Lecture Notes in Computer Science 4391

Commenced Publication in 1973
Founding and Former Series Editors:
Gerhard Goos, Juris Hartmanis, and Jan van Leeuwen

Editorial Board

Yannis Stylianou Marcos Faundez-Zanuy
Anna Esposito (Eds.)

Progress in Nonlinear Speech Processing

Springer

Volume Editors

Yannis Stylianou
University of Crete
Computer Science Department
Heraklion, Crete, Greece, 71409
E-mail: yannis@csd.uoc.gr

Marcos Faundez-Zanuy
Escola Universitària Politècnica de Mataró
Barcelona, Spain
E-mail: faundez@eupmt.es

Anna Esposito
Seconda Università di Napoli
Dipartimento di Psicologia
Via Vivaldi 43, 81100 Caserta, Italy
E-mail: iiass.annaesp@tin.it

Library of Congress Control Number: 2007922930

CR Subject Classification (1998): H.5.2, H.5, I.2.6-7, I.4-5, I.2.10, F.4.3

LNCS Sublibrary: SL 6 – Image Processing, Computer Vision, Pattern Recognition, and Graphics

ISSN 0302-9743
ISBN-10 3-540-71503-7 Springer Berlin Heidelberg New York
ISBN-13 978-3-540-71503-0 Springer Berlin Heidelberg New York

Typesetting: Camera-ready by author, data conversion by Scientific Publishing Services, Chennai, India
Printed on acid-free paper SPIN: 12038923 06/3142 5 4 3 2 1 0

Preface

The last meeting of the Management Committee of the COST Action 277: "Nonlinear Speech Processing" was held in Heraklion, Crete, Greece, September 20–23, 2005 during the Workshop on Nonlinear Speech Processing (WNSP). This was the last event of COST Action 277. The Action started in 2001. During the workshop, members of the Management Committee and invited speakers presented overviews of their work during these four years (2001–2005) of research combining linear and nonlinear approaches for processing the speech signal. In this book, 13 contributions summarize part of this (mainly) European effort in this field. The aim of this book is to provide an additional and/or an alternative way to the traditional approach of linear speech processing to be used by researchers working in the domain. For all the chapters presented here, except Chaps. 4, 5, and 12, there is audiovisual material available at http://www.ics.forth.gr/wnsp05/index.html, where corresponding lectures and Power Point presentations are available by the authors.

The contributions cover the following areas:

1. Speech analysis for speech synthesis, speech recognition, speech–non-speech discrimination and voice quality assessment
2. Speaker recognition/verification from a natural or modified speech signal
3. Speech recognition
4. Speech enhancement
5. Emotional state detection

Speech Analysis

Although in many speech applications the estimation of the glottal waveform is very useful, the estimation of this signal is not always robust and accurate. Given a speech signal, the glottal waveform may be estimated through inverse filtering. For certain speech processing areas like the analysis of pathologic voices, where voice quality issues are closely related to the vocal folds activity, an accurate inverse filtering is highly desired. The chapter by Jacqueline Walker and Peter Murphy provides and extensive review on the estimation and analysis of the glottal waveform. The presentation starts with analog inverse filtering approaches and extends to approaches based on nonlinear least squares estimation methods.

Analysis of speech signals provides many features for efficient voice quality assessment. Peter Murphy presents a tool based on the rahmonic analysis of speech for detecting irregularities in synthetic and natural human voice signals. The cepstrum is decomposed into two areas; the low quefrency and the high quefrency. The first rahmonic in the high-quefrency region provides information about the periodicity of a signal. In this chapter, a new measure taking into account all rahmonics in the cepstrum is proposed, and results using synthetic

and real speech data are presented providing therefore an additional measure to the usual harmonic-to-noise ratio-related measures for voice quality assessment.

A new tool for spectral analysis of speech signals referred to as chirp group delay is presented by Baris Bozkurt, Thierry Dutoit and Laurent Couvreur. With this tool a certain number of group delay functions are computed in the z-plane on circles different from the usual unit circle. Two important applications of this tool are presented by the authors; formant tracking and a new set of features for the automatic speech recognition task.

Two applications of the so-called neurocomputational speech and sounds processing are presented in the chapter prepared by Jean Rouat, Stéphane Loiselle and Ramin Pichevar. There is evidence that for both the visual and auditory systems the sequence order of firing is crucial to perform recognition tasks (rank order coding). In the first application a speech recognition system based on rank order coding is presented and it is compared against a conventional HMM-based recognizer. In the second application the acoustical source separation is addressed where simultaneous auditory images are combined with a network of oscillatory spiking neurons to segregate and bind auditory objects.

In the chapter presented by Maria Markaki, Michael Wohlmayer and Yannis Stylianou a speech–non-speech classifier is developed based on modulation spectrograms. The information bottleneck method is used for extracting relevant speech modulation frequencies across time and frequency dimensions creating therefore a set of characteristic modulation spectra for each type of sound. An efficient and simple classifier based on the similarity of a sound to these characteristic modulation spectra is presented.

The chapter by Yannis Pantazis and Yannis Stylianou deals with the automatic detection of audible discontinuities in concatenative speech synthesis. Both linear and nonlinear features are extracted at the boundaries of connected speech segments and a list of distances is evaluated. For the evaluation purposes, results from a subjective test for the same task have been taken into account. Among the most promising features for this task are the amplitude and frequency modulations occuring in the spectra of the continuous speech when two perceptually incompatible speech segments are joined. The Fisher linear discriminator seems to perform a high detection score.

Speaker Recognition/Verification

A review and perspectives on voice disguise and its automatic detection are given in the chapter prepared by Patrick Perrot, Guido Aversano and Gérard Chollet. A list of ways for modifying the quality of voice is presented along with the different techniques proposed in the literature. A very difficult topic is that of automatic detection of disguised voice. The authors describe a list of main indicators that can be used for this task.

Bouchra Abboud, Hervé Bredin, Guido Aversano, and Gérard Chollet present an overview on audio-visual identity verification tasks. Face and voice transformation techniques are reviewed for the face, speaker and talking face verification.

It is shown that rather a limited amount of information can be modified for troubling state-of-the-art audiovisual identity verification systems. An explicit talking face modeling is then proposed to overcome the weak points of these systems.

State-of-the-art systems and challenges in the text-independent speaker verification task are presented in the chapter prepared by Dijana Petrovska-Delacrétaz, Asma El Hannani and Gérard Chollet. Speakers' variability and variabilities on the transmission channel are discussed along with the possible choices of speech parameterization and speaker models. The use of speech recognition for the speaker verification task is also discussed, showing that a development of new services based on speaker and speech recognition is possible.

Marcos Faundez-Zanuy and Mohamed Chetouani present an overview of nonlinear predictive models and their application in speaker recognition. Challenges and possibilities in extracting nonlinear features towards this task are provided along with the various strategies that one can follow for using these nonlinear features. Both nonparametric (e.g., codebook based) and parametric approaches (e.g., Volterra series) are described. A nonlinear extension of the well-known linear prediction theory is provided, referred to as neural predictive coding.

Speech Recognition

Although hidden markov models dominate the speech recognition area, the support vector machine(SVM) is a powerful tool in machine learning and the chapter by R.Solera-Ureña, J.Padrell-Sendra, D. Martín-Iglesias, A. Gallardo-Antolín, C.Peláez-Moreno and F.Díaz-de-María is an overview of the application of SVMs for automatic speech recognition, for isolated word recognition and for continuous speech recognition and connected digit recognition.

Speech Enhancement

Considering single and multichannel-based solutions for the speech enhancement task, A. Hussain, M. Chetouani, S. Squartini, A. Bastari and F. Piazza present an overview of the noise reduction approaches focusing on the additive independent noise case. The non-Gaussian properties of the involved signals and the lack of linearity in the related processes provide a motivation for the development of nonlinear algorithms for the speech enhancement task. A very useful table summarizing the advantages and drawbacks of the currently proposed nonlinear techniques is presented at the end of the chapter.

Emotional State Detection

The use of visual and auditory information for predicting the emotional state of humans is discussed in the chapter by Anna Esposito. Results from subjective tests using a single channel (audio or visual) and combined channels (using both visual and auditory information) are provided. Based on these results, auditory channels outperform visual channels of information in predicting the emotional

state of a human being. An information theoretic model is then proposed to support these results.

The editors are grateful to the colleagues who contributed to this book, and to the reviewers for their willingness to review the chapter and provide useful feedback to the authors. We would like also to thank the COST office that provided financial support for the organization of the WNSP in Crete, the University of Crete and the Institute of Computer Science at FORTH for technically supporting the event. Especially, we would like to thank Theodosia Bitzou for designing the logo of the meeting and the folder given to the participants, and Manolis Zouraris, Andreas Holzapfel and Giorgos Kafentzis for organizing the audio-visual material. Finally, we would like to thank Springer, and particularly Alfred Hofmann and Ursula Barth for their help in publishing this post-conference book.

January 2007

Yannis Stylianou
Marcos Faundez-Zanuy
Anna Esposito

Organization

WNSP 2005 was organized by the department of Computer Science, Univeristy of Crete, the Institute of Computer Science of the Foundation for Research and Technology Hellas (FORTH), and the COST (European Cooperation in the field of Scientific and Technical Research) office.

Scientific Committee

Gérard Chollet	ENST Paris, France
Thierry Dutoit	FPMS, Mons, Belgium
Anna Esposito	2nd University of Napoli, Italy
Marcos Faundez-Zanuy	EUPMT, Barcelona, Spain
Eric Keller	University of Lausanne, Switzerland
Gernot Kubin	TUG, Graz, Austria
Petros Maragos	NTUA, Athens, Greece
Jean Schoentgen	University Libre Bruxelles, Belgium
Yannis Stylianou	University of Crete, Greece

Program Committee

Conference Chair	Yannis Stylianou (University of Crete, Greece)
Organizing Chair	Yannis Agiomyrgiannakis (University of Crete, Greece)
Audio-Video Material	Manolis Zouraris (University of Crete, Greece)
Local Arrangements	Maria Markaki (University of Crete, Greece)

Sponsoring Institutions

COST Office, Brussels, Belgium
University of Crete, Heraklion, Crete, Greece
Institute of Computer Science, FORTH, Heraklion, Crete, Greece

Table of Contents

Progress in Nonlinear Speech Processing