frac{1}{2\pi} \int_{-\infty}^{\infty} \sin(\omega y) \phi_y(\omega) d\omega \quad -\infty \leq y \leq \infty$$

$$(F-40d) \quad p_Y(y) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \cos(\omega y) \phi_y(\omega) d\omega = \frac{1}{\pi} \int_0^{\infty} \cos(\omega y) \phi_y(\omega) d\omega \quad -\infty \leq y \leq \infty$$

$$(F-40e) \quad p_Y(y) = \frac{a^{2N}}{\pi} \int_0^{\infty} \frac{\cos(\omega y)}{(\omega^2 + a^2)^N} d\omega \quad -\infty \leq y \leq \infty$$

In Equation (F-40c), the integrand of the integral in the imaginary part is an odd function of ω , which integrates to zero over the real line. This implies that the PDF is real-valued, in accordance with theory. The integrand of the integral in the real part is an even function of ω , which implies that the integral of the real part can be evaluated as twice the integral over the positive real line, as indicated in Equation (F-40d). The last expression, Equation (F-40e) is evaluated by referring to equation no. 3.737.1 of (Gradshteyn and Ryzhik, 1980), which results in

$$(F-41) \quad p_Y(y) = \frac{a}{2^{2N-1}(N-1)!} \left[\sum_{k=0}^{N-1} \frac{(2N-2-k)! (2a)^k}{k! (N-1-k)!} |y|^k \right] e^{-a|y|} \quad -\infty \leq y \leq \infty$$

with parameter $a > 0$ as in Equation (F-39c). Since y is the sum of N identically-distributed, zero-mean, Laplace random variables, then herein y is said to be SL-distributed, and $p_Y(y)$ is the PDF associated with the SL distribution. For large N the SL distribution approximates the Gaussian distribution, as expected based on the Central Limit Theorem. In particular, the approximation is very good for $N > 30$, as verified by Michels (1996) via software-based analyses.

F.5 ACS Estimator Lags

The SL distribution describes the probabilistic behaviour of unscaled real and imaginary components of complex-valued random

variables for four distinct cases of interest in radar systems and other applications. Specifically,

- (A) ACS lags $m \neq 0$ for the time-average class (circular; biased; unbiased) of ACS estimators for a scalar white noise sequence;
- (B) ACS lags $m \neq 0$ for ensemble-averaged ACS estimators of a scalar white noise process;
- (C) off-diagonal elements in the covariance matrix for the time-average class of estimators for a vector white noise sequence; and
- (D) off-diagonal elements in the covariance matrix for ensemble-averaged estimators of a vector white noise process.

Each case (and sub-cases, for time-average estimators) differs from the others on the basis of the scaling factor used. Therefore, it is important to determine the PDF of a generic scale transformation on random variable y . Let L denote a positive-valued integer constant, and define a real-valued random variable r as

$$(F-42) \quad r = \frac{y}{L} \quad -\infty \leq r \leq \infty$$

The mean and variance of r are obtained simply as

$$(F-43) \quad \mu_r = E[r] = \frac{1}{L} E[y] = 0$$

$$(F-44a) \quad \sigma_r^2 = E[(r - \mu_r)^2] = E[r^2] = \frac{1}{L^2} E[y^2]$$

$$(F-44b) \quad \sigma_r^2 = \frac{1}{L^2} \sigma_y^2 = \frac{N}{L^2} \sigma_v^2 = \frac{2N}{L^2} \sigma_z^2 = \frac{2N}{L^2} \sigma_u^4$$

And the PDF of r is determined via the transformation of variables method. Specifically,

$$(F-45) \quad p_R(r) = p_Y(y(r)) \left| \frac{dy(r)}{dr} \right| \quad -\infty \leq r \leq \infty$$

From Equation (F-42),

$$(F-46) \quad y(r) = y = Lr$$

$$(F-47) \quad \frac{dy(r)}{dr} = L$$

and from Equations (F-41) and (F-46),

$$(F-48) \quad p_Y(y(r)) = \frac{a}{2^{2N-1}(N-1)!} \left[\sum_{k=0}^{N-1} \frac{(2N-2-k)! (2a)^k |L|^k}{k! (N-1-k)!} |r|^k \right] e^{-a|L||r|}$$

Combining Equations (F-47) and (F-48) leads to the desired result,

$$(F-49a) \quad p_R(r) = \frac{a|L|}{2^{2N-1}(N-1)!} \left[\sum_{k=0}^{N-1} \frac{(2N-2-k)! (2a)^k |L|^k}{k! (N-1-k)!} |r|^k \right] e^{-a|L||r|} \quad -\infty \leq r \leq \infty$$

$$(F-49b) \quad a^2 = \frac{2N}{L^2 \sigma_r^2} = \frac{2N}{\sigma_y^2} = \frac{2}{\sigma_v^2} = \frac{1}{\sigma_z^2} = \frac{1}{\sigma_u^4} \quad a > 0$$

In Equation (F-49) it is important to preserve the absolute value operator on L to emphasize that L is positive-valued. This PDF is referred to herein as the scaled SL distribution. Values of the parameters N and L for the cases of interest in this report are listed on Table F-1. For the ensemble-average cases in Table F-1,

K denotes the number of realizations averaged. Recall that in all cases N represents the total number of independent Laplace-distributed variables combined together, whereas L represents the normalizing factor applied to the sum. Also, M_c represents the number of unique lags for the circular estimator, as defined in Equation (E-7).

N	L	CASE DESCRIPTION
N	N	Time-average, circular, scalar, real and imaginary ACS lags $1 \leq m \leq M_c$
$\frac{N}{2}$	$\frac{N}{2}$	Time-average, circular, scalar, real ACS lag $m = N/2$ for N even only
N-m	N	Time-average, biased, scalar, real and imaginary ACS lag m for $1 \leq m \leq N-1$
N	N	Time-average, unbiased, scalar, real and imaginary ACS lags $1 \leq m \leq N-1$
KN	KN	Ensemble-average of time-averaged, circular, scalar, real and imaginary ACS lags $1 \leq m \leq M_c$
$\frac{KN}{2}$	$\frac{KN}{2}$	Ensemble-average of time-averaged, circular, scalar, real ACS lag $m = N/2$ for N even only
K(N-m)	KN	Ensemble-average of time-averaged, biased, scalar, real and imaginary ACS lag m for $1 \leq m \leq N-1$
KN	KN	Ensemble-average of time-averaged, unbiased, scalar, real and imaginary ACS lags $1 \leq m \leq N-1$
N	N	Time-average covariance matrix real and imaginary parts of off-diagonal elements (lag $m = 0$)
K	K	Ensemble-average covariance matrix real and imaginary parts of off-diagonal elements (lag $m = 0$)

Table F-1. Values of scaled SL PDF parameters N and L for cases of interest.

APPENDIX G. MULTIPLE HYPOTHESES TESTING FOR ECG DIAGNOSIS

Michels (1991) extended the innovations-based generalized likelihood ratio test for binary hypotheses involving scalar, complex-valued sequences to the multichannel signal case, with the final test expressions as summarized in Section 5.0. The derivation is based on the Neyman-Pearson criterion, which is the appropriate criterion for radar detection applications (target detection in clutter, interference, and noise). However, ECG diagnosis differs from radar detection in two important ways. First, the vector of ECG traces (the channel output) is real-valued. Second, the general ECG diagnosis formulation involves multiple hypotheses. Thus, an alternative approach based on the Bayes criterion is formulated herein for ECG trace discrimination.

Multiple hypotheses testing is a well-established procedure, and is discussed in several texts. The brief discussion presented herein is adopted from the text by Srinath and Rajasekaran (1979), with some minor modifications and convenient notational changes. The formulation based on the Bayes criterion is adopted, wherein the average cost of making a decision is minimized. Consider an $(M+1)$ -hypotheses problem with H_0 as the null hypothesis, and M alternative hypotheses $\{H_1, H_2, \dots, H_M\}$. Let C_{ij} denote the cost associated with selecting hypothesis H_i when hypothesis H_j is true, and let $\mathcal{P}[H_j]$ denote the prior probability corresponding to the occurrence of hypothesis H_j . Finally, denote the data to be tested as a JN -element vector $\underline{\mathbf{x}}$, composed by the concatenation of the real-valued residual vector sequence $\{\underline{\mathbf{x}}(n) | n=0, 1, \dots, N-1\}$,

$$(G-1) \quad \underline{\mathbf{x}} = \begin{bmatrix} \underline{\mathbf{x}}(0) \\ \underline{\mathbf{x}}(1) \\ \vdots \\ \underline{\mathbf{x}}(N-1) \end{bmatrix}$$

As shown by Srinath and Rajasekaran (1979), the Bayes criterion for hypothesis selection leads to a decision rule based on the values of $M+1$ functions $\{f_0(\underline{x}), f_1(\underline{x}), \dots, f_M(\underline{x})\}$, where each function is of the form

$$(G-2) \quad f_i(\underline{x}) = \sum_{\substack{j=0 \\ j \neq i}}^M (C_{ij} - C_{jj}) \mathcal{P}[H_j] p(\underline{x}|H_j) \quad i = 0, 1, \dots, M$$

In Equation (G-2), $p(\underline{x}|H_j)$ represents the posterior probability density function for hypothesis H_j . Hypothesis H_i is selected if the corresponding function $f_i(\underline{x})$ attains the minimum value among the $M+1$ functions. Implied in the formulation that led to Equation (G-2) is the assumption that

$$(G-3) \quad C_{ij} > C_{jj}$$

This assumption states that the cost of making an incorrect decision is larger than the cost associated with a correct decision, which is a reasonable posture in most applications, including radar systems and ECG diagnosis. Notice that with the constraint (G-3), each term in the summation of Equation (G-2) is non-negative, and consequently, each function $f_i(\underline{x})$ is non-negative also.

A special case of Equation (G-2) is of practical and theoretical interest. Let

$$(G-4a) \quad C_{ij} = 1 \quad i \neq j$$

$$(G-4b) \quad C_{jj} = 0 \quad j = 0, 1, \dots, M$$

With these conditions, Equation (G-2) becomes

$$(G-5a) \quad f_i(\underline{\xi}) = \sum_{\substack{j=0 \\ j \neq i}}^M \mathcal{P}[\mathbf{H}_j] p(\underline{\xi}|\mathbf{H}_j) = p(\underline{\xi}) \sum_{\substack{j=0 \\ j \neq i}}^M \mathcal{P}[\mathbf{H}_j|\underline{\xi}] \quad i = 0, 1, \dots, M$$

$$(G-5b) \quad f_i(\underline{\xi}) = [1 - \mathcal{P}[\mathbf{H}_i|\underline{\xi}]] p(\underline{\xi}) \quad i = 0, 1, \dots, M$$

where Bayes' rule has been invoked in the second equality of Equation (G-5a). As before, the hypothesis that corresponds with the minimum-valued function $f_i(\underline{\xi})$ is selected. This case is referred to as the minimum probability of error criterion.

Consider now the minimum probability of error case when all hypotheses are equally likely a priori. That is, all prior probabilities are the same,

$$(G-6) \quad \mathcal{P}[\mathbf{H}_0] = \mathcal{P}[\mathbf{H}_1] = \dots = \mathcal{P}[\mathbf{H}_M] = P_p$$

For conditions (G-4) and (G-6), Equation (G-2) reduces to

$$(G-7) \quad f_i(\underline{\xi}) = P_p \sum_{\substack{j=0 \\ j \neq i}}^M p(\underline{\xi}|\mathbf{H}_j) = [1 - p(\underline{\xi}|\mathbf{H}_i)] P_p \quad i = 0, 1, \dots, M$$

Since P_p is a fixed constant, the minimum-valued function $f_i(\underline{\xi})$ is that one for which $p(\underline{\xi}|\mathbf{H}_i)$ is a maximum.

Michels (1991) has derived the multivariate PDF $p(\underline{\xi}|\mathbf{H}_j)$ for the Gaussian-distributed, zero-mean, complex-valued, residual vector. The PDF for the real-valued case is obtained as a simple modification of the complex-valued case PDF. Additionally, since the natural logarithm is a monotonic function, it is equivalent (and convenient) to consider the natural logarithm of the

multivariable PDF $p(\underline{\varepsilon}|\mathbf{H}_i)$. It follows from the results obtained by Michels (1991) that the log-likelihood under the i th hypothesis is

$$(G-8a) \quad \mathcal{L}(\underline{\varepsilon}|\mathbf{H}_i) = \ln[p(\underline{\varepsilon}|\mathbf{H}_i)] = \sum_{n=0}^{N-1} \left[-\frac{J}{2} \ln[2\pi] - \ln[|\Omega(\mathbf{H}_i)|] - \underline{\varepsilon}^T(n|\mathbf{H}_i) \Omega^{-1}(\mathbf{H}_i) \underline{\varepsilon}(n|\mathbf{H}_i) \right]$$

$$(G-8b) \quad \mathcal{L}(\underline{\varepsilon}|\mathbf{H}_i) = -\frac{JN}{2} \ln[2\pi] - N \ln[|\Omega(\mathbf{H}_i)|] - \sum_{n=0}^{N-1} \underline{\varepsilon}^T(n|\mathbf{H}_i) \Omega^{-1}(\mathbf{H}_i) \underline{\varepsilon}(n|\mathbf{H}_i)$$

The maximum-valued log-likelihood function, $\mathcal{L}(\underline{\varepsilon}|\mathbf{H}_i)$, is that one for which the sum of the second and third terms in the right-hand-side of Equation (G-8b) is a minimum (due to the negative sign), since the first term on the right-hand-side is a fixed constant for all hypotheses. In fact, for both applications considered herein the second term, $-N \ln[|\Omega(\mathbf{H}_i)|]$, ends up positive-valued upon evaluation because the determinant of the covariance matrix $\Omega(\mathbf{H}_i)$ is less than unity (and the natural logarithm of a quantity less than one is negative-valued). The finite sum of weighted quadratic terms is the normalized residual sequence power, where the normalization factor is the true covariance matrix of the residual vector under the i th hypothesis. This term is negative-valued always. In summary, for the minimum probability of error criterion with equal prior probabilities, the selected hypothesis is the one which corresponds to the maximum-valued log-likelihood function.

Without loss of generality, the log-likelihood function $\mathcal{L}(\underline{\varepsilon}|\mathbf{H}_i)$ can be replaced by a simplified log-likelihood function of the form

$$(G-9) \quad \mathcal{L}(\underline{\varepsilon}|\mathbf{H}_i) = N \ln[|\Omega(\mathbf{H}_i)|] + \sum_{n=0}^{N-1} \underline{\varepsilon}^T(n|\mathbf{H}_i) \Omega^{-1}(\mathbf{H}_i) \underline{\varepsilon}(n|\mathbf{H}_i)$$

The decision rule must be modified accordingly since the simplification includes a sign change. Specifically, the selected hypothesis is the one which corresponds to the minimum-valued log-likelihood function $l(\underline{\xi}|\mathbf{H}_j)$. Figure G-1 is a block diagram for the multiple hypotheses test based on the minimum probability of error criterion with equal prior probabilities; that is, the decision rule using the log-likelihood function in Equation (G-9).

For complex-valued residual vector sequences, the corresponding result involves two modifications to Equation (G-8b). First, the constant term becomes $-JN \ln[\pi]$. Second, the transpose operator is replaced by the Hermitian operator.

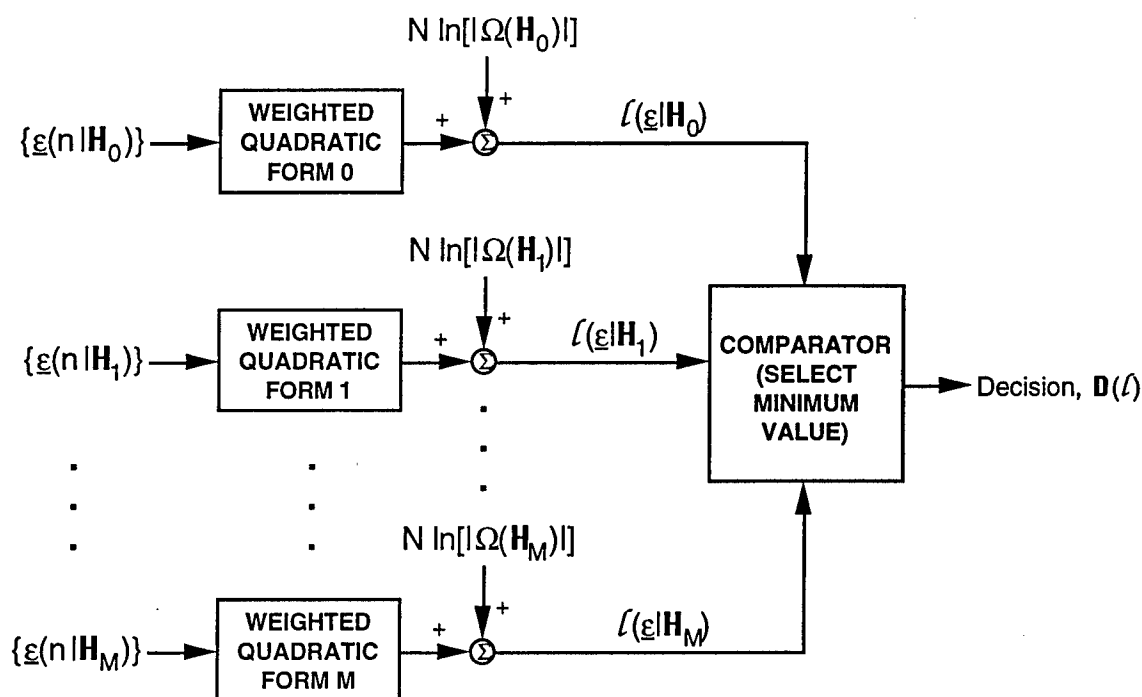


Figure G-1. Multiple hypotheses test block diagram (minimum probability of error criterion with equal prior probabilities).

REFERENCES

H. Akaike

- (1974) "Stochastic Theory of Minimal Realization," IEEE Transactions on Automatic Control, Vol. AC-19, No. 6 (December), pp. 667-674.
- (1975) "Markovian Representation of Stochastic Processes by Canonical Variables," SIAM Journal on Control, Vol. 13, No. 1 (January), pp. 162-173.

B. D. O. Anderson and J. B. Moore

- (1979) Optimal Filtering, Prentice-Hall, Englewood Cliffs, NJ.

K. S. Arun and S. Y. Kung

- (1990) "Balanced Approximation of Stochastic Systems," SIAM Journal on Matrix Analysis and Applications, Vol. 11, No. 1 (January), pp. 42-68.

P. Beckmann

- (1967) Probability in Communication Engineering, Harcourt, Brace & World, Inc., New York, NY.

J. V. Candy

- (1976) Realization of Invariant System Descriptions From Markov Sequences, Ph. D. Dissertation, Department of Electrical Engineering, University of Florida, Gainesville, FL.

T. C. Chou

- (1986) Electrocardiography in Clinical Practice (second edition), Grunpe and Stratton, Inc., Orlando, FL.

G. R. Cooper and C. D. McGillem

- (1971) Probabilistic Methods of Signal and System Analysis, Holt, Rinehart and Winston, Inc., New York, NY.

V. G. Dávila-Román

(1994) Private communication, Washington University School of Medicine, Cardiovascular Division, St. Louis, MO.

D. W. Davis and J. R. Román

(1996) Output Data Technique Model Identification Software, SSC Technical Report No. SSC-TR-96-01, Scientific Studies Corporation, Palm Beach Gardens, FL.

E. J. Davison and S. H. Wang

(1974) "Properties and calculation of transmission zeros of linear multivariable systems," Automatica, Vol. 10, pp. 643-658.

(1976) "Remark on multiple transmission zeros of a system," Automatica, Vol. 12, p. 195.

U. B. Desai, D. Pal, and R. D. Kirkpatrick

(1985) "A realization approach to stochastic model reduction," International Journal on Control, Vol. 42, No. 4, pp. 821-838.

J. J. Dongarra, C. B. Moler, J. R. Bunch, and G. W. Stewart

(1979) LINPACK Users' Guide, Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA.

A. Emami-Naeini and P. Van Dooren

(1982) "Computation of zeros of linear multivariable systems," Automatica, Vol. 18, No. 4, pp. 415-430.

P. L. Faure

(1976) "Stochastic realization algorithms," in System Identification: Advances and Case Studies, R. K. Mehra and D. G. Laniotis (eds.), Academic Press, New York, NY.

B. R. Frieden

(1983) Probability, Statistical Optics, and Data Testing, Springer-Verlag, New York, NY.

I. M. Gelfand and A. M. Yaglom

(1959) "Calculation of the amount of information about a random function contained in another such function," American Mathematical Society Translations (2), Vol. 12, pp. 199-246.

I. S. Gradshteyn and I. M. Ryzhik

(1980) Table of Integrals, Series, and Products, Academic Press, New York, NY.

A. C. Guyton

(1991) Textbook of Medical Physiology, Eighth Edition, W. B. Saunders Co., Philadelphia, PA.

N. A. J. Hastings and J. B. Peacock

(1975) Statistical Distributions: A Handbook for Students and Practitioners, John Wiley & Sons, New York, NY.

A. G. Jaffer, M. H. Baker, W. P. Ballance, and J. R. Staub

(1991) Adaptive Space-Time Processing Techniques for Airborne Radars, RL Technical Report No. RL-TR-91-162, Rome Laboratory, Griffiss AFB, NY.

R. E. Kalman, P. L. Falb, and M. A. Arbib

(1969) Topics in Mathematical System Theory, McGraw-Hill Book Co., New York, NY.

S. Y. Kung

- (1974) "A new identification and model reduction algorithm via singular value decomposition," Proceedings of the 12th Asilomar Conference on Circuits, Systems, and Computers, Pacific Grove, CA, pp. 705-714.

A. J. Laub and B. C. Moore

- (1978) "Calculation of transmission zeros using QZ techniques," Automatica, Vol. 14, pp. 557-566.

P. A. S. Metford and S. Haykin

- (1985) "Experimental analysis of an innovations-based detection algorithm for surveillance radar," IEE Proceedings, Vol. 132, Pt. F, No. 1 (February), pp. 18-26.

J. H. Michels

- (1990) Synthesis of Multichannel Autoregressive Random Processes and Ergodicity Considerations, RL Technical Report No. RADC-TR-90-211, Rome Laboratory, Rome, NY.
- (1991) Multichannel Detection Using the Discrete-Time Model-Based Innovations Approach, RL Technical Report No. RL-TR-91-269, Rome Laboratory, Rome, NY.
- (1992a) "Detection of Partially Correlated Signals in Clutter Using a Multichannel Model-Based Approach," presented at the 1992 National Telesystems Conference, May 19-20, The George Washington University, Ashburn, VA.
- (1992b) Considerations of the Error Variances of Time-Averaged Estimators for Correlated Processes, RL Technical Report No. RL-TR-92-339, Rome Laboratory, Rome, NY.
- (1996) Private communication.

B. C. Moore

- (1981) "Principal component analysis in linear systems: Controllability, observability, and model reduction,"

IEEE Transactions on Automatic Control, Vol. AC-26, No. 1 (February), pp. 17-31.

A. H. Nuttall

(1976) "Multivariate Linear Predictive Spectral Analysis Employing Weighted Forward and Backward Averaging: A Generalization of Burg's Algorithm," Naval Underwater Systems Center Tech. Report No. TR-5501, New London, CT.

C. C. Paige and M. A. Saunders

(1981) "Towards a Generalized Singular Value Decomposition," SIAM Journal on Numerical Analysis, Vol. 18, No. 3 (June), pp. 398-405.

M. C. Pease

(1965) Methods of Matrix Algebra, Academic Press, New York, NY.

M. Rangaswamy and J. H. Michels

(1996) Private communication.

M. Rangaswamy, P. Chakravarthi, D. Weiner, L. Cai, H. Wang, and A. Ozturk

(1993) Signal Detection in Correlated Gaussian and Non-Gaussian Radar Clutter, RL Technical Report No. RL-TR-93-79, Rome Laboratory, Griffiss AFB, NY.

M. Rangaswamy, D. D. Weiner, and J. H. Michels

(1993) "Multichannel Detection for Correlated Non-Gaussian Random Processes Based on Innovations," presented at the SPIE International Symposium on Optical Engineering and Photonics in Aerospace and Remote Sensing (Conference 1955), April 12-16, Orlando, FL.

J. R. Román and D. W. Davis

- (1993a) Multichannel System Identification and Detection Using Output Data Techniques, RL Technical Report No. RL-TR-93-141, Rome Laboratory, Rome, NY.
- (1993b) State-Space Models for Multichannel Detection, RL Technical Report No. RL-TR-93-146, Rome Laboratory, Rome, NY.
- (1994) "Multichannel Processing for Biomedical Applications," presented at the Fourth Annual IEEE Dual Use Technologies and Applications Conference, May 23-26, Utica, NY.
- (1996) Multichannel System Identification and Detection Using Output Data Techniques - Phase II, Vol. II, RL Technical Report, Rome Laboratory, Rome, NY.

J. R. Román, D. W. Davis, J. H. Michels, and V. G. Dávila-Román

- (1996a) "Model-Based Multichannel Detection of Cardiac Conduction Abnormalities," presented at the American College of Cardiology 45th Annual Scientific Session, March 24-27, Orlando, FL.
- (1996b) "Model-Based Multichannel Diagnosis of Cardiac Conduction Abnormalities," presented at the 23rd Annual Computers in Cardiology Conference, September 8-11, Indianapolis, IN.

R. O. Schmidt

- (1979) "Multiple emitter location and signal parameter estimation," Proceedings of the RADC Spectrum Estimation Workshop, Griffiss AFB, Rome, NY, pp. 243-258; also in IEEE Transactions on Antennas and Propagation, Vol. AP-34, No. 3 (March 1986), pp. 276-280.
- (1981) A Signal Subspace Approach to Multiple Emitter Location and Spectral Estimation, Ph. D. Dissertation, Department of Electrical Engineering, Stanford Univ., Stanford, CA.

- M. D. Srinath and P. K. Rajasekaran
(1979) An Introduction to Statistical Signal Processing With Applications, John Wiley & Sons, Inc., New York, NY.
- O. N. Strand
(1977) "Multichannel Complex Maximum Entropy (Auto-Regressive) Spectral Analysis," IEEE Transactions on Automatic Control, Vol. AC-22, No. 4 (August), pp. 634-640.
- C. W. Therrien
(1983) "On the relation between triangular matrix decomposition and linear prediction," Proceedings of the IEEE, Vol. 71, No. 12 (December), pp. 1459-1460.
- P. Van Overschee and B. De Moor
(1991) "Subspace Algorithms for the Stochastic Identification Problem," ESAT Report, Dept. of Electrical Engineering, Katholieke Universiteit Leuven, Haverlee, Belgium.
(1993) "Subspace Algorithms for the Stochastic Identification Problem," Automatica, Vol. 29, No. 3, pp. 649-660.
- G. S. Wagner
(1994) Marriott's Practical Electrocardiography, Ninth Edition, Williams & Wilkins Publishing Co., Baltimore, MD.
- H. Wang and L. Cai
(1994) "On Adaptive Spatial-Temporal Processing for Airborne Surveillance Radar Systems," IEEE Transactions on Aerospace and Electronic Systems, Vol. 30, No.3 (July), pp. 660-670.

J. Ward

- (1994) Space-Time Adaptive Processing for Airborne Radar, Technical Report No. TR-1015 (December), contract no. F19628-95-C-0002, Lincoln Laboratory, Massachusetts Institute of Technology, Lexington, MA.

R. A. Wiggins and E. A. Robinson

- (1965) "Recursive Solution to the Multichannel Filtering Problem," Journal of Geophysical Research, Vol. 70, No.8 (April 15), pp. 435-441.

J. L. Willems

- (1990) Common Standards for Quantitative Electrocardiography - 10th and Final Progress Report, CSE Coordinating Center, Division of Medical Informatics, University Hospital Gasthuisberg, Leuven, Belgium.
- (1994) Private communication, CSE Coordinating Center, Division of Medical Informatics, University Hospital Gasthuisberg, Leuven, Belgium.

J. L. Willems, C. Abreu-Lima, P. Arnaud, C. R. Brohet, B. Denis, J. Gehring, I. Graham, G. van Herpen, H. Machado, J. Michaelis, and S. D. Moulopoulos

- (1990) "Evaluation of ECG Interpretation Results Obtained by Computer and Cardiologists," Methods of Information in Medicine, Vol. 29, No. 4 (September), pp. 308-316.

J. L. Willems, C. Abreu-Lima, P. Arnaud, J. H. van Bommel, C. R. Brohet, R. Degani, B. Denis, J. Gehring, I. Graham, G. van Herpen, H. Machado, P. W. Macfarlane, J. Michaelis, S. D. Moulopoulos, P. Rubel, and C. Zywiets

- (1991) "The diagnostic performance of computer programs for the interpretation of electrocardiograms," New England

Journal of Medicine, Vol. 325, December 19, pp. 1767-1773.

H. P. Zeiger and A. J. McEwen

(1974) "Approximate Linear Realizations of Given Dimension via Ho's Algorithm," IEEE Transactions on Automatic Control, Vol. AC-19, No. 2 (April), pg. 153.

C. Zywietz, J. L. Willems, P. Arnaud, J. H. van Bommel, R. Degani, P. W. Macfarlane

(1990) "Stability of Computer ECG Amplitude Measurements in the Presence of Noise", Computers and Biomedical Research, Vol. 23, pp. 10-31.

MISSION OF ROME LABORATORY

Mission. The mission of Rome Laboratory is to advance the science and technologies of command, control, communications and intelligence and to transition them into systems to meet customer needs. To achieve this, Rome Lab:

- a. Conducts vigorous research, development and test programs in all applicable technologies;
- b. Transitions technology to current and future systems to improve operational capability, readiness, and supportability;
- c. Provides a full range of technical support to Air Force Material Command product centers and other Air Force organizations;
- d. Promotes transfer of technology to the private sector;
- e. Maintains leading edge technological expertise in the areas of surveillance, communications, command and control, intelligence, reliability science, electro-magnetic technology, photonics, signal processing, and computational science.

The thrust areas of technical competence include: Surveillance, Communications, Command and Control, Intelligence, Signal Processing, Computer Science and Technology, Electromagnetic Technology, Photonics and Reliability Sciences.



DEPARTMENT OF THE AIR FORCE
AIR FORCE RESEARCH LABORATORY (AFRL)

15 Jun 04

MEMORANDUM FOR DTIC-OCQ

ATTN: Larry Downing
Ft. Belvoir, VA 22060-6218

FROM: AFRL/FOIP

SUBJECT: Distribution Statement Change

1. The following documents (previously limited by SBIR data rights) have been reviewed and have been approved for Public Release; Distribution Unlimited:

ADB226867, "Multichannel System Identification and Detection Using Output Data Techniques", RL-TR-97-5, Vol 1.

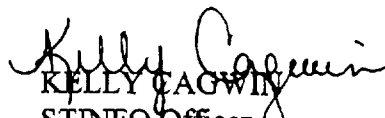
ADB176689, "Multichannel System Identification and Detection Using Output Data Techniques", RL-TR-93-141.

ADB198116, "Multichannel Detection Using Higher Order Statistics", RL-TR-95-11.

ADB232680, "Two-Dimensional Processing for Radar Systems", RL-TR-97-127.

ADB276328, "Two-Dimensional Processing for Radar Systems", AFRL-SN-RS-TR-2001-244.

2. Please contact the undersigned should you have any questions regarding this memorandum. Thank you very much for your time and attention to this matter.


KELLY CAGWIN
STINFO Officer
Information Directorate
315-330-7094/DSN 587-7094