Preface

This volume contains invited and contributed papers presented at the 9th International Summer School "*Neural Nets E.R. Caianiello*" on Nonlinear Speech Processing: Algorithms and Analysis, held in Vietri sul Mare, Salerno, Italy, during September 13–18, 2004.

The aim of this book is to provide primarily high-level tutorial coverage of the fields related to nonlinear methods for speech processing and analysis, including new approaches aimed at improving speech applications.

Fourteen surveys are offered by specialists in the field. Consequently, the volume may be used as a reference book on nonlinear methods for speech processing and analysis. Also included are fifteen papers that present original contributions in the field and complete the tutorials.

The volume is divided into five sections: Dealing with Nonlinearities in Speech Signal, Acoustic-to-Articulatory Modeling of Speech Phenomena, Data Driven and Speech Processing Algorithms, Algorithms and Models Based on Speech Perception Mechanisms, and Task-Oriented Speech Applications.

Dealing with Nonlinearities in Speech Signals is an introductory section where nonlinear aspects of the speech signal are introduced from three different points of view. The section includes three papers. The first paper, authored by Anna Esposito and Maria Marinaro, is an attempt to introduce the concept of nonlinearity revising several nonlinear phenomena observed in the acoustics, the production and the perception of speech. Also discussed is the engineering endeavor to model these phenomena.

The second paper, by Marcos Faundez-Zanuy, gives an overview of nonlinear predictive models, with special emphasis on neural nets, and discusses several well-known nonlinear strategies, such as multistart random weights initialization, regularization, early stop with validation, committees of neural nets, and neural net architectures.

The third paper, by Simon Haykin, faces the problem of processing nonlinear, non-Gaussian, and nonstationary signals describing the mathematical implications derived by these assumptions. The topic has important practical implications of its own, not only in speech but also in the field of signal processing.

Acoustic-to-Articulatory Modeling of Speech Phenomena deals with problems related to the acoustic-phonetic theory in which basic speech sounds are characterized according to both their articulatory features and the associated acoustic measurements. Fundamental and innovative ideas in speech production are covered. This section contains three papers. The first paper, authored by Eric Keller, discusses voice quality within a large predictive and methodological framework. Voice quality phenomena are reviewed at two levels: (1) at the level of independent variables, topic-, affective-, attitude-, emotion-, gender-, articulation-, language-, sociolect-, gender-, age- and speaker-related predictors; and (2) at the level of dependent variables, where the empirical identification of voice quality parameters in the speech signal are summarized. Specifically, Fant's original and revised source-filter models are reviewed.

The second paper, by Gernot Kubin, Claudia Lainscsek, and Erhard Rank, discusses the identification of nonlinear oscillator models for speech analysis and synthesis. The paper, starting from the first successful application of a nonlinear oscillator model to high-quality speech signal processing, reviews the numerous developments that have been initiated to turn nonlinear oscillators into a standard tool for speech technology, and compares several of these attempts with a special emphasis on adaptive model identification from data.

The third paper, by Jean Schoentgen, revises speech modeling based on acoustic-to-articulatory mapping. The acoustic-articulatory mapping is the inference of acoustic equivalents of a speaker's vocal tract. This mapping involves the computation of models of the vocal tract whose eigenfrequencies are identical to the speaker's formant frequencies. The usefulness of such a transformation is in the idea that formant data may be interpreted and manipulated more easily in the transform domain (i.e., the geometric domain) and therefore acoustic-to-geometric mapping would be of great use in the framework of automatic speech and speaker recognition.

Data Driven and Speech Processing Algorithms deals with new and standard techniques used to provide speech features valuable for related speech applications. This section contains five papers.

The first, by Alessandro Bastari, Stefano Squartini, and Francesco Piazza, reports on the problem of separating a speech signal from a set of observables when the mixing system is undetermined. A common way to face this task is to see it as a Blind Source Separation (BSS) problem. The paper revises several approaches to solve different formulations of the blind source separation problem and also suggests the use of alternative time-frequency transforms such as the discrete wavelet transform (DWT) and the Stockwell transform (ST). The second paper, by Gerard Chollet, Kevin McTait, and Dijana Petrovska-Delacretaz, reviews experiments exploiting the automatic language independent speech processing (ALISP) approach to the development of speech processing applications driven by data, and how this strategy could be particularly useful for low-rate speech coding, recognition, translation and speaker verification.

The third and the fifth papers, by Peter Murphy and Olatunji Akande, and Yannis Stylianou, respectively, describe time-domain and frequency-domain techniques to estimate the harmonic-to-noise ratio as an indicator of the aperiodicity of a voice signal. New algorithms are proposed and applications to continuous speech recognition are also envisaged.

The fourth paper, on a predictive connectionist approach to speech recognition, by Bojan Petek, describes a context-dependent hidden control neural network (HCNN) architecture for large-vocabulary continuous-speech recognition. The basic building element of the proposed architecture, the context-dependent HCNN model, is a connectionist network trained to capture the dynamics of speech sub-word units. The HCNN model belongs to a family of Hidden Markov model/multi-layer perceptron (HMM/MLP) hybrids, usually referred to as predictive neural networks.

Algorithms and Models Based on Speech Perception Mechanisms includes three papers. The first, by Anna Esposito and Guido Aversano, discusses speech segmentation methods that do not use linguistic information and proposes a new segmentation algorithm based on perceptually processed speech features. A performance study is also

reported through performance comparisons with standard speech segmentation methods such as temporal decomposition, Kullback–Leibler distances, and spectral variation functions.

The second paper, by Amir Hussain, Tariq Durrani, John Soraghan, Ali Aikulaibi, and Nhamo Mterwa, reports on nonlinear adaptive speech enhancement schemes inspired by features of early auditory processing, which allows for the manipulation of several factors that may influence the intelligibility and perceived quality of the processed speech. In this context it is shown that stochastic resonance might be a general strategy employed by the central nervous system for the improved detection of weak signals and that the effects of stochastic resonance in sensory processing might extend past an improvement in signal detection.

The last paper, by Jean Rouat, Ramin Pichevar, and Stéphanie Loiselle, presents potential solutions to the problem of sound separation based on computational auditory scene analysis (CASA), by using nonlinear speech processing and spiking neural Networks. The paper also introduces the reader to the potential use of spiking neurons in signal and spatiotemporal processing.

Task Oriented Speech Applications includes the papers of 15 contributors which propose original and seminal works on speech applications and suggest new principles by means of which task oriented applications may be successful.

The editors would like to thank first of all the COST European Cooperation in the field of Scientific and Technical Research, the oldest and most widely used system for research networking in Europe. COST provided full financial support for a significant number of attendants plus some financial contributions for two outstanding speakers (Simon Haykin and José Príncipe), for which we are very grateful. COST 277 set up the first summer school on the a COST framework in the last 33 years. Thus, this work can be considered historic, and we hope to repeat this successful event in the near future. COST is based on an inter-governmental framework for cooperation research agreed following a ministerial conference in 1971. The mission of COST is to strengthen Europe in scientific and technical research through the support of European cooperation and interaction between European researchers. Its aims are to strengthen noncompetitive and prenormative research in order to maximize European synergy and added value.

The keynote presentations reported in this book are mostly from speakers who are part of the Management Committee of COST Action 277, "Nonlinear Speech Processing," which has acted as a catalyst for research on nonlinear speech processing since June 2001.

The editors are extremely grateful to the International Society of Phonetic Sciences (ISPHS), and in particular Prof. Ruth Bahr, the International Institute for Advanced Scientific Studies "E.R. Caianiello," the Università di Salerno, Dipartimento di Fisica, the Seconda Università di Napoli, in particular the Dean of the Facoltà di Psicologia, Prof. Maria Sbandi, the Regione Campania, and the Provincia di Salerno for their support in sponsoring, financing, and organizing the school. Special thanks are due to Tina Nappi and Michele Donnarumma for their editorial and technical support, and to Guido Aversano, Marinella Arnone, Antonietta M. Esposito, Antonio Natale, Luca Pugliese, and Silvia Scarpetta for their help in the local organization of the school.

In addition, the editors are grateful to the contributors of this volume and the keynote speakers whose work stimulated an extremely interesting interaction with the attendees, who in turn shall not be forgotten – they are highly motivated and bright.

This book is dedicated to those who recognize the nonsense of wars, and to children's curiosity. Both are needed to motivate our research.

September 2004

Gerard Chollet
Anna Esposito
Marcos Faundez-Zanuy
Maria Marinaro

Scientific Committee

Gerard Chollet (ENST – CNRS URA, Paris, France)
Anna Esposito (Seconda Università di Napoli and IIASS, Vietri sul Mare, Italy)
Marcos Faundez-Zanuy (Escola Universitaria Politecnica de Mataro, Barcelona, Spain)
Maria Marinaro (Università di Salerno and IIASS, Vietri sul Mare, Italy)
Eric Moulines (ENST – CNRS URA, Paris, France)
Isabel Troncoso (INESC-ID/IST, Portugal)

Sponsors and Supporters

European Commission COST Action 277: Nonlinear Speech Processing
International Institute for Advanced Scientific Studies “E.R. Caianiello” (IIASS), Italy
International Speech Communication Association (ISCA)
Seconda Università di Napoli, Facoltà di Psicologia (Italy)
International Society of Phonetic Sciences (ISPhS)
Università di Salerno, Dipartimento di Scienze Fisiche E.R. Caianiello (Italy)
Regione Campania (Italy)
Provincia di Salerno (Italy)