PROCESS DISTORTION MEASURES AND SIGNAL PROCESSING

A DISSERTATION

SUBMITTED TO THE DEPARTMENT OF ELECTRICAL ENGINEERING

AND THE COMMITTEE ON GRADUATE STUDIES

OF STANFORD UNIVERSITY

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

By

Yasuo Matsuyama
August, 1978

G81

STANFORD UNIVERSITY
Committee on Graduate Studies

Notice of Appointment of Ph.D. Dissertation Reading Committee

This form should be submitted to the Graduate Study Office as soon as appropriate during a candidate's Ph.D. program and no later than the date on which the candidate submits his or her request for a University Oral Examination to be scheduled (Form G21).

The minimum membership of a Ph.D. reading committee is (1) the principal dissertation adviser; (2) a second member from within the major department; and (3) a third member chosen from the major or another department. The principal adviser and at least one of the other committee members must be Academic Council members.

In cases where the dissertation topic makes advice from outside the major department useful, the appointment of an appropriate outside reader should be made early, and he or she should be encouraged to follow and advise on the progress of the dissertation research and preparation.

To the University Committee on Graduate Studies:

The following Ph.D. dissertation reading committee has been approved for Yasuo Matsuyama (Candidate's name),

whose dissertation title will be Process distortion measures and signal processing
(Type or print carefully, as this will now be the title of record.)

If field of investigation has changed, it now is ______.

Principal adviser's signature: Robert M. Gray — Assoc. Prof. — Electrical Eng. — Jul 3, 1978
Typed name: Robert M. Gray (Title) (Department) (Date)

Signature: M. Morf — Assist. Prof. — Electrical. Eng. — Jul 3, 1978
Typed name: Martin Morf (Title) (Department) (Date)

Signature: J. W. Goodman — Professor — Electrical Eng. — Jul 11 1978
Typed name: J.W. Goodman (Title) (Department) (Date)

Signature: ______
Typed name: (Title) (Department) (Date)

It is understood that each member of the reading committee will certify by signature on the final copies of the dissertation that he has read the dissertation, and that in his opinion it is of a scope and quality acceptable in fulfillment of this requirement. At least one member of the reading committee will read the dissertation in its final submitted form and certify its acceptability on Form G82.

Signature: J. G. Linvill
Chairman of Major Department

Major Department: Electrical Engineering

7/17/78
Date

Form G81 3M-9-74-39727

I certify that I have read this thesis and that in my opinion it is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

R. M. Gray

I certify that I have read this thesis and that in my opinion it is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

M. Morf

I certify that I have read this thesis and that in my opinion it is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

Joseph W. Goodman

Approved for the University Committee
on the Graduate Division:

Dean of Graduate Studies

Oral Exam

INFORMATION SYSTEMS LABORATORY
Department of Electrical Engineering
STANFORD UNIVERSITY Stanford, CA 94305

SPECIAL SEMINAR

PROCESS DISTORTION MEASURES AND

SIGNAL PROCESSING

by

Yasuo Matsuyama
Information Systems Laboratory
Stanford University

A process distortion determines the degree of similarity between two stochastic processes with respect to an underlying distortion measure by considering the entire process. Two types of measures -- the $\bar{\rho}$ and Prohorov distances -- are considered. These distortion measures are shown to be topologically equivalent, in specific instances, and can be reduced to spectral distortion equivalents.

One application of such process distortion measures is to the mismatch problem of linear predictors and interpolators. It will be shown that variants of these measures can serve as indicators of the robustness against possible mismatch. In addition, these spectral distortions and their relatives appear to be useful in the mathematical approximation of human perception of speech distortion. In particular, topological and coding equivalences as well as approximations and numerical computations will be discussed.

Finally, to give a practical application of the above theory, the problem of speech compression is discussed. In particular, a speech waveform coder is designed using a simple and direct application of the spectral distortion theory. This coder is shown to be easily implemented, provide high intelligibility of decoded speech, and yet provide a compression of speech into a 1 bit/sample bit stream. Of particular interest is an off-line LPC approach combined with an interesting application of universal coding used in the design. A demonstration of system performance will be provided.

Date: Thursday, July 20, 1978

Time: 2:00 p.m.

Place: Durand 450

FORM TX

UNITED STATES COPYRIGHT OFFICE

CERTIFICATE OF COPYRIGHT REGISTRATION

REGISTRATION NUMBER
TX 247-016

TX TXU

This certificate, issued under the seal of the Copyright Office in accordance with the provisions of section 410(a) of title 17, United States Code, attests that copyright registration has been made for the work identified below. The information in this certificate has been made a part of the Copyright Office records.

Barbara Ringer

79-05,904

Register of Copyrights
United States of America

EFFECTIVE DATE OF REGISTRATION

APR 17 1979 (Month) (Day) (Year)

DO NOT WRITE ABOVE THIS LINE. IF YOU NEED MORE SPACE, USE CONTINUATION SHEET

1 Title

TITLE OF THIS WORK: PROCESS DISTORTION MEASURES AND SIGNAL PROCESSING.

PREVIOUS OR ALTERNATIVE TITLES:

If a periodical or serial give: Vol....... No....... Issue Date........

PUBLICATION AS A CONTRIBUTION: (If this work was published as a contribution to a periodical, serial, or collection, give information about the collective work in which the contribution appeared.)

Title of Collective Work:........ Vol....... No....... Date........ Pages........

2 Author(s)

IMPORTANT: Under the law, the "author" of a "work made for hire" is generally the employer, not the employee (see instructions). If any part of this work was "made for hire" check "Yes" in the space provided, give the employer (or other person for whom the work was prepared) as "Author" of that part, and leave the space for dates blank.

1

NAME OF AUTHOR: YASUO MATSUYAMA

Was this author's contribution to the work a "work made for hire"? Yes...... No. X

DATES OF BIRTH AND DEATH: Born 1947 (Year) Died.......... (Year)

AUTHOR'S NATIONALITY OR DOMICILE: Citizen of JAPAN (Name of Country) or Domiciled in........ (Name of Country)

WAS THIS AUTHOR'S CONTRIBUTION TO THE WORK: Anonymous? Yes...... No. X Pseudonymous? Yes...... No. X

If the answer to either of these questions is "Yes, see detailed instructions attached.

AUTHOR OF: (Briefly describe nature of this author's contribution) ENTIRE TEXT

2

NAME OF AUTHOR:

Was this author's contribution to the work a "work made for hire"? Yes...... No......

DATES OF BIRTH AND DEATH: Born.......... (Year) Died.......... (Year)

AUTHOR'S NATIONALITY OR DOMICILE: Citizen of........ (Name of Country) or Domiciled in........ (Name of Country)

WAS THIS AUTHOR'S CONTRIBUTION TO THE WORK: Anonymous? Yes...... No...... Pseudonymous? Yes...... No......

If the answer to either of these questions is "Yes, see detailed instructions attached.

AUTHOR OF: (Briefly describe nature of this author's contribution)

3

NAME OF AUTHOR:

Was this author's contribution to the work a "work made for hire"? Yes...... No......

DATES OF BIRTH AND DEATH: Born.......... (Year) Died.......... (Year)

AUTHOR'S NATIONALITY OR DOMICILE: Citizen of........ (Name of Country) or Domiciled in........ (Name of Country)

WAS THIS AUTHOR'S CONTRIBUTION TO THE WORK: Anonymous? Yes...... No...... Pseudonymous? Yes...... No......

If the answer to either of these questions is "Yes, see detailed instructions attached.

AUTHOR OF: (Briefly describe nature of this author's contribution)

3 Creation and Publication

YEAR IN WHICH CREATION OF THIS WORK WAS COMPLETED: Year 1978....

(This information must be given in all cases.)

DATE AND NATION OF FIRST PUBLICATION: Date MARCH 15, 1979 * (Month) (Day) (Year)

Nation U.S.A. (Name of Country)

(Complete this block ONLY if this work has been published.)

4 Claimant(s)

NAME(S) AND ADDRESS(ES) OF COPYRIGHT CLAIMANT(S):

YASUO MATSUYAMA
1-14 ISHIKAWACHO
NAKAKU, YOKOHAMA, JAPAN

TRANSFER: (If the copyright claimant(s) named here in space 4 are different from the author(s) named in space 2, give a brief statement of how the claimant(s) obtained ownership of the copyright.)

- *Complete all applicable spaces (numbers 5-11) on the reverse side of this page*
- *Follow detailed instructions attached*
- *Sign the form at line 10*

DO NOT WRITE HERE

Page 1 of 2 pages

year date in copyright notice 1978 TX 247-016	EXAMINED BY: DW CHECKED BY:	APPLICATION RECEIVED: APR 17 79	FOR COPYRIGHT OFFICE USE ONLY
	CORRESPONDENCE: ☐ Yes	DEPOSIT RECEIVED: APR 17 79 ~~APR 17 79~~	
	DEPOSIT ACCOUNT FUNDS USED: ☑	REMITTANCE NUMBER AND DATE	

DO NOT WRITE ABOVE THIS LINE. IF YOU NEED ADDITIONAL SPACE, USE CONTINUATION SHEET (FORM TX/CON)

5 Previous Registration

PREVIOUS REGISTRATION:

- Has registration for this work, or for an earlier version of this work, already been made in the Copyright Office? Yes........ No X.....
- If your answer is "Yes," why is another registration being sought? (Check appropriate box)
  - ☐ This is the first published edition of a work previously registered in unpublished form.
  - ☐ This is the first application submitted by this author as copyright claimant.
  - ☐ This is a changed version of the work, as shown by line 6 of this application.
- If your answer is "Yes," give: Previous Registration Number .... Year of Registration ....

6 Compilation or Derivative Work

COMPILATION OR DERIVATIVE WORK: (See instructions)

PREEXISTING MATERIAL: (Identify any preexisting work or works that this work is based on or incorporates.)

MATERIAL ADDED TO THIS WORK: (Give a brief, general statement of the material that has been added to this work and in which copyright is claimed.)

7 Manufacturing

MANUFACTURERS AND LOCATIONS: (If this is a published work consisting preponderantly of nondramatic literary material in English, the law may require that the copies be manufactured in the United States or Canada for full protection. If so, the names of the manufacturers who performed certain processes, and the places where these processes were performed *must* be given. See instructions for details.)

NAMES OF MANUFACTURERS	PLACES OF MANUFACTURE
UNIVERSITY MICROFILMS INTERNATIONAL	ANN ARBOR, MICHIGAN 48106

8 License For Handicapped

REPRODUCTION FOR USE OF BLIND OR PHYSICALLY-HANDICAPPED PERSONS: (See instructions)

- Signature of this form at space 10, and a check in one of the boxes here in space 8, constitutes a non-exclusive grant of permission to the Library of Congress to reproduce and distribute solely for the blind and physically handicapped and under the conditions and limitations prescribed by the regulations of the Copyright Office: (1) copies of the work identified in space 1 of this application in Braille (or similar tactile symbols); or (2) phonorecords embodying a fixation of a reading of that work; or (3) both.

a ☒ Copies and phonorecords b ☐ Copies Only c ☐ Phonorecords Only

9 Fee and Correspondence

DEPOSIT ACCOUNT: (If the registration fee is to be charged to a Deposit Account established in the Copyright Office, give name and number of Account.)

Name: UNIVERSITY MICROFILMS INTERNATIONAL

Account Number: 313130

CORRESPONDENCE: (Give name and address to which correspondence about this application should be sent.)

Name: UNIVERSITY MICROFILMS INTERNATIONAL

Address: 300 N. ZEEB RD. (Apt.)

ANN ARBOR, (City) MICHIGAN (State) 48106 (ZIP)

10 Certification (Application must be signed)

CERTIFICATION: * I, the undersigned, hereby certify that I am the: (Check one)

☐ author ☐ other copyright claimant ☐ owner of exclusive right(s) ☒ authorized agent of: YASUO MATSUYAMA
(Name of author or other copyright claimant, or owner of exclusive right(s))

of the work identified in this application and that the statements made by me in this application are correct to the best of my knowledge.

Handwritten signature: (X) Barbara S. Neeley

Typed or printed name BARBARA S. NEELEY Date MAR 30 1979

11 Address For Return of Certificate

YASUO MATSUYAMA
(Name)

1-14 ISHIKAWACHO
(Number, Street and Apartment Number)

NAKAKU, YOKOHAMA, JAPAN
(City) (State) (ZIP code)

MAIL CERTIFICATE TO

20 JUN 1979

(Certificate will be mailed in window envelope)

* 17 U.S.C. § 506(e): Any person who knowingly makes a false representation of a material fact in the application for copyright registration provided for by section 409, or in any written statement filed in connection with the application, shall be fined not more than $2,500.

ABSTRACT

A process distortion measure is a measure of the "distance" or "badness of approximation" between two random processes. Examples are the $\bar{\rho}$ or Ornstein distance, a limiting Prokhorov distance, and several spectral distortion measures introduced in the speech literature. The first part of this dissertation is devoted to the development of several properties, interrelations, and interpretations of these distortion measures. In particular, the measures are grouped by topological and coding equivalence.

The second part of this work develops two applications of process distortion measures to signal processing. One application is to the mismatch problem of linear predictors and interpolators. It is shown that variants of these measures can serve as indicators of the robustness against the possible mismatch resulting when a system designed for a particular source model is applied to another source. The second application is to speech compression systems. A speech waveform coder is designed using a universal coding approach for a tree-encoder fake-process decoder system. Several distortion measures are utilized both for system design and quality measurement. The system has only moderate complexity, yet provides intelligible speech at a rate of less than one bit per sample.

ACKNOWLEDGEMENT

I would like to thank Professor Robert M. Gray for his superb guidance on research and many helpful comments on this dissertation.

I am also grateful to Professors Martin Morf and Joseph W. Goodman for reading this dissertation. Drs. Robert Mathews, Kenneth Cummins and Mr. Mat Herndon receive my thanks for their help in experiments. Thanks also go to my office-mates Drs. Luis A. Buzo and Robert J. Fontana for fruitful discussions.

It is a pleasant duty for me to acknowledge my appreciation of the Japan Society for the Promotion of Science, Murata Overseas Scholarship Foundation and the U.S. Air Force Office of Scientific Research for financial support.

Ms. Patricia Bastein's quick and excellent typing made possible the timely submission of this thesis.

Finally, I express my deep indebtedness to my wife Machiko for her love, encouragement and support.

TABLE OF CONTENTS

TABLE OF CONTENTS (CONT)

LIST OF ILLUSTRATIONS

LIST OF TABLES

Chapter 1

INTRODUCTION

A process distortion measure determines the degree of similarity between two stochastic processes with respect to an underlying alphabet distortion measure. The first explicit use of such a process distortion measure was by Ornstein [8] in his proof of the Isomorphism Theorem of Ergodic Theory. Subsequently, Moser, Phillips and Varadhan [45], and Weiss [4] introduced a process distance related to the Prokhorov distance between random variables [5]. For finite alphabets both distances are easily seen to be equivalent. An extension of Ornstein's distance to complete separable metric space alphabets such as the real line was made by Gray, Neuhoff and Shields [3]. It was there shown that this measure -- called the $\bar{\rho}$-distance -- could often be evaluated or bounded in terms of the power spectral densities of two processes. In such cases the $\bar{\rho}$-distance can be viewed as a spectral distortion measure, that is, a measure of the distortion between two spectra.

Spectral distortion measures have also been proposed for use in system design and quality evaluation in speech signal processing studies [24], [27], [28]. One goal of this dissertation is to develop the properties, interrelations and interpretations of these various distortion measures. Chapter 2 is devoted to a preliminary discussion of stochastic processes, process distortion measures, and topologies defined by such distortion measures. In Chapter 3, conditions are given for the topological equivalence of the $\bar{\rho}$-distance and the Prokhorov process-distance. In certain cases, these measures become equivalent to spectral distortion measures. Next, a "cookbook" of spectral distortion measures is developed giving various topological and coding equivalences, properties, approximations and computational aspects of the various measures.

The second goal of this work is the application of process distortion measures to signal processing. The first application is to the mismatch problem of linear predictors and interpolators. It is shown in Chapter 4 that variants of the process distortion measures can serve as indicators of the robustness of such systems against the possible mismatch of input sources. The second application is to the design and quality evaluation of speech compression systems. This work takes advantage of the fact that such spectral distortion measures appear to be subjectively meaningful measures of speech quality. In Chapter 5, one such compression system is developed and evaluated. The method adopted is a parallel tree search encoder with a fake-process decoder. The system does not use any on-line linear predictive coding (LPC) techniques, but such techniques are used off-line to find a finite codebook of autoregressive processes used to model the speech. Since the system does not use on-line LPC, the implementation complexity is medium. The reproduced voice is highly intelligible at 5-8 Kbps.

In the final chapter, another possible speech encoder is proposed based on the predictor mismatch idea of Chapter 4. Spectral distortion measures are crucial tools for the design.

The experimental work described here was carried out on the Information Systems Laboratory PDP-11 UNIX multi-user system.

Chapter 2

PRELIMINARIES

2.1 STOCHASTIC PROCESSES

For simplicity, we consider only discrete time stochastic processes since these are the most natural models for digital systems and since discrete time processes can be used to approximate continuous time processes.

Let A be a complete separable metric space and $\mathcal{A}$ the Borel σ-field of subsets of A. Let A_k and $\mathcal{A}_k$ denote replicas of A and $\mathcal{A}$ respectively. Define a sequence measurable space:

$$(A^{\infty},\mathcal{A}^{\infty}) = \bigtimes_{k=-\infty}^{\infty} (A_k,\mathcal{A}_k) \qquad .$$

In other words, A^{∞} is the set of all doubly infinite sequences $x = (\ldots,x_{-2},x_{-1},x_0,x_1,x_2,\ldots)$, where $x_k \in A_k$ for all k, and $\mathcal{A}^{\infty}$ is the σ-field generated by all cylinder sets of the form:

$$C_i^j = \ldots \times A_{i-2} \times A_{i-1} \times C_i \times C_{i+1} \times \ldots \times C_j \times A_{j+1} \times A_{j+2} \times \ldots ,$$

with $C_k \in \mathcal{A}_k$ and $i \leq j$. Let μ_{α} be a probability measure on $(A^{\infty},\mathcal{A}^{\infty})$. The triplet $[A^{\infty},\mathcal{A}^{\infty},\mu_{\alpha}]$ is a probability space and is called a sequence space. The sequence of random variables $X_n : A^{\infty} \to A$ defined by $X_n(x) = x_n$ forms a stochastic process which is denoted by $\alpha, [A,\mu_{\alpha}]$, $[A,\mu_{\alpha}X]$ or, simply, $\{X_n\}$.

Let $T: A^{\infty} \to A^{\infty}$ be the shift transformation which maps a sequence x into the sequence Tx where $(Tx)_n = x_{n+1}$. If μ_{α} is shift invariant, i.e., if $\mu_{\alpha}(TE) = \mu_{\alpha}(E)$ for all $E \in \mathcal{A}^{\infty}$, then $[A,\mu_{\alpha},X]$ is said to be stationary. If $E(X_{n+k}) = E(X_n)$ and $E(X_{n+k}X_{m+k}) = E(X_n X_m)$, all n,k,m hold, the process $[A,\mu_{\alpha},X]$ is ergodic if $\mu_{\alpha}(E)=0$ or 1 for all invariant sets, that is, sets for which TE=E. Stationarity and ergodicity make possible the estimation of moments using relative frequencies. If the second moment is finite, then the process is called a second order process.

In the prediction and interpolation problems, we shall use a mapping:

$$f: \quad A^{\infty} \rightarrow \hat{A} \tag{2.1.1}$$

to form a sequence of estimates $\hat{x}_k = f(T^k x)$, where $\hat{A}$ is another complete separable metric space (in most engineering problems, $A=\hat{A}=R$., the real line). In this case, the probability measure $\mu_{\hat{\alpha}}$ for $\hat{x} = (\ldots,\hat{x}_{-1},\hat{x}_0,\hat{x}_1,\ldots)$ is specified by

$$\mu_{\hat{\alpha}}(\hat{E}) = \mu_{\alpha}(f^{-1}(\hat{E})) \qquad , \quad \hat{E} \in \hat{\mathcal{a}}^{\infty}$$

Here $[\hat{A}^{\infty},\hat{\mathcal{a}}^{\infty},\mu_{\hat{\alpha}}]$ is the sequence space for $\hat{x}$.

A weakly stationary process $[A,\mu_{\alpha},X]$ with a spectral distribution $F(\lambda)$ which satisfies

$$\int_{-\pi}^{\pi} \log F'(\lambda)\, d\lambda > -\infty \tag{2.1.2}$$

is said to be nondeterministic, where $F'(\lambda)$ is the differentiation of the absolute continuous part of $F(\lambda)$. If $F(\lambda)$ itself is absolutely continuous, then (2.1.2) is reduced to

$$\int_{-\pi}^{\pi} \log f(\lambda)\, d\lambda > \infty \qquad , \tag{2.1.3}$$

where $f(\lambda)$ is the spectral density function of $\{X_n\}$. In this case the process is said to be purely nondeterministic. The spectral factorization theorem states that $\{X_n\}$ is purely nondeterministic if and only if the spectral density f has the following form [46]:

$$
\left.
\begin{aligned}
f(\lambda) &= |f^+(e^{i\lambda})|^2 \\
\text{where} \quad f^+(e^{i\lambda}) &= \sigma_f D(e^{i\lambda}) , \\
D(z) &= \sum_{k=0}^{\infty} d_k z^{-k} \neq 0 \quad , \quad |z| > 1 , \\
d_0 &= 1 , \\
\sum_{k=0}^{\infty} |d_k|^2 &< \infty , \\
\sigma_f^2 &= \exp\{(2\pi)^{-1} \int_{-\pi}^{\pi} \log f(\lambda) d\lambda\} ,
\end{aligned}
\right\} \tag{2.1.4}
$$

that is, $D(z)$ is a monic polynomial which is analytic in the open unit circle, and f^+ is a causal filter with a gain σ_f. Equation (2.1.4) is equivalent to a one-sided moving average (MA) representation [46]:

$$X_n = \sigma_f \sum_{k=0}^{\infty} b_k \xi_{n-k} , \tag{2.1.5}$$

where $\{\xi_n\}$ is white with $E \xi_n^2 = 1$. The white process $\{\sigma_f \xi_n\}$ is called the innovations process of $\{X_n\}$. Since $\sigma_f < \infty$ for weakly stationary processes, $f(\lambda)$ can also be expressed in factored form with

$$
\left.
\begin{aligned}
f^+(\lambda) &= \sigma_f / A(e^{i\lambda}) , \\
A(z) &= \sum_{k=0}^{\infty} a_k z^{-k} \neq 0 \quad , \quad |z| > 1 , \\
a_0 &= 1 , \\
\sum_{k=0}^{\infty} |a_k|^2 &< \infty .
\end{aligned}
\right\} \tag{2.1.6}
$$

with σ_f as before. This yields a one-sided autoregressive (AR) model of the form:

$$X_n = \sum_{k=1}^{\infty} a_k X_{n-k} + \sigma_f \xi_n \quad . \tag{2.1.7}$$

In the following chapters, the concept of L_p space of spectral densities will be used. The L_p norm is defined by

$$\|f\|_p = \{(2\pi)^{-1} \int_{-\pi}^{\pi} |f(\lambda)|^p d\lambda\}^{1/p} < \infty \quad . \tag{2.1.8}$$

Two spectral densities $f(\lambda)$ and $g(\lambda)$ are identified if $f = g$ almost everywhere. In this case, the class of spectral densities forms an L_p normed linear space with the norm (2.1.8).

2.2 AUTOREGRESSIVE MODELS AND LADDER FORMS

Given a second order stochastic process $\{X_n\}$ with a purely nondeterministic spectral density $f(\lambda)$, an m^{th} order autoregressive model $f_m(\lambda)$ of $f(\lambda)$ is defined by

$$f_m(\lambda) = \sigma_f(m)^2/|A_m(e^{i\lambda})|^2 \tag{2.2.1}$$

where

$$A_m(z) = \sum_{k=0}^{m} a_{mk}z^{-k} \quad , \quad a_{m0} = 1 \ , \tag{2.2.2}$$

and $\{a_{mk}\}_{k=1}^{m}$ and $\sigma_f(m)^2$ are obtained by the minimization of

$$E|\sum_{k=0}^{m} a_{mk}X_{n-k}|^2 = \frac{1}{2\pi}\int_{-\pi}^{\pi} f(\lambda)|A_m(e^{i\lambda})|^2 d\lambda \ .$$

The resulting minimum mean square error is given by $\sigma_f(m)^2$ [18]. The solution is given by Levinson's equation [20]:

$$\sigma_f(m)^2 = \frac{1}{2\pi}\int_{-\pi}^{\pi} f(\lambda)A_m(e^{-i\lambda})d\lambda = \sum_{k=0}^{m} a_{mk}r_f(k)$$

$$0 = \frac{1}{2\pi}\int_{-\pi}^{\pi} f(\lambda)A_m(e^{-i\lambda})e^{-ij\lambda}d\lambda = \sum_{k=0}^{m} a_{mk}r_f(k-j)$$

$$j=1,2,\ldots,m \tag{2.2.3}$$

where $\{r_f(k)\}$ is the autocorrelation sequence of $\{X_n\}$. Note that (2.2.3) is equivalent to the Yule-Walker [35] or correlation matching equations:

$$r_f(k) = \frac{1}{2\pi}\int_{-\pi}^{\pi} f(\lambda)e^{ik\lambda}d\lambda = \frac{1}{2\pi}\int_{-\pi}^{\pi} \frac{\sigma_f(m)^2}{|A_m(e^{i\lambda})|^2} e^{ik\lambda}d\lambda \tag{2.2.4}$$

$$k = 0,1,\ldots,m$$

One obtains from (2.2.3) that for any m^{th} order polynomial $C_m(z) = 1+c_1 z^{-1} + \dots + c_m z^{-m}$,

$$\frac{1}{2\pi}\int_{-\pi}^{\pi} f(\lambda)|C_m(e^{i\lambda})|^2 d\lambda = \frac{1}{2\pi}\int_{-\pi}^{\pi} \frac{\sigma_f(m)^2}{|A_m(e^{i\lambda})|^2} |C_m(e^{i\lambda})|^2 d\lambda \tag{2.2.5}$$

and

$$\sigma_f(m)^2 = \frac{|T_{m+1}(f)|}{|T_m(f)|}$$

where

$$T_m(f) = \left\{\frac{1}{2\pi}\int_{-\pi}^{\pi} f(\lambda)e^{i(k-j)\lambda}d\lambda;\ k,j=0,1,\dots m-1\right\}$$

$$= \{r_f(k-j);k,j=0,1,\dots,m-1\}$$

is the Toeplitz autocorrelation matrix for the spectral density f. It is well-known that $\sigma_m^2(f) \downarrow \sigma_\infty^2(f) = \sigma_f^2$ [18]. By solving equation (2.2.3) or (2.2.4), one obtains the autoregressive coefficients $\{a_{mk}\}_{k=1}^{m}$.

A relative of the autoregressive coefficients is the set of partial autocorrelation coefficients [34], [35]. They are also called reflection coefficients because of their meaning in the internal ladder structure [36].

Levinson's equation (2.2.3) can be rewritten as

$$\begin{bmatrix} r_f(0) & r_f(1) & \dots & r_f(m) \\ r_f(1) & r_f(0) & & \vdots \\ \vdots & & \ddots & \vdots \\ r_f(m) & \dots & \dots & r_f(0) \end{bmatrix} \begin{bmatrix} 1 \\ a_{m1} \\ \vdots \\ a_{mm} \end{bmatrix} = \begin{bmatrix} \sigma_f(m)^2 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \tag{2.2.6}$$

The reflection coefficients are then given by

$$\{k_i^a\} = \{a_{ii}\} \quad \text{for} \quad i = 0,1,\ldots,m \quad . \tag{2.2.7}$$

Note that $k_0^a = a_{00} = 1$. From (2.2.6), one obtains

$$\begin{bmatrix} r_f(0) & r_f(1) & \cdots & r_f(m) & r_f(m+1) \\ r_f(1) & r_f(0) & & \vdots & \vdots \\ \vdots & & \ddots & \vdots & \vdots \\ r_f(m) & & & \ddots & \\ r_f(m+1) & \cdots & & \cdots & r_f(0) \end{bmatrix} \begin{bmatrix} 1 \\ a_{m1} \\ \vdots \\ a_{mm} \\ 0 \end{bmatrix} = \begin{bmatrix} \sigma_f(m)^2 \\ \vdots \\ \vdots \\ 0 \\ -\beta_m \end{bmatrix} \tag{2.2.8}$$

and

$$\begin{bmatrix} r_f(0) & r_f(1) & \cdots & r_f(m) & r_f(m+1) \\ r_f(1) & r_f(0) & & \vdots & \vdots \\ \vdots & & \ddots & & \\ r_f(m) & & \cdots & & r_f(0) \\ r_f(m+1) & & \cdots & & r_f(0) \end{bmatrix} \begin{bmatrix} 0 \\ a_{mm} \\ \vdots \\ a_{m1} \\ 1 \end{bmatrix} = \begin{bmatrix} -\beta_m \\ 0 \\ \vdots \\ 0 \\ \sigma_f(m)^2 \end{bmatrix} \tag{2.2.9}$$

where

$$-\beta_m^a = \sum_{i=0}^{m} a_{mi}\, r_f(m+1-i) \quad . \tag{2.2.10}$$

The operation, k_m^a x (2.2.9) + (2.2.8), yields

$$\begin{bmatrix} r_f(0) & \cdots & r_f(m+1) \\ \vdots & & \vdots \\ r_f(m+1) & \cdots & r_f(0) \end{bmatrix} \begin{bmatrix} 1 \\ a_{m1} + k_m^a\, a_{mn} \\ \vdots \\ a_{mm} + k_m^a a_{m1} \\ k_m^a \end{bmatrix} = \begin{bmatrix} \sigma_f(m)^2 - k_m^a \beta_m^a \\ 0 \\ \vdots \\ 0 \\ -\beta_m^a + k_m^a \sigma_f(m)^2 \end{bmatrix} \tag{2.2.11}$$

From equations (2.2.6) and (2.2.11), the following set of equations is obtained:

$$\left.\begin{aligned}
\sigma_f(m+1)^2 &= \sigma_f(m)^2 - k_m^a \beta_m^a \\
\beta_m^a &= k_m^a \sigma_f(m)^2 \\
a_{m+1,0} &= a_{m0} = k_0^a = 1 \\
a_{m+1,i} &= a_{mi} + k_m^a a_{m,m+1-i} \\
a_{m+1,m+1} &= k_m^a a_{m0} = k_m^a \quad .
\end{aligned}\right\} \qquad (2.2.12)$$

Consider

$$\left.\begin{aligned}
A_m^a(z) &= \sum_{i=0}^{m} a_{mi} z^{-i} \quad , \quad a_{m0} = 1 \\
B_m^a(z) &= z^{-(m+1)} A_m^a(1/z)
\end{aligned}\right\} \qquad (2.2.13)$$

Then, by equation (2.2.12), one obtains

$$\left.\begin{aligned}
A_{m+1}^a(z) &= A_m^a(z) + k_m^a B_m^a(z) \\
zB_m^a(z) &= k_m^a A_{m-1}^a(z) + B_{m-1}^a(z) \quad .
\end{aligned}\right\} \qquad (2.2.14)$$

Equation (2.2.14) provides an intriguing realization of an autoregressive process depicted in Fig. 2.2.1. The structure is called the ladder from realization of an autoregressive process. Because of (2.2.12) and $\sigma_f(m)^2 \leqq \sigma_f(m+1)^2$,

$$-1 \leqq k_i^a \leqq 1 \quad , \; i=1,2,\ldots,m \quad . \qquad (2.2.15)$$

This property is important since $\{k_i^a\}_{i=1}^m$ can be quantized without violating stability as long as (2.2.15) holds [20]. This ladder form will be used in our applications on actual speech signals.

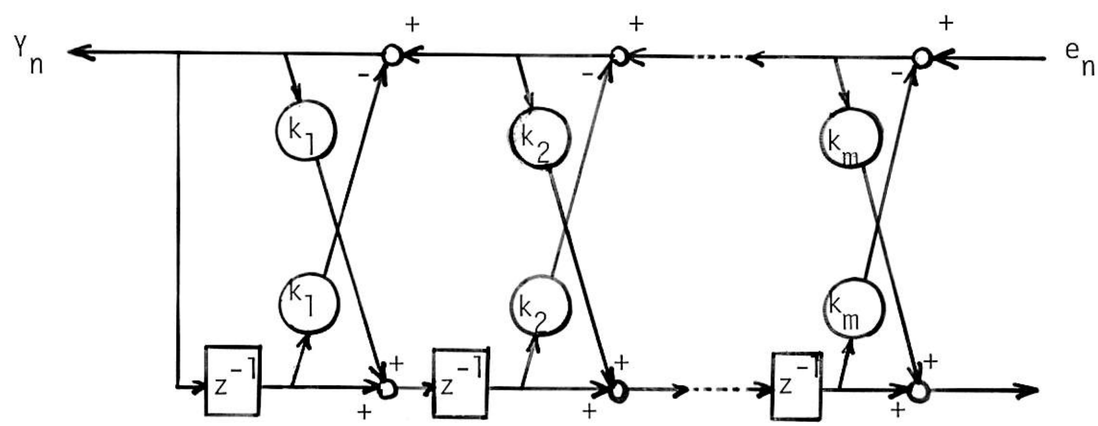


Fig. 2.2.1 Ladder realization of autoregressive process

2.3 $\bar{\rho}$-DISTANCE AND PROKHOROV PROCESS-DISTANCE

First, the definitions of "distortion measure," "distance" and "metric" are given. Let $\mathcal{P}_A$ be a class of stochastic processes with alphabet A. Then:

Definition 2.3.1: A distortion measure is an assignment $d(\alpha,\beta)$ to each $\mu_\alpha, \mu_\beta \in \mathcal{P}_A$ such that

(i) $d(\alpha,\beta) \geq 0$,

(ii) $d(\alpha,\alpha) = 0$.

Terminologies: "ρ-distance" and "Prokhorov process-distance" are used even if they are not distances but distortion measures. If the following properties are satisfied, "metric" will be used.

(iii) $d(\alpha,\beta) = 0$ implies $\alpha = \beta$,

(iv) $d(\alpha,\beta) = d(\beta,\alpha)$,

(v) $d(\alpha,\beta) + d(\beta,\gamma) \geq d(\alpha,\gamma)$ for any $\mu_\alpha, \mu_\beta, \mu_\gamma \in \mathcal{P}_A$.

Note that $\alpha = \beta$ means $\mu_\alpha(E) = \mu_\beta(E)$ for all events $E \in \mathcal{A}^\infty$.

The $\bar{\rho}$-distance [3] is a generalization of the Ornstein's $\bar{d}$-distance [8], and is defined as follows:

Definition 2.3.3: Let A be a complete separable metric space. Given two stationary stochastic processes $[A,\mu_\alpha,X]$ and $[A,\mu_\beta,Y]$, and a family of distortion measures $\rho_n(\cdot,\cdot)$ on $A^n \times A^n$, n=1,2,..., then the $\bar{\rho}$-distance $\bar{\rho}(\alpha,\beta)$ is defined by

$$\bar{\rho}(\alpha,\beta) = \sup_n \bar{\rho}_n(\alpha,\beta) ,$$

where

$$\bar{\rho}_n(\alpha,\beta) = \inf_{p^n \in \mathcal{P}_n} E_{p^n}[\rho_n(X^n,Y^n)]$$

and

$$E_{p^n}[\rho_n(X^n,Y^n)] = \int_{A^n x A^n} \rho_n(x^n,y^n)\,dp^n(x^n,y^n) \quad ,$$

and $\mathcal{P}_n$ is the class of all joint probability measures p^n on $(A^n x\, A^n,\ \mathcal{A}^n x\, \mathcal{A}^n)$ with marginals μ_α^n and μ_β^n, that is

$$\mathcal{P}_n = \{p^n : p^n(ExA^n) = \int_{ExA^n} dp^n(x^n,y^n) = \mu_\alpha^n(E) \quad ,$$

$$p^n(A^n xF) = \int_{A^n xF} dp^n(x^n,y^n) = \mu_\beta^n(F) \quad ,$$

$$\forall\ E,F \in \mathcal{A}^n\} \ .$$

The following property will be used in later sections.

<u>Theorem 2.3.4 [3]</u>: Given a single-letter distortion measure, i.e.,

$$\rho_n(x^n,y^n) = \frac{1}{n}\sum_{i=0}^{n-1} \rho(x_i,y_i) \quad ,$$

then,

$$\bar{\rho}(\alpha,\beta) = \inf_{p\in Q_{\alpha,\beta}} E_p[\rho(X_0,Y_0)] \quad ,$$

where $Q_{\alpha,\beta}$ is the collection of all stationary probability measures on the sequence space $(A^\infty xA^\infty,\ \mathcal{A}^\infty x\mathcal{A}^\infty)$ having marginals μ_α and μ_β, i.e.,

$$Q_{\alpha,\beta} = \{p : p(ExA^\infty) = \mu_\alpha(E),$$
$$p(A^\infty xF) = \mu_\beta(F),\ \forall\ E,\ F\in\mathcal{A}^\infty\} \ .$$

Some other important properties [3] for single letter distortion measures are:

(i) $\sup_n \bar{\rho}_n(\alpha,\beta) = \lim_{n \to \infty} \bar{\rho}_n(\alpha,\beta)$,

(ii) $\bar{\rho}$ is a metric if $\rho(\cdot,\cdot)$ is a metric.

Another process distortion measure is related to the Prokhorov distance [5]. There are several versions of the definition. We adopt the following one:

<u>Definition 2.3.5</u>: Under the same condition as the Definition 2.3.3, the Prokhorov process-distance is defined by

$$\bar{\Pi}_\rho(\alpha,\beta) = \sup_n \Pi_\rho^{(n)}(\alpha,\beta)$$

where

$$\Pi_\rho^{(n)}(\alpha,\beta) = \inf_{p^n \in \mathcal{P}_n} \inf_{\gamma_n} [\gamma_n : p^n\{(x^n,y^n):\rho_n(x^n,y^n) > \gamma_n\} \leq \gamma_n] ,$$

where $\mathcal{P}_n$ is defined as previously.

Note that another definition:

$$\bar{\Pi}_\rho^*(\alpha,\beta) = \sup_n \{\inf_{k \geq n} \Pi_\rho^{(k)}(\alpha,\beta)\}$$

$$= \varliminf_{n \to \infty} \Pi_\rho^{(n)}(\alpha,\beta)$$

is used in [4]. Clearly, $\bar{\Pi}_\rho^*(\alpha,\beta) \leq \bar{\Pi}_\rho(\alpha,\beta)$. In either case $0 \leq \bar{\Pi}_\rho(\alpha,\beta) \leq 1$.

The preceeding definition is related to the Prokhorov distance random variables. The original definition on random variables [5] is:

$$\Pi_\rho^{(n)}(\alpha,\beta) = \inf_\epsilon \{\epsilon : \mu_\alpha^{(n)}(F) \leq \mu_\beta^{(n)}(F^\epsilon) + \epsilon \;, \forall \text{ closed } F \subset A^n\}$$

$$\vee \inf_\epsilon \{\epsilon : \mu_\beta^{(n)}(F) \leq \mu_\alpha^{(n)}(F^\epsilon) + \epsilon, \forall \text{ closed } F \subset A^n\}$$

where

$\vee$ = the maximum and

$F^\epsilon = \{a^n \in A^n : \rho_n(a^n, F) < \epsilon\}$

and $\rho_n(\cdot,\cdot)$ is a metric for A^n. If A^n is a complete separable metric space with the metric $\rho_n(\cdot,\cdot)$, an important equivalent definition [6] is made. That is,

$$\Pi_\rho^{(n)}(\alpha,\beta) = \inf_\gamma [\gamma : \exists \text{ a joint measure } p^n \text{ on } A^n x A^n \text{ with marginals } \mu_\alpha^{(n)}, \mu_\beta^{(n)} \text{ such that } p^{(n)}\{(x^n,y^n) : \rho_n(x^n,y^n) > \gamma\} \leq \gamma]$$

If $\rho_n(\cdot,\cdot)$ is a metric then $\overline{\Pi}_\rho(\alpha,\beta)$ is also a metric. We emphasize that if $\rho_n(\cdot,\cdot)$ is not a metric, then the above two definitions need not coincide. The limiting Prokhorov distortion of Definition 2.3.5 is a limiting form of the second definition. The distance $\overline{\Pi}_\rho$ is not a true Prokhorov distance between processes since convergence with respect to $\overline{\Pi}_\rho$ is not equivalent to weak convergence.

Examples of alphabet distortion measures are as follows. When A is countable, the most popular choice is

$$d_n(x^n,y^n) = \frac{1}{n}\sum_{i=0}^{n-1} d_H(x_i,y_i)$$

$$= \frac{1}{n}\,(\text{number of } i \text{ for } x_i \neq y_i) \qquad (2.3.1)$$

This is the Hamming distance. The $\bar{\rho}$-distance in this case is reduced to the Ornstein's $\bar{d}$-distance [8]. There are many properties found for $\bar{d}$ [3], [8], [9]. A distance measure similar to Hamming's is the Lee distance [47] defined by

$$\ell_n(x^n, y^n) = \frac{1}{n}\sum_{i=0}^{n-1} \ell(x_i, y_i) = \frac{1}{n}\sum_{i=0}^{n-1} \min\{|x_i - y_i|, a - |x_i - y_i|\} \tag{2.3.2}$$

with $A = \{0,1,2,\ldots,a-1\}$. It will be seen that $\bar{\rho}$ and $\bar{\Pi}_\rho$ for d_H and ℓ are equivalent.

Another class of an alphabet distortion measure when $A = \{0,1,2,\ldots, a-1\}$ is

$$s_n(x^n, y^n) = 1 - \frac{k}{n} \tag{2.3.3}$$

where k is the maximum integer for which one can find subsequences $i_1 < i_2 < \cdots < i_k$ and $j_1 < j_2 < \cdots < j_k$ with $x_{i_r} = y_{i_r}$ for $1 \leq r \leq k$. This is Levenshtein's distance [4], [10]. Note that $s_n(x^n, y^n) \leq d_H(x^n, y^n)$. A generalization of $s_n(x^n, y^n)$ has been made in [11]. A weighted Levenshtein distance for $x^n, y^n \in A^n$, $A = \{0,1,\ldots,a-1\}$ is given by

$$t_n(x^n, y^n) = \frac{1}{n}\min(ku + mv + nw) \tag{2.3.4}$$

where u, v, w are nonnegative weightings, and

k = number of substitution

m = number of insertion

n = number of deletion

to make x^n match to y^n. Note that $u = v = w = 1$ is the case that Levenshtein adopted, $u = 1$, $v = \infty$, $w = \infty$ correspond to the Hamming distance, and $u = \infty$, $v = \infty$, $w = 1$ reflect the case of $s_n(\cdot,\cdot)$. The distortion measure $t_n(x^n,y^n)$ satisfies the requirements for a metric except for symmetry.

For $A=R$, $\rho_n(x^n,y^n) = \frac{1}{n}\sum_{i=0}^{n-1} |x_i-y_i|^2$, i.e., the square error alphabet distortion measure, we have the following:

<u>Proposition 2.3.1:</u> [3] Let $[R,\mu_\alpha,X]$ and $[R,\mu_\beta,Y]$ be zero-mean, second order stationary processes having spectral densities $f_\alpha(\lambda)$ and $f_\beta(\lambda)$. Then, for $\rho_n(x^n,y^n) = \frac{1}{n}\sum_{i=0}^{n-1} |x_i-y_i|^2$, we have that

$$\bar{\rho}(\alpha,\beta) \geqq \frac{1}{2\pi}\int_{-\pi}^{\pi} \left| \sqrt{f_\alpha(\lambda)} - \sqrt{f_\beta(\lambda)}\right|^2 d\lambda = \|\sqrt{f_\alpha} - \sqrt{f_\beta}\|_2^2 . \tag{2.3.5}$$

if μ_α and μ_β are Gaussian measures, then

$$\bar{\rho}(\alpha,\beta) = \| \sqrt{f_\alpha} - \sqrt{f_\beta} \|_2^2 . \tag{2.3.6}$$

2.4 DISTORTION MEASURES AND TOPOLOGY

If a metric d is assigned to a class of stochastic processes $\mathcal{P}_A$, then a pair $\langle \mathcal{P}_A/d, d \rangle$ forms a topological metric space. Here, $\mathcal{P}_A/d$ denotes the quotient space wherein $\mu_\alpha \in \mathcal{P}_A$ and $\mu_\beta \in \mathcal{P}_A$ with $d(\alpha,\beta) = 0$ are considered equal. In this section, we generalize several standard metric space ideas [2] to a distortion space where d need not be a metric.

Consider a class $\mathcal{O}_d$ of subsets (called open sets) of $\mathcal{P}_A/d$ such that

(i) $\mathcal{P}_A/d \in \mathcal{O}_d$, $\emptyset \in \mathcal{O}_d$,

(ii) $O_1 \in \mathcal{O}_d$ and $O_2 \in \mathcal{O}_d$ imply $O_1 \cap O_2 \in \mathcal{O}_d$,

(iii) $O_\nu \in \mathcal{O}_d$ implies $\bigcup_\nu O_\nu \in \mathcal{O}_d$,

(iv) for a given $\epsilon > 0$ and $\mu_\alpha \in \mathcal{P}_A/d$

$$B_\epsilon(\alpha) \triangleq \{ \mu_\beta : d(\alpha,\beta) < \epsilon,\ \mu_\beta \in \mathcal{P}_A/d\} \in \mathcal{O}_d .$$

Let τ_d be the minimal class of open sets which satisfies (i)-(iv). In other words, τ_d is the intersection of all classes $\mathcal{O}_d$ satisfying (i)-(iv).

<u>Definition 2.4.1</u>: The pair $\langle \mathcal{P}_A/d, \tau_d \rangle$ is called the topological distortion space induced by a distortion measure d.

The main purpose of this section is to develop criteria for the comparison of topologies induced by different distortion measures.

Definition 2.4.2: Let τ and σ be two topologies, i.e., two classes of open sets for $\wp_A$. τ is said to be coarser (or weaker) than σ if $\tau \subset \sigma$, that is, every member of τ is also a member of σ. An equivalent statement is that σ is finer (or stronger) than τ.

The comparison of topologies is discussed through the concept of convergence of elements. A generalization of the convergence in a topological space becomes possible using filters [2].

Definition 2.4.3: A nonempty family of sets $\mathfrak{F}$ is said to be a filter if

(i) $\emptyset \not\subset \mathfrak{F}$,

(ii) $A, B \in \mathfrak{F}$ imply $A \cap B \in \mathfrak{F}$ where $A, B \subset \wp_A$,

(iii) $A \in \mathfrak{F}$ and $A \subset B$ imply $B \in \mathfrak{F}$.

If all open sets containing μ_α belong to $\mathfrak{F}$, then $\mathfrak{F}$ is said to converge to μ_α. It is shown in standard topology texts (e.g., [2]), that convergence of the filter $\mathfrak{F} = \{U : U \supset A_m \text{ for some } m\}$ where $A_m = \{\mu_{\alpha_m}, \mu_{\alpha_{m+1}}, \ldots\}$ is an extension of the notion of pointwise convergence of $\{\mu_{\alpha_n}\}$ in $\langle \wp_A/d, \tau_d \rangle$. The following proposition can be shown [2]:

Proposition 2.4.4 [2] Let τ_1 and τ_2 be two topologies for $\wp_A$. Then, the following two conditions are equivalent:

(i) $\tau_1 \subset \tau_2$,

(ii) if a filter converges with respect to τ_2, that filter also converges with respect to τ_1.

An immediate corollary is obtained:

Corollary 2.4.5: Let d_1 and d_2 be two distortion measures. If $d_2(\alpha,\alpha_n) \to 0$ as $n \to \infty$ implies $d_1(\alpha,\alpha_n) \to 0$ for $\alpha, \alpha_n \in \wp_A/d_2$ then two topologies defined by these distortion measures have the relation

$$\tau_{d_1} \subset \tau_{d_2} \quad \text{for } \wp_A/d_2$$

We note that if $d_1(\alpha,\beta) \leq d_2(\alpha,\beta)$ for all $\mu_\alpha, \mu_\beta \in \wp_A/d_2$, then Corollary 2.4.5 is satisfied.

$\tau_{d_1} \subset \tau_{d_2}$ will be simply denoted by $d_2 \Rightarrow d_1$. The coarsest topology is the so-called trivial topology $\langle \wp_A, \tau_w \rangle$ with $\tau_w = \{\emptyset, \wp_A\}$. On the other hand, the discrete topology $\langle \wp_A, \tau_s \rangle$ with $\tau_s = \{$all subsets of $\wp_A\}$ is the finest one.

Another form of implication and equivalence of distortion measure is the following.

Definition 2.4.6: If for each $\mu_\beta, \mu_\alpha, \mu_\gamma \in \wp_A$, $d_1(\alpha,\beta) \leq d_1(\beta,\gamma)$ implies $d_2(\alpha,\beta) \leq d_2(\alpha,\gamma)$, then d_1 is said to be stronger in a coding sense, and $d_1 \Rrightarrow d_2$ is used for notation.

The name and application of this concept arises in the following coding or quantization problem. Given a distortion measure d_i and a finite codebook Γ of indices γ for which $\mu_\gamma \in \wp_A$, define the minimum distortion quantizer (or coder), $\hat{C}_i$, such that $\hat{C}_i(\alpha) = \beta$ if $d_i(\alpha,\beta) \leq d_i(\alpha,\gamma)$ for all $\gamma \in \Gamma$ with some tie-breaking rule. If $d_1 \Rrightarrow d_2$, then

$$d_2(\alpha,\hat{C}_2(\alpha)) = d_2(\alpha,\hat{C}_1(\alpha)) \qquad ,$$

that is, the closest reproduction symbol $\hat{C}_1(\alpha)$ under d_1 is also the closest reproduction under d_2. If $d_1 \Leftrightarrow d_2$ then, the tie-breaking rule can be chosen so that $\hat{C}_1(\alpha) = \hat{C}_2(\alpha)$ for all α and hence the codes yield identical outputs for a given input.

Chapter 3

PROCESS DISTORTION MEASURES

3.1 CONDITIONS FOR TOPOLOGICAL EQUIVALENCE OF $\bar{\rho}$ AND $\bar{\Pi}_\rho$

It is useful to know when topologies defined by the $\bar{\rho}$ and $\bar{\Pi}_\rho$ distances, say $\tau_{\bar{\rho}}$ and $\tau_{\bar{\Pi}_\rho}$, coincide since each possesses different properties and implications. The following propositions lead to sufficient conditions for $\tau_{\bar{\rho}} = \tau_{\bar{\Pi}_\rho}$.

Proposition 3.1.1: [38] For $\mu_\alpha, \mu_\beta \in \mathcal{P}_A$, a class of stationary processes, we have

$$\{\Pi_\rho^{(n)}(\alpha,\beta)\}^2 \leqq \bar{\rho}_n(\alpha,\beta) \quad , \tag{3.1.1}$$

$$\{\bar{\Pi}_\rho(\alpha,\beta)\}^2 \leqq \bar{\rho}(\alpha,\beta) \quad . \tag{3.1.2}$$

Proposition 3.1.2: [38] If ρ_n is bounded, that is,

$$\rho_n^* \stackrel{\Delta}{=} \sup_{x^n, y^n \in A^n} \rho_n(x^n,y^n) < \infty \quad , \tag{3.1.3}$$

then

$$\bar{\rho}_n(\alpha,\beta) \leqq (1+\rho_n^*)\Pi_\rho^{(n)}(\alpha,\beta) \quad . \tag{3.1.4}$$

If

$$\rho^* = \sup_n \rho_n^* < \infty \quad , \tag{3.1.5}$$

i.e., ρ_n is uniformly bounded, then

$$\bar{\rho}(\alpha,\beta) \leqq (1+\rho^*)\bar{\Pi}_\rho(\alpha,\beta) \quad . \tag{3.1.6}$$

In the next proposition, boundedness is not required, but the form of the alphabet distortion ρ_n is restricted in another sense.

This proposition is a generalization of a result given in [38] in the sense that fewer restrictions are placed on the alphabet distortion ρ_n. In [52], a similar bound for $\bar{\rho}_n$ is given. That bound blows up to infinity as $n \to \infty$, however, and hence cannot be used for the process distance.

<u>Proposition 3.1.3</u>: Let $\rho(x,y) = d(x,y)^q$ where $d(\cdot,\cdot)$ is a metric and $0 < q < \infty$. Let $\rho_n(x^n,y^n) = \frac{1}{n}\sum_{i=0}^{n-1} \rho(x_i,y_i)$. If there exist reference letters $a^*, b^* \in A$ and a constant $r > 1$ such that

$$\max[E_{\mu_\alpha}\{\rho(X_0,a^*)\}^r,\ E_{\mu_\beta}\{\rho(b^*,Y_0)\}^r] = \rho^* < \infty \tag{3.1.7}$$

then

$$\bar{\rho}_n(\alpha,\beta) \leq \Pi_\rho^{(n)}(\alpha,\beta) + K\{\Pi_\rho^{(n)}(\alpha,\beta)\}^{r-1/r} \tag{3.1.8}$$

where K is a nonnegative finite number independent of n.

Moreover,

$$\bar{\rho}(\alpha,\beta) \leq \bar{\Pi}_\rho(\alpha,\beta) + K\{\bar{\Pi}_\rho(\alpha,\beta)\}^{r-1/r} . \tag{3.1.9}$$

<u>Proof</u>.

The following three inequalities are used frequently in the proof:

(i) $\left(\frac{1}{n}\sum_{i=0}^{n-1} |a_i|\right)^r \leq \frac{1}{n}\sum_{i=0}^{n-1} |a_i|^r$, for $1 \leq r < \infty$,

a variant of the Hölder's inequality.

(ii) $|a+b|^q \leq C_q(|a|^q + |b|^q)$, for $0 < q < \infty$ (3.1.10)

where $C_q = \max(1, 2^{q-1})$. Note that $q < 1$ is allowed [7].

(iii) $\rho_n(x^n,y^n) = \frac{1}{n}\sum_{i=0}^{n-1} \rho(x_i,y_i) \leq C_q\{\rho_n(x^n,z^n) + \rho_n(z^n,y^n)\}$.

The inequality is obtained from the triangle inequality on $d(\cdot,\cdot)$ and (ii).

(iv) $$\{\rho_n(x^n,y^n\}^r = \{\frac{1}{n}\sum_{i=0}^{n-1}\rho(x_i,y_i)\}^r \leqq \frac{1}{n}\sum_{i=0}^{n-1}\rho(x_i,y_i)^r$$

This is the same as (i).

Observe that

$$E_{p^n}\rho_n(X^n,Y^n) \leqq \Pi_\rho^{(n)}(\alpha,\beta) + \int_{G_n}\rho_n(x^n,y^n)\,dp^n(x^n,y^n)$$

where

$$G_n = \{(x^n,y^n):\rho_n(x^n,y^n) \geqq \Pi_\rho^{(n)}(\alpha,\beta)\}$$

One obtains

$$\int_{G_n}\rho_n(x^n,y^n)\,dp^n(x^n,y^n) = E\,p_n\{\rho_n(X^n,Y^n)\cdot 1_{G_n}\}$$

$$\leqq \{E_{p^n}\rho_n(X^n,Y^n)^r\}^{1/r}(E_{p^n}1_{G_n}{}^{r/r-1})^{r-1/r} \qquad \text{(Hölder's inequality)}$$

$$\leqq (E_{p^n}[C_q\{\rho_n(X^n,a^{*n}) + \rho_n(a^{*n},Y^n)\}]^r)^{1/r}(E_{p^n}1_{G_n})^{r-1/r} \qquad \text{(by (iii))}$$

$$\leqq [E_{p^n}C_q^rC_r\{\rho_n(X^n,a^{*n})^r+\rho_n(a^{*n},Y^n)^r\}]^{1/r}(E_{p^n}1_{G_n})^{r-1/r} \qquad \text{(by (ii))}$$

$$\leqq [E_{p^n}C_q^rC_r^2\{\rho_n(X^n,a^{*n})^r+\rho_n(a^{*n},b^{*n})+\rho_n(b^{*n},Y^n)^r\}]^{1/r}(E_{p^n}1_{G_n}) \qquad \text{(by (iii))}$$

$$\leqq [E_{p^n}C_q^rC_r^2\{\frac{1}{n}\sum_{i=0}^{n-1}\rho(X_i,a^*)^r + \frac{1}{n}\sum_{i=0}^{n-1}\rho(a^{*n},b^{*n})^r$$

$$+\frac{1}{n}\sum_{i=0}^{n-1}\rho(b^{*n},Y^n)^r\}]^{1/r}(E_{p^n}1_{G_n})^{r-1/r} \qquad \text{(by (iv))}$$

$$= [C_q^r C_r^2 E_p\{\rho(X_0,a^*)^r+\rho(a^*,b^*)^r+\rho(b^*,Y_0)^r\}]^{1/r}(E_{p^n} 1_{G_n})^{r-1/r}$$

(by stationarity)

$$\leq C_q C_r^{2/r}\{2\rho^*+\rho(a^*,b^*)^r\}^{1/r}(E_{p^n} 1_{G_n})^{r-1/r}$$

$$\stackrel{\Delta}{=} K(E_{p^n} 1_{G_n})^{r-1/r}$$

$$\leq K\{\Pi_\rho^{(n)}(\alpha,\beta)\}^{r-1/r}$$

where

$$K = C_q C_r^{2/r}\{2\rho^* + \rho(a^*,b^*)^r\}^{1/r} \in [0,\infty)$$

is independent of n. Now

$$\bar{\rho}_n(\alpha,\beta) \leq \Pi_\rho^{(n)}(\alpha,\beta) + K\{\Pi_\rho^{(n)}(\alpha,\beta)\}^{r-1/r}$$

was obtained. Moreover,

$$\bar{\rho}(\alpha,\beta) \leq \bar{\Pi}_\rho(\alpha,\beta) + K\{\bar{\Pi}_\rho(\alpha,\beta)\}^{r-1/r}$$

because K is independent of n. □

If either Proposition 3.1.2 or 3.1.3 hold, we have $\tau_{\bar{\rho}} = \tau_{\bar{\Pi}_\rho}$ from Corollary 2.4.5 and Proposition 3.1.1. In Proposition 3.1.2, a nonsymmetric alphabet distortion is included provided it is uniformly bounded. In Proposition 3.1.2, the boundedness is replaced by a weaker condition: the existence of reference letters. The alphabet distortion must be symmetric, however, since $\rho(\cdot,\cdot)$ is a power of a metric $d(\cdot,\cdot)$. We note that $\rho(x,y)$ can be $\sqrt{|x-y|}$ if equation (3.1.7) holds for an appropriate r.

As an immediate result of Propositions 2.3.1, 3.1.1 and 3.1.3, one obtains:

Proposition 3.1.4: Let $\mathcal{G}_R \subset \mathcal{P}_R$ be a class of stationary Gaussian processes whose variances are uniformly bounded. Then the following equivalence holds:

$$\bar{\Pi}_\rho(\alpha,\beta) \Leftrightarrow \bar{\rho}(\alpha,\beta) = \left\|\sqrt{f_\alpha(\lambda)} - \sqrt{f_\beta(\lambda)}\right\|_2^2$$

where $\rho_n(x^n,y^n) = \frac{1}{n}\sum_{i=0}^{n-1}|x_i-y_i|^2$ and $x_i,y_i \in R$.

Proof.

Choose $d(x,y) = |x-y|$, $q = 2$, $r = 2$ and $a^* = b^* = 0$ in Proposition 3.1.3. □

Other properties on the topological equivalence of process distortion measures are:

(i) $\bar{d}(\alpha,\beta) \Leftrightarrow \bar{\ell}(\alpha,\beta) \Leftrightarrow \bar{\Pi}_\ell(\alpha,\beta) \Leftrightarrow \bar{\Pi}_d(\alpha,\beta)$ because

$d_H(x,y) \leq \ell(x,y) \leq \lfloor a/2 \rfloor\, d_H(x,y)$, and equations (3.1.2) and (3.1.6) hold.

(ii) $\bar{s}(\alpha,\beta) \leq \bar{d}(\alpha,\beta)$ which follows from $s_n(x^n,y^n) \leq d_n(x^n,y^n)$.

Therefore $\bar{d} \Rightarrow \bar{s}$.

(iii) $\bar{s}(\alpha,\beta) \Leftrightarrow \bar{\Pi}_s(\alpha,\beta)$, $\bar{t}(\alpha,\beta) \Leftrightarrow \bar{\Pi}_t(\alpha,\beta)$

because

$$\{\bar{\Pi}_s(\alpha,\beta)\}^2 \leq \bar{s}(\alpha,\beta) \leq 2\bar{\Pi}_s(\alpha,\beta)$$

and

$$\{\bar{\Pi}_t(\alpha,\beta)\}^2 \leq \bar{t}(\alpha,\beta) \leq \{1+\max(u,v,w)\}\bar{\Pi}_t(\alpha,\beta)$$

hold by virtue of Propositions 3.1.1 and 3.1.2.

3.2 SPECTRAL DISTORTION MEASURES

3.2.1 General Comments

In the previous section it was shown in the Gaussian case that $\bar{\rho}$ and $\bar{\pi}_\rho$ are equivalent to a distortion measure on spectral densities.

Such a spectral distortion measure is a kind of a process distortion measure which depends on the processes only through their second order properties. Spectral properties are often all that is needed for signal processing in practical cases.

We consider a class of spectral densities:

$$\mathcal{N} = \{f : f(\lambda) \geqq 0,\ \|f\|_1 < \infty,\ \|1/f\|_1 < \infty\} \quad . \tag{3.2.1}$$

$1/f \in L_1$ is assumed to ensure finite average power in inverse filters. Note that for $f \in \mathcal{N}$, f and $1/f$ have both causal moving average and autoregressive models since from Jensen's inequality [30]

$$\int_{-\pi}^{\pi} \log \frac{1}{f(\lambda)}\, d\lambda \leqq \log \{\int_{-\pi}^{\pi} \frac{1}{f(\lambda)}\, d\lambda\} < \infty \quad ,$$

$$\int_{-\pi}^{\pi} \log f(\lambda) d\lambda \leqq \log\{ \int_{-\pi}^{\pi} f(\lambda) d\lambda\} < \infty \quad .$$

Let $[R,\mu_\alpha,X]$ and $[R,\mu_\beta,Y]$ be two stochastic processes having spectral densities $f(\lambda)$ and $g(\lambda)$ which belong to the class $\mathcal{N}$. The autoregressive representations are

$$X_n = - \sum_{k=1}^{\infty} a_k X_{n-k} + \sigma_f \xi_n^\alpha \quad , \tag{3.2.2}$$

$$Y_n = - \sum_{k=1}^{\infty} b_k X_{n-k} + \sigma_g \xi_n^\beta \quad ,$$

where $\{\xi_n^{\alpha}\}$ and $\{\xi_n^{\beta}\}$ are independent and identically distributed processes with zero mean and unit variance. Their spectral densities are

$$\left.\begin{aligned} f(\lambda) &= |f^+(e^{i\lambda})|^2 \\ \text{with} \quad f^+(e^{i\lambda}) &= \sigma_f/A(e^{i\lambda}) \\ \text{where} \quad A(e^{i\lambda}) &= \sum_{k=0}^{\infty} a_k z^{-k} \quad , \quad a_0 = 1 \quad , \end{aligned}\right\} \qquad (3.2.4)$$

and

$$\left.\begin{aligned} g(\lambda) &= |g^+(e^{i\lambda})|^2 \\ \text{with} \quad g^+(e^{i\lambda}) &= \sigma_g/B(e^{i\lambda}) \\ \text{where} \quad B(e^{i\lambda}) &= \sum_{k=0}^{\infty} b_k z^{-k} \quad , \quad b_0 = 1 \quad . \end{aligned}\right\} \qquad (3.2.5)$$

A common class of distortion measures on $\eta \cap L_p$ is that of difference distortion measures of the form

$$d(f,g) = \|\varphi(f-g)\|_p \quad ,$$

where $\varphi: (-\infty,\infty) \to [0,\infty)$, $\varphi(0) = 0$. Usually $\varphi(|x|)$ is assumed to be nondecreasing or convex $\cup$. An alternate class has the form

$$d'(f,g) = \varphi(\|f-g\|_p)$$

and is called a norm-difference distortion measure.

Most spectral distortion measures appearing in speech signal processing literature, however, are not of the above form. Instead they are ratio distortion measures having the form

$$d_\varphi(f,g) = \|\varphi(f/g)\|_p$$

with $\varphi:[0,\infty) \to [0,\infty)$, $\varphi(1) = 0$. The subscript φ will often be replaced by a mnemonic. We can also have a norm-ratio distortion measure of the form

$$d'_\varphi(f,g) = \varphi(\|f/g\|_p) \quad .$$

Variations on the ratio distortion measures that occur in speech processing are the gain-normalized and gain-optimized distortion measures. The gain-normalized distortion measure is given by

$$d_{n\varphi}(f,g) = d_\varphi\left(\frac{f}{\sigma_f^2}, \frac{g}{\sigma_g^2}\right) .$$

Several important spectral distortion measures may be expressed as a weighted sum of gain distortion and gain-normalized distortion measure. This property is computationally advantageous. A gain-optimized distortion measure has the form

$$d_\varphi^o(f,g) = \inf_{\sigma^2>0} d_\varphi\left(f, \sigma^2 \frac{g}{\sigma_g^2}\right) = \inf_{\sigma^2>0} d_\varphi\left(\frac{f}{\sigma^2}, \frac{g}{\sigma_g^2}\right) \quad .$$

If the infimum is a minimum, the optimum σ^2 is denoted by σ_o^2 and called the optimum reproduction gain. The gain optimization is done for one of two reasons. First, we may ignore the original gain of a reproduction symbol and replace it by a gain chosen to minimize the given distortion measure. Second, by removing dependence of d_φ on a reproduction parameter such as the gain, it allows us the freedom of using a different distortion measure on the gains. Since $d_\varphi(f,g) \geq d_\varphi^o(f,g)$, Corollary 2.4.5 implies

$$d_{\varphi}(f,g) \Rightarrow d_{\varphi}^{o}(f,g) \qquad .$$

It is often useful to symmetrize distortion measures so that the distortion between f and g is the same as that between g and f. Given two spectral ratio distortion measures, $d_1(f,g) = \|\varphi_1(f/g)\|_p$, $d_2(f,g) = \|\varphi_2(f/g)\|_p$, we can form a new distortion measure $d^{(q)}(f,g)$ by

$$d^{(q)}(f,g) = \{\|\varphi_1(f/g)\|_p^q + \|\varphi_2(f/g)\|_p^q\}^{1/q}$$

for $q \geqq 1$. Alternatively, we may form instead

$$d^{*}(f,g) = \frac{1}{2}\|\varphi_1(f,g) + \varphi_2(f,g)\|_p \qquad ,$$

where the 1/2 is used for convenience. Note that one obtains

$$d^{*} \iff d^{(q)} \iff d^{(1)} \quad \text{for } q \geqq 1 \qquad (3.2.6)$$

by use of

$$a^q + b^q \leqq (a+b)^q \leqq 2^{q-1}(a^q+b^q) \qquad (3.2.7)$$

for $a,\ b \geqq 0$, $q \geqq 1$, where the righthand side is a special case of eq. (3.1.10). Equation (3.2.6) states that $d^{(q)}(f,g)$ and $d^{*}(f,q)$ are topologically equivalent symmetrizations.

3.2.2 Examples of Spectral Distortion Measures

In this section several examples of spectral distortion measures are discussed.

1) $\bar{\rho}$-Spectral Distortion Measure

$$d_{\rho}(f,g) = \frac{1}{2\pi}\int_{-\pi}^{\pi} \left| \sqrt{f(\lambda)} - \sqrt{g(\lambda)}\right|^2 d\lambda = \|\sqrt{f} - \sqrt{g}\|_2^2 \; . \quad (3.2.8)$$

This is a $\bar{\rho}$-distance with a square-metric alphabet distortion under the Gaussian assumption.

2) Itakura-Saito Distortion Measure

$$d_{IS}(f,g) = \|f/g - 1 - \log(f/g)\|_1 \quad (3.2.9)$$

This is a ratio distortion measure with $\varphi(x) = x-1-\ell n\, x \geqq 0$. This distance was introduced by Itakura and Saito [24], and has the property that for fixed f and a class

$$\eta_m \triangleq \{f_m \in \eta,\ f_m(\lambda) = \sigma_f^2(m)/\left|\sum_{k=0}^{m} a_{mk} e^{-k\lambda}\right|^2,\ a_{m0}=1\} \quad ,$$

the spectrum $f_m \in \eta_m$ which minimizes $d_{IS}(f,f_m)$ is given by equation (2.2.3). Itakura and Saito also showed that if the underlying process was assumed Gaussian, then, the maximum log-likelihood approach to waveforms [22] led to the minimization of d_{IS}. In Section 4.1, it will be seen that $2d_{IS}(f,g)$ is the I-divergence rate of two stationary Gaussian processes. That is, for two Gaussian processes $[R,\mu_\alpha]$ and $[R,\mu_\beta]$,

$$d_{IS}(f,g) = \lim_{N\to\infty} \frac{2}{N} I_N(\alpha|\beta)$$

where

$$I_N(\alpha|\beta) = \int dz^N p_\alpha(z^N) \log \frac{p_\alpha(z^N)}{p_\beta(z^N)} \quad .$$

A detailed explanation is relegated to that section since this problem is related to predictor mismatch. We note that

$$d_{IS}(f,g) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{f}{g} \, d\lambda - 1 - \log \frac{\sigma_f^2}{\sigma_g^2}$$

$$= r_{f/g}(0) - 1 - \log \frac{\sigma_f^2}{\sigma_g^2} \qquad (3.2.10)$$

where $r_{f/g}(0)$ is the corresponding integral or the autocorrelation of the process with the spectral density $f(\lambda)/g(\lambda)$.

3) Itakura Distortion Measure

$$d_I(f,g) = \log \frac{1}{2\pi} \left\{ \int_{-\pi}^{\pi} \frac{f(\lambda)/\sigma_f^2}{g(\lambda)/\sigma_g^2} \, d\lambda \right\}$$

$$= \log \left\{ \frac{\sigma_g^2}{\sigma_f^2} \, r_{f/g}(0) \right\} \quad . \qquad (3.2.11)$$

This distortion was introduced by Itakura [25] as the gain optimized Itakura-Saito distortion, that is,

$$d_I(f,g) = d_{IS}^{o}(f,g) \quad .$$

In this case, the optimal gain is $\sigma_o^2 = \sigma_g^2 \, r_{f/g}(0)$. The Itakura distortion is a gain-normalized norm-ratio distortion measure with $\varphi(x) = \ell n \, x$ for $x \geq 1$ since

$$d_I(f,g) = \log \left(\left\| \frac{f/\sigma_f^2}{g/\sigma_g^2} \right\|_1 \right) \quad .$$

4) Model Distortion Measures

There are two classes of model distortion measures: noncausal model distortion measures and causal model distortion measures. The noncausal model distortion measure is defined by

$$
\begin{aligned}
d_{nm}(f,g)^2 &= \| \sqrt{f/g} - 1\|_2^2 \\
&= \| \, |f^+/g^+| - 1\|_2^2 \\
&= r_{f/g}(0) - 2 \cdot \frac{1}{2\pi} \int_{-\pi}^{\pi} \sqrt{\frac{f(\lambda)}{g(\lambda)}} \, d\lambda + 1 \, . \qquad (3.2.12)
\end{aligned}
$$

This is a ratio measure with $\varphi(x) = |\sqrt{x} - 1|$. Since a new input to the associated digital filter $|f^+(e^{i\lambda})/g^+(e^{i\lambda})|$ affects the past outputs, the word "noncausal" is used. On the other hand, the causal model distortion measure is defined by

$$
\begin{aligned}
d_{cm}(f,g)^2 &= \|f^+/g^+ - 1\|_2^2 \\
&= \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{f(\lambda)}{g(\lambda)} \, d\lambda - 2\mathrm{Re} \left\{ \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{f^+(e^{i\lambda})}{g^+(e^{i\lambda})} \, d\lambda \right\} + 1 \\
&= r_{f/g}(0) - 2 \frac{\sigma_f}{\sigma_g} + 1 \, . \qquad (3.2.13)
\end{aligned}
$$

In this case, an input to the digital filter $f^+(e^{i\lambda})/g^+(e^{i\lambda})$ does not affect past outputs.

We also have a gain-normalized model distortion measure:

$$d_{ncm}(f,g)^2 = d_{cm}(f/\sigma_f^2,\ g/\sigma_g^2)^2$$

$$= \left\| \frac{f^+/\sigma_f}{g^+/\sigma_g} - 1 \right\|_2^2$$

$$= \frac{\sigma_g^2}{\sigma_f^2}\, r_{f/g}(0) - 1 \quad . \tag{3.2.14}$$

From (3.2.11) and (3.2.14), one obtains

$$d_{ncm}(f,g)^2 = e^{d_I(f,g)} - 1 \quad . \tag{3.2.15}$$

The gain-normalized causal model distortion measure was introduced by Itakura [25] as an approximation to the Itakura distortion measure for small values:

$$d_{ncm}(f,g)^2 \simeq d_I(f,g) \quad .$$

The distortion measure d_{ncm} has the property of the Itakura-Saito distance that for fixed f and the class η_m, a minimum distortion $g \in \eta_m$ will have the form $f_m(\lambda) = \sigma^2/|A_m(e^{i\lambda})|^2$, where A_m is the same as that of the d_{IS} minimization, but σ^2 is arbitrary. Chaffee [26] used the gain-normalized causal model distortion measure in his rate-distortion approach to select speech models from a finite codebook of monic autoregressive filters. Chaffee used an alternate criterion on the gain. Note that one can also consider a gain optimized causal model distortion:

$$d^o_{cm}(f,g)^2 = 1 - \frac{\sigma_f^2/\sigma_g^2}{r_{f/g}(0)} = 1-e^{-d_I(f,g)} \tag{3.2.16}$$

In this case, $\sigma_o = \sigma_g^2 r_{f/g}(0)/\sigma_f$.

5) L_1 Spectral Ratio Distortion Measure

This is a spectral ratio measure with $\varphi(x) = |x-1|$.

$$d_1(f,g) = \frac{1}{2\pi} \int_{-\pi}^{\pi} |f(\lambda)/g(\lambda) - 1| \, d\lambda$$

$$= r_{f/g}(0) - \frac{1}{\pi} \int_{-\pi}^{\pi} \min\left\{ \frac{f(\lambda)}{g(\lambda)}, 1 \right\} d\lambda + 1 \quad . \qquad (3.2.17)$$

All of the preceeding distortion measures are nonsymmetric except for $d_\rho(f,g)$. We next consider several symmetric distortion measures.

6) Log Spectral Deviation

$$d_{\log}(f,g) = \|\log(f/g)\|_p \qquad (3.2.18)$$

This is a spectral ratio measure with $\varphi(x) = |\ell n \, x|$ and is one of the most frequently proposed distortion measures for speech [27], [28], [29]. Common choices for p are p = 1,2 and ∞. Note that $d_{\log}(f,h) \leq d_{\log}(f,g) + d_{\log}(g,h)$ so that $d_{\log}(f,g)$ is a metric distortion measure on $\mathcal{M}$.

The remaining measures are all symmetrized versions of measures 2)-5). The simplest and most useful method of the symmetrization in Section 3.2.1 is chosen.

7) <u>Cosh Distortion Measure</u>

$$d_{cosh}(f,g) = d^*_{IS}(f,g)$$
$$= \frac{1}{2} \| \sqrt{f/g} - \sqrt{g/f} \|_2^2$$
$$= \frac{1}{2} \| r_{f/g}(0) - 2 + r_{g/f}(0) \|_1$$
$$= \frac{1}{2} d_{IS}^{(1)}(f,g) \quad . \qquad (3.2.19)$$

This is simply an algebraic mean of $d_{IS}(f,g)$, and was introduced by Gray and Markel [27]. The name "cosh" comes from the relation:

$$d_{cosh}(f,g) = \| \cosh\{\log(f/g)\} - 1 \|_1 \quad .$$

This measure has some interesting interpretations. In statistics, detection theory and information theory [15], [16], [17], one often uses the symmetrized J-divergence defined as

$$J_N(\alpha|\beta) = I_N(\alpha|\beta) + I_N(\beta|\alpha) \quad .$$

Then, we have under the Gaussian assumption [19], [21]

$$\lim_{N\to\infty} \frac{1}{N} J_N(\alpha|\beta) = \frac{1}{2} \{ d_{IS}(f,g) + d_{IS}(g,f) \}$$
$$= d_{cosh}(f,g) \quad .$$

The second interpretation comes from the $\bar{\rho}$-distance. Equation (2.3.6) of Proposition 2.3.1 gives

$$d_{cosh}(f,g) = \frac{1}{2} \bar{\rho}(\alpha/\beta, \beta/\alpha)$$

if μ_α and μ_β are Gaussian measures. Here $\mu_{\alpha/\beta}$ and $\mu_{\beta/\alpha}$ denote the measures of the output processes of mismatched inverse filters. The mismatch of inverse filters is explained in detail in Chapter 4. Furthermore, if the variances of the Gaussian output processes are uniformly bounded, this is equivalent to $\bar{\pi}_\rho(\alpha/\beta,\ \beta/\alpha)$ by virtue of Proposition 3.1.4. We emphasize that even if the processes are not Gaussian,

$$\frac{1}{2}\,\bar{\rho}(\alpha/\beta,\ \beta/\alpha) \geqq d_{\cosh}(f,g)$$

holds according to Proposition 2.3.1.

8) Gain-Optimized cosh Measure

From

$$d_{\cosh}(f,\sigma^2 g/\sigma_g^2) = \frac{1}{2}\{(\sigma_g^2/\sigma^2)r_{f/g}(0)-2+(\sigma^2/\sigma_g^2)r_{g/f}(0)\} \quad ,$$

one obtains the minimizing gain

$$\sigma_o^2 = \{r_{f/g}(0)r_{g/f}(0)\}^{1/2}$$

yielding

$$\begin{aligned} d^o_{\cosh}(f,g) &= d_{\cosh}(f,\sigma_o^2 g/\sigma_g^2) \\ &= 2\{r_{f/g}(0)^{1/2}r_{g/f}(0)^{1/2} - 1\} \quad . \qquad (3.2.20) \end{aligned}$$

9) Symmetrized Itakura Distortion

$$\begin{aligned} d_I^*(f,g) &= \frac{1}{2}\{d_I(f,g) + d_I(g,f)\} \\ &= \frac{1}{2}\log\{r_{f/g}(0)r_{g/f}(0)\} \quad . \qquad (3.2.21) \end{aligned}$$

Note that

$$d^{o}_{cosh}(f,g) = 2\left\{e^{d^*_I(f,g)} -1\right\} \quad . \qquad (3.2.22)$$

10) <u>Symmetrized Model Distortion</u>

$$d^{(q)}_{cm}(f,g) = \left\{\left\| \frac{f^+}{g^+} - 1\right\|^q_2 + \left\| \frac{g^+}{f^+} - 1 \right\|^q_2\right\}^{1/q} \quad , \qquad (3.2.23)$$

$$d^{(q)}_{ncm}(f,g) = d^{(q)}_{cm}(f/\sigma^2_f,\ g/\sigma^2_g) \quad , \qquad (3.2.24)$$

and

$$d^{(q)}_{nm}(f,g) = \{ \| \sqrt{f/g} -1 \|^q_2 + \| \sqrt{g/f} - 1 \|^q_2 \}^{1/q} . \qquad (3.2.25)$$

Note that if $q = 2$,

$$d^{(2)}_{nm}(f,g)^2 = \| \sqrt{f/g} -1 \|^2_2 + \| \sqrt{g/f} -1\|^2_2$$

$$\leq \bar{\rho}(\alpha/\beta,\ \beta/\beta) + \bar{\rho}(\beta/\alpha,\ \alpha/\alpha)$$

with the equality when μ_α and μ_β are Gaussian measures.

11) <u>Symmetrized L_1 Ratio Distortion</u>

$$d^{(1)}_1(f,g) = \|f/g - 1\|_1 + \|g/f - 1\|_1$$

$$= \|f/g - g/f\|_1 \qquad (3.2.26)$$

which follows from $|x-1| + |1/x-1| = |x-1/x|$ for $x > 0$.

Many other distortion measures can be defined by the combinations in Section 3.2.1, but the distortion measures listed above are the basic distortions considered here.

3.3 PROPERTIES AND INTERRELATIONS OF SPECTRAL DISTORTION MEASURES

1) Separation of Gain

Several distortion measures can be expressed by the sum of the gain-normalized distortion and the distortion on two white processes with the same gain as f and g, i.e., σ_f^2 and σ_g^2.

We have

$$d_{cm}(f,g)^2 = \frac{\sigma_f^2}{\sigma_g^2} d_{ncm}(f,g)^2 + \left(\frac{\sigma_f}{\sigma_g} - 1\right)^2$$

$$= \frac{\sigma_f^2}{\sigma_g^2} d_{ncm}(f,g)^2 + d_{cm}(\sigma_f^2,\sigma_g^2)^2 \quad , \qquad (3.3.1)$$

$$d_{IS}(f,g) = d_{cm}(f,g)^2 + 2\left(\frac{\sigma_f}{\sigma_g} - 1 - \log\frac{\sigma_f}{\sigma_g}\right)$$

$$= \frac{\sigma_f^2}{\sigma_g^2} d_{ncm}(f,g)^2 + \frac{\sigma_f^2}{\sigma_g^2} - 1 - \log\frac{\sigma_f^2}{\sigma_g^2}$$

$$= \frac{\sigma_f^2}{\sigma_g^2} d_{ncm}(f,g)^2 + d_{IS}(\sigma_f^2,\sigma_g^2) \qquad (3.3.2)$$

One also obtains from (3.2.10) and (3.2.11) that

$$d_{IS}(f,g) = d_I(f,g) + d_{IS}(\sigma_o^2,\sigma_g^2) \qquad (3.3.3)$$

where $\sigma_o^2 = \sigma_g^2\, r_{f/g}(0)$ minimizes $d_{IS}\left(f, \frac{\sigma^2}{\sigma_g^2} g\right)$ over σ^2. Other gain separation properties are

$$d_{cosh}(f,g) = \frac{\sigma_f^2}{2\sigma_g^2} d_{ncm}(f,g)^2 + \frac{\sigma_g^2}{2\sigma_f^2} d_{ncm}(g,f)^2$$

$$+ d_{cosh}(\sigma_f^2,\sigma_g^2) \qquad (3.3.4)$$

and

$$d_{\log}(f,g)^2 = d_{\log}\left(\frac{f}{\sigma_f^2}, \frac{g}{\sigma_g^2}\right)^2 + d_{\log}(\sigma_f^2,\sigma_g^2)^2 . \qquad (3.3.5)$$

When the distortion between autoregressive models is calculated, equations (3.3.1)-(3.3.5) are important. This is because the gain term and the spectral term are separated and we only need to make subroutines to calculate the gain-normalized spectral distortion measures. Equations (3.3.1)-(3.3.5) imply that $d_{cm}(f,g) \geqq d_{cm}(\sigma_f^2,\sigma_g^2)$, $d_{IS}(f,g) \geqq d_{IS}(\sigma_f^2,\sigma_g^2)$. These relations are called the innovations property [21]. Note that, from (3.3.1) and (3.3.2), one obtains that

$$\begin{aligned} d_{IS}(f,g)-d_{IS}(\sigma_f^2,\sigma_g^2) &= d_{cm}(f,g)^2-d_{cm}(\sigma_f^2,\sigma_g^2)^2 \\ &= \frac{\sigma_f^2}{\sigma_g^2} d_{ncm}(f,g)^2 . \end{aligned} \qquad (3.3.6)$$

This means that the additivity removing the innovations distortion makes d_{IS} and d_{cm}^2 the same.

2) Cascade Relations

Consider three spectral densities $f(\lambda)$, $\hat{f}(\lambda)$ and $g_m(\lambda)$ where $f(\lambda)$ is considered as an original spectrum, $\hat{f}(\lambda)$ a reproduction of f, and $g_m(\lambda)$ is any m^{th} order spectral density. A useful property of some distortion measures in communication systems is that small distortions $d(f,\hat{f})$ and $d(\hat{f},g_m)$ imply small overall distortion $d(f,g_m)$. For example, metric distortion measures such as d_{log} have this property. $d_{IS}(f,g)$ and $d_{cm}(f,g)^2$ also have this property in the special case considered next.

Proposition 3.3.1: Let $\hat{f}_m(\lambda)$ be an m^{th} order estimate of $f(\lambda)$ obtained by Levinson's equation (2.2.3), i.e., by the minimization of d_{IS}. Then for any m^{th} order spectrum $g_m(\lambda)$ of the form

$$g_m(\lambda) = \sigma_g^2/|B_m(e^{i\lambda})|^2$$

with

$$B_m(z) = \sum_{k=0}^{m} b_{mk} z^{-k}, \quad b_{m0} = 1 \quad ,$$

we have that

$$d_{IS}(f,g_m) = d_{IS}(f,\hat{f}_m) + d_{IS}(\hat{f}_m,g_m) \quad . \tag{3.3.7}$$

If $\tilde{f}_m \in \eta_m$ is an estimate by the minimization of d_{cm}^2, then

$$d_{cm}(f,g_m)^2 = d_{cm}(f,\tilde{f}_m)^2 + \frac{\sigma_f^2}{\sigma_f(m)^2} d_{cm}(\tilde{f}_m,g_m)^2$$

$$\leq d_{cm}(f,\tilde{f}_m)^2 + d_{cm}(\tilde{f}_m,g_m)^2 . \tag{3.3.8}$$

Proof

(i) $$d_{IS}(f,g_m) = \frac{1}{2\pi}\int_{-\pi}^{\pi} \frac{f(\lambda)}{g_m(\lambda)}\, d\lambda - 1 - \log \frac{\sigma_f^2}{\sigma_g^2}$$

$$= \frac{1}{2\pi}\int_{-\pi}^{\pi} \frac{\hat{f}_m(\lambda)}{g_m(\lambda)}\, d\lambda - 1 - \log \frac{\sigma_f(m)^2}{\sigma_g^2} + \log \frac{\sigma_f(m)^2}{\sigma_f^2}$$

$$= d_{IS}(\hat{f}_m, g_m) + d_{IS}(f, \hat{f}_m) \qquad ,$$

where (2.2.3) was used.

(ii) Let $\tilde{f}_m(\lambda) = \sigma_{\tilde{f}}(m)^2/|\tilde{A}_m|^2$ be the m^{th} order spectrum which yields the minimum of $d_{cm}(f,g_m)$ Then $\tilde{A}_m$ satisfies Levinson's equation because of (3.3.6). Then

$$d_{cm}(f,\tilde{f}_m)^2 = \frac{\sigma_f(m)^2}{\sigma_{\tilde{f}}(m)^2} - 2\frac{\sigma_f}{\sigma_{\tilde{f}}(m)} + 1$$

$$= 1 - \frac{\sigma_f^2}{\sigma_f(m)^2}$$

since $\sigma_{\tilde{f}}(m) = \sigma_f(m)^2/\sigma_f$ minimizes $d_{cm}(f,\tilde{f}_m)$. Using the above relations, one obtains

$$d_{cm}(f,g_m) = \frac{1}{2\pi}\int_{-\pi}^{\pi} f/g_m d\lambda - 2\sigma_f/\sigma_{\tilde{f}} + 1$$

$$= \frac{1}{2\pi}\int_{-\pi}^{\pi} \frac{\sigma_f^2}{|A|^2}\,\frac{|B_m|^2}{\sigma_g^2}\, d\lambda - 2\sigma_f/\sigma_g + 1$$

$$= \frac{1}{2\pi}\int_{-\pi}^{\pi} \frac{\sigma_f(m)^2}{|\tilde{A}_m|^2}\,\frac{|B_m|^2}{\sigma_g^2}\, d\lambda - 2\sigma_f/\sigma_g + 1$$

$$= \frac{\sigma_f(m)^2}{\sigma_{\tilde{f}}(m)^2} \frac{1}{2\pi} \int_{-\pi}^{\pi} \tilde{f}_m / g_m d\lambda - 2\sigma_f/\sigma_g + 1$$

$$= \frac{\sigma_f(m)^2}{\sigma_{\tilde{f}}(m)^2} d_{cm}(\tilde{f}_m, g_m)^2 - 2 \frac{\sigma_f(m)^2}{\sigma_{\tilde{f}}(m)^2} \frac{\sigma_{\tilde{f}}(m)}{\sigma_g} - \frac{\sigma_f(m)^2}{\sigma_{\tilde{f}}(m)^2}$$

$$- 2 \frac{\sigma_f}{\sigma_g} + 1$$

$$= \frac{\sigma_f^2}{\sigma_f(m)^2} d_{cm}(\tilde{f}_m, g_m)^2 + 1 - \frac{\sigma_f(m)^2}{\sigma_{\tilde{f}}(m)^2}$$

$$= 1 - \frac{\sigma_f^2}{\sigma_f(m)^2} + \frac{\sigma_f^2}{\sigma_f(m)^2} d_{cm}(\tilde{f}_m, g_m)^2$$

$$= d_{cm}(f, \tilde{f}_m)^2 + \frac{\sigma_f^2}{\sigma_f(m)^2} d_{cm}(\tilde{f}_m, g_m)^2$$

$$\leq d_{cm}(f, \tilde{f}_m)^2 + d_{cm}(\tilde{f}_m, g_m)^2 \quad . \qquad \square$$

The main significance of (3.3.7) and (3.3.8) is for the quantization or coding of the model spectra $\hat{f}_m$ or $\tilde{f}_m$. Under d_{IS} or d_{cm}^2, the spectrum g which is obtained by the quantization or coding of $\hat{f}_m$ or $\tilde{f}_m$ is still close to the unknown original spectrum f provided $d_{IS}(\hat{f}_m, g_m)$ or $d_{cm}(\tilde{f}_m, g_m)^2$ is small. This property is desirable for the codebook design of Section 5.1.

3) Topologies Generated by Nonsymmetric Distortion Measures

First we have

$$d_{cm}(f,g) \geqq d_{nm}(f,g) \qquad (3.3.9)$$

because $|f^+/g^+| \geqq f^+/g^+$. Hence $d_{cm} \Rightarrow d_{cm} \Rightarrow d_{nm}$. Next by the fact that $x \to 1 \Rightarrow x-1-\ell n\ x \to 0$, we have that $d_{cm}(f,g_n) \to 0$ implies $d_{IS}(f,g_n) \to 0$, because $d_{cm}(f,g) \to 0 \Rightarrow |\sigma_f/\sigma_{g_n} -1| \to 0 \Rightarrow \sigma_f/\sigma_g - 1 - \log(\sigma_f/\sigma_g) \to 0$. Thus $d_{IS} \Leftrightarrow d_{cm} \Rightarrow d_{nm}$ is obtained. We have from (3.2.15) and (3.2.16) that $d_{ncm} \Leftrightarrow d_I \Leftrightarrow d^o_{cm}$ because $e^y - 1 \to 0 \Leftrightarrow 0 < y \to 0 \Leftrightarrow 1-e^{-y} \to 0$. One obtains from (3.3.1) and the innovation property of d_{cm} that

$$d_{ncm}(f,g)^2 \leqq \frac{\sigma_g^2}{\sigma_f^2} d_{cm}(f,g)^2 \leqq \frac{d_{cm}(f,g)^2}{\{1-d_{cm}(f,g)\}^2}$$

which implies $d_{cm} \Rightarrow d_{ncm}$. One more implication is that $d_1 \Rightarrow d_{ncm}$ because $d_{nm}(f,g)^2 = \| \sqrt{f/g} -1\|_2^2 \leq \|f/g-1\|_1$ by virtue of $|x-1| \geqq |\sqrt{x} -1|^2$. On the relation of $\bar{\rho}$-distance,

$$d_{ncm}(f,g)^2 \leqq \bar{\rho}(\alpha/\beta,\ \beta/\beta) \qquad (3.3.10)$$

is obtained by Proposition 2.3.1 where $\bar{\rho}$ uses square difference distortion and $\mu_{\alpha/\beta}$, $\mu_{\beta/\beta}$ are the probability measures of mismatch inverse filtered and white processes respectively. A detailed explanation of the inverse filtering is relegated to Chapter 4.

On the coding equivalence, one obtains $d_{ncm} \Leftrightarrow d_I \Leftrightarrow d^o_{cm}$ since they are monotically related by (3.2.15) and (3.2.16).

We have proved the following proposition:

Proposition 3.3.2: The topological and coding equivalencies of non-symmetric distortion measures are:

$$d^{o}_{cm} \Leftrightarrow d_I \Leftrightarrow d_{ncm}$$

$$\Uparrow$$

$$d_{IS} \Leftrightarrow d_{cm}$$

$$\Downarrow$$

$$\bar{\rho} \Rightarrow d_{nm} \Leftarrow d_1$$

where $\bar{\rho}$ means $\bar{\rho}(\alpha/\beta, \beta/\beta)$ as in (3.3.10), and

$$d_{ncm} \Leftrightarrow d_I \Leftrightarrow d^{o}_{cm} \quad .$$

4) Topologies Generated By Symmetric Distortions

The cosh measure and the log spectral deviation are the most commonly used symmetric distortions. Their relations to other measures are here developed. First one obtains

$$d_{log}(f,g)^2 \leqq 2d_{cosh}(f,g) \tag{3.3.11}$$

by simply comparing the integrands. Equations (3.2.13), (3.2.19) and (3.3.9) yield the following chain of inequalities:

$$
\begin{aligned}
2d_{\cosh}(f,g)^2 &= r_{f/g}(0) + r_{g/f} - 2 \\
&= d_{cm}(f,g)^2 + d_{cm}(g,f)^2 + 2\left(\sqrt{\sigma_f/\sigma_g} - \sqrt{\sigma_g/\sigma_f}\right)^2 \\
&= d_{cm}^{(2)}(f,g)^2 + 2\left(\sqrt{\sigma_f/\sigma_g} - \sqrt{\sigma_g/\sigma_f}\right)^2 \\
&\geqq d_{cm}^{(2)}(f,g)^2 \\
&\geqq d_{nm}^{(2)}(f,g)^2 \qquad , \qquad (3.3.12)
\end{aligned}
$$

and therefore $d_{\cosh} \Rightarrow d_{cm}^{(2)} \Rightarrow d_{nm}^{(2)}$. By the Minkowski inequality and the parallelogram law,

$$
\begin{aligned}
\{2d_{\cosh}(f,g)\}^{1/2} &= \|\sqrt{f/g} - \sqrt{g/f}\|_2 \leqq \|\sqrt{f/g} - 1\|_2 + \|\sqrt{g/f} - 1\|_2 \\
&= d_{nm}^{(1)}(f,g) \quad \leqq \sqrt{2}\, d_{nm}^{(2)}(f,g)^2 \quad .
\end{aligned}
$$

Combining the above inequalities, we have that

$$
\frac{1}{2} d_{nm}^{(2)}(f,g)^2 \leqq \frac{1}{2} d_{cm}^{(2)}(f,g)^2 \leqq d_{\cosh}(f,g) \leqq d_{nm}^{(2)}(f,g)^2 \leqq d_{cm}^{(2)}(f,g)^2
$$

which yields $d_{\cosh} \Leftrightarrow d_{cm}^{(q)} \Leftrightarrow d_{nm}^{(q)}$ for $q \geqq 1$.

By an analogous way to the development for nonsymmetric distortions, the following relations are easily obtained:

$$
\overset{\circ}{d}_{\cosh}(f,g) \Leftrightarrow d_I^{*}(f,g)
$$

and

$$
\overset{\circ}{d}_{\cosh}(f,g) \Leftrightarrow d_I^{*}(f,g) \Leftrightarrow d_{ncm}^{(2)} \quad .
$$

Other implications are $d_I^{*} \Leftrightarrow d_{ncm}^{(q)}$, $d_{cm}^{(q)} \Rightarrow d_{nm}^{(q)}$ and $d_1^{(q)} \Rightarrow d_{nm}^{(q)}$ which are obtained by $d_I \Leftrightarrow d_{ncm}$, $d_{cm} \Rightarrow d_{nm}$ and $d_I \Rightarrow d_{nm}$, respectively. On the relations of $\bar{\rho}$-distance to spectral distortion measures, we have

$$d_{cosh} \Leftarrow \bar{\rho}(\alpha/\beta,\beta/\alpha) \Leftarrow \bar{\rho}(\alpha/\beta,\beta/\beta) + \bar{\rho}(\beta/\alpha,\alpha/\alpha) \Rightarrow d_{nm}^{(2)} \Leftrightarrow d_{nm}^{(q)} ,$$

where the implication between the $\bar{\rho}$-distances are obtained by the parallelogram law: $(x-z)^2 \leq 2\{(x-y)^2+(y-z)^2\}$. To summarize we have:

<u>Proposition 3.3.3:</u> The topological and coding equivalences of symmetric distortion measures are:

$$\bar{\rho}(\alpha/\beta,\beta/\alpha) \Leftarrow \bar{\rho}(\alpha/\beta,\beta/\beta) + \bar{\rho}(\beta/\alpha,\alpha/\alpha)$$

$$\Downarrow$$

$$d_{log} \Leftarrow \overbrace{d_{cosh} \Leftrightarrow d_{cm}^{(q)} \Leftrightarrow d_{nm}^{(q)}} \Leftarrow d_{1}^{(q)}$$

$$\Downarrow$$

$$\overbrace{d_{cosh}^{o} \Leftrightarrow d_{I}^{*} \Leftrightarrow d_{ncm}^{(q)}}$$

and

$$d_{cosh}^{o} \Leftrightarrow d_{I}^{*} \Leftrightarrow d_{ncm}^{(q)} .$$

In addition to the above relations, all symmetrized distortions are stronger than their unsymmetrized versions.

5) <u>A Condition for Topological Equivalence of All Un-normalized Spectral Distortions</u>

In Propositions 3.3.2 and 3.3.3, the topological implications by distortion measures of the class η of spectral densities were considered. One might pose a question: How large a portion of the class of spectral densities will be lost if we require the topological equivalence of all original spectral distortion measures within a subclass?

The following proposition provides sufficient conditions for such a subclass. Uniform integrability [30] is the key for the equivalence.

Definition 3.3.4: A class of functions $\{h_\alpha(\lambda), \alpha \in \Lambda\}$ ($\lambda \in D$ where D is the whole domain) is uniformly integrable if

(i) $\int_D |h_\alpha(\lambda)| d\lambda \leqq K < \infty$ for all $\alpha \in \Lambda$

and

(ii) for all $\epsilon > 0$, there exists $\delta(\epsilon) > 0$ such that for $E \subset D$

$$\lambda(E) < \delta \text{ implies } \int_E |h_\alpha(\lambda)| d\lambda < \epsilon \text{ for all } \alpha \in \Lambda .$$

Proposition 3.3.5: Let $\eta^{(1)}$ be a subclass of η such that the class

$$\{f : f \in \eta^{(1)}\} \cup \{1/f : f \in \eta^{(1)}\}$$

is uniformly integrable. Then, for spectral densities of the class $\eta^{(1)}$, the topologies generated by the following all distortion measures are all equivalent:

$$d_{IS},\ d_{cm},\ d_{nm},\ d_1,\ d_\rho,\ d_{\log},\ d_{\cosh},\ d_{cm}^{(q)},\ d_{nm}^{(q)} \text{ and } d_1^{(q)}$$

for $q \geq 1$.

Proof.

Since the proof consists of repetitions of the same arguments, only one case $d_\rho \Rightarrow d_{nm}$ is shown to see how the conditions on the class $\eta^{(1)}$ are used.

Let $f(\lambda)$ and $g_n(\lambda)$ be elements of $\eta^{(1)}$ such that

$d_\rho(f,g_n) = \|\sqrt{f} - \sqrt{g_n}\|_2^2 \to 0$ as $n \to \infty$. Then $\delta > 0$, there exists N_1 such that $n \geqq N_1$ implies $\lambda(|\sqrt{f} - \sqrt{g_n}|^2 > \delta) < \delta$ because L_1-convergence implies the convergence in measure [30]. Define the set $F = \{\lambda : |\sqrt{f} - \sqrt{g_n}|^2 > \delta\}$. Then for all $\epsilon > 0$ there exists N_2 such that $n \geqq N_2$ implies

$$d_{nm}(f,g_n)^2 = \frac{1}{2\pi}\int_F |\sqrt{f/g_n} - 1|^2 d\lambda + \frac{1}{2\lambda}\int_{F^c} |\sqrt{f/g_n} - 1|^2 d\lambda$$

$$\leqq \epsilon + \frac{1}{2\pi}\int_{F^c} |\sqrt{f/g_n} - 1|^2 d\lambda$$

by the uniform integrability. Note that $|\sqrt{f} - \sqrt{g_n}|^2 \leqq \delta$ in F^c. Then $|\sqrt{f/g_n} - 1|^2 \leqq \delta/g_n$ λ-a.e. Therefore,

$$d_{nm}(f,g_n)^2 \leqq \epsilon + \delta \frac{1}{2\pi} \int_{F^c} \frac{1}{g_n} d\lambda \leqq \epsilon + \delta K .$$

Since δ can be taken arbitrarily small for large n, $d_{nm}(f,g_n) \to 0$ as $n \to \infty$. Hence $d_\rho \Rightarrow d_{nm}$. Other implications are proven under minor modifications of the above method. □

We note that

$$\mathcal{N}^{(1)} \supset \mathcal{N}^{(2)} = \{f(\lambda) : 0 < K_1 \leqq f(\lambda) \leqq K_2 < \infty\} .$$

6) Approximation of Small Deviations

Several spectral distortion measures have linear relationships to other distortion measures when distortions are small.

Proposition 3.3.6: If $f^{+}(e^{i\lambda}) \simeq g^{+}(e^{i\lambda})$ where both $f^{+}(e^{i\lambda})$ and $g^{+}(e^{i\lambda})$ are stable causal filters, then the following relationships hold:

(i) $$d_{\log}(f,g)^2 \simeq 2d_{IS}(f,g) \simeq 4d_{nm}(f,g)^2 \quad , \qquad (3.3.13)$$

(ii) $$d_{\log}(f,g)^2 \simeq 2d_{\cosh}(f,g) \simeq 2d_{nm}^{(2)}(f,g)^2 \quad , \qquad (3.3.14)$$

(iii) $$d_{\log}\left(\frac{f}{\sigma_f^2}, \frac{g}{\sigma_g^2}\right)^2 \simeq 2d_{ncm}(f,g)^2 \simeq 2d_{cm}^{o}(f,g)^2 \quad , \qquad (3.3.15)$$

(iv) $$d_{\log}\left(\frac{f}{\sigma_f^2}, \frac{g}{\sigma_g^2}\right)^2 \simeq d_{ncm}^{(2)}(f,g)^2 \simeq 2d_I^*(f,g) \simeq d_{\cosh}^{o}(f,g) \ . \qquad (3.3.16)$$

Proof.

(i) $f/g - 1 - \log(f/g) = e^{\log(f/g)} - 1 - \log(f/g) \simeq \frac{1}{2}\left(\log \frac{f}{g}\right)^2$ which is obtained by the Taylor series expansion of the exponential. The right equality is obtained in the same way.

(ii) This simply follows from symmetrization.

(iii) The leftmost equality comes from:

$$\begin{aligned} d_{\log}\left(\frac{f}{\sigma_f^2}, \frac{g}{\sigma_g^2}\right) &= d_{\log}\left(\frac{1}{|A|^2}, \frac{1}{|B|^2}\right)^2 \\ &= \frac{1}{2\pi}\int_{-\pi}^{\pi} \{\log|B/A|^2\}^2 d\lambda \\ &= \frac{1}{2\pi}\int_{-\pi}^{\pi} \left(\log\frac{B}{A} + \log\frac{B^*}{A^*}\right)^2 d\lambda \\ &= \frac{1}{2\pi}\int_{-\pi}^{\pi} \left(\log\frac{B}{A}\right)^2 d\lambda + 2\cdot\frac{1}{2\pi}\int_{-\pi}^{\pi} \left|\log\frac{B}{A}\right|^2 d\lambda \\ &+ \frac{1}{2\pi}\int_{-\pi}^{\pi}\left(\log\frac{B^*}{A^*}\right)^2 d\lambda \end{aligned}$$

$$= 2 \cdot \frac{1}{2\pi} \int_{-\pi}^{\pi} |\log \frac{B}{A}|^2 \, d\lambda$$

where the last equality follows from the fact that the other terms are equal to zero. This is because the summation of the first and third terms which are conjugates of each other must be twice the real part, which is zero since A and B are monic [31]. Furthermore, $\log(1+\delta) \simeq \delta$ gives

$$\begin{aligned} d_{\log}\left(\frac{f}{\sigma_f^2}, \frac{g}{\sigma_g^2}\right) &= 2 \cdot \frac{1}{2\pi} \int_{-\pi}^{\pi} \left|\log\left(1 + \frac{B-A}{A}\right)\right|^2 d\lambda \\ &\simeq 2 \cdot \frac{1}{2\pi} \int_{-\pi}^{\pi} \left|\frac{B-A}{A}\right|^2 d\lambda \\ &= 2 d_{ncm}(f,g)^2 \quad . \end{aligned}$$

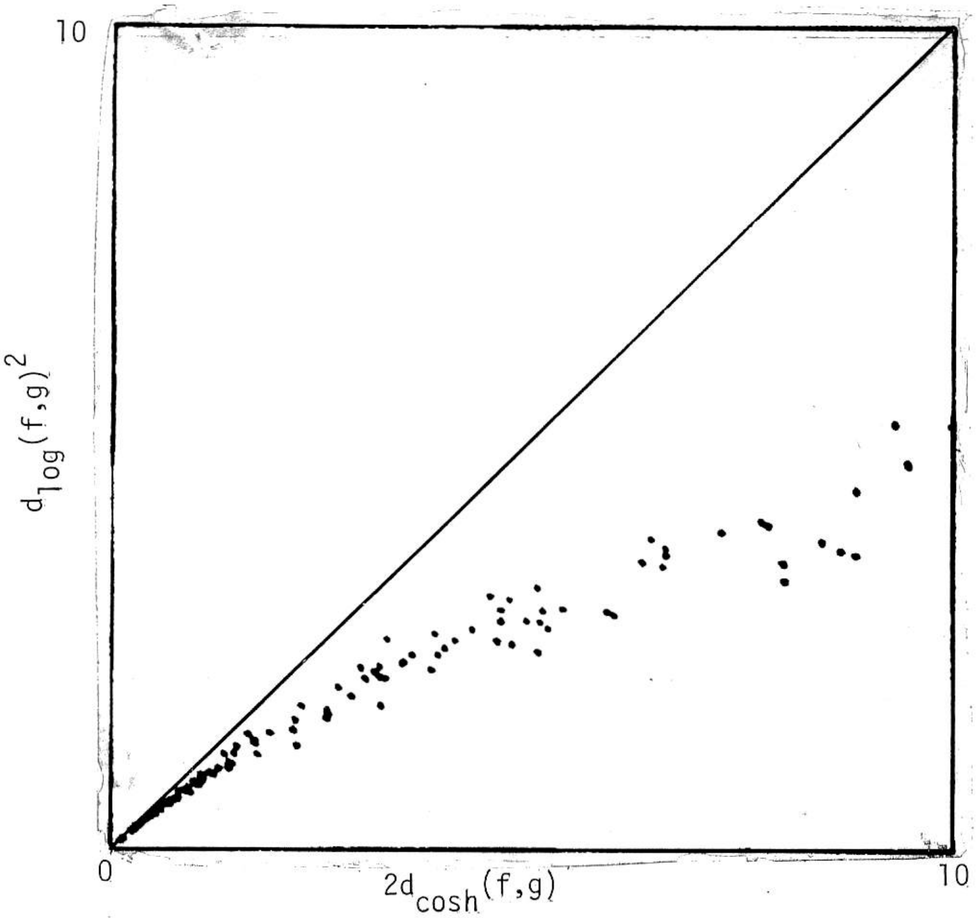
Other equalities come from (3.2.15) and (3.2.16).

(iv) The result easily follows from the symmetrization of (iii) and (3.2.22).

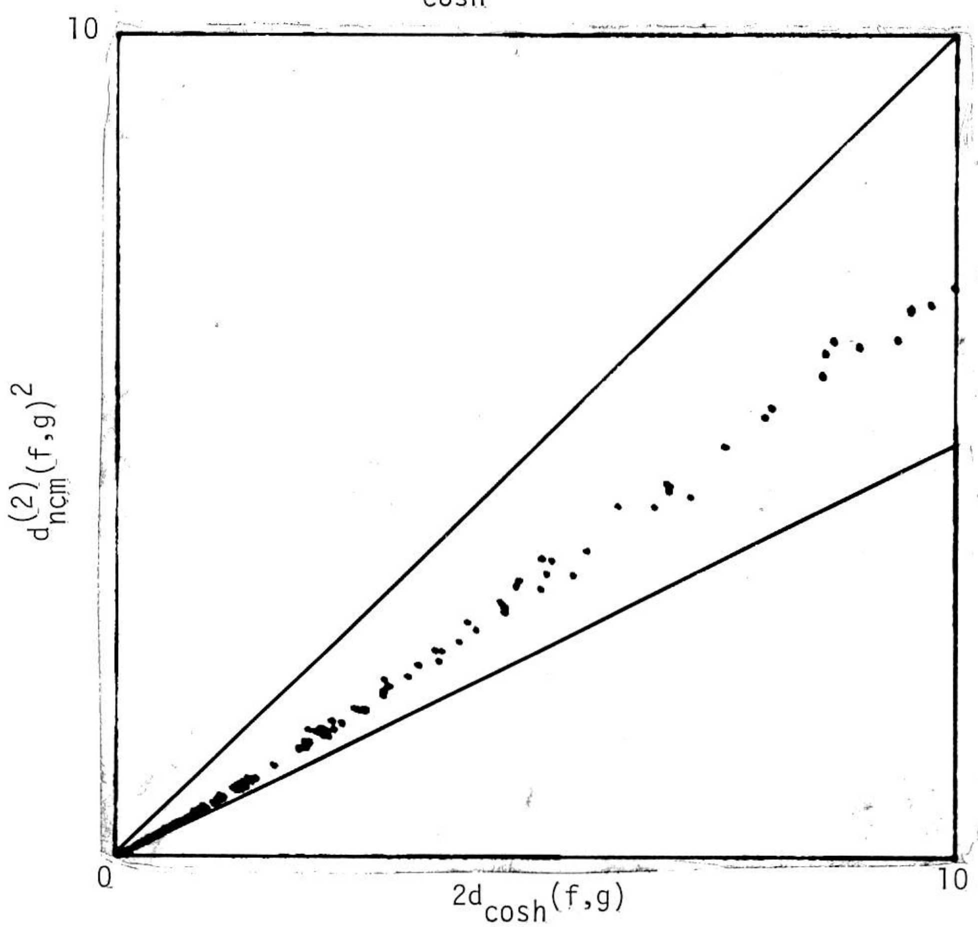
Proposition 3.3.6, which was obtained by simple calculus, has an important implication: When these spectral distortion measures are used in small deviation cases, all give nearly the same result. We checked a few relations in Proposition 3.3.6 by obtaining scatter plots from actual speech: "The pipe began to rust while new," spoken by a female and, "Thieves who rob friends deserve jail," spoken by a male. Figure 3.3.1 compares several distortion values. The computation methods are explained in the next section. The degree of coincidence, i.e., the linear relationship depends upon distortion measures. The bifurcation of Fig. 3.3.1 arises because $d_{IS}(\sigma_f^2, \sigma_g^2)$ is a function of $\pm\, d_{cm}(\sigma_f^2, \sigma_g^2)^2$. That is,

$$d_{IS}(\sigma_f^2,\sigma_g^2) = \{1 \pm d_{cm}(\sigma_f^2,\sigma_g^2)^2\}^2 - 1 - \log\{1 \pm d_{cm}(\sigma_f^2,\sigma_g^2)^2\} \quad .$$

And the gain separation properties explain scattering around the gain curves. Figure 3.4.1.d is the most important one, since the approximation of $d_{log}\left(\frac{f}{\sigma_f^2}, \frac{g}{\sigma_g^2}\right)$ by $d_{ncm}(f,g)$ is frequently used because of its computational advantage. This is used in the quantization of reflection coefficients [31]. Gray and Markel [27] suggest that 2 dB of d_{log} is the barely perceptible difference limen, where $4.34d_{log}$ gives the dB scale. Figure 3.3.1d shows that within the limen, the two distortion measures are linearly related. Therefore, the quantization of spectra using both measures is equivalent in the coding sense in that region.

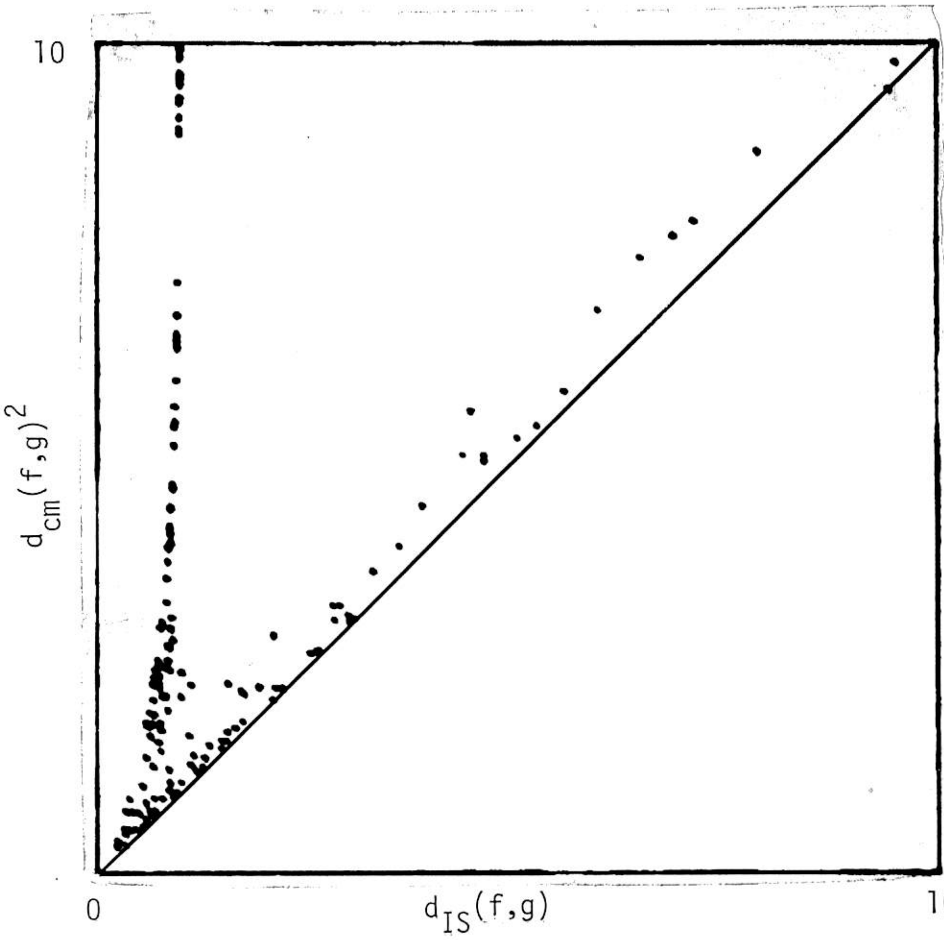


(a)

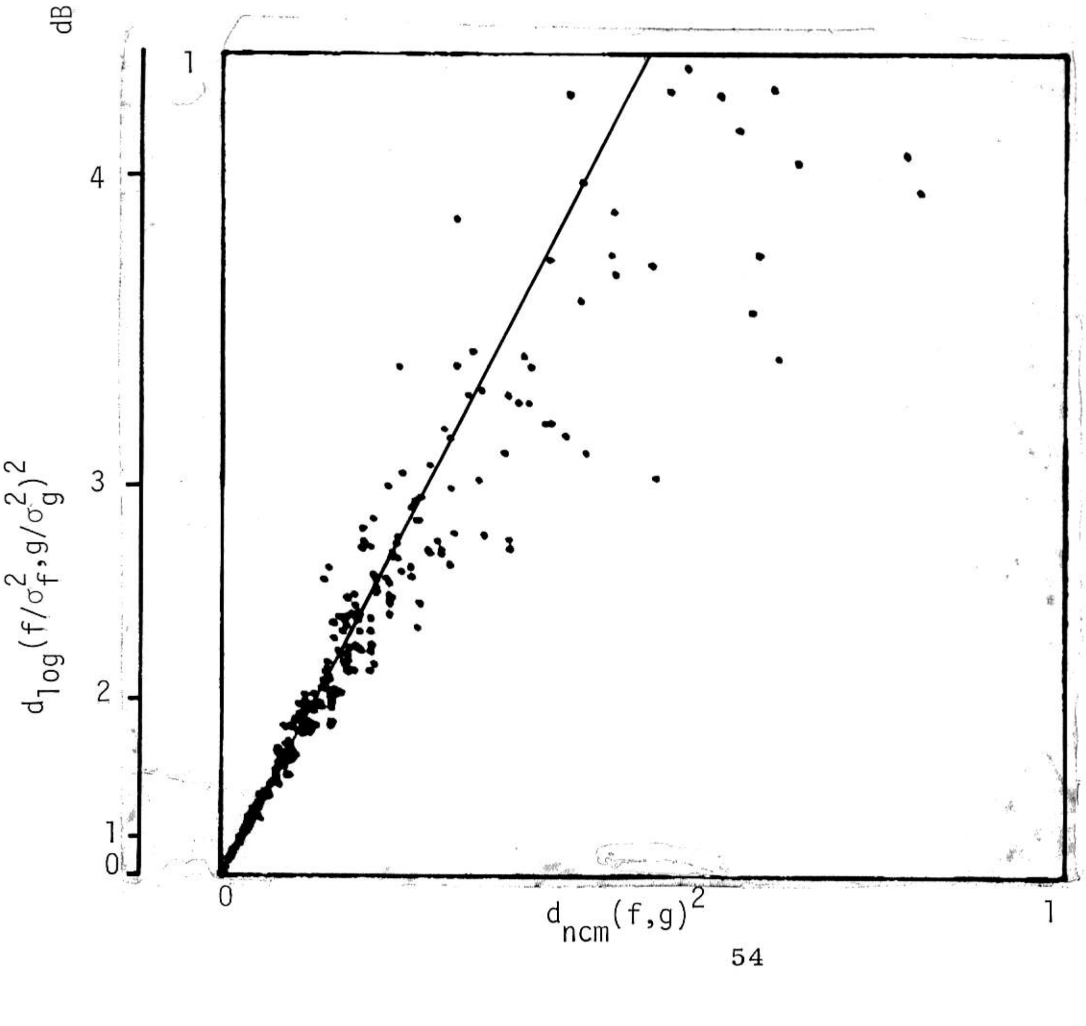


(b)

Fig. 3.3.1 Scatter plots of spectrel distortion measures


10
$d_{cm}(f,g)^2$
0
$d_{IS}(f,g)$
10

(c)


dB
1
4
3
2
1
0
$d_{log}(f/\sigma_f^2, g/\sigma_g^2)^2$
0
$d_{ncm}(f,g)^2$
1

(d)

3.4 Computation and Related Matters

There are two prevalent methods of estimating spectra from time series: direct computation by FFT and indirect computation from parameter estimation of AR models. The first method is useful for hardware such as a spectrum analyzer, but it takes considerable calculation time if computer software is used, and the entire spectrum needs to be stored. Moreover, FFT produces a very noisy spectrum. The second method may take a longer time than the first, but it has a software-computational advantage and saves memory space when a specific set of spectra is used as a reproduction class. It turns out that the model estimation method fits our laboratory experiments which mainly rely on s software. In what follows, a minimum necessary discussion on computational aspects is given.

1) Minimum Variance Versus Total Residual Error

Since only a sample path is given, a method using the mathematical expectation $E(\cdot)$ cannot be used. Instead, assuming ergodicity, we make use of the relation

$$E\left(\sum_{k=0}^{m} b_{mk} Y_{n-k}\right)^2 = \lim_{N \to \infty} \left\{ \frac{1}{N} \sum_{n=n_0}^{n_0+N-1} \left(\sum_{k=0}^{m} b_{mk} Y_{n-k}\right)^2 \right\}$$

so that for large enough N we should have

$$E(\cdot) \simeq \frac{1}{N} \sum_{n=n_0}^{n_0+N-1} (\cdot) \quad . \qquad (3.4.1)$$

Using $\frac{1}{N} \sum_{n=n_0}^{n_0+N-1} (\cdot)$ instead of $E(\cdot)$, an estimate of autocorrelation sequence $\{r_f(k)\}_{k=0}^{m}$ is obtained [35]. By solving (2.2.3), an m^{th}

order AR estimate $f_m(\lambda)$ of $f(\lambda)$ is obtained. There is another method called recursive estimation [32], [33] which sometimes gives a better estimate of $f(\lambda)$ since the method is an unbiased estimation. However, if a relatively small number of samples are treated as in speech applications, the method needs a great deal more computation because of the recursive, sample by sample operation.

2) Computation of the Normalized Causal Model Distortion

For two polynomials $C(z)$ and $D(z)$, the inner product is defined by

$$\langle C(z), D(z)\rangle_f = \frac{\sigma_f^2}{2\pi}\int_{-\pi}^{\pi} \frac{C(e^{i\lambda})D^*(e^{i\lambda})}{|A(e^{i\lambda})|^2}\, d\lambda \tag{3.4.2}$$

for $f(\lambda) = \sigma_f^2/|A(e^{i\lambda})|^2$, $A(z) = \sum_{k=0}^{m} a_k z^{-k}$ with $a_0 = 1$. Then

$$\begin{aligned} d_{ncm}(f,g)^2 &= \left\| \frac{B(e^{i\lambda})-A(e^{i\lambda})}{A(e^{i\lambda})} \right\|_2^2 \\ &= \frac{1}{\sigma_f^2}\langle B-A,\ B-A\rangle_f \\ &= \frac{\|B\|_f^2 - 2\langle A,B\rangle_f + \|A\|_f^2}{\sigma_f^2} \\ &= \frac{\|B\|_f^2}{\|A\|_f^2} - 1 \end{aligned} \tag{3.4.3}$$

since

$$\langle A,B\rangle_f = \langle A,1\rangle_f = \langle A,A\rangle_f = \|A\|_f^2 = \sigma_f^2$$

by (2.2.8). When segments of time series are given, this is calculated using the correspondence (3.4.1). In that case, we have

$$d_{ncm}(f,g)^2 = \frac{(\underset{\sim}{b}-\underset{\sim}{a})^T R_f (\underset{\sim}{b}-\underset{\sim}{a})}{\underset{\sim}{a}^T R_f \underset{\sim}{a}}$$

$$= \frac{\sum_{n=-m}^{m} r_b(n) r_f(n)}{\sum_{n=-m}^{m} r_a(n) r_f(n)} - 1 \qquad (3.4.4)$$

where r_a, r_b are autocorrelations of $\{a_{mi}\}_{i=0}^{m}$ and $\{b_{mi}\}_{i=0}^{m}$. This holds also for $m = \infty$ [21]. From (3.4.4) and the gain separation properties (3.3.1)-(3.3.4), the causal distortion and their relatives are calculated. The computation of other distortion measures relies on FFT methods except for d_{log}. The log spectral deviation d_{log} can be approximated by cepstral coefficients [27].

4) Computation of $d_{ncm}(f,g)^2$ from Reflection Coefficients

Equation (3.4.4) is usable when both $\{a_{mi}\}_{i=1}^{m}$ and $\{r_f(i)\}_{i=0}^{m}$ are stored in the memory as references (or, equivalently $\{k_i^a\}_{i=1}^{m}$ and $\{r_f(i)\}_{i=0}^{m}$ are memorized, because there are algorithms to convert $\{k_i^a\}_{i=1}^{m}$ to $\{a_{mi}\}_{i=1}^{m}$ and vice versa [20]). However, $\{r_f(i)\}_{i=0}^{m}$ can be derived from $\{a_{mi}\}_{i=0}^{m}$ using (2.2.6). In order to save memory space only the LPC coefficients are stored. Because of the remarkable property (2.2.15), the form of the reflection coefficients $\{k_i^a\}_{i=1}^{m}$, is chosen in LPC communication systems [20] [29]. In this case, $d_{ncm}(f,g)^2$ is calculated as follows. Since the method essentially calculates the righthand side of (3.4.4), it takes much more computational time. The following relationships are the basic equations.

By the definition of $\{k_i^a\}_{i=1}^m$, we have

$$\underset{\sim}{a}_m = A_{m-1} \underset{\sim}{k}_m^a \tag{3.4.5}$$

where

$$\underset{\sim}{a}_m = (a_{m1}, a_{m2}, \ldots, a_{mn})^T \tag{3.4.6}$$

$$A_{m-1} = \begin{bmatrix} 1 & a_{11} & a_{22} & \cdots & a_{m-1,m-1} \\ & 1 & a_{12} & & \vdots \\ & & 1 & \ddots & \vdots \\ & & & \ddots & a_{m-1,1} \\ & 0 & & & 1 \end{bmatrix}, \tag{3.4.7}$$

$$\underset{\sim}{k}_m^a = (k_1^a, \ldots, k_m^a)^T \quad . \tag{3.4.8}$$

By (2.2.6), one obtains

$$A_{m-1}^T \begin{bmatrix} r_f(0) & \cdots & r_f(m-1) \\ \vdots & \ddots & \vdots \\ r_f(m-1) & \cdots & r_f(0) \end{bmatrix} A_{m-1} = \Lambda_{m-1} \tag{3.4.9}$$

where

$$\Lambda_{m-1} = \mathrm{diag}\{\sigma_f(i)^2\}_{i=0}^{m-1} \tag{3.4.10}$$

Note that

$$\sigma_f(i)^2 = \sigma_f(0)^2 \prod_{j=1}^{i} \{1-(k_j^a)^2\} \tag{3.4.11}$$

holds because of (2.2.12) where $\sigma_f(0)^2 = r_f(0)$. Finally a formula is obtained:

$$d_{ncm}(f,g)^2 = \frac{(\underset{\sim}{b}_m - \underset{\sim}{a}_m)^T R_f (\underset{\sim}{b}_m - \underset{\sim}{a}_m)}{\sigma_f(m)^2}$$

$$= \frac{(\underset{\sim}{b}_m - \underset{\sim}{a}_m)^T A_{m-1}^{-T} \Lambda_{m-1} A_{m-1} (\underset{\sim}{b}_m - \underset{\sim}{a}_m)}{\sigma_f(m)^2}$$

$$= \frac{1}{\sigma_f(m)^2} \sum_{i=1}^{m} \{k_i^b(a) - k_i^a\}^2 \sigma_f(i)^2$$

$$= \sum_{i=1}^{m} \{k_i^b(a) - k_i^a\}^2 / \prod_{j=i+1}^{m} \{1 - (k_j^a)^2\} \qquad (3.4.12)$$

where $\underset{\sim}{k}^b(a) = A_{m-1}^{-1} \underset{\sim}{b}_m$.

Chapter 4

PREDICTOR AND INTERPOLATOR MISMATCH

4.1 MISMATCH OF ONE-STEP PREDICTORS ON AUTOREGRESSIVE PROCESS

Linear least squares prediction has a long history and has been made use of in various areas because of its intuitive principle and the handiness of the resulting calculations [14], [18], [44]. In actual cases, however, it is rare that the complete statistics of the underlying stochastic processes are given. A predictor designed without the true statistics is essentially "mismatched" to the true underlying stochastic processes. Considering the least squares criterion, one might expect that the resulting performance would be good enough if the guessed process is sufficiently "close" in some sense to the true process. This concept can be regarded as the robustness of the predictor against the mismatch.

This mismatch problem can be found in coding and in the recognition of speech. Therein, several predictors are designed using typical reference sounds or clusters of time series. Inverse filters which correspond to each predictor are implemented. The residual processes of inverse filters with an input speech segment are compared in terms of the previous spectral distortion measures. If the distortion is considerably large then the predictor is regarded as mismatched to the speech. Now, speech coding can be viewed as finding the predictor which gives the smallest mismatch distortion measure. More consideration of the above idea will be discussed in Chapter 6.

Let $[R,\mu_\alpha,X]$ and $[R,\mu_\beta,Y]$ be second order stochastic processes governed by autoregressive processes (3.2.2) and (3.2.3) of class η. We consider the problem of approximating X_n by the linear combination of $\{X_{n-1}, X_{n-2}, \ldots\}$, say $\hat{X}_n(X_{n-1}, X_{n-2}, \ldots)$, which minimizes the mean square error:

$$E|X_n - \hat{X}_n(X_{n-1}, X_{n-2}, \ldots)|^2 \quad .$$

The predictor $\hat{X}_n$ encodes all past values. If the predictor uses only a finite number of past values remote samples are weighted by zero. Suppose all past values are stored in an analog shift register in order. At each time the predictor obtains one new sample, and shifts it into the register, and predicts the value of the next sample. This is an example of sliding-block coding [1]. The mapping $f(\cdot)$ of equation (2.1.1) corresponds to the predictor. It is well-known [13] that a unique solution is given as

$$\hat{X}_n(X_{n-1}, X_{n-2}, \ldots) = -\sum_{k=1}^{\infty} a_k X_{n-k} \tag{4.1.1}$$

where $\{a_k\}$ are the autoregressive coefficients of (3.2.4), and the minimum mean square error is

$$E|X_n - \hat{X}_n(X_{n-1}, X_{n-2}, \ldots)|^2 = \sigma_f^2 \tag{4.1.2}$$

where f is the spectrum of $\{X_n\}$. Consider the case where one designs the predictor based on a process $\{Y_n\}$ and applies it to a process $\{X_n\}$. The mismatched prediction is then

$$\hat{Y}_n(X_{n-1}, X_{n-2}, \ldots) = -\sum_{k=1}^{\infty} b_k X_{n-k} \tag{4.1.3}$$

The mean square mismatch error is given by

$$\epsilon(\beta|\alpha)^2 \triangleq E|X_n - \hat{Y}_n(X_{n-1}, X_{n-2}, \ldots)|^2$$

$$= E|\sum_{k=0}^{\infty} b_k X_{n-k}|^2$$

$$= \sum_{k=0}^{\infty} \sum_{\ell=0}^{\infty} b_k b_\ell \frac{1}{2\pi} \int_{-\pi}^{\pi} f(\lambda) e^{-(k-\ell)\lambda} d\lambda$$

$$= \frac{1}{2\pi} \int_{-\pi}^{\pi} (\sum_{k=0}^{\infty} b_k e^{-ik\lambda})(\sum_{\ell=0}^{\infty} b_\ell e^{i\ell\lambda}) f(\lambda) d\lambda$$

$$= \frac{\sigma_g^2}{2\pi} \int_{-\pi}^{\pi} \frac{f(\lambda)}{g(\lambda)} d\lambda \qquad (4.1.4)$$

There are three kinds of mismatch criteria which are free from probability densities. The first one is the mean square error criterion:

$$C_A = |\epsilon(\beta|\alpha)^2 - \epsilon(\beta|\beta)^2| / \epsilon(\beta|\beta)^2 \quad . \qquad (4.1.5)$$

The second one is the root mean square error criterion:

$$C_B^2 = |\epsilon(\beta|\alpha) - \epsilon(\beta|\beta)|^2 / \epsilon(\beta|\beta)^2 \quad . \qquad (4.1.6)$$

Thirdly, one can compare prediction residual processes:

$$C_C^2 = \frac{E|\{X_n - \hat{Y}_n(X_{n-1}, X_{n-2}, \ldots)\} - \{X_n - \hat{X}_n(X_{n-1}, X_{n-2}, \ldots)\}|^2}{\epsilon(\alpha|\alpha)^2} \qquad (4.1.7)$$

It is easily seen that $\epsilon(\beta|\alpha) \geqq \epsilon(\alpha|\alpha)$ always holds; however, $\epsilon(\beta|\alpha) \geqq \epsilon(\beta|\beta)$ need not be true.

We define "mismatch-robustness" as follows.

Definition 4.1.1: A predictor is said to be mismatch-robust under a distortion measure d if given $\epsilon > 0$ there exists $\delta > 0$ such that $d(\alpha,\beta) < \delta$ implies (mismatch criterion) $< \epsilon$.

By (4.1.4),

$$C_A = \frac{1}{2\pi} \left| \int_{-\pi}^{\pi} \left\{ \frac{f(\lambda)}{g(\lambda)} - 1 \right\} d\lambda \right| \leq \frac{1}{2\pi} \int_{-\pi}^{\pi} \left| \frac{f(\lambda)}{g(\lambda)} - 1 \right| d\lambda$$

$$= d_1(f,g) \quad . \tag{4.1.8}$$

Equations (2.3.5) and (4.1.4) yield

$$C_B^2 = \frac{1}{2\pi} \left| \sqrt{\int_{-\pi}^{\pi} f/g \; d\lambda} - 1 \right|^2$$

$$\leq \frac{1}{2\pi} \int_{-\pi}^{\pi} \left| \sqrt{f/g} - 1 \right|^2 d\lambda$$

$$= d_{nm}(f,g)^2$$

$$\leq \bar{\rho}(\alpha/\beta,\beta/\beta) \tag{4.1.9}$$

where $\bar{\rho}(\alpha/\beta,\beta/\beta)$ means the $\bar{\rho}$-distance between the mismatch inverse filtered process $\sum_{k=0}^{\infty} b_k X_{n-k}$ and the true inverse filtered process $\sum_{k=0}^{\infty} b_k Y_{n-k} = \sigma_g \xi_n^{\beta}$. Using a similar derivation as (4.1.8), we obtain

$$C_C^2 = \frac{1}{2\pi} \int_{-\pi}^{\pi} \left| f^+/g^+ - 1 \right|^2 d\lambda = d_{nm}(f,g)^2 \quad . \tag{4.1.10}$$

Proposition 4.1.2: One-step predictors for autoregressive processes in the class η are mean square error mismatch-robust under d_1. Root mean square error and residual process mismatch-robustness hold under d_{nm} and d_{cm}, respectively. Since $d_{nm}(f,g)^2 \leq \bar{\rho}(\alpha/\beta,\beta/\beta)$, the root mean square error mismatch-robustness is guaranteed if $\bar{\rho}(\alpha/\beta,\beta/\beta)$ is small.

The preceeding mismatch discussion provides an interpretation for the two model distortion measures, $d_{nm}(f,g)^2$ and $d_{cm}(f,g)^2$. In Fig. 4.1.1, $\Phi(e^{i\lambda})$ is either $f^+(e^{i\lambda})$, which is associated with the predictor $\hat{X}_n(X_{n-1},X_{n-2},\ldots)$, or $|f^+(e^{i\lambda})| = \sqrt{f(\lambda)}$. $\Psi(e^{i\lambda})$ is a digital filter either $g^+(e^{i\lambda})$ associated with $\hat{Y}_n(X_{n-1},X_{n-2},\ldots)$, or $|g^+(e^{i\lambda})| = \sqrt{g(\lambda)}$. Figure 4.1.1a shows that $\|\Phi(e^{i\lambda})/\Psi(e^{i\lambda})-1\|_2^2$ measures how similarly the filters Φ and Ψ behave by comparing the inverse filtered spectrum with the white spectrum. On the other hand, in Fig. 4.1.1b, $1/\Phi$ is the whitening filter for $\{X_n\}$ and $1/\Psi$ is a mismatched inverse filter. It measures the error power between the true and the mismatch-whitened processes:

$$E|W_n - \xi_n|^2 = \|\Phi/\Psi - 1\|_2^2 \quad .$$

That is, the model distortion measure $\|\Phi/\Psi-1\|_2^2$ compares the error powers between the true whitened process and the mismatch whitened process.

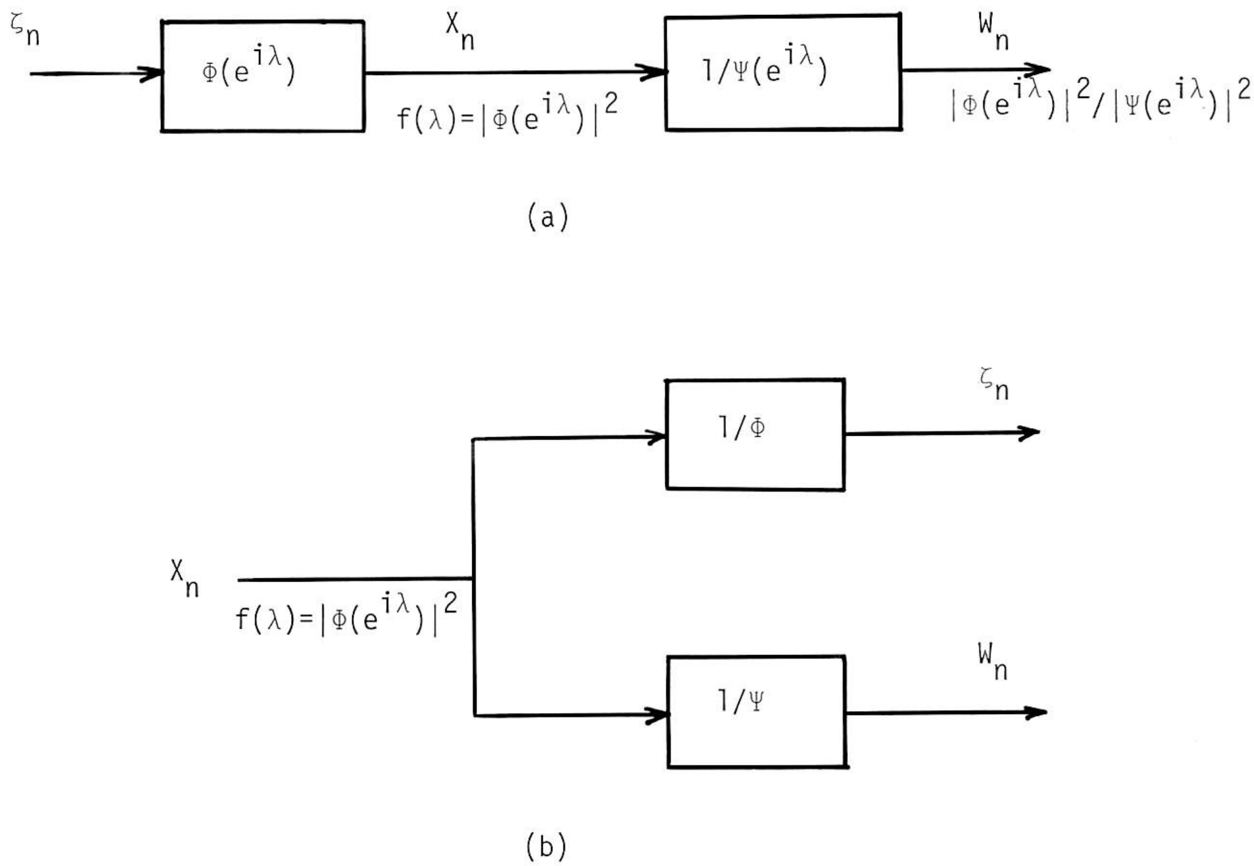


Fig. 4.1.1 Mismatch of an inverse filter

4.2 INTERPOLATOR MISMATCH

Another important example for the noncausal sliding block structure is an interpolator, $\hat{X}_n(\ldots,X_{n-2},X_{n-1},X_{n+1},X_{n+2},\ldots)$, which minimizes

$$E|X_n-\hat{X}_n(\ldots,X_{n-2},X_{n-1},X_{n+1},X_{n+2},\ldots)|^2 \quad .$$

In this section, spectral densities of class $\mathcal{N}$ are considered again. Here, the spectral representation [14] is used. Suppose

$$\hat{Y}_n(\ldots,Y_{n-1},Y_{n+1},\ldots) = \int_{-\pi}^{\pi} K_\beta(\lambda)e^{in\lambda}Z_Y(d\lambda) \quad .$$

Then the projection property [13] gives

$$E(Y_n-\hat{Y}_n)Y_m = \frac{1}{2\pi}\int_{-\pi}^{\pi}\{1-K_\beta(\lambda)\}e^{in\lambda}g(\lambda)d\lambda = 0 \quad \text{for } m \neq n.$$

Hence

$$K_\beta(\lambda) = 1 - \frac{c}{g(\lambda)} \quad .$$

Since

$$E|Y_n-\hat{Y}_n|^2 = \frac{1}{2\pi}\int_{-\pi}^{\pi}|1-K_\beta(\lambda)|^2 g(\lambda)d\lambda$$

$$= \frac{c^2}{2\pi}\int_{-\pi}^{\pi}\frac{1}{g(\lambda)}\, 1\lambda$$

$$= \frac{1}{2\pi}\int_{-\pi}^{\pi}\{1-K_\beta(\lambda)\}g(\lambda)d\lambda$$

$$= c \quad ,$$

one obtains

$$c = \frac{1}{\frac{1}{2\pi}\int_{-\pi}^{\pi}\frac{1}{g(\lambda)}d\lambda} \tag{4.2.1}$$

and

$$K_\beta(\lambda) = 1 - \frac{1}{\frac{g(\lambda)}{2\pi} \int_{-\pi}^{\pi} \frac{1}{g(\lambda)}\, d\lambda}$$

The mismatch error is

$$\delta(\alpha|\beta)^2 = E|X_n - \hat{Y}_n(\ldots, X_{n-1}, X_{n+1}, \ldots)|^2$$

$$= \frac{\frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{f(\lambda)}{g^2(\lambda)}\, d\lambda}{\left\{\frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{1}{g(\lambda)}\, d\lambda\right\}^2} .$$

Then

$$C_A = \frac{|\delta^2(\alpha|\beta) - \delta^2(\beta|\beta)|}{\delta^2(\beta|\beta)}$$

$$= \frac{\left|\frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{1}{g(\lambda)} \left\{\frac{f(\lambda)}{g(\lambda)} - 1\right\} d\lambda\right|}{\frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{1}{g(\lambda)}\, d\lambda} \leqq \frac{\frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{1}{g(\lambda)} \left| \frac{f(\lambda)}{g(\lambda)} - 1\right| d\lambda}{\frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{1}{g(\lambda)}\, d\lambda} \qquad (4.2.2)$$

$$C_B^2 = \frac{|\delta(\alpha|\beta) - \delta(\beta|\beta)|^2}{\delta^2(\beta|\beta)}$$

$$= \left| \sqrt{\frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{f(\lambda)}{g^2(\lambda)}\, d\lambda} - \sqrt{\frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{1}{g(\lambda)}} \right|^2 \Big/ \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{1}{g(\lambda)}\, d\lambda$$

$$\leqq \frac{\frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{1}{g(\lambda)} \left\{\sqrt{\frac{f(\lambda)}{g(\lambda)}} - 1\right\}^2 d\lambda}{\frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{1}{g(\lambda)}\, d\lambda} \qquad (4.2.3)$$

<u>Proposition 4.2.1:</u> The linear least squares interpolators are mean square error mismatch-robust according to the spectral distortion measure of (4.2.2). The root mean square mismatch-robustness is guaranteed under the spectral distortion measure of (4.2.3).

It is important to point out that (4.2.1) implies that

$$\frac{1}{2\pi}\int_{-\pi}^{\pi}\frac{1}{f(\lambda)}\,d\lambda < \infty$$

is the condition for a non-deterministic process for the linear least squares interpolation, but

$$\frac{1}{2\pi}\int_{-\pi}^{\pi}\log\left\{\frac{1}{f(\lambda)}\right\}d\lambda < \infty$$

is the corresponding condition for the prediction case. Note that the Gaussian assumption ensures the optimality of the predictors and interpolators designed by the linear least squares error principle. Even in this case, Jensen's inequality

$$\frac{1}{2\pi}\int_{-\pi}^{\pi}\log\left\{\frac{1}{f(\lambda)}\right\}d\lambda \leqq \log\left\{\frac{1}{2\pi}\int_{-\pi}^{\pi}\frac{1}{f(\lambda)}\,d\lambda\right\}$$

says that a process for which perfect interpolation is possible need not be deterministic from the prediction view.

4.3 AN INFORMATION THEORETIC PREDICTOR MISMATCH MEASURE

The prediction error processes played the main role in studying mismatch robustness. Under the Gaussian assumption, the predictor mismatch can be measured by an information theoretic quantity.

Let $[R,\mu_\alpha,X]$ and $[R,\mu_\beta,Y]$ be stationary, zero-mean Gaussian processes with spectral densities $f(\lambda),\ g(\lambda) \in \eta$. Let $\{Z_n\}$ be an unknown process: either $\{X_n\}$ or $\{Y_n\}$. Define two hypotheses:

H_α - observed sample Z_{n-1}^N is from $\{X_n\}$

H_β - observed sample Z_{n-1}^N is from $\{Y_n\}$

where $Z_{n-1}^N = (Z_{n-N}, Z_{n-N+1},\ldots,Z_{n-1})$. If a linear least squares one-step predictor $\hat{Y}_n$ is used under H_β, then $\hat{Z}_n = \hat{Y}_n(Z_{n-1},Z_{n-2},\ldots)$. A mismatch occurs when H_α is true since

$$\hat{Y}_n(Z_{n-1},Z_{n-2},\ldots) = \hat{Y}_n(X_{n-1},X_{n-2},\ldots)$$

under H_α. By Bays rule,

$$P(H_\alpha|Z_{n-1}^N) = \frac{p(Z_{n-1}^N|H_\alpha)P(H_\alpha)}{p(Z_{n-1}^N|H_\alpha)P(H_\alpha)+p(Z_{n-1}^N|H_\beta)P(H_\beta)}$$

and a similar relation interchanging α and β holds. One then obtains

$$\log\frac{p(Z_{n-1}^N|H_\alpha)}{p(Z_{n-1}^N|H_\beta)} = \log\frac{P(H_\alpha|Z_{n-1}^N)}{P(H_\beta|Z_{n-1}^N)} - \log\frac{P(H_\alpha)}{P(H_\beta)} .$$

The first term of the righthand side is a posteriori information that H_α is favored, and the second term is a priori information. The lefthand side is therefore the information that H_α is favored resulting from the observation of Z_{n-1}^N. Taking the expectation yields then

$$E\left[\log \frac{P(Z_{n-1}^N | H_\alpha)}{P(Z_{n-1}^N | H_\beta)}\right] = \int_{-\infty}^{\infty} \cdots \int p_\alpha(z^N) \log \frac{p_\alpha(z^N)}{p_\beta(z^N)} \, dz^N$$

$$= I_N(\alpha|\beta) \qquad , \qquad (4.3.1)$$

where $I_N(\alpha|\beta)$ is called the Kullbuck-Leibler number, I-divergence or relative entropy [15], [16], [17]. Since (4.3.1) depends on N, and since in signal processing N can often be taken to be large, we define a predictor mismatch criterion as

$$\frac{1}{2} C_D = \lim_{N \to \infty} \frac{1}{N} I_N(\alpha|\beta) \quad . \qquad (4.3.2)$$

By the Gaussian assumption one obtains [15], [16], [17]

$$C_D = \lim_{N \to \infty} \frac{1}{N} \left\{ \mathrm{trace}(R_N Q_N^{-1} - I_N) - \log \frac{|R_N|}{|Q_N|} \right\} \qquad (4.3.3)$$

where R_N and Q_N are covariance matrices of X_n^N and Y_n^N. The following lemma provides a means of calculating the asymptotic value of (4.3.3).

<u>Lemma 4.3.1 [18]</u>: Let $\{\lambda_k^{(N)}, k=1,2,\ldots,N\}$ be eigenvalues of the NxN Toeplitz autocorrelation matrices $R_N = \{r(i-j)\}$. Define

$$f(\lambda) = \sum_{k=-\infty}^{\infty} r(k) e^{ik\lambda} \quad .$$

Let

$$M = \operatorname*{ess\,sup}_{\lambda} f(\lambda) < \infty$$

If $G(x)$ is continuous on $[0,M]$, then

$$\lim_{N \to \infty} \frac{1}{N} \sum_{k=1}^{N} G(\lambda_k^{(N)}) = \frac{1}{2\pi} \int_{-\pi}^{\pi} G\{f(\lambda)\}\, d\lambda \quad . \tag{4.3.4}$$

As a direct consequence of the lemma, one obtains

$$\lim_{N \to \infty} \frac{1}{N} \log \frac{|R_N|}{|Q_N|} = \frac{1}{2\pi} \int_{-\pi}^{\pi} \log \frac{f(\lambda)}{g(\lambda)}\, d\lambda \quad . \tag{4.3.5}$$

If $\operatorname{ess\,inf}_{\lambda}\{f(\lambda)/g(\lambda)\} > 0$,

$$\lim_{N \to \infty} \frac{1}{N} \operatorname{trace} R_N Q_N^{-1} = \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{f(\lambda)}{g(\lambda)}\, d\lambda \tag{4.3.6}$$

is obtained [19]. This yields the following.

<u>Proposition 4.3.2</u>: For Gaussian stationary processes of class η one-step predictors are hypothesis test mismatch-robust according to

$$C_D = \frac{1}{2\pi} \int_{-\pi}^{\pi} \left\{ \frac{f(\lambda)}{g(\lambda)} - 1 - \log \frac{f(\lambda)}{g(\lambda)} \right\} d\lambda$$

$$= d_{IS}(f,g) \quad . \tag{4.3.7}$$

$d_{IS}(f,g)$ is the Itakura-Saito distortion measure which was introduced in the Section 3.2.2.

Looking at the list of the spectral distortion measures in Section 3.2.2, one realizes, by virtue of (4.3.4), that

$$\frac{1}{2\pi} \int_{-\pi}^{\pi} \varphi(f/g)\, d\lambda = \lim_{N \to \infty} \frac{1}{N} \sum_{k=1}^{N} \varphi(\eta_k^{(N)})$$

where $\{\eta_k^{(N)}\}_{k=1}^N$ are the eigenvalues of $R_N Q_N^{-1}$. That is, all of the spectral ratio distortion measures are small $\Longleftrightarrow$ $\{\eta_k^N\}_{k=1}^N$ are close to 1 $\Longleftrightarrow$ $R_N Q_N^{-1}$ is close to a unit matrix asymptotically. This is a general criterion and applies to $d_{IS}(f,g)$ and $d_{cosh}(f,g)$, where the Gaussian assumption was needed for the derivation. By the choice of a proper φ, many other distortion measures can be presented. We note that, in the case of $d_\rho(f,g)$, square roots of the eigenvalues of R_N and Q_N are compared by the square difference distortion.

As a final comment of this section, we given Fig. 4.3.1 which shows the nonsymmetry of the spectral weightings of $d_1(f,g)$, $d_{nm}(f,g)$ and $d_{IS}(f,g)$. The nonsymmetry puts much weight when $f(\lambda)$ is larger than $g(\lambda)$. That is, peaks and an envelope of $f(\lambda)$ are emphasized when the order of $f(\lambda)$ is relatively low. This is regarded to be associative to human perception of speech distortion [22].

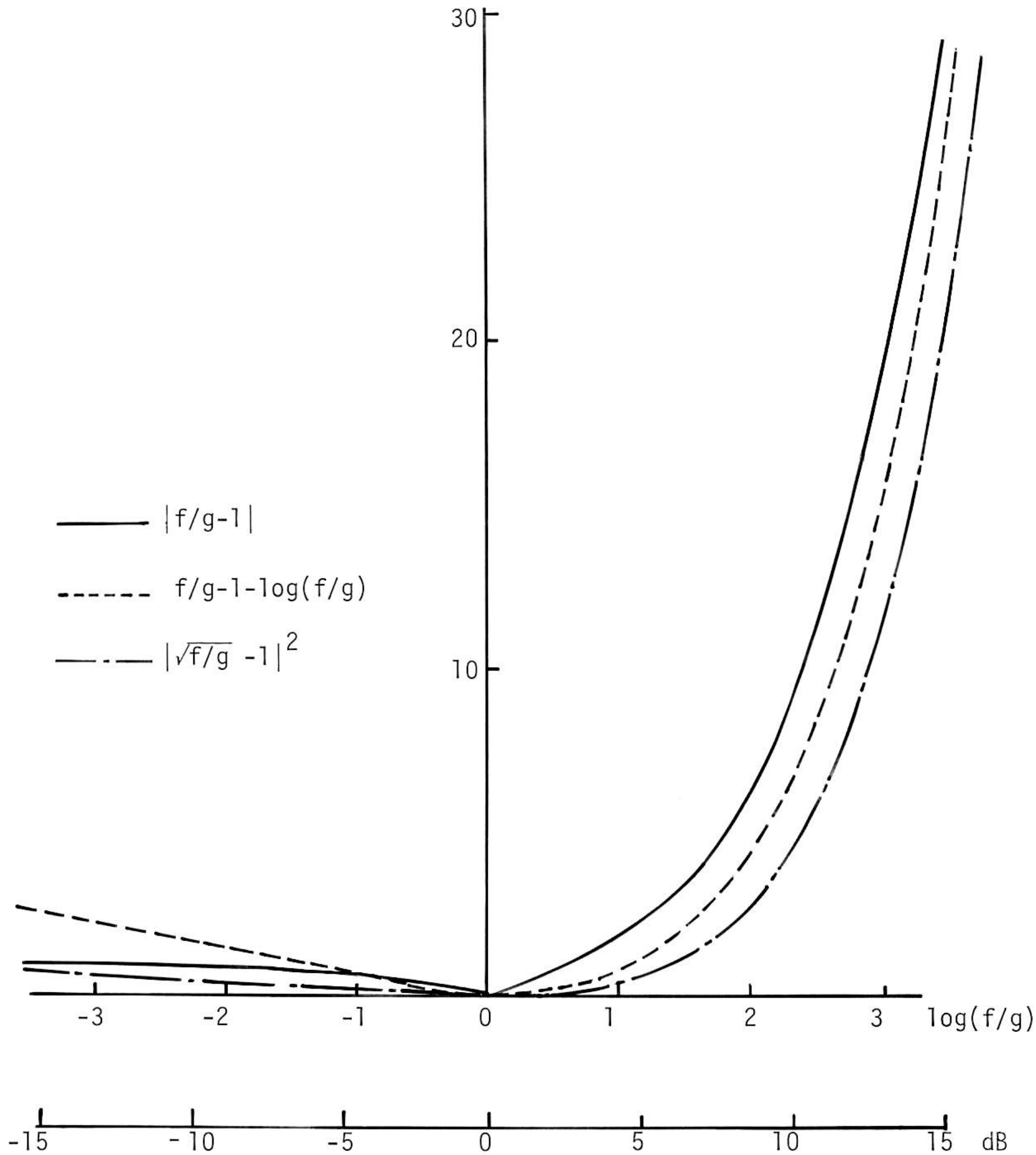


Fig. 4.3.1 Nonsymmetric spectral weighting

4.4 THE GENERAL CASE OF PREDICTOR MISMATCH

In the previous sections, $\{X_n\}$ and $\{Y_n\}$ belonged to the class $\mathcal{N}$. For the predictor mismatch problem, the class can be expanded to the weakly stationary processes which satisfy (2.1.2). k-step prediction can be treated here. For the k-step predictor mismatch, the mean square error criterion

$$C_A = \frac{|\epsilon_k(\beta|\alpha)^2 - \epsilon_k(\beta|\beta)^2|}{\epsilon_1(\beta|\beta)^2}$$

and the root mean square error criterion

$$C_B^2 = \frac{|\epsilon_k(\beta|\alpha) - \epsilon_k(\beta|\beta)|^2}{\epsilon_1(\beta|\beta)^2}$$

will be considered. Here,

$$\epsilon_k(\beta|\alpha)^2 = E|X_n - \hat{Y}_n(X_{n-k}, X_{n-k-1}, \ldots)|^2 \quad .$$

First, the spectral expression $H_\beta^{(k)}(\lambda)$ for the k-step predictor $\hat{Y}_n(Y_{n-k}, Y_{n-k-1}, \ldots)$ is derived following [12] and [13]. Define a normalized prediction process $\{\xi_n^\beta\}$ by

$$Y_n - \hat{Y}_n(Y_{n-1}, Y_{n-2}, \ldots) \triangleq \epsilon_1(\beta|\beta)\xi_n^\beta \quad .$$

The unique solution which minimizes $\epsilon_1(\beta|\beta)^2$ satisfies the projection property [12], [13]:

$$E(\xi_n^\beta \xi_m^\beta) = \delta_{n-m} \quad . \tag{4.4.1}$$

Define a number c_m^β by

$$c_m^\beta = E(Y_n \xi_{n-m}^\beta) \qquad .$$

Now, introduce random variables:

$$U_n^\beta = \sum_{m=0}^{\infty} c_m^\beta \xi_{n-m}^\beta \qquad ,$$

$$V_n^\beta = Y_n - U_n^\beta \qquad . \tag{4.4.2}$$

Note that (4.4.2) corresponds to the moving average representation.

Then, one obtains

$$E(\xi_m^\beta V_n^\beta) = E(U_m^\beta V_n^\beta) = 0 \qquad \forall\ m,n$$

$$E(\xi_n^\beta X_m^\beta) = 0 \qquad m < n \ . \tag{4.4.3}$$

Hence

$$E[\{\sum_{m=k}^{\infty} c_m^\beta \xi_{n-m}^\beta + V_n^\beta\}\{X_n^\beta - (\sum_{m=k}^{\infty} c_m^\beta \xi_{n-m}^\beta + V_n^\beta)\}] = 0 \qquad .$$

Then, the projection property implies

$$\hat{Y}_n(Y_{n-k}, Y_{n-k-1}, \ldots) = \sum_{m=k}^{\infty} c_m^\beta \xi_{n-m}^\beta + V_n^\beta \qquad . \tag{4.4.4}$$

Since $\{Y_n\}$ is weakly stationary, so are $\{\xi_n^\beta\}$, $\{U_n^\beta\}$ and $\{V_n^\beta\}$.

Therefore, the spectral representations [14] can be introduced.

$$
\left.
\begin{aligned}
Y_n &= \int_{-\pi}^{\pi} e^{in\lambda} Z_Y(d\lambda) \\
\xi_n^\beta &= \int_{-\pi}^{\pi} e^{in\lambda} Z_\xi^\beta(d\lambda) \\
U_n^\beta &= \int_{-\pi}^{\pi} e^{in\lambda} Z_U^\beta(d\lambda) \\
V_n^\beta &= \int_{-\pi}^{\pi} e^{in\lambda} Z_V^\beta(d\lambda)
\end{aligned}
\right\}
$$

where

$$
\left.
\begin{aligned}
E|Z_Y(d\lambda)|^2 &= \frac{1}{2\pi} G(d\lambda) \\
E|Z_\xi^\beta(d\lambda)|^2 &= \frac{1}{2\pi} d\lambda \\
E|Z_U^\beta(d\lambda)|^2 &= \frac{1}{2\pi} G_U(d\lambda) \\
E|Z_V^\beta(d\lambda)|^2 &= \frac{1}{2\pi} G_V(d\lambda)
\end{aligned}
\right\} \quad . \tag{4.4.5}
$$

That is, $G(\lambda)$, $G_U(\lambda)$ and $G_V(\lambda)$ are spectral distribution functions of $\{Y_n\}$, $\{U_n^\beta\}$ and $\{V_n^\beta\}$ respectively. From (4.4.3) and the spectral representations, one obtains

$$G(\lambda) = G_U(\lambda) + G_V(\lambda) \tag{4.4.6}$$

and

$$G_U(d\lambda) = |C_\beta(e^{i\lambda})|^2 d\lambda \tag{4.4.7}$$

where

$$C_\beta(e^{i\lambda}) \triangleq \sum_{m=0}^{\infty} c_m^\beta e^{-im\lambda} \quad . \tag{4.4.8}$$

The convergence is guaranteed in the mean square since $\{U_n^\beta\}$ is a second order process. From (4.4.7), $G_U(\lambda)$ is an absolutely continuous part of $G(\lambda)$ and

$$g(\lambda) = |C_\beta(e^{i\lambda})|^2 \qquad (4.4.9)$$

is the spectral density for $\{U_n^\beta\}$. This procedure is Wald's orthogonal decomposition of Y_n into U_n^β and V_n^β, whose spectral distribution functions are the absolutely continuous part, $G_U(\lambda)$, and the remaining part, $G_V(\lambda)$, of $G(\lambda)$. Equation (4.4.7) and the spectral representations imply that

$$Z_U^\beta(d\lambda) = C_\beta(e^{i\lambda})Z_\xi^\beta(d\lambda) \quad .$$

Since $V_n^\beta \perp U_n^\beta$, one obtains

$$Z_Y(d\lambda) = C_\alpha(e^{i\lambda})Z_\xi^\beta(d\lambda) + Z_V^\beta(d\lambda) \quad .$$

Let ξ_n^α correspond to $\Xi^\alpha(\lambda)$ in the spectral representation, then

$$\xi_n^\beta = \int_{-\pi}^{\pi} \Xi_n^\beta(\lambda)\{C_\beta(e^{i\lambda})Z_\xi^\beta(d\lambda) + Z_V^\beta(d\lambda)\} \ . \qquad (4.4.10)$$

By equating (4.4.5) and (4.4.10), one obtains

$$\Xi_n^\beta(\lambda) = \begin{cases} \dfrac{e^{in\lambda}}{C_\beta(e^{i\lambda})} & \text{a.e. wrt Lebesque measure } \lambda \\ 0 & \text{a.e. wrt } G_V \text{ measure} \end{cases}$$

Hence G_V increases only on $S_\beta \subset [-\pi,\pi)$ with $\lambda(S_\beta) = 0$. If $\hat{Y}_0(Y_{-k},Y_{-k-1},\ldots)$ corresponds to $H_\beta^{(k)}(\lambda)$, then

$$\hat{Y}_n(Y_{n-k},Y_{n-k-1},\ldots) = \int_{-\pi}^{\pi} e^{i(n-k)}H_\beta^{(k)}(\lambda)Z_Y(d\lambda). \qquad (4.4.11)$$

Now one obtains, from (4.4.4) and (4.4.11) with the aid of the spectral representations, that

$$Y_n-\hat{Y}_n(Y_{n-k},Y_{n-k-1},\ldots) = \sum_{m=0}^{k-1} c_m^{\beta}\xi_{n-m}^{\beta}$$

$$= \int_{-\pi}^{\pi} e^{i(n-k)\lambda}\{e^{ik\lambda}-H_{\beta}^{(k)}(\lambda)\}Z_Y(d\lambda)$$

$$= \int_{-\pi}^{\pi} e^{i(n-k)\lambda}\{e^{ik\lambda}-H_{\beta}^{(k)}(\lambda)\}Z_U^{\beta}(d\lambda) \quad .(4.4.12)$$

Then,

$$\epsilon_k^2(\beta|\beta) = \sum_{m=0}^{k-1} |c_m^{\beta}|^2$$

$$= \frac{1}{2\pi}\int_{-\pi}^{\pi} |e^{ik\lambda}-H_{\beta}^{(k)}(\lambda)|\,G_U(d\lambda)$$

$$+ \frac{1}{2\pi}\int_{-\pi}^{\pi} |e^{ik\lambda}-H_{\beta}^{(k)}(\lambda)|\,G_V(d\lambda)$$

$$= \frac{1}{2\pi}\int_{-\pi}^{\pi} |e^{ik\lambda}-H_{\beta}^{(k)}(\lambda)|^2 G_U(d\lambda) \quad . (4.4.13)$$

Hence

$$H_{\beta}^{(k)}(\lambda) = e^{ik\lambda} \quad \text{for } \lambda \in S_{\beta} \quad .$$

By (4.4.7), (4.4.8) and (4.4.13),

$$H_{\beta}^{(k)}(\lambda) = e^{ik\lambda}\,\frac{\sum_{m=k}^{\infty} c_m^{\beta}e^{-im\lambda}}{C_{\beta}(e^{i\lambda})} \quad \text{for } \lambda\in[-\pi,\pi)-S_{\beta} \quad .$$

The mean square mismatch error is then obtained as:

$$\epsilon_k(\beta|\alpha)^2 = E|X_n-\hat{Y}_n(X_{n-k},X_{n-k-1},\ldots)|^2$$

$$= \frac{1}{2\pi}\int_{-\pi}^{\pi} |e^{ik\lambda}- H_{\beta}^{(k)}(\lambda)|^2 F(d\lambda) \quad .$$

Then

$$C_A = \frac{|\epsilon_k^2(\beta|\alpha) - \epsilon_k^2(\beta|\beta)|}{|c_0^\beta|^2}$$

$$= \frac{1}{2\pi} \left| \int_{-\pi}^{\pi} \left| \frac{K_\beta^{(k)}(\lambda)}{c_0^\beta} \right|^2 F(d\lambda) - \int_{-\pi}^{\pi} \left| \frac{K_\beta^{(k)}(\lambda)}{c_0^\beta} \right|^2 G(d\lambda) \right| \qquad (4.4.14)$$

and

$$C_B^2 = \frac{|\epsilon_k(\beta|\alpha) - \epsilon_k(\beta|\alpha)|^2}{|c_0^\beta|^2}$$

$$= \frac{1}{2\pi} \left| \sqrt{\int_{-\pi}^{\pi} \left| \frac{K_\beta^{(k)}(\lambda)}{c_0^\beta} \right|^2 F(d\lambda)} - \sqrt{\int_{-\pi}^{\pi} \left| \frac{K_\beta^{(k)}(\lambda)}{c_0^\beta} \right|^2 G(d\lambda)} \right|^2 \qquad (4.4.15)$$

where

$$K_\beta^{(k)}(\lambda) = e^{ik\lambda} - H_\beta^{(k)}(\lambda)$$

$$= \begin{cases} 0 & \lambda \in S_\beta \\ e^{ik\lambda} \dfrac{\sum_{m=0}^{k-1} c_m^\beta e^{-m\lambda}}{c_\beta(e^{i\lambda})} & \lambda \in [-\pi, \pi) - S_\beta \end{cases} . \qquad (4.4.16)$$

One can interpret $K_\beta^{(k)}(\lambda)$ as a spectral representation of an inverse filter for $[R, \mu_\beta, Y]$ which corresponds to the k-step prediction. We now have the following:

Proposition 4.4.1: k-step predictors are mean square error mismatch-robust under the spectral distortion measure (4.4.14). The root mean square error mismatch-robustness holds according to the spectral distortion measure (4.4.15).

To close this chapter, the following comments are given:

(i) For purely non-deterministic processes of either an auto-regressive type (3.3.2) or a moving average type (4.4.2),

$$(c_0^{\alpha})^2 = \sigma_f^2 = \exp\left\{\frac{1}{2\pi}\int_{-\pi}^{\pi} \log f(\lambda)\,d\lambda\right\} \tag{4.4.17}$$

(ii) For the k-step predictor mismatch of the same purely non-deterministic processes, $d_1(f,g)$, $d_{nm}(f,g)$ and $\bar{\rho}(\alpha/\beta,\beta/\beta)$ can ensure the mismatch robustness in a modified sense. That is stated in the next corollary.

Corollary 3.1.2: For purely non-deterministic processes, k-step predictors are mismatch robust according to

$$C_A' = \frac{|\epsilon_k(\alpha|\beta)^2 - \epsilon_k(\beta|\beta)^2|}{k\sum_{m=0}^{k-1}|c_m^{\beta}|^2} \leq d_1(f,g)$$

and

$$(C_B')^2 \triangleq \frac{|\epsilon_k(\alpha|\beta) - \epsilon_k(\beta|\beta)|^2}{k\sum_{m=0}^{k-1}|c_m^{\beta}|^2} \leq d_{nm}(f,g) \leq \bar{\rho}(\alpha/\beta,\beta/\beta).$$

Chapter 5

SPEECH COMPRESSION

5.1 LPC CODEBOOK

This section considers a direct application of spectral distortion measures to Linear Predictive Coding (LPC) speech compression systems. An LPC speech system extracts three sets of parameters for each $\simeq$ 20 ms frame of speech data: filter coefficients, a gain, and a pitch (depending on the classification of "voiced" or "unvoiced") [20]. Since the filter coefficients, which are usually in the form of reflection coefficients, correspond to the shape of human vocal tract [36], the parameters seem to have considerable redundancy from a communication engineering viewpoint. In other words, a small deviation of reflection coefficients which corresponds to a little variation of the acoustic tube area will yield barely perceptible differences of sound. This evidence leads to the idea of making a finite codebook for LPC parameter vectors. If a simple yet rich enough codebook can be made, the communication system needs to use only a low data rate by specifying an index to a "codeword" consisting of a set of filter coefficients.

An early trial in this line is found in Chaffee's work [26]. Therein, he made an initial guess from twenty-five phonemes, and then added a new parameter set if a given frame in a training sequence had a minimum distortion over the codebook of $d^2_{ncm} > 0.6$. The total number or the cardinality of the codebook was 256 i.e., 8 bits were needed to specify a set of filter coefficients. Instead of this covering argument, another way of finding a codebook exists [37], [48]. The following discussion is based on [48] and the method uses an optimum quantization without a statistical model [49]. A detailed explanation of the codebook design will be found in Buzo's work [37].

We are given a set of vector reflection coefficients $\{\underset{\sim}{k}_i\}_{i=1}^{M}$ where $\underset{\sim}{k}_i = (k_{1i}, k_{2i}, \ldots, k_{mi})^T$, and i corresponds to the frame number. We wish to design an N-level quantizer q or a codebook $\mathcal{C}$ with the cardinality N. The quantizer makes a partition $\{P_j; j=0,1,\ldots,N-1\}$ of the m-dimentional interval $(-1,1)^m$ and the collection of reproduced reflection coefficient vector $\{\hat{\underset{\sim}{k}}_j\}$, $j=0,1,\ldots,N-1$. The resulting quantization is

$$q(\underset{\sim}{k}_i) = \hat{\underset{\sim}{k}}_j \quad \text{if} \quad \underset{\sim}{k}_i \in P_j \quad .$$

The quantizer is designed to minimize $E\{d(\underset{\sim}{k}_i, q(\underset{\sim}{k}_i))\}$, where $d(\cdot,\cdot)$ is a spectral distortion measure. Since we are given only a sample sequence $\{\underset{\sim}{k}_i\}_{i=1}^{M}$, the following training procedure [49] is adopted by assuming stationarity and ergodicity.

(1) An initial guess $q^{(0)}$ is made (0^{th} iteration).

(2) Given the quantizer $q^{(n)}$ of the n^{th} iteration, define a new partition

$$P_j^{(n+1)} = \{\underset{\sim}{k} \in (-1,1)^m : d(\underset{\sim}{k}, P_j^{(n)}) \leq d(\underset{\sim}{k}, P_\ell^{(n)}), j \neq \ell\}$$

with some tie-breaking rule.

(3) Find the optimum output levels $\{\hat{\underset{\sim}{k}}_j^{(n+1)}\}_{j=0}^{N-1}$ that minimize

$$\sum_{\substack{nL \leq i < (n+1)L \\ \underset{\sim}{k}_i \in P_j^{(n+1)}}} d(\underset{\sim}{k}_i, \hat{\underset{\sim}{k}}_j^{(n+1)}) \quad , \quad j = 0,1,\ldots,N-1 \quad .$$

(4) Calculate the sample average distortion

$$\Delta_{n+1} = \sum_{i=nL}^{(n+1)L-1} d(\underset{\sim}{k}_i, q^{(n+1)}(\underset{\sim}{k}_i)) \quad .$$

If $|\Delta_{n+1} - \Delta_n| \leqq \epsilon$, then quit, otherwise increment n and go to (2).

The LPC codebook approach provides a subjective test for spectral distortion measures. Suppose we have several codebooks of the same cardinality, $\mathcal{C}_{IS}$, $\mathcal{C}_{cm}$, $\mathcal{C}_{log}$ etc., which are made using the corresponding distortion measures. If a specific codebook gives the best reproduction of speech, then the corresponding distortion might be more consistent with the human perception of speech. It is easily conceived, however, that for the comparison of large cardinality codebooks, no significant difference will appear because of the linear relationships of Proposition 3.3.6.

A codebook of this type is used in the next section and Chapter 6.

5.2 TREE CODING BY PARALLEL TREE SEARCH

5.2.1 Universal Coding and Parallel Tree Search

As was explained in Section 5.1, an LPC coding system usually makes models and sends parameters blockwise. On the other hand, there are other data compression systems which encode waveforms directly. Such waveform coding systems need a higher data rate than the LPC system, yet some are still in the low bit rate range. The advantages of the waveform encoding are:

(i) the possibility of on-line usage because of direct coding on waveforms;

(ii) implementation complexity is usually much less than the on-line LPC.

Anderson and Bodie [39] were the first to use a tree coding of speech waveforms. A tree coding consists of finding the best path map if a tree structure for an input under a given criterion. A good introduction is given in [49], and discussions on parallel tree coding which include the usual tree coding as a special case will be given in this section. Since their method [39] used a tree search directly applied to speech waveforms yielding piecewise linear reproduced waveforms, a data rate of 2 bits/sample was needed.

Subsequently, Wilson and Husain [40] added an LPC adaptive scheme to tree coding and achieved 1.048 bits/sample coding of speech. In their system, an adaptive method is employed by sending a few LPC parameters which are estimated on-line. This scheme is a deviation from traditional waveform encoders since the system uses a reduced on-line LPC algorithm. By using only a portion of LPC, a good code can be obtained with nearly the same implementation complexity, yet yielding a lower bit rate.

Looking at speech waveforms, one easily realizes that there are distinctive modes. For example, voiced parts are of high amplitude and considerably regular. On the contrary, unvoiced parts are of low level and no periodicity is observed. One then realizes that if one uses several trees in parallel, each of which is responsible for the coding of the specific mode, then the coding system covers the entire class of speech within reasonable distortion. This idea was motivated by the concept of universal coding [50] which works well for classes of possible input processes. We design a parallel tree encoder to be universal to input speech waveforms. A parallel tree search is possible since the increase of the number of trees will grow the implementation complexity only linearly.

Figure 5.2.1 explains the idea of the parallel tree search, where a two tree case is shown. A mapping $F(\cdot)$ maps a binary sequence which corresponds to the path to a reproduced waveform. A tree encoding finds the path whose mapping gives the minimum error.

Since the number of branches blow up exponentially, all possible paths to be chosen cannot be memorized. Hence a suboptimal method must be used. Currently, the M-algorithm [42] is the simplest effective method. The M-algorithm looks ahead up to M steps and memorizes 2^M paths. If one future branch generates the smallest error, the current trunk corresponding to the path is chosen. The algorithm then proceeds one step ahead. This is a kind of a sliding-block code [1] called an incremental tree encoding. The parallel tree search is understood in the following way. For convenience, let tree 0 correspond to unvoiced waveform segments and the tree 1 to voiced parts (in the actual system,

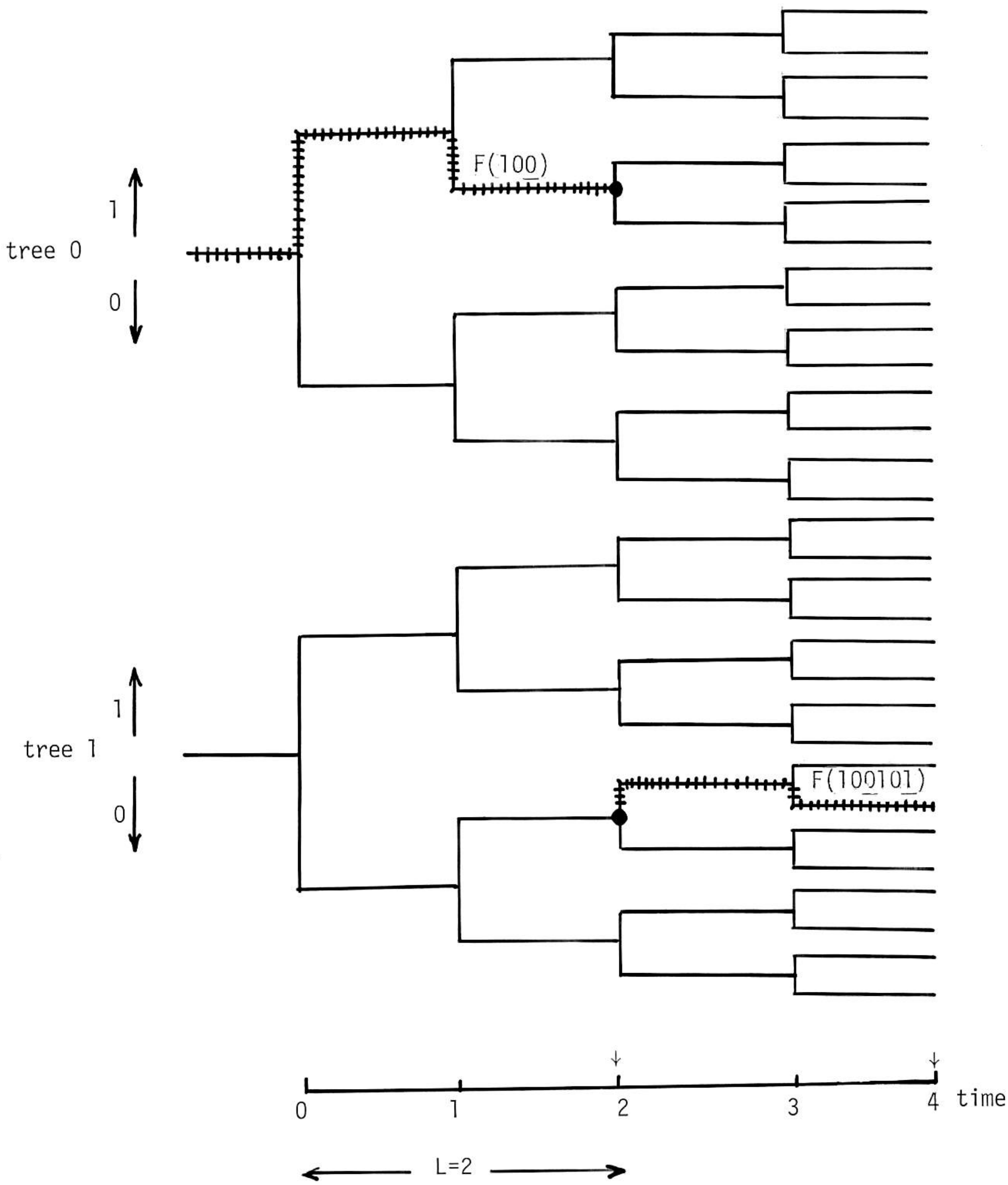


Fig. 5.2.1 Two-tree coding

eight trees are used). In the first block of length L = 2, tree 0 gives a path map 10 and tree 1 gives a path map, say 01. Let tree 0 with a codeword 10<u>0</u> give a smaller distortion than a codeword 01<u>1</u> which is generated by tree 1. Then, 10<u>0</u> become the coded word, and the input waveform is regarded as unvoiced. In the next block, 10<u>1</u> is chosen giving the decision that the second block is voiced. At the end of each block, one bit side information is used to specify the chosen tree. The side information bit is underscored.

We note in our application tree coding is a better choice than trellis coding [51] since digital filters with feedback having infinite impulse responses are used in our system.

5.2.2 Fake Process Decoder

Since we avoid using on-line LPC adaptation, a good decoder for the tree search is needed since only a small number of trees will be allowed in practical cases. A fake process decoder proposed by Linde and Gray [41] is a promising candidate. By digital simulations, this decoder with a matched tree search encoder has been shown to outperform predictive quantizers when the waveforms are first order AR processes. It is also believed that the fake process approach will show good performance on higher order AR processes.

The fake process data compresser is a combination of a shift register, a scrambler, a non-linear mapping that performs a probability distribution transformation, and memory devices such as AR filters. All of these components are driven by a binary sequence obtained by the parallel tree search. Figure 5.2.2 shows the total system of the parallel tree search encoder/fake process decoder. An important attribute is that the encoder contains a replica of the decoder. The function of each part is explained as follows. The weighted summation and the scrambler generate uncorrelated uniform random numbers on a finite subset of $[0,1)$. The probability distribution transformer $F_{\xi_i}^{-1}$ is used to shape the uniform random number to a desired distribution. The transformed random numbers are input to AR filters of ladder forms to color the pseudo-white processes. According to the parallel tree search's side information, the best reproduction is selected at the switch.

5.2.3 Specifications of the System

The followings are the actual selection of the system specification.

(i) shift register length: $k = 3$.

(ii) tree search depth: $M = 4$.

(iii) scrambler: period 1/3 scrambler [41].

(iv) $F_{\xi_i}(\cdot)$: Laplace distribution (double exponential density).

(v) (number of ladder forms) = (number of trees) = 8.

(vi) (order of ladder forms) = 5.

(vii) quantization of reflection coefficients: uniform, 5 bits.

The reasons for the above choices (i)-(vii) are as follows: On (i) and (ii), we have to choose k and M as small as possible so that $k \leq M$, since less complex implementation is desired. On the scrambler, the 1/3 period scrambler is reported sufficient in [41] yet it is simple.

On the decision of the inverse distribution, experiments on actual

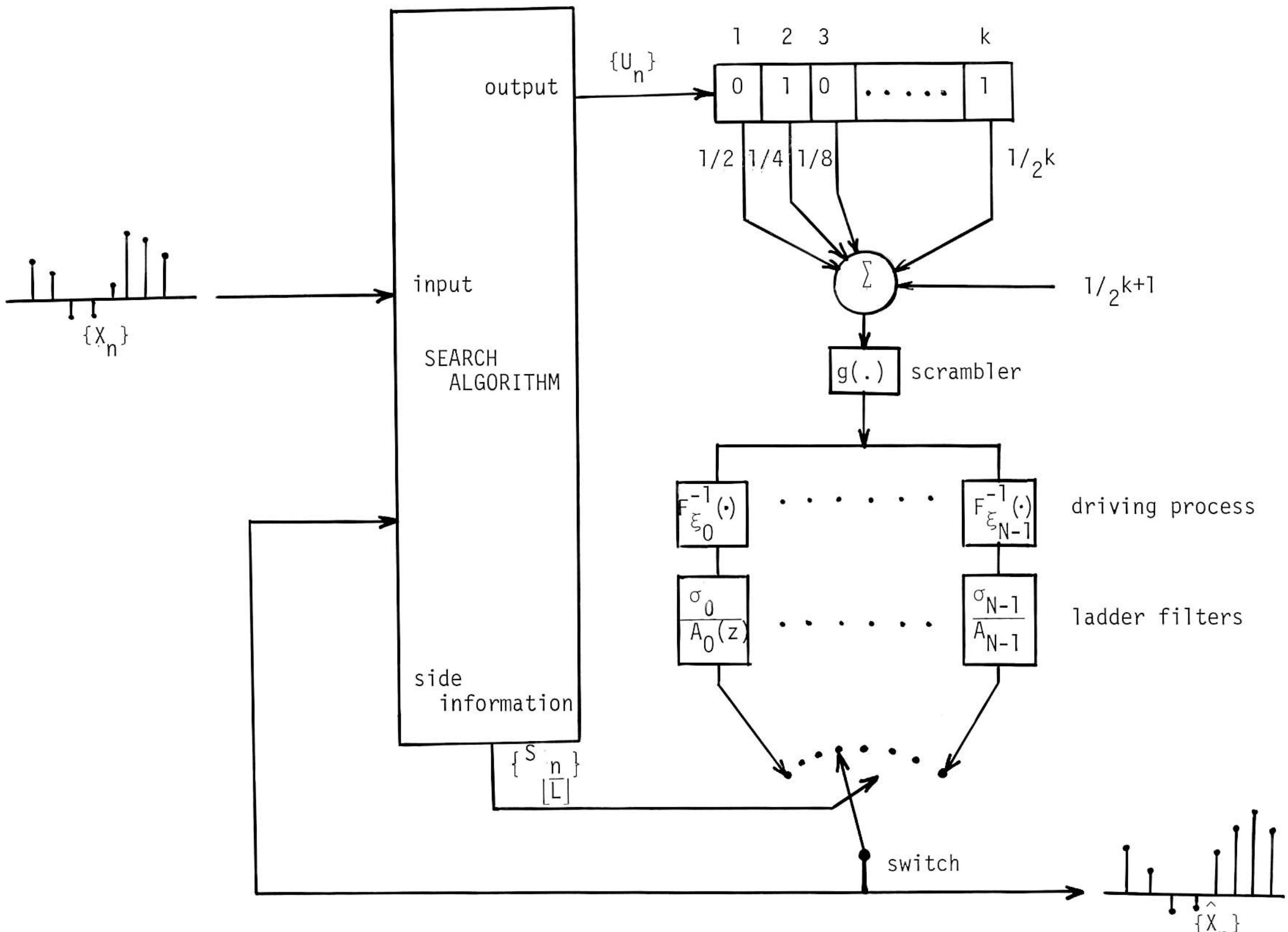


Fig. 5.2.2 Parallel tree search encoder/fake-process decoder

speech were performed. First, the speech was parsed into segments of 8 ms. Each chunk was then categorized into eight classes according to the innovation gains and the second reflection coefficients, and the representing ladder forms for each class were found. Each class of waveform segments was then inverse-filtered and the histogram was taken to estimate the probability density functions. An important point is that the probability density is obtained from the output of a mismatched inverse filter. If the chunk of waveform is inverse-filtered by its own exact inverse filter, then the output must be nearly white or very periodic. This is not our case since the filters are fixed, i.e., on-line LPC is not used. Figure 5.2.3 shows the probability density function obtained from "Thieves who rob friends deserve jail," with a male speaker. They can be well-approximated by the Laplace distribution, i.e., the double-sided exponential density. Examples of other sentences are omitted since their behavior is similar. One reason for the Laplace density might be as follows: In the speech case, pitch pulse trains with fluctuations appear in voiced sounds. Since their amplitudes are larger than unvoiced noisy-looking parts, the decay is slower than Gaussian densities. The sharp spike around the origin is formed by low level unvoiced parts.

The number of ladder forms was chosen as eight because of the limitations of PDP-11 UNIX (multi-user) system. When a four tree case was tried, the reproduced voice was sometimes unsatisfactory. The eight tree (or eight ladder form) system is the smallest satisfactory one that produces a highly intelligible reproduction.

The spectral distortion measure is applied to the selection of reference ladder filters. Using the method explained in Section 5.1,

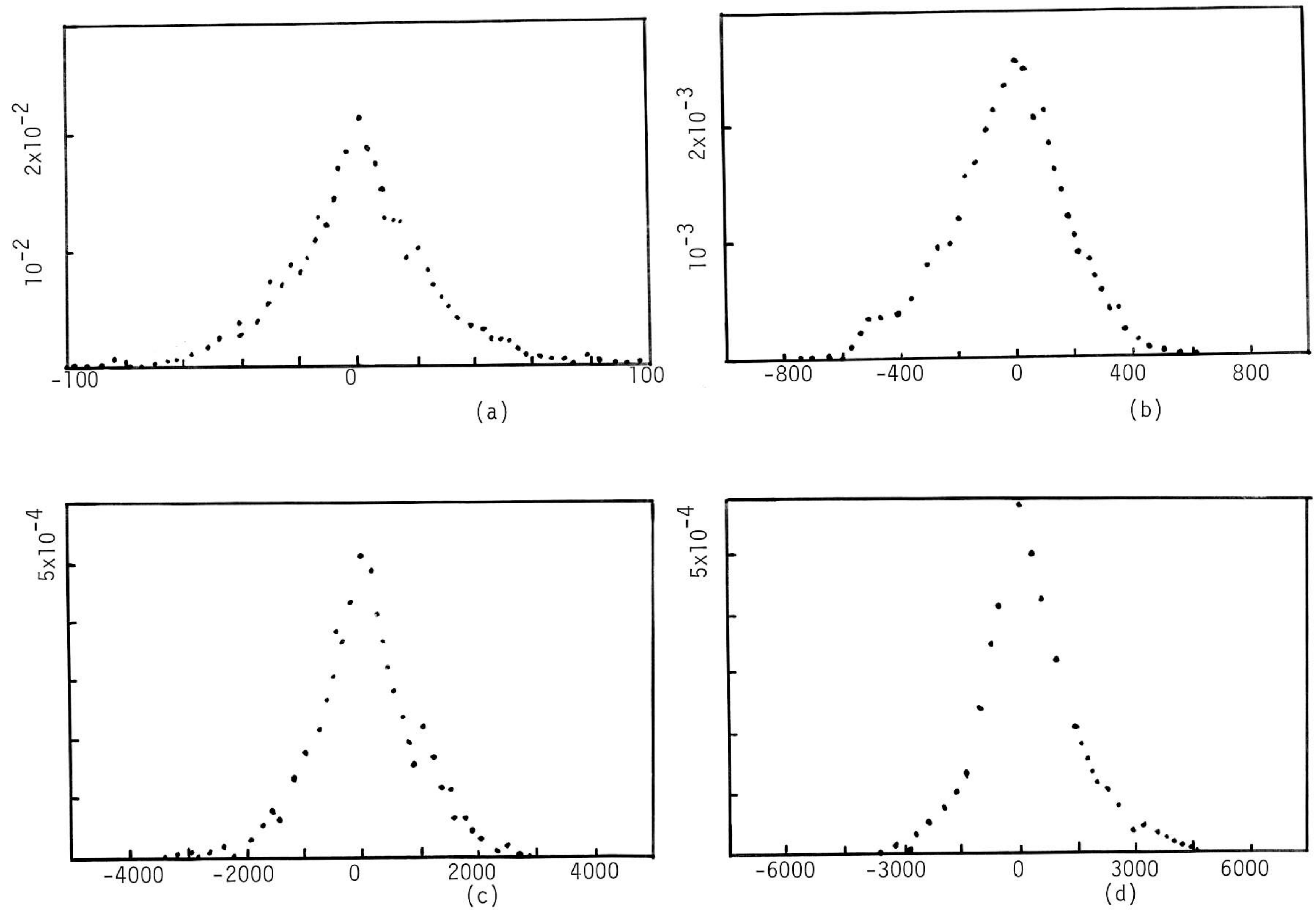
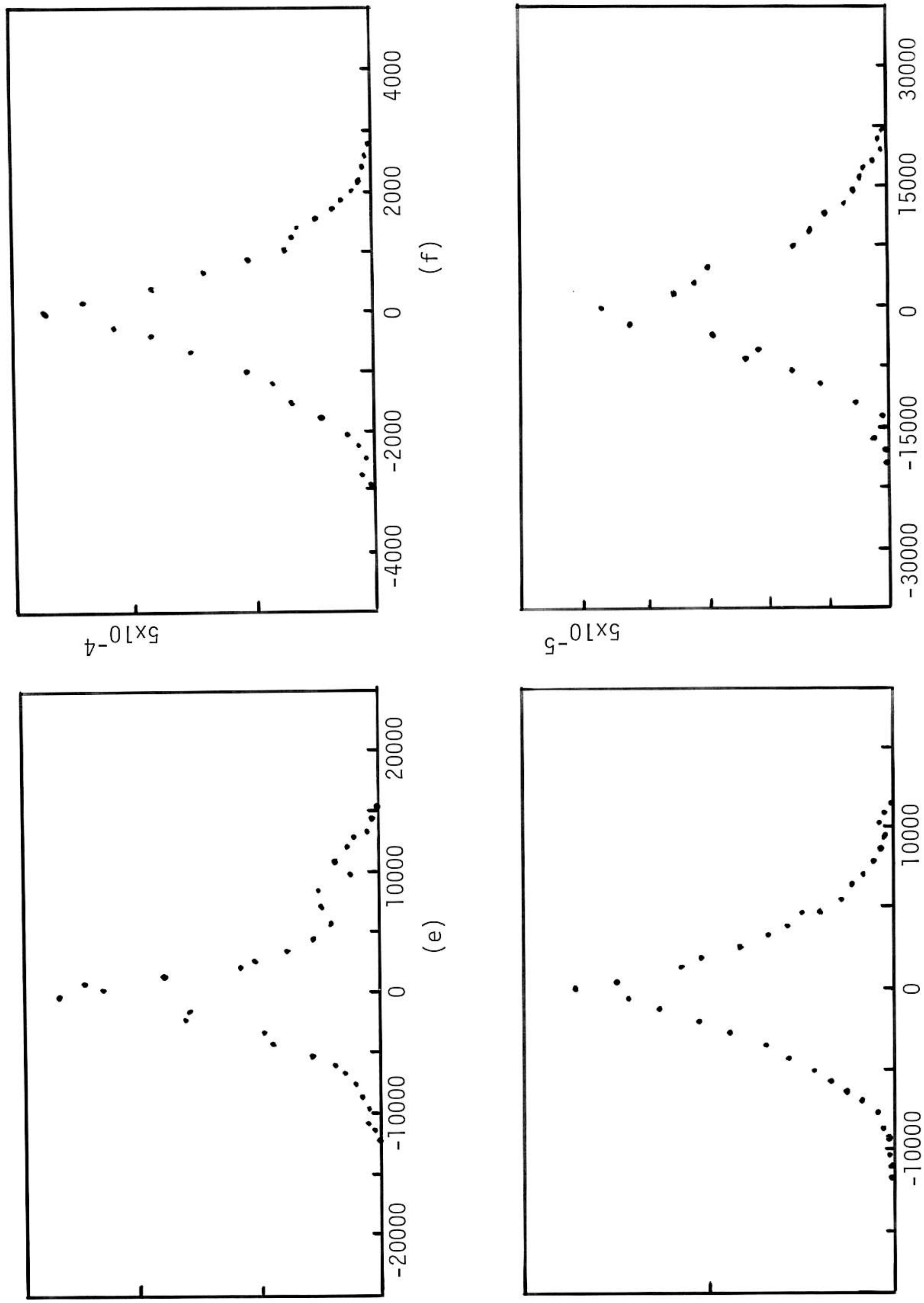


Fig. 5.2.3 Probability densities of mismatch inverse filtered processes



(e)
-20000
-10000
0
10000
20000
10^-4
(f)
-4000
-2000
0
2000
4000
5x10^-4
(g)
-10000
0
10000
10^-4
(h)
-30000
-15000
0
15000
30000
5x10^-5

sixteen reference ladders were chosen. It was found that there is a class of sounds which can be reproduced in a good quality by several subtrees. Such examples are /s/, /th/ and /shh/, etc. The reason is that any subsystem can generate noisy sounds easily. The reference ladders corresponding to these specific phonemes are eliminated, and eight out of the sixteen candidates are adopted. Since there is a report [43] that eighth order ladder filters are enough, the order was reduced to fifth because a lower complexity than the LPC system is desired here. The five bit quantization seems the lowest possible when the number of filters is eight and the uniform quantization is adopted.

5.2.4 Error Criteria

Since an incremental tree search by M-algorithm is adopted, the square error of the original and the reproduced voice is the simplest criterion for the implementation. However, this measure is subjectively inadequate. The square error of waveform is used merely because of the simplicity. Wilson and Husain [40] introduced a weighted square error criterion by realizing that low frequency errors are not very destructive on speech quality. The criterion is

$$\sum_n h(X_n - \hat{X}_n)^2$$

where X_n and $\hat{X}_n$ are the original and the reproduction, and $h(\cdot)$ is an impulse response of the high-pass filter

$$H(z^{-1}) = 1 + a_1 z^{-1} + a_2 z^{-2} \quad . \tag{5.2.1}$$

In their work, $a_1 = -0.59$ and $a_2 = -0.39$ were chosen.

Other possible measures are the spectral distortions. That experiment is under investigation.

5.2.5 A Variable-Length Coding

There are many silent parts in normal speech. Moreover, even if we listen to continuous sounds, there are considerable durations of very low level parts. For these portions, information on the waveform is totally useless since normal listeners do not perceive it. In our selection of the eight ladder forms, there is a ladder filter which corresponds to this silent mode. If the tree search finds that the input waveform corresponds to such a silent part, then the encoder omits sending the waveform information, but sends only the side information which tells that silence is sent. This scheme can be easily added to the original system, reducing the data rate via simple variable-length coding.

5.2.6 Results

Experiments were performed on "The pipe began to rust while new," and "Add the sum to the product of these three," of female voice as well as "Cats and dogs each hate the other," of male voice. Both simple square error and the frequency weighted squared error criteria worked well yielding highly intelligible reproductions. However, the parameters chosen in [40] yielded worse quality than the simple square error criterion because low pass filtering occurs. Figure 5.2.4a is the original speech. Figure 5.2.4b is the reproduction using the square error criterion. There are parts corrupted by high frequency noise. This is the

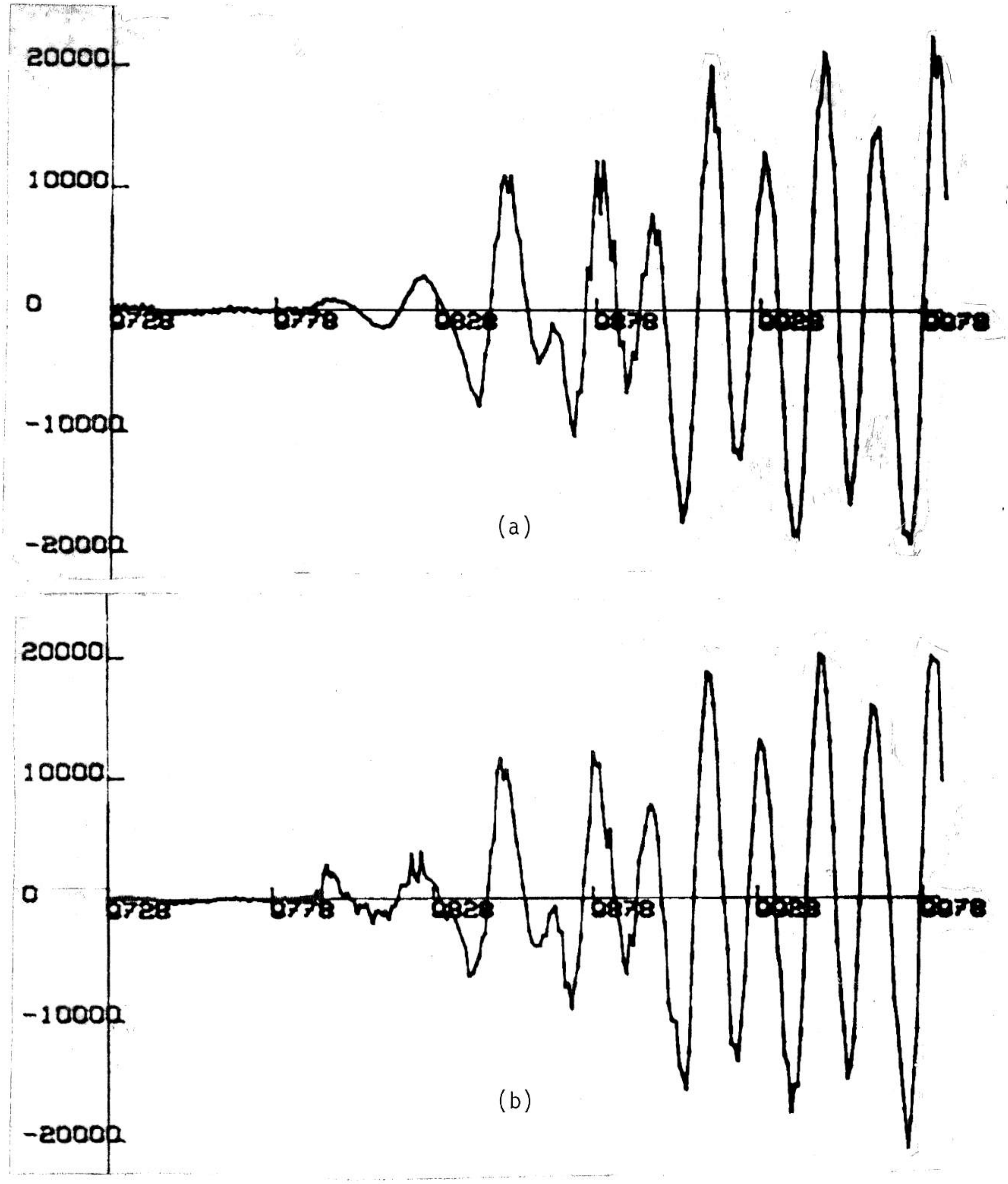
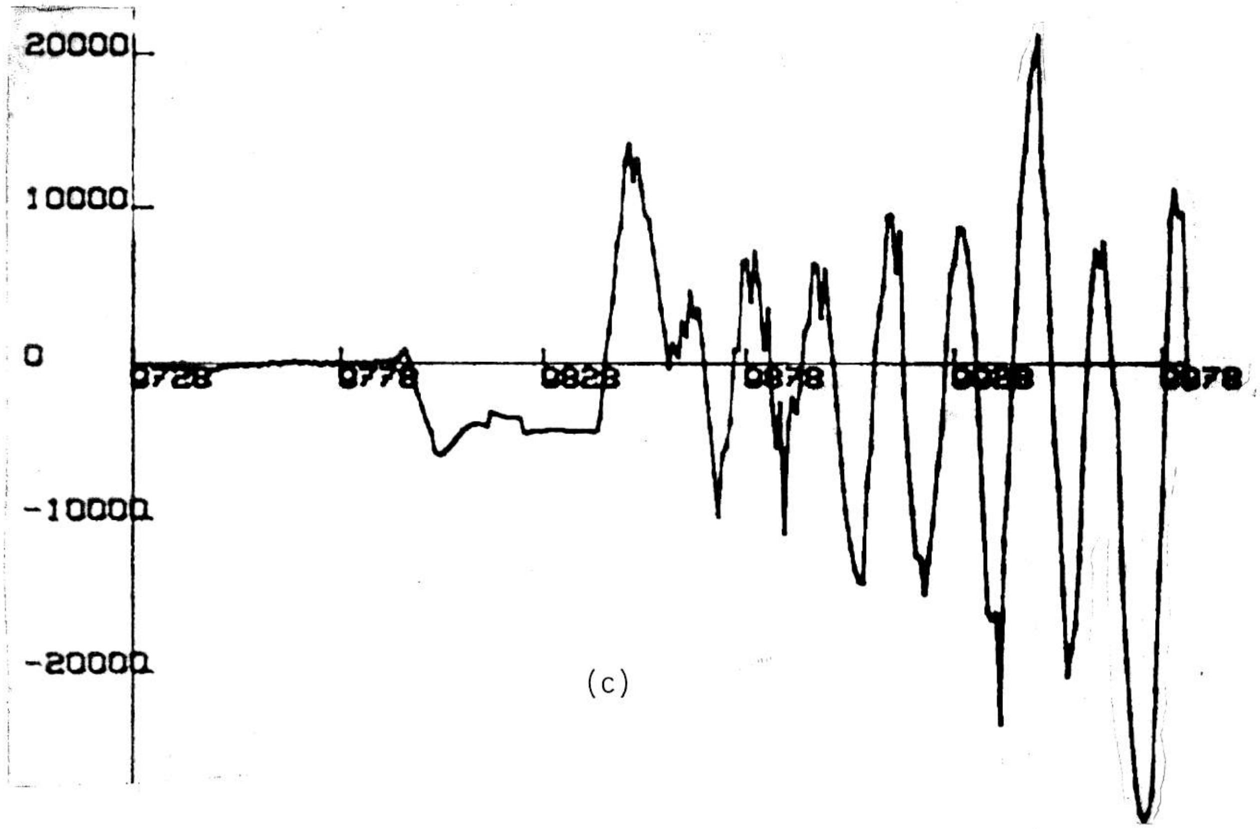
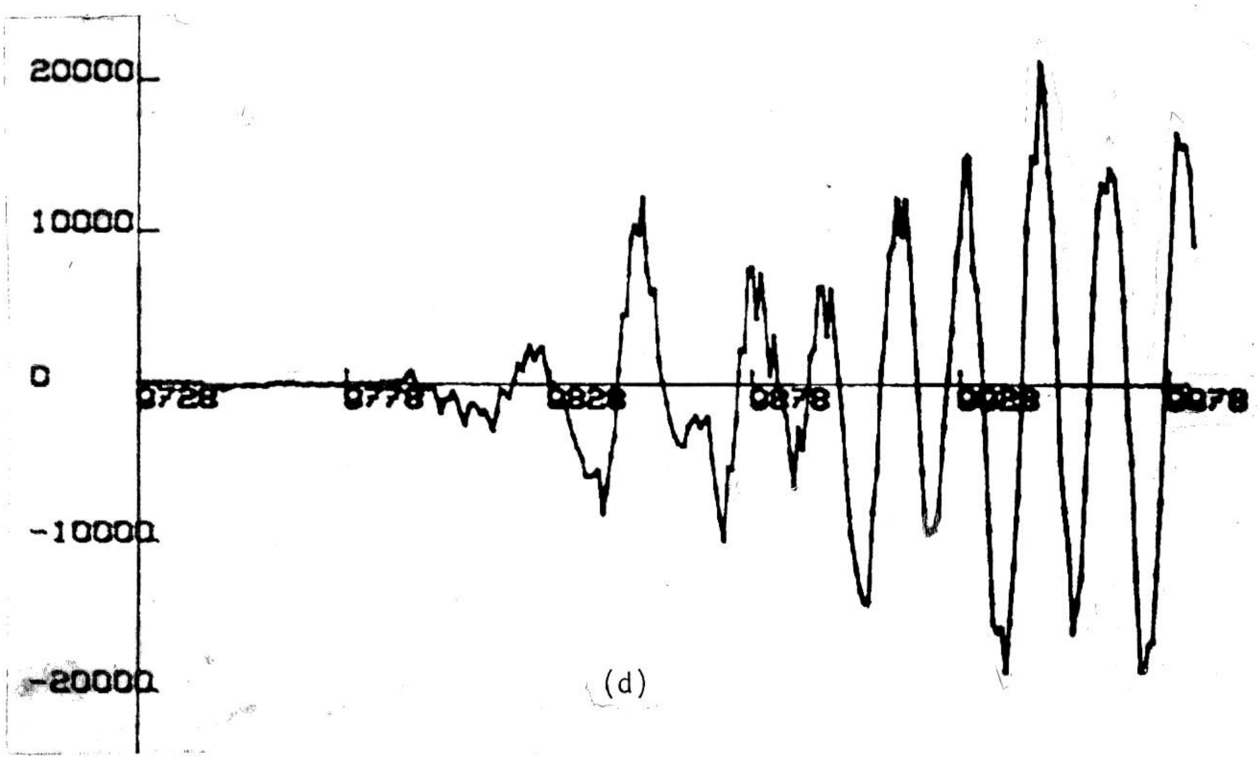


Fig. 5.2.4 Original and reproduced speech waveforms



(c)



(d)

defect of that criterion. Fig. 5.2.4c is a reproduction using the frequency weighted square error with parameters chosen in [40]. The transient part is insufficiently reproduced. Hence, for our system, $a_1 = -0.59$, $a_2 = -0.39$ are not good choices. By parameter adjustment comparing the original and the reproduced spectra, $a_1 = -0.5$, $a_2 = -0.1$ are found to be better. The final criterion was judged the best by listening tests among the waveform error criteria. Figure 5.2.4d shows the reproduced waveform from this criteion.

Figure 5.2.5 shows scatter plots of distortion measures between the original and the reproduction of "The pipe began to rust while new," of female voice. The figures show total, normalized and gain term of distortion measures respectively. d_{IS} and d^2_{log} are compared there. One finds a couple of tendencies of these distortion measures from the figures. The bifurcation is found in Fig. 5.2.5a and Fig. 5.2.5c. Since

$$d_{log}(\sigma_f^2, \sigma_g^2)^2 = \left(\log \frac{\sigma_f^2}{\sigma_g^2}\right)^2$$

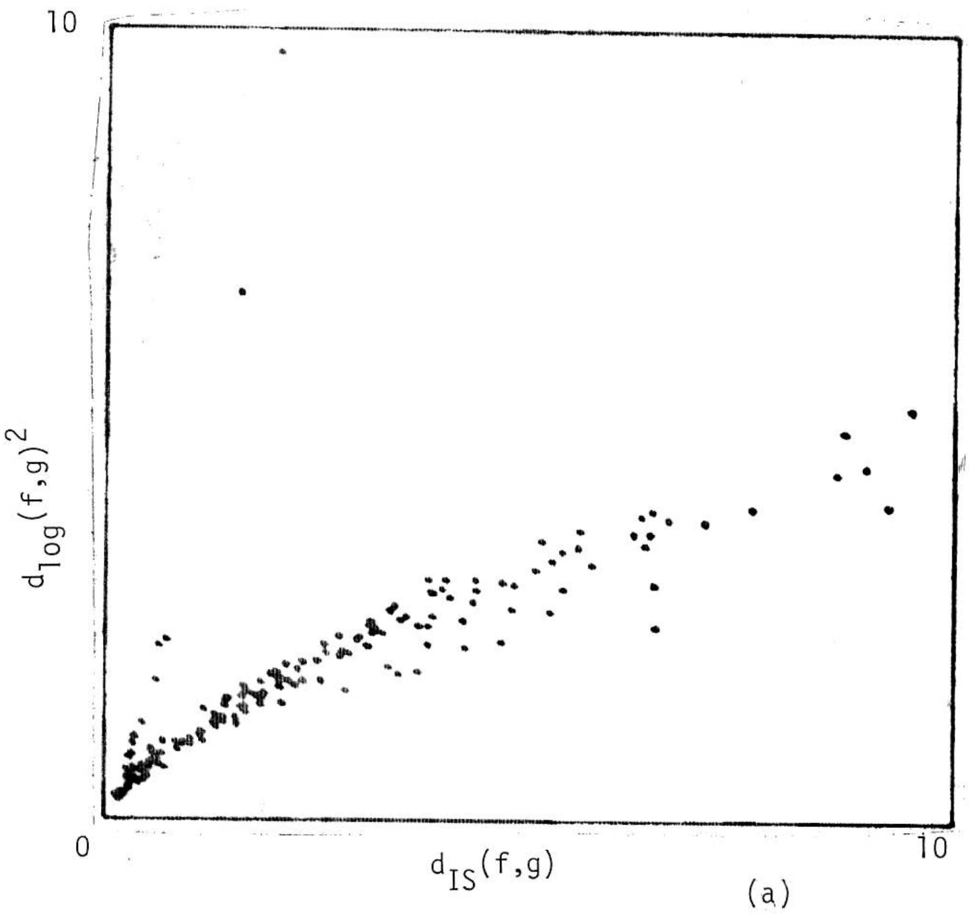
and

$$d_{IS}(\sigma_f^2, \sigma_g^2) = \frac{\sigma_f^2}{\sigma_g^2} - 1 - \log \frac{\sigma_f^2}{\sigma_g^2} ,$$

one obtains

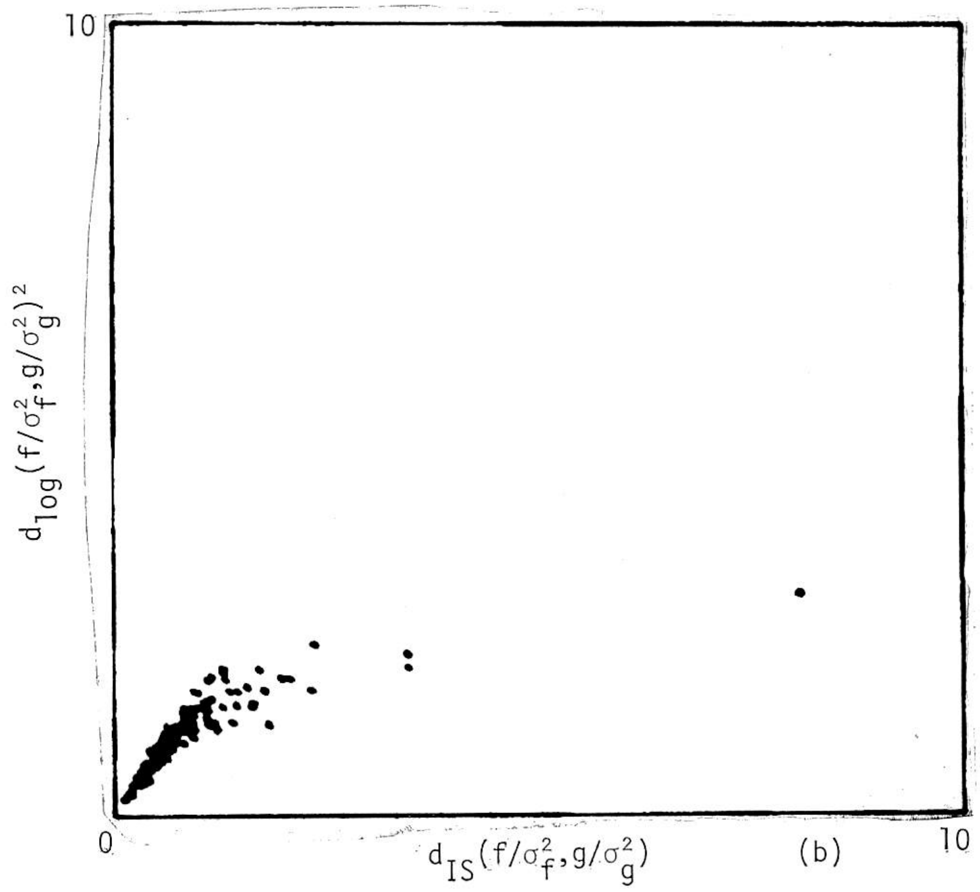
$$d_{IS}(\sigma_f^2, \sigma_g^2) = \exp\{\pm\, d_{log}(\sigma_f^2, \sigma_g^2)\} - 1 \mp d_{log}(\sigma_f^2, \sigma_g^2). \quad (5.2.2)$$

Therefore, the bifurcation appears. If d_{IS} is symmetrized yielding d_{cosh}, such a bifurcation does not appear. From Fig. 5.2.5b, it is observed that the gain normalized d_{IS} (which is equal to d^2_{ncm}) is more sensitive to spectral deviations than d^2_{log}. The points of large

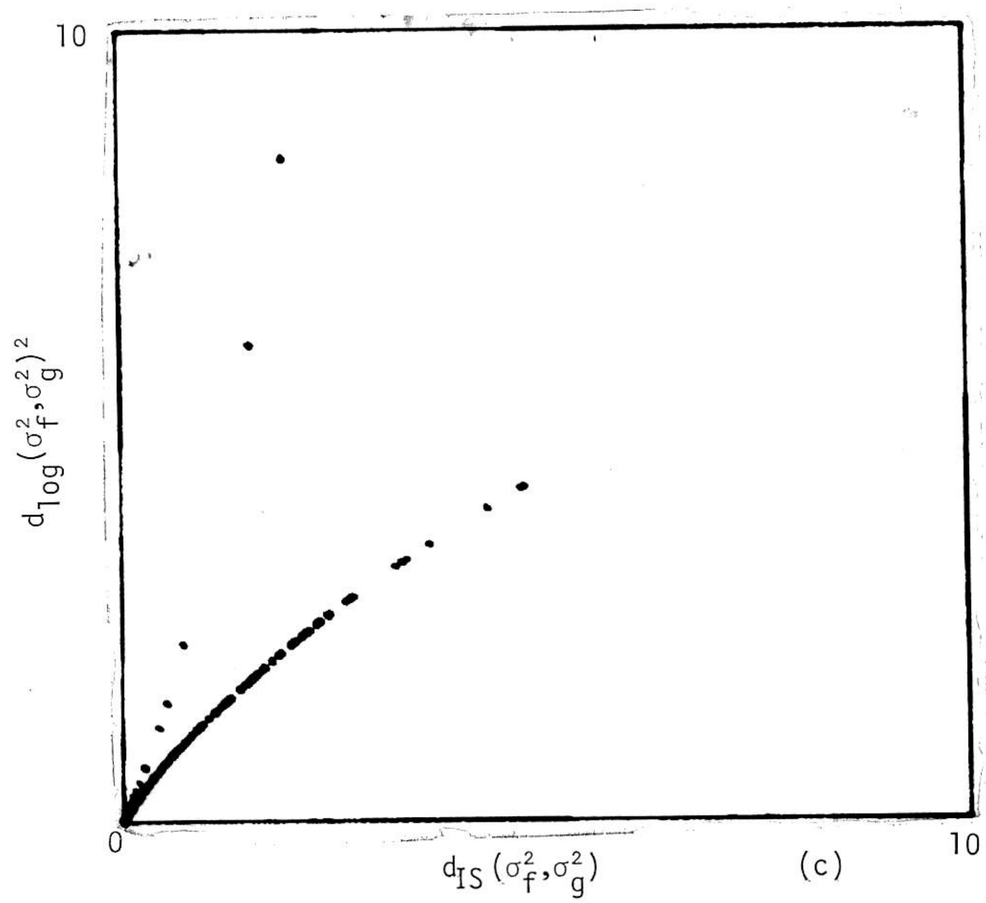


(a)

Fig. 5.2.5 Scatter plots of distortion measures between original and reproduced speeches


10
$d_{log}(f/\sigma_f^2, g/\sigma_g^2)^2$
0
$d_{IS}(f/\sigma_f^2, g/\sigma_g^2)$
10

(b)


10
$d_{log}(\sigma_f^2, \sigma_g^2)^2$
0
$d_{IS}(\sigma_f^2, \sigma_g^2)$
10

(c)

deviation in Fig. 5.2.5b correspond to points of very small gain distortion in Fig. 5.2.5c. This means that our speech compresser yields large spectral distortion only when the gains are very small, i.e., mostly at the silent parts which are not perceptible.

Finally we give comparisons between the system in [40] and ours. Both reproduction qualities were compared using experimental tapes at the Lake Tahoe 1978 Information Theory Workshop. Each system reproduced highly intelligible decoded voices. A comparison table is given in Table 5.2.1.

TABLE 5.2.1

Comparison of Specification and Performance

Wilson & Husain	The System as of August 1978	
incremental tree	hybrid of incremental tree and block structure, fake process approach, Laplace p.d.f.	
on-line identification of LPC's, scale adaptation	off-line design of 8 reference ladders (implementation complexity is less)	
order of filter = 3	5, reflection coefficients are uniformly quantized into 5 bits	
8 kHz sampling rate, 8000 + 384 = 8384 bps	8 kHz sampling rate, 8000 x (1+3/128) = 8187.5 bps (side information is half)	
****	can omit information for silence parts (variable length coding) 8000 x (percentage) = 5500 bps	
mode block length is 31 ~ 62 ms	8 ~ 24 ms	
frequency-weighted square error	frequency weighted square error (coefficients are different)	spectral distortion measures
good reproduction quality	good reproduction quality	under investigation

Chapter 6

CONCLUDING REMARKS

In this work, we have investigated the theory of process distortion measures and their applications. Among such measures, spectral distortions provided simple and direct engineering applications, especially to speech encoding systems. Our main interest in the previous chapter was waveform encoding using a parallel tree search encoder and fake process decoder. Waveform error criteria were used to find the best path map of the parallel tree. It is believed that error criteria using spectal distortion measures give different path maps which possibly yield better reproductions. Experiments based on this idea were tried. If spectral error criteria are used blockwise without any overlap, the quality of decoded sounds is not good because the pitches are not correctly reproduced. Several possibilities for the pitch reproduction have been tried. One easy way is a sliding-blockwise summation of spectral distortions. Preliminary experiments showed that the method is a possible candidate.

It may be useful for future studies to comment on a side-information only reproduction. In this case, ladder filters are driven by pseudo-white processes which are obtained by the Laplace distribution transformation of binary i.i.d. numbers. Since the cardinality of the codebook used in the previous chapter was eight, the reproduced sound was unsatisfactory. It is easily conceived, however, that reproduced speech will sound like whispering voice if a codebook with much larger cardinality is used.

A practical application of the predictor mismatch idea can be found in the following speech compression system which avoids on-line LPC estimation, yet achieves low data rates. The system is an inverse-filter matching encoder which observes the mismatch errors. From another viewpoint, the system will be seen to use only the side information of the parallel tree search system. Figure 6.1.1 explains the total system, which is essentially in the same line as Chaffee's approach [26]. Buzo suggests the same system [37] from an optimum LPC codebook viewpoint.

The operation principle is as follows. By virtue of (4.1.4), the mean square mismatch error is

$$\epsilon(\beta|\alpha)^2 = E\left|\sum_{k=0}^{m} b_k X_{n-k}\right|^2 = \frac{\sigma_f^2}{2\pi}\int_{-\pi}^{\pi} \frac{|B(e^{i\lambda})|^2}{|A(e^{i\lambda})|^2}\, d\lambda \tag{6.1}$$

where

$$\left.\begin{aligned} A(z) &= 1 + a_1 z^{-1} + a_2 z^{-2} + \dots + a_m z^{-m} \\ B(z) &= 1 + b_1 z^{-1} + b_2 z^{-2} + \dots + b_m z^{-m} \end{aligned}\right\} \tag{6.2}$$

Let the unknown input be $\{X_n\}$ with spectral density $f(\lambda) = \sigma_f^2/|A(e^{i\lambda})|^2$. A filter $B(z)$ which minimizes

$$C_E = \epsilon(\beta|\alpha)^2 \tag{6.3}$$

is selected from a given codebook. Since only one sample path of $\{X_n\}$ is given, the correspondence (3.4.1) is relied on. That is,

$$C_E' = \sum_{n=n_0}^{n_0+N-1} \left|\sum_{k=0}^{m} b_k X_{n-k}\right|^2 \tag{6.4}$$

is used instead of C_E. If the codebook is rich enough to inverse-filter most inputs almost optimally, then one obtains

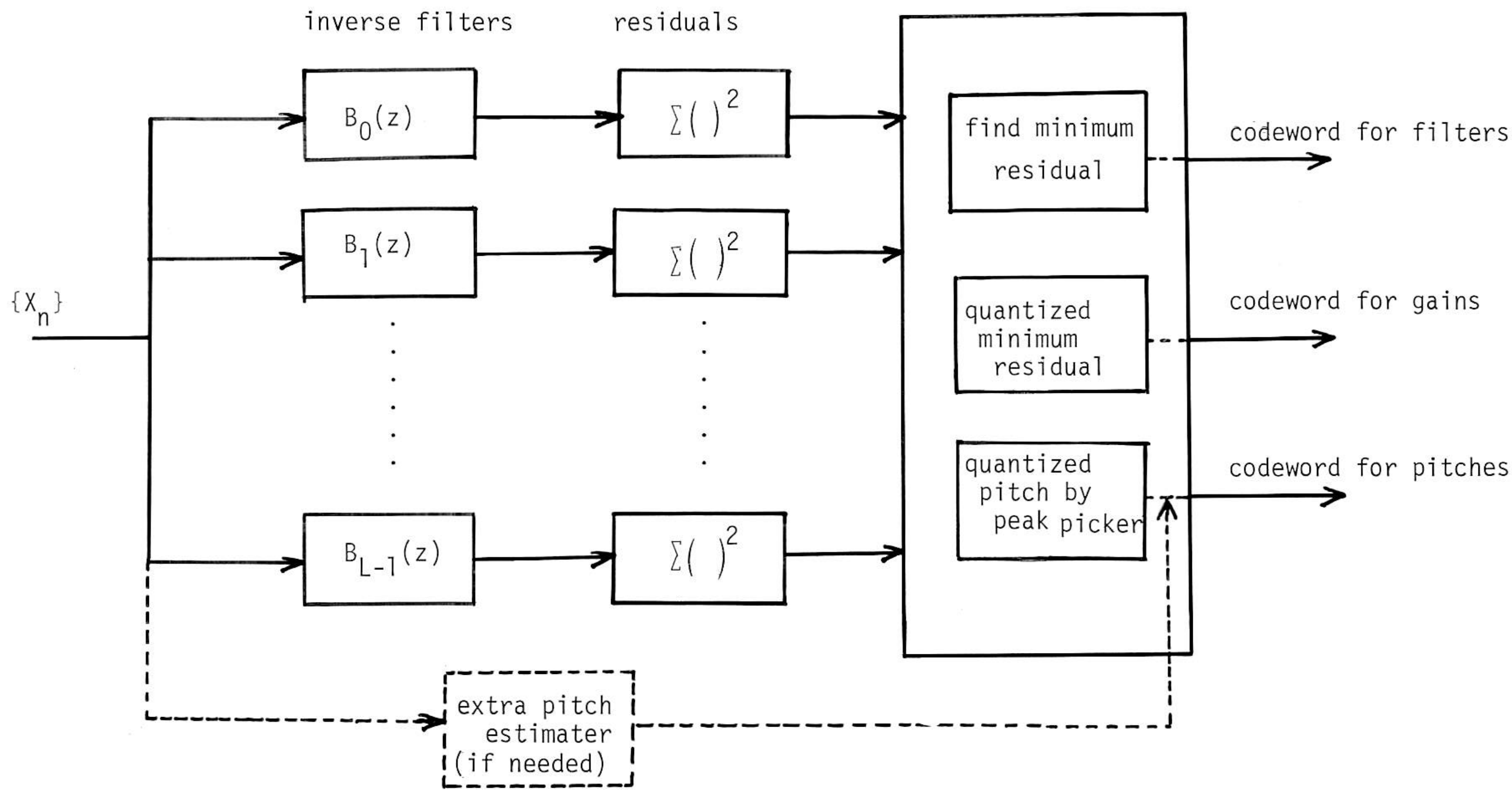


Fig. 6.1 Filter matching speech encoder

$$(C_E)_{opt} = \epsilon(\hat{\alpha}|\alpha)^2 \simeq \epsilon(\alpha|\alpha)^2 = \sigma_f^2 \qquad (6.5)$$

where $\hat{\alpha}$ corresponds to $\hat{A}(e^{i\lambda})$ which minimizes the criterion C_E.

Note that

$$(C'_E)_{opt} = \sum_{n=n_0}^{n_0+N-1} \left| \sum_{k=0}^{m} a_k X_{n-k} \right|^2 . \qquad (6.6)$$

By this method, we have

$$\frac{1}{2\pi} \int_{-\pi}^{\pi} \hat{f}(\lambda)\,d\lambda = \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{\epsilon(\hat{\alpha}|\alpha)^2}{|\hat{A}|^2}\,d\lambda$$

$$\simeq \frac{1}{2\pi} \int_{-\pi}^{\pi} f(\lambda)\,d\lambda \qquad (6.7)$$

Therefore, this method nearly achieves power matching. By the following, the system can be regarded as an encoder under the Itakura-Saito distortion measure d_{IS}.

Proposition 6.1: Finding the spectrum $\hat{f}(\lambda) = \hat{\sigma}_f^2/|\hat{A}|^2$ which minimizes C_E with the power matching

$$\frac{1}{2\pi} \int_{-\pi}^{\pi} f(\lambda) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \hat{f}(\lambda)\,d\lambda$$

is equivalent to the minimization of $d_{IS}(f,g)$ over g.

The proof is given in the appendix. In [37], a different proof is given.

For pitch estimation, sharp spikes of the chosen inverse filter's output might be used. By picking up the spikes, an estimate of a pitch period would be possible. If a more precise estimate of the pitch is needed an extra specific device is attached to the system. The study in this line has begun. We comment that the configuration of the system is appropriate for digital filters and registers.

We have shown that process distortion measures, especially spectral distortion measures, can be good tools for statistical signal processing systems. We suspect that many other such applications are possible.

Appendix A

PROOF OF PROPOSITION 6.1

Lemma A.1: Let $\hat{A}(e^{i\lambda}) = \sum_{k=0}^{m} \hat{a}_k e^{-ik\lambda}$ with $\hat{a}_0 = 1$ minimizes $\sum_{k=0}^{m} |b_k X_{n-1}|^2$, then

$$\langle 1,1\rangle_{\hat{A}} \langle 1,z^{-k}\rangle_A = \langle 1,1\rangle_A \langle 1,z^{-k}\rangle_{\hat{A}} \tag{A.1}$$

for $k = -m,-m+1,\ldots,0,1,\ldots,m$.

Proof.

This is a direct result of (3.4.11) and (3.5.2). □

Proof of Proposition 6.1:

Since $$d_{IS}(f,g) = d_{cm}(f,g)^2 + 2\left(\frac{\sigma_f}{\sigma_g} - 1 - \log\frac{\sigma_f}{\sigma_g}\right)$$

by (3.4.2), the minimization of $d_{IS}(f,g)$ on $\{b_1,b_2,\ldots,b_m\}$ is equivalent to that of C_E. When $b_k = \hat{a}_k$, $k=1,\ldots,m$,

$$d_{IS}(f,\sigma^2/|\hat{A}|^2) = \frac{\sigma_f^2}{\sigma^2}\{1+d_{cm}(1/|A|^2,1/|\hat{A}|^2)^2\}-1-\log\frac{\sigma_f^2}{\sigma^2}$$

$d_{IS}(f,\sigma^2/|\hat{A}|^2)/\partial\sigma^2 = 0$ gives

$$\begin{aligned}
\sigma^2 &= \sigma_f^2\{1+d_{ncm}(1/|A|^2,1/|\hat{A}|^2)^2\} \\
&= \sigma_f^2\{\|A\|_A^2 + \|\hat{A}-A\|_A^2\} \\
&= \sigma_f^2\|\hat{A}\|^2 \\
&= \sigma_f^2 \frac{\|1\|_A^2}{\|1\|_{\hat{A}}^2} .
\end{aligned}$$

Therefore,

$$\hat{\sigma}_f^2 \|1\|_{\hat{A}}^2 = \sigma_f^2 \|1\|_A^2 .$$

That is,

$$\frac{1}{2\pi} \int_{-\pi}^{\pi} \hat{f}(\lambda)\,d\lambda = \frac{1}{2\pi} \int_{-\pi}^{\pi} f(\lambda)\,d\lambda$$

by Lemma A.1. □

Appendix B

OUTLINE OF THE EXPERIMENTAL SYSTEM

The purpose of this appendix is to give an outline of the experimental speech waveform encoder/decoder system from a software simulation viewpoint.

(1) Input Speech: sampled at 8 kHz and in 16 bit binary form whose reference names are f1, f2, ..., f6.

(2) Programs: all programs are written in the Language C;

(i) a main program (main.c) which contains parallel tree searches;

(ii) Subroutines for shift registers (shreg.c), scramblers (scrbl.c), inverse Laplace distributions (tlap.c), ladder filters (slad.c), and error criteria (wdisfn.c for waveform errors and sdisfn.c for spectral distortion measures);

(iii) a D/A driver.

Compilation and execution methods are described in the UNIX manual.

(3) Output: After the execution, a reproduced waveform is obtained in 16 bit binary form. Higher 8 bits are D/A converted by the AR-11 D/A. The converted analog waveform is low pass filtered by a 12 pole Butterworth filter with $f_c = 3$ kHz. After 100:1 attenuation, the waveform is input to a loud voice speaker.

Soft copies and detailed explanations of all programs are stored in a portable disk of Stanford ISL named /yasuo.

REFERENCES

[1] R.M. Gray, "Sliding Block Source Coding," IEEE Trans. on Info. Theory, IT-21 (1975), pp. 357-368.

[2] O. Takenouchi, Topology, Hirokawa Publishing Co., Tokyo, Japan, 1962.

[3] R.M. Gray, D.L. Neuhoff and P.C. Shields, "A Generalization of Ornstein's $\bar{d}$-distance with Applications to Information Theory," Ann. Probability, 3(1975), pp. 315-328.

[4] B. Weiss, "Equivalence of Measure Preserving Transformations," The Institute of Advanced Studies, The Hebrew University of Jerusalem, 1976.

[5] Y.V. Prokhorov, "Convergence of Random Processes and Limit Theorems in Probability Theory," Theory of Probability and Its Application, 1(1956), pp. 157-214.

[6] V. Strassen, "The Existence of Probability Measures with Given Marginals," Ann. Math. Statis., 36(1965), pp. 423-429.

[7] W. Rudin, Real and Complex Analysis, p. 76, McGraw-Hill, N.Y., 1974.

[8] D.S. Ornstein, "An Application of Ergodic Theory to Probability Theory," Ann. Prob., 1(1973), pp. 43-58.

[9] D.L. Neuhoff, "Source Coding and Distance Measures on Random Processes," Ph.D. Dissertation, Stanford University, 1974.

[10] V.I. Levenshtein, "Binary Codes Capable of Correcting Deletions, Insertions and Reversals," Soviet Phys. Doklady, 10(1966), pp. 707-710.

[11] T. Okuda, E. Tanaka and T. Kasai, "A Method for the Correction of Garbled Words Based on the Levenshtein Metric," IEEE Trans. on Computers, C-25(1976), pp. 172-177.

[12] J.L. Doob, Stochastic Processes, John Wiley & Sons, N.Y., 1953.

[13] U. Grenander and M. Rosenblatt, Statistical Analysis of Stationary Time Series, John Wiley & Sons, N.Y., 1957.

[14] Y. Rozanov, Stationary Random Processes, Holden-Day, San Francisco, 1967.

[15] S. Kullbuck, Information Theory and Statistics, John Wiley & Sons, New York, 1959.

[16] T. Kailath, "The Divergence and Battacharyya Distance Measures in Signal Selection," IEEE Trans. on Comm. Tech., COM-15(1967), pp. 52-60.

[17] M.S. Pinsker, Information and Information Stability of Random Variables and Processes, Holden-Day, San Francisco, 1964.

[18] U. Grenander and G. Szegö, Toeplitz Forms and Their Applications, Univ. of California Press, Berkeley, 1958.

[19] P. Papantoni-Kazakos, R. Li and D. Kazakos, "Linear Dimentionality Reduction of Stationary Vector Gaussian Processes," submitted for publication, 1977.

[20] J.D. Markel and A.H. Gray, Jr., Linear Prediction of Speech, Springer, N.Y., 1976.

[21] Y. Matsuyama, A. Buzo and R.M. Gray, "Spectral Distortion Measures for Speech Compression," Tech. Rept. No. 6504-3, SEL-78-015, Stanford University, 1978.

[22] F. Itakura and S. Saito, "A Statistical Method for Estimation of Speech Spectral Density and Formant Frequencies," Trans. Inst. Electron. & Comm. Eng. (Japan), 53(1970), pp. 35-42.

[23] N. Levinson, "The Wiener RMS (Root Mean Square) Error Criterion in Filter Design and Prediction," J. Math. Phys., 25(1947), pp. 261-278.

[24] F. Itakura and S. Saito, "Analysis Synthesis Telephony Based Upon the Maximum Likelihood Method," Reports of the 6th Internat'l. Congr. Acoust., Y. Kohashi, ed., C-5-5, C17-20, Tokyo, Japan, 1968.

[25] F. Itakura, "Minimum Prediction Residual Principle Applied to Speech Recognition," IEEE Trans. on Acoust. Speech & Signal Proc., ASSP-23 (1975), pp. 67-72.

[26] D.L. Chaffee, "Applications of Rate Distortion Theory to Bandwidth Compression of Speech Signals," Ph.D. Dissertation, Univ. of Calif., Los Angeles, 1975.

[27] A.H. Gray, Jr. and J.D. Markel, "Distance Measures for Speech Processing," IEEE Trans. on Acoust., Speech & Signal Proc., ASSP24 (1976), pp. 380-391.

[28] R. Viswanathan and J. Makhoul, "Quantization Properties of Transmission Parameters in Linear Predictive Systems," IEEE Trans. on Acoust., Speech & Signal Proc., ASSP-23(1975), pp. 309-321.

[29] A.H. Gray, Jr., R.M. Gray and J.D. Markel, "Comparison of Optimal Quantizations of Speech Reflection Coefficients," IEEE Trans. on Acoust., Speech & Signal Proc., ASSP-25(1977), pp. 9-23.

[30] E. Hewitt and K. Stronberg, Real and Abstract Analysis, Springer, N.Y., 1965.

[31] A.H. Gray, Jr. and J.D. Markel, "Quantization and Bit Allocation in Speech Processing," IEEE Trans. on Acoust., Speech & Signal Proc., ASSP-24(1976), pp. 459-473.

[32] M. Anderson, "On the Calculation of Filter Coefficients for Maximum Entropy Spectral Analysis," Geophysics 39(1974), pp. 69-72.

[33] M. Morf, B. Dickinson, T. Kailath and A. Vieira, "Efficient Solution of Covariance Equations for Linear Prediction," IEEE Trans. on Acoustics, Speech & Signal Proc., ASSP-25(1977), pp. 429-433.

[34] F. Itakura and S. Saito, "Speech Analysis-Synthesis System Based on the Partial Autocorrelation Coefficient," Acoust. Society of Japan Meeting (1969).

[35] G.E.P. Box and G.M. Jenkins, Time Series Analysis; Forecasting and Control, Holden-Day, San Francisco, 1970.

[36] H. Wakita, "Direct Estimation of the Vocal Tract Shape by Inverse Filtering of Acoustic Speech Waveforms," IEEE Trans. on Audio & Electron., AU-21(1973), pp. 417-427.

[37] L.A. Buzo, "Optimal Vector Quantization for Linear Predicted Coded Speech," Ph.D. Dissertation, Stanford Univ., EE Dept., 1978.

[38] P. Pantoni-Kazakos and R.M. Gray, "Robustness of Estimators on Stationary Observations," submitted for publication, 1977.

[39] J.B. Anderson and J.B. Bodie, "Tree Encoding of Speech," IEEE Trans. on Info. Theory, IT-21(1975), pp. 379-387.

[40] S.G. Wilson and S. Husain, "Tree Encoding of Speech at 8000 bps with a Frequency-Weighted Error Criterion," submitted for publication, 1977.

[41] Y. Linde and R.M. Gray, "A Fake Process Approach to Data Compression," to appear IEEE Trans. on Comm., 1978.

[42] F. Jelinek and J.B. Anderson, "Instrumentable Tree Encoding of Information Sources," IEEE Trans. on Info. Theory, IT-17,(1971), pp. 118-119.

[43] F. Itakura and S. Saito, "The Quality of Speech Analysis-Synthesis System by Partial Correlation Coefficients," Joint Meeting of Electrical Engineers, Japan (1970).

[44] J. Makhoul, "Linear Prediction: A Tutorial Review," Proc. IEEE, 63 (1975), pp. 561-580.

[45] J. Moser, E. Phillips and S. Varadhan, "Ergodic Theory," Seminar Note, Courant Inst. of Math. Sciences, N.Y. University, N.Y. (1975).

[46] J.L. Doob, Stochastic Processes, John Wiley & Sonss, N.Y., 1953

[47] W.W. Peterson and E.J. Weldon, Jr., Error-Correcting Codes, 2nd edition, MIT Press, Cambridge, MA, 1972.

[48] R.M. Gray, L.A. Buzo, Y. Matsuyama, A.H. Gray and J.D. Markel, "Source Coding and Speech Compression" to be presented at the Internat. Telecommunication Conf., Los Angeles, CA 1978.

[49] T. Berger, Rate Distortion Theory, Prentice Hall, New Jersey, 1971.

[50] L.D. Davisson, "Universal Noiseless Coding," IEEE Trans. Inform Theory, IT-19 (1973), pp. 783-795.

[51] J.W. Mark, "Adaptive Trellis Encoding of Discrete-Time Sources with a Distortion Measure," IEEE Trans. Comm., COM-25 (1977), pp. 408-417.

[52] R.J. Fontana, "A Class of Composite Sources and Their Ergodic and Information Theoretic Properties," Ph.D. Dissertation, Stanford University, 1978.