Digital Processing of Speech Signals

Dig
Proce
o
Spe
Sig

L.R. Rabiner /

PRENTICE-HALL
SIGNAL PROCESSING SERIES
ALAN V. OPPENHEIM,
SERIES EDITOR

0-13-213603-1

PRENTICE-HALL SIGNAL PROCESSING SERIES

Alan V. Oppenheim, Editor

ANDREWS and HUNT *Digital Image Restoration*
BRIGHAM *The Fast Fourier Transform*
BURDIC *Underwater Acoustic System Analysis*
CASTLEMAN *Digital Image Processing*
CROCHIERE and RABINER *Multirate Digital Signal Processing*
DUDGEON and MERSEREAU *Multidimensional Digital Signal Processing*
HAMMING *Digital Filters, 2e*
HAYKIN, ED. *Array Signal Processing*
LEA, ED. *Trends in Speech Recognition*
LIM, ED. *Speech Enhancement*
MCCLELLAN and RADER *Number Theory in Digital Signal Processing*
OPPENHEIM, ED. *Applications of Digital Signal Processing*
OPPENHEIM, WILLSKY, with YOUNG *Signals and Systems*
OPPENHEIM and SCHAFER *Digital Signal Processing*
RABINER and GOLD *Theory and Applications of Digital Signal Processing*
RABINER and SCHAFER *Digital Processing of Speech Signals*
ROBINSON and TREITEL *Geophysical Signal Analysis*
TRIBOLET *Seismic Applications of Homomorphic Signal Processing*

AT&T

DIGITAL PROC
OF
SPEECH SIG

Lawrence R. Rabiner

Acoustics Research Laboratory
Bell Telephone Laboratories
Murray Hill, New Jersey

Ron

School of
Georgia I
A

Prentice-Hall, Inc., Englewood Clif

Library of Congress Cataloging in Publication Data

Rabiner, Lawrence R (date)
Digital processing of speech signals.

(Prentice-Hall signal processing series)
Includes bibliographies and index.
1. Speech processing systems. 2. Digital electronics. I. Schafer, Ronald W., joint author.
II. Title.
TK7882.S65R3 621.38'0412 78-8555
ISBN 0-13-213603-1

Printed in the United States of America

19 18 17 16 15 14

PRENTICE-HALL INTERNATIONAL, INC., *London*
PRENTICE-HALL OF AUSTRALIA PTY., LIMITED, *Sydney*
PRENTICE-HALL OF CANADA, LTD., *Toronto*
PRENTICE-HALL OF INDIA PRIVATE LIMITED, *New Delhi*
PRENTICE-HALL OF JAPAN, INC., *Tokyo*
PRENTICE-HALL OF SOUTHEAST ASIA PTE. LTD., *Singapore*
WHITEHALL BOOKS LIMITED, *Wellington, New Zealand*

To our parents,

Dr. and Mrs. Nathan Rabiner

and

Mr. and Mrs. William Schafer,

for instilling within us the thirst for knowledge
and the quest for excellence;

and to our families,

Suzanne, Sheri, and Wendi Rabiner

and

Dorothy, Bill, John, and Kate Schafer,

for their love, encouragement, and support.

8

Linear Predictive Coding of Speech

8.0 Introduction

One of the most powerful speech analysis techniques is the method of linear predictive analysis. This method has become the predominant technique for estimating the basic speech parameters, e.g., pitch, formants, spectra, vocal tract area functions, and for representing speech for low bit rate transmission or storage. The importance of this method lies both in its ability to provide extremely accurate estimates of the speech parameters, and in its relative speed of computation. In this chapter, we present a formulation of the ideas behind linear prediction, and discuss some of the issues which are involved in using it in practical speech applications.

The basic idea behind linear predictive analysis is that a speech sample can be approximated as a linear combination of past speech samples. By minimizing the sum of the squared differences (over a finite interval) between the actual speech samples and the linearly predicted ones, a unique set of predictor coefficients can be determined. (The predictor coefficients are the weighting coefficients used in the linear combination.)

The philosophy of linear prediction is intimately related to the basic speech synthesis model discussed in Chapter 3 in which it was shown that speech can be modelled as the output of a linear, time-varying system excited

Linear predictive techniques have already been discu
the waveform quantization methods of Chapter 5. There
a linear predictor could be applied in a differential qu
reduce the bit rate of the digital representation of the
fact, the mathematical basis for an adaptive high ord
DPCM waveform coding is identical to the analysis that w
chapter. In adaptive DPCM coding the emphasis is on fi
will reduce the variance of the difference signal so that
also be reduced. In this chapter we take a more genera
how the basic linear prediction idea leads to a set of analys
be used to estimate parameters of a speech model. This
predictive analysis techniques is often referred to as linea
LPC.

The techniques and methods of linear prediction h
the engineering literature for a long time. The ideas of
been in use in the areas of control, and information theor
system estimation and system identification. The term s
particularly descriptive of LPC methods in that once the
have been obtained, the system has been uniquely identif
it can be modelled as an all-pole linear system.

As applied to speech processing, the term linear
variety of essentially equivalent formulations of the prob
speech waveform [1-18]. The differences among these f
those of philosophy or way of viewing the problem.
differences concern the details of the computations used t
coefficients. Thus as applied to speech, the various (ofte
tions of linear prediction analysis have been:

1. the covariance method [3]
2. the autocorrelation formulation [1,2,9]
3. the lattice method [11,12]
4. the inverse filter formulation [1]
5. the spectral estimation formulation [12]
6. the maximum likelihood formulation [4,6]
7. the inner product formulation [1]

In this chapter we will examine in detail the similarities a
only the first three basic methods of analysis listed abov
formulations are equivalent to one of these three.

The importance of linear prediction lies in the acc
basic model applies to speech. Thus a major part of this c
discussion of how a variety of speech parameters can be re
linear prediction methods. Furthermore some typical exa

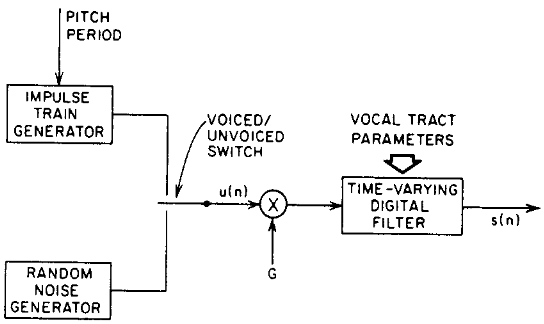

Fig. 8.1 Block diagram of simplified model for speech production.

8.1 Basic Principles of Linear Predictive Analysis

Throughout this book we have repeatedly referred to the basic discrete-time model for speech production that was developed in Chapter 3. The particular form of this model that is appropriate for the discussion of linear predictive analysis is depicted in Fig. 8.1. In this case, the composite spectrum effects of radiation, vocal tract, and glottal excitation are represented by a time-varying digital filter whose steady-state system function is of the form

$$H(z) = \frac{S(z)}{U(z)} = \frac{G}{1 - \sum_{k=1}^{p} a_k z^{-k}} \tag{8.1}$$

This system is excited by an impulse train for voiced speech or a random noise sequence for unvoiced speech. Thus, the parameters of this model are: voiced/unvoiced classification, pitch period for voiced speech, gain parameter G, and the coefficients $\{a_k\}$ of the digital filter. These parameters, of course, all vary slowly with time.

The pitch period and voiced/unvoiced classification can be estimated using one of the many methods already discussed in this book or by methods based on linear predictive analysis to be discussed later in this chapter. As discussed in Chapter 3, this simplified all-pole model is a natural representation of non-nasal voiced sounds, but for nasals and fricative sounds, the detailed acoustic theory calls for both poles and zeros in the vocal tract transfer function. We shall see, however, that if the order p is high enough, the all-pole model provides a good representation for almost all the sounds of speech. The major advantage of this model is that the gain parameter, G, and the filter coefficients $\{a_k\}$ can be estimated in a very straightforward and computationally efficient manner by the method of linear predictive analysis.

For the system of Fig. 8.1, the speech samples $s(n)$ are related to the

A linear predictor with prediction coefficients, α_k is
output is

$$\tilde{s}(n) = \sum_{k=1}^{p} \alpha_k s(n-k)$$

Such systems were used in Chapter 5 to reduce the
signal in differential quantization schemes. The syst
linear predictor is the polynomial

$$P(z) = \sum_{k=1}^{p} \alpha_k z^{-k}$$

The prediction error, $e(n)$, is defined as

$$e(n) = s(n) - \tilde{s}(n) = s(n) - \sum_{k=1}^{p} \alpha$$

From Eq. (8.5) it can be seen that the prediction er
of a system whose transfer function is

$$A(z) = 1 - \sum_{k=1}^{p} \alpha_k z^{-k}$$

It can be seen by comparing Eqs. (8.2) and (8.5) that
the model of Eq. (8.2) exactly, and if $\alpha_k = a_k$, then
prediction error filter, $A(z)$, will be an *inverse filter* for
(8.1), i.e.,

$$H(z) = \frac{G}{A(z)}$$

The basic problem of linear prediction analysis
predictor coefficients $\{\alpha_k\}$ directly from the speech si
to obtain a good estimate of the spectral properties of
the use of Eq. (8.7). Because of the time-varying na
the predictor coefficients must be estimated from sho
signal. The basic approach is to find a set of predi
minimize the mean-squared prediction error over a sho
waveform. The resulting parameters are then *assume*
the system function, $H(z)$, in the model for speech pr

That this approach will lead to useful results
obvious, but it can be justified in several ways. Firs
then $e(n) = Gu(n)$. For voiced speech this means th
a train of impulses; i.e., $e(n)$ would be small most of
α_k's that minimize prediction error seems consistent
second motivation for this approach follows from the f
erated by Eq. (8.2) with non-time-varying coefficients

DOCKET ALARM

Explore Litigation Insights

Docket Alarm provides insights to develop a more informed litigation strategy and the peace of mind of knowing you're on top of things.

Real-Time Litigation Alerts

Keep your litigation team up-to-date with **real-time alerts** and advanced team management tools built for the enterprise, all while greatly reducing PACER spend.

Our comprehensive service means we can handle Federal, State, and Administrative courts across the country.

Advanced Docket Research

With over 230 million records, Docket Alarm's cloud-native docket research platform finds what other services can't. Coverage includes Federal, State, plus PTAB, TTAB, ITC and NLRB decisions, all in one place.

Identify arguments that have been successful in the past with full text, pinpoint searching. Link to case law cited within any court document via Fastcase.

Analytics At Your Fingertips

Learn what happened the last time a particular judge, opposing counsel or company faced cases similar to yours.

Advanced out-of-the-box PTAB and TTAB analytics are always at your fingertips.

API

Docket Alarm offers a powerful API (application programming interface) to developers that want to integrate case filings into their apps.

LAW FIRMS

Build custom dashboards for your attorneys and clients with live data direct from the court.

Automate many repetitive legal tasks like conflict checks, document management, and marketing.

FINANCIAL INSTITUTIONS

Litigation and bankruptcy checks for companies and debtors.

E-DISCOVERY AND LEGAL VENDORS

Sync your system to PACER to automate legal marketing.