New Systems and Architectures for Automatic Speech Recognition and Synthesis

Edited by

Renato De Mori and Ching Y. Suen

New Systems and Architectures for Automatic Speech Recognition and Synthesis

Edited by

Renato De Mori and Ching Y. Suen

Department of Computer Science, Concordia University
Montréal, Québec H3G 1M8, Canada

Springer-Verlag Berlin Heidelberg New York Tokyo

Published in cooperation with NATO Scientific Affairs Division

Proceedings of the NATO Advanced Study Institute on New Systems and Architectures for Automatic Speech Recognition and Synthesis held at Bonas, Gers, France, 2–14 July 1984

ISBN 3-540-15177-X Springer-Verlag Berlin Heidelberg New York Tokyo
ISBN 0-387-15177-X Springer-Verlag New York Heidelberg Berlin Tokyo

Library of Congress Cataloging in Publication Data. NATO Advances Study Institute on New Systems and Architecture for Automatic Speech Recognition and Synthesis (1984 : bonas France) New systems and architecture for automatic speech recognition and synthesis. (NATO ASI series. Series F, Computer and system sciences ; vol. 16) "Proceedings of the NATO Advanced Study Institute on New Systems and Architecture for Automatic Speech Recognition and Synthesis held at Bonas, Gers, France, 2–14 July 1984" — T.p. verso. "Published in cooperation with NATO Scientific Affairs Division." Includes index. 1. Automatic speech recognition — Congresses. 2. Speech synthesis — Congresses. I. De Mori, Renato. II. Suen, Ching Y. III. North Atlantic Treaty Organization. Scientific Affairs Division. IV. Title. V. Series: NATO ASI series. Series F, Computer and system sciences ; no. 16. TK7895.S65N375 1984 629.8'92 85-17228
ISBN 0-387-15177-X (U.S.)

Printing: Beltz Offsetdruck, Hemsbach; Bookbinding: J. Schäffer OHG [illegible]unstadt
2145/3140-543210

List of Participants

Mr. A. Aggoun
C.N.E.T.
B.P. 40
22301 Lannion
France

Dr. W. A. Ainsworth
Dept. of Communication and Neuroscience
University of Keele
Keele, Staffs. ST5 5BG
England

Mr. J. M. Benedi Ruiz
Departamento de Electronicae Informatica
Universidad de Valencia
Burjasot — Valencia
Spain

Prof. R. Bisiani
Carnegie-Mellon University
Dept. of Computer Science
Schenley Park
Pittsburgh, Pa 15213
U.S.A.

Dr. A. Cappelli
Istituto di Linguistica Computazionale
Via della Faggiola, 32
56100 Pisa
Italy

Dr. G. Chollet
E.N.S.T.
46, Rue Barrault
75634 Paris Cedex 13
France

Dr. A. Ciaramella
CSELT
Via G. Reiss Romoli 274
Torino 10148
Italy

Prof. G. Danserau
Université de Quebec à Hull
Dept. Informatique
CP 1250 succ. B
Hull, Quebec J8X 3X7
Canada

Prof. R. De Mori
Dept. of Computer Science
Concordia University
Montreal, Quebec H3G 1M8
Canada

Mr. Di Martino
C.R.I.N.
B.P. 239
54506 Vandoeuvre
France

Dr. D. Dours
U.E.R. de Mathématiques - Informatiques-Gestion
Université Paul Sabatier
118, Rue de Narbonne
31062 Toulouse Cedex
France

Mr. D. Fohr
C.R.I.N.
B.P. 239
54506 Vandoeuvre
France

Mr. N.J.A. Forse
British Telecom Research Laboratories
Martlesham Heath
Ipswich IP5 7RE
England

Dr. P. Frison
I.R.I.S.A.
Campus de Beaulieu
35042 Rennes Cedex
France

Prof. K. S. Fu
School of Electrical Engineering
Purdue University
West Lafayette, Ind. 47907
U.S.A.

Dr. R. Gubrynowicz
Université de Provence, Institut de Phonétique
29, Avenue Robert Schuman
13621 Aix en Provence Cedex
France

Prof. J. P. Haton
C.R.I.N.
Université de Nancy 1
B.P. 239
54506 Vandoeuvre
France

Dipl. Ing. H. Heckl
Siemens AG,
ZTI INF
Otto-Hahn-Ring 6
D 8000 München 83
West Germany

Dr. W. Hess
Lehrstuhl Für Datenverarbeitung
Technische Universität München
Arcisstr. 21
8000 München 2
West Germany

Prof. J. L. Houle
Ecole Polytechnique de Montréal
C.P. 6079 Succ "A"
Montreal, Que. H3C 3A7
Canada

Dr. A. Ingegnoli
C/O FACRC
Via Della Magione
00040 Pomezia (Rome)
Italy

Ing. B. Kammerer
Siemens AG, ZTI INF 111
Otto-Hahn-Ring 6
D 8000 München 83
West Germany

Dipl. Ing. W. Kulas
Lehrstuhl Fuer Allgemeine Elektronik
und Akustik
Ruhr-Universität Bochum
Universität Str. 150
D-4630 Bochum
West Germany

Prof. M. Kunt
Laboratoire de Traitement de Signaux
Ecole Polytechnique Fédérale de Lausanne
16, Chemin de Bellerive
1007 Lausanne
Switzerland

Prof. P. Laface
Departimento di Automatica e Informatica
CORSO Duca Degli Abruzzi 24
10129 Torino
Italy

Mr. L. Leguennec
C.N.E.T.
B.P. 40
22301 Lannion Cedex
France

Prof. H. Leich
Faculté Polytechnique e Mons
31 Boulevard Dolez
7000 Mons
Belgium

Mr. P. Lockwood
Laboratoire de Marcoussis
91460 Marcoussis
France

Mr. U. Macdonald
Dept. of Computational
University of Saint Andrews
North Haugh, St. Andrews
Scotland

Dr. J. Manley
Centre for Research on Perception
and Cognition
University of Sussex
Brighton BN1 9QG
England

Dr. J. P. Martens
Laboratorium voor Electronica
St-Pietersnieuwstraat 41
9000 GENT
Belgium

Dr. G. Mercier
C.N.E.T.
B.P. 40
22301 Lannion Cedex
France

Dr. R. K. Moore
Royal Signals and Radar Establishment
St. Andrews Road
Great Malvern, Worcs. WRI4 3PS
England

Mr. J. Mudler
Institut für Nachrichtentechnik
TU Braunschweig
Postfach 3329
West Germany

Mr. R. L. Muhlfeld
IMMD 5
Universität Erlangen
Martensstr. 3
8520 Erlangen
West Germany

Dr. E. Mumolo
FACE RC
Via Della Magione 10
00040 Pomezia (Rome)
Italy

Dr. H. Ney
PHILIPS Gmbh
Forschungslaboratorium Hamburg
P.O. Box 540840
2000 Hamburg 54
West Germany

Prof. H. Niemann
Universität Erlangen-Nurnberg, IMMD
Martensstr. 3
8520 Erlangen
West Germany

Prof. J. Ohala
Phonology Laboratory
Department of Linguistics
University of California
Berkeley, CA 94720
U.S.A.

Dr. M. Ohala
Linguistics Program
California State
University at San Joseé
San Jose, CA 95192
U. S. A.

Mr. W. Ronnebrinck
Institut Für Nachrichtentechnik
Technische Universität Braunschweig
Schleinitzstr. 23
3300 Braunschweig
West Germany

Dipl. Ing. H. W. Rühl
Lehrsuhl für Allg. Elektrotechnik
und Akustik
Ruhr-Universität Bochum
Postfach 102148
4630 Bochum 1
West Germany

Dr. G. Ruske
Lehrstuhl für Datenverarbeitung
Technische Universität München
Franz-Joseph Str. 38
8000 München 40
West Germany

Prof. B. Sankur
Bogasici University
Dept. of Electrcal Engineering
Bebek, Istanbul
Turkey

Mr. E. G. Schukat
Universität Erlangen-Nurnberg
IMMD 5
Martensstr. 3
8520 Erlangen
West Germany

Dr. D. Sciarra
Electronica San Giorgio
Via Puccini 2
16154 Genova Sestri Ponente
Italy

Mr. N. Sedgwick
Acorn Computers Ltd.
Fulbourn Road
Cherry Hinton
Cambridge CB1 4JN
England

Dr. D. Sinclair
IBM Science Center
Athelstan House
St. Clement Street
Winchester, S023 9DR
England

Prof. W. Steenaart
Electr. Eng. Dept.
University of Ottawa
Ottawa, Ontario K1N 6N5
Canada

Prof. Ching Y. Suen
Dept. of Computer Science
Concordia University
1455 de Maisonneuve Blvd., West
Montreal, Quebec H3G 1M8
Canada

Mr. A. Tassy
E.N.S.T.
46, Rue Barrault
75013 Paris
France

Mr. E. Verdonck
Katholieke Universitat Leuven
CME — ESAT
De Croylaan 52
3030 Heverlee
Belgium

Mr. J. Verhoeven
University of Antwerp (UIA)
Dept. of Didactics & Criticism
Universiteitsplein 1
2610 Wilrijk
Belgium

Dr. E. Vidal Ruis
Centro de Informatica
Universidad de Valencia
Burjasot — Valencia
Spain

Mlle. Nadine Vigouroux
Laboratoire CERFIA
(mr. Perennou)
118, Rte de Narbonne
31062 Toulouse, Cedex
France

Mr. D. Wood
GEC Research Laboratories
East Lane
Wembley, Middx
England

Prof. N. Yalabik
Computer Engineering Dept.
Middle East Technical University
Ankara
Turkey

FOREWORD

It is an established tradition that researchers from many countries get together on the average every three years for a two week Advanced Studies Institute on Automatic Speech Recognition and Synthesis. According to ASI policies the Institute is financed by NATO. This book contains the texts of lectures and papers contributed by the attendees of the ASI which was held July 2 — 14, 1984, at Bonas, Gers, France. Focussed on New Systems and Architectures for Automatic Speech Recognition and Synthesis, this book is divided into 4 parts:

(a) *Review of basic algorithms*

(b) *System architecture and VLSI for automatic Speech*

(c) *Software systems for automatic speech recognition,*

(d) *Speech synthesis and phonetics.*

Due to the international nature of the Institute, the readers will find in this book different styles, different points of view and applications to different languages. This reflects also some characteristics of the International Association for Pattern Recognition (IAPR) whose technical committee on Speech Recognition has organized this ASI.

Proposed contributions have been reviewed by an Editorial Committee composed of W. Ainsworth (Kent), R. Bisiani (Pittsburgh), J. P. Haton (Nancy), W. Hess (Munich), J. L. Houle (Montréal), P. Laface (Turin), R. Moore (Malvern), H. Niemann (Erlangen) and J. Ohala (Berkeley).

Typesetting of the book was performed using SYMSET facilities developed entirely by the Department of Computer Science at Concordia University. Special thanks are due to L. Lam, H. Monkiewicz and L. Thiel.

Montreal, Canada
May 1985

R. De Mori and C. Y. Suen

TABLE OF CONTENTS

AN OVERVIEW OF DIGITAL TECHNIQUES FOR PROCESSING SPEECH SIGNALS

Murat Kunt

Signal Procesing Laboratory

Swiss Federal Institute of Technology

16 Ch. de Bellerive

CH - 1007 Lausanne, Switzerland

and

Heinz Hugli

Mircrotechnique Institute

University of Neuchatel

Rue de la Maladière 71

CH - 2007 Neuchatel, Switzerland

ABSTRACT

This paper discusses major digital signal processing methods used in processing speech signals. Basic tools, such as the discrete Fourier transform, the z transform and linear filter theory are briefly introduced first. A general view of fast transformation algorithms and most widely used particular fast transformations are given. Linear prediction is then described with a particular emphasis on its lattice structure. A brief introduction to homomorphic processing for multiplied and convolved signals and to its applications in speech processing is given. Recalling some fundamentals of the speech signal, various speech analysis and synthesis models are described, showing which kind of processing methods are

involved. Finally, two aspects of speech recognition are presented: feature traction and pattern matching using dynamic time warping.

1. INTRODUCTION

Because of its multidisciplinary character, digital signal processing became increasingly important in a number of scientific and technical areas. Continuous interaction between the methods and the particular applications have led to an avalanche on both sides. Increasingly sophisticated methods are developed to fulfil wider needs of a large number of applications. There is no doubt that one of the major application areas of digital signal processing is speech signals. Over the last two decades, considerable effort has been devoted to analyse, code, model, synthesize and recognize speech signals. A dozen of books are already available, presenting various aspects of digital speech processing.

This paper attempts to give a tutorial review of major digital signal processing methods used in processing speech signals. Because of space limitations and the wide range of the subject, in depth treatments are omitted. Essence of the methods and insight for the interpretation of the results are indicated whenever possible. In section two, basic methods are defined such as the discrete Fourier transform, correlation functions, the z transform, the convolution, and the linear system theory. A general view of fast transformation algorithms is given, showing structures for hardware and software. Commonly used fast transformations are also briefly indicated. The last part of this section presents the linear prediction models and tools for one dimensional signals and introduces its lattice structure, a structure that is modular and hence suitable for various implementations. In section three, homomorphic processing of multiplied and convolved signals is discussed with particular emphasis on its applications to speech signals, particularly for deconvolution. Section four gives an overview of the speech analysis and synthesis methods using previously defined tools. Speech recognition is summarized in section five with a particular emphasis on pattern matching. The objective, in these last two sections, is to point out particular digital signal processing methods used for reaching the goals.

2.0 BASIC METHODS

In this section basic signal processing methods are defined and their use in speech processing are discussed. Analysis and synthesis tools for digital signals, such as the discrete Fourier transform and the correlation function, and for systems, such as the z transform and the convolution are described first. A brief discussion on linear filters and fast transformations is presented next. The section ends with a rather detailed description of linear prediction. For more detail, the reader may consult [1] and [2].

THE DISCRETE FOURIER TRANSFORM

The discrete Fourier transform of a digital signal x(k) is a complex series defined by:

$$X(n) = \sum_{k=k_0}^{k_0+N-1} x(k)\exp(-j2\pi kn/N) \qquad (1)$$

with $n = -N/2, \ldots, N/2-1$

In this definition, only N consecutive samples of the signal are used starting at k = k0. The series X(n) is periodical in n with a period of N. The integer variable n represents discrete frequencies. For example n = 0 is the DC component and n = N/2 is the folding frequency, i.e. half of the sampling rate.

The inverse transform is given by:

$$x(k) = (1/N) \sum_{n=-N/2}^{N/2-1} X(n)\ \exp(j2\pi nk/N) \qquad (2)$$

with $k = k0, \ldots, k0 + N - 1$

Eq. (1) is referred to as the analysis of the signal, whereas eq. (2) is used to synthesize the signal from its Fourier Transform. From the complex numbers $X(n)$ two real sequences are obtained. The magnitude $X(n)$ plotted as a function of n is the magnitude spectrum. The argument $\arg[X(n)]$ is the phase spectrum. They inform on the frequency distributions of complex exponential signals composing the analysed signal $x(k)$. If the number of samples N is small compared to the total length of the signal, these spectra are called short term spectra. On a long signal, such as a speech signal, several short term spectra can be computed. Sections of the signal used in these computations may partially overlap or may be apart. If these spectra are plotted in three dimensions as a function of the frequency n and of the time (for example time instants corrresponding to the beginning of each signal section), the resulting surface is called spectrogram. It is usually represented as a black-and-white two level image on the (n,k) plane. Additional grey levels, if available, give more precise and detailed information on the frequency variations of various components of the signal. In section 1.6 fast algorithms for computing spectrograms will be discussed.

1.2 CORRELATION FUNCTIONS

The similarity of two signals $x(k)$ and $y(k)$ is measured by their cross correlation function defined by:

$$\varphi_{xy}(k) = \sum_{l=-\infty}^{+\infty} x(l)\ y(k+l) \tag{3}$$

For a given delay k of the second signal $y(k)$ with respect to the first signal $x(k)$, the cross correlation function is just the integral of the product of these two signals. It reaches its maximum value for the greatest similarity. If $x(k)$ is identical to $y(k)$, the crcss correlation function is called autocorrelation function. It is given by:

$$\varphi_x(k) = \sum_{l=-\infty}^{+\infty} x(l)\ x(k+l) \tag{4}$$

Its maximum is at the origin $k = 0$. If this function is normalized by dividing it by the variance of the signal $x(k)$, the result is called correlation coefficient. Its values lie between +1 and -1.

An equivalent way of computing correlation functions is obtained by taking the discrete Fourier transform of both side of eq. (3) or eq. (4). One obtains respectively;

$$\Phi_{xy}(n) = X^*(n)\ Y(n) \tag{5}$$

and

$$\Phi_x(n) = X^*(n)\ X(n) = |X(n)|^2 \tag{6}$$

These results can be proved easily. They are left as exercises to the reader.

2.3 THE z TRANSFORM

The discrete Fourier transform is a very powerful tool for analysing and synthesising signals. It is not, however, suitable for studying signal processing systems. A more general transformation is needed. The z transform fulfils this need and becomes identical to Fourier transform in a particular case. The z transform of a signal is defined by:

$$X(z) = \sum_{k=-\infty}^{+\infty} x(k)\ z^{-k} \tag{7}$$

where z is a complex variable. A power series, such as this one, may not converge for all the possible values of z. The area of the complex plane z containing all the values for which eq. (7) converges is called convergence region.