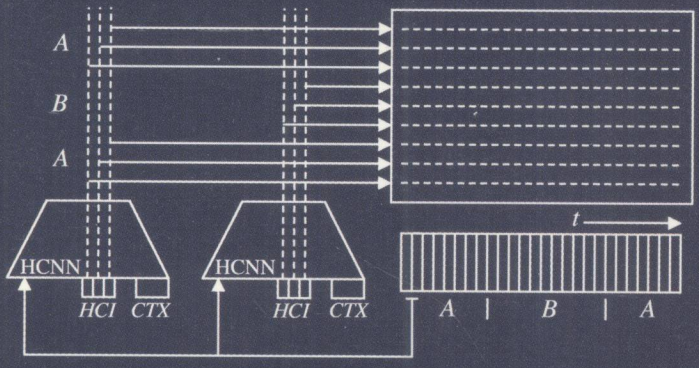
Tutorial

LNAI 3445

Gérard Chollet
Anna Esposito
Marcos Faundez-Zanuy
Maria Marinaro (Eds.)

Nonlinear Speech Modeling and Applications

Advanced Lectures and Revised Selected Papers



Springer

Gérard Chollet Anna Esposito
Marcos Faundez-Zanuy Maria Marinaro (Eds.)

Nonlinear Speech Modeling and Applications

Advanced Lectures and Revised Selected Papers

Springer

Series Editors

Jaime G. Carbonell, Carnegie Mellon University, Pittsburgh, PA, USA
Jörg Siekmann, University of Saarland, Saarbrücken, Germany

Volume Editors

Gérard Chollet
CNRS URA-820, ENST, Dept. TSI
46 rue Barrault, 75634 Paris cedex 13, France
E-mail: chollet@tsi.enst.fr

Anna Esposito
Second University of Naples, Department of Psychology
and
IIASS, International Institute for Advanced Scientific Studies
Via Pellegrino 19, Vietri sul Mare (SA), Italy
E-mail: iiass.annaesp@tin.it

Marcos Faundez-Zanuy
Escola Universitaria Politecnica de Mataro
Avda. Puig i Cadafalch 101-111, 08303 Mataro (Barcelona), Spain
E-mail: faundez@eupmt.es

Maria Marinaro
University of Salerno "E.R.Caianiello", Dept. of Physics
Via Salvatore Allende, Baronissi, 84081 Salerno, Italy
and
IIASS, International Institute for Advanced Scientific Studies
Via Pellegrino 19, Vietri sul Mare (SA), Italy
E-mail: iiass.vietri@tin.it

Library of Congress Control Number: 2005928448

CR Subject Classification (1998): I.2.7, J.5, C.3

ISSN 0302-9743
ISBN-10 3-540-27441-3 Springer Berlin Heidelberg New York
ISBN-13 978-3-540-27441-4 Springer Berlin Heidelberg New York

Typesetting: Camera-ready by author, data conversion by Olgun Computergrafik
Printed on acid-free paper SPIN: 11520153 06/3142 5 4 3 2 1 0

Lecture Notes in Artificial Intelligence 3445

Edited by J. G. Carbonell and J. Siekmann

Subseries of Lecture Notes in Computer Science

Lecture Notes in Artificial Intelligence (LNAI)

Vol. 3587: P. Perner, A. Imiya (Eds.), Machine Learning and Data Mining in Pattern Recognition. XVII, 695 pages. 2005.

Vol. 3575: S. Wermter, G. Palm, M. Elshaw (Eds.), Biomimetic Neural Learning for Intelligent Robots. IX, 383 pages. 2005.

Vol. 3571: L. Godo (Ed.), Symbolic and Quantitative Approaches to Reasoning with Uncertainty. XVI, 1028 pages. 2005.

Vol. 3559: P. Auer, R. Meir (Eds.), Learning Theory. XI, 692 pages. 2005.

Vol. 3554: A. Dey, B. Kokinov, D. Leake, R. Turner (Eds.), Modeling and Using Context. XIV, 572 pages. 2005.

Vol. 3533: M. Ali, F. Esposito (Eds.), Innovations in Applied Artificial Intelligence. XX, 858 pages. 2005.

Vol. 3528: P.S. Szczepaniak, J. Kacprzyk, A. Niewiadomski (Eds.), Advances in Web Intelligence. XVII, 513 pages. 2005.

Vol. 3518: T.B. Ho, D. Cheung, H. Liu (Eds.), Advances in Knowledge Discovery and Data Mining. XXI, 864 pages. 2005.

Vol. 3508: P. Bresciani, P. Giorgini, B. Henderson-Sellers, G. Low, M. Winikoff (Eds.), Agent-Oriented Information Systems II. X, 227 pages. 2005.

Vol. 3505: V. Gorodetsky, J. Liu, V. A. Skormin (Eds.), Autonomous Intelligent Systems: Agents and Data Mining. XIII, 303 pages. 2005.

Vol. 3501: B. Kégl, G. Lapalme (Eds.), Advances in Artificial Intelligence. XV, 458 pages. 2005.

Vol. 3492: P. Blache, E. Stabler, J. Busquets, R. Moot (Eds.), Logical Aspects of Computational Linguistics. X, 363 pages. 2005.

Vol. 3488: M.-S. Hacid, N.V. Murray, Z.W. Raś, S. Tsumoto (Eds.), Foundations of Intelligent Systems. XIII, 700 pages. 2005.

Vol. 3476: J. Leite, A. Omicini, P. Torroni, P. Yolum (Eds.), Declarative Agent Languages and Technologies II. XII, 289 pages. 2005.

Vol. 3464: S.A. Brueckner, G.D.M. Serugendo, A. Karageorgos, R. Nagpal (Eds.), Engineering Self-Organising Systems. XIII, 299 pages. 2005.

Vol. 3452: F. Baader, A. Voronkov (Eds.), Logic for Programming, Artificial Intelligence, and Reasoning. XI, 562 pages. 2005.

Vol. 3446: T. Ishida, L. Gasser, H. Nakashima (Eds.), Massively Multi-Agent Systems I. XI, 349 pages. 2005.

Vol. 3445: G. Chollet, A. Esposito, M. Faundez-Zanuy, M. Marinaro (Eds.), Nonlinear Speech Modeling and Applications. XIII, 433 pages. 2005.

Vol. 3438: H. Christiansen, P.R. Skadhauge, J. Villadsen (Eds.), Constraint Solving and Language Processing. VIII, 205 pages. 2005.

Vol. 3430: S. Tsumoto, T. Yamaguchi, M. Numao, H. Motoda (Eds.), Active Mining. XII, 349 pages. 2005.

Vol. 3419: B. Faltings, A. Petcu, F. Fages, F. Rossi (Eds.), Constraint Satisfaction and Constraint Logic Programming. X, 217 pages. 2005.

Vol. 3416: M. Böhlen, J. Gamper, W. Polasek, M.A. Wimmer (Eds.), E-Government: Towards Electronic Democracy. XIII, 311 pages. 2005.

Vol. 3415: P. Davidsson, B. Logan, K. Takadama (Eds.), Multi-Agent and Multi-Agent-Based Simulation. X, 265 pages. 2005.

Vol. 3403: B. Ganter, R. Godin (Eds.), Formal Concept Analysis. XI, 419 pages. 2005.

Vol. 3398: D.-K. Baik (Ed.), Systems Modeling and Simulation: Theory and Applications. XIV, 733 pages. 2005.

Vol. 3397: T.G. Kim (Ed.), Artificial Intelligence and Simulation. XV, 711 pages. 2005.

Vol. 3396: R.M. van Eijk, M.-P. Huget, F. Dignum (Eds.), Agent Communication. X, 261 pages. 2005.

Vol. 3394: D. Kudenko, D. Kazakov, E. Alonso (Eds.), Adaptive Agents and Multi-Agent Systems II. VIII, 313 pages. 2005.

Vol. 3392: D. Seipel, M. Hanus, U. Geske, O. Bartenstein (Eds.), Applications of Declarative Programming and Knowledge Management. X, 309 pages. 2005.

Vol. 3374: D. Weyns, H.V.D. Parunak, F. Michel (Eds.), Environments for Multi-Agent Systems. X, 279 pages. 2005.

Vol. 3371: M.W. Barley, N. Kasabov (Eds.), Intelligent Agents and Multi-Agent Systems. X, 329 pages. 2005.

Vol. 3369: V.R. Benjamins, P. Casanovas, J. Breuker, A. Gangemi (Eds.), Law and the Semantic Web. XII, 249 pages. 2005.

Vol. 3366: I. Rahwan, P. Moraitis, C. Reed (Eds.), Argumentation in Multi-Agent Systems. XII, 263 pages. 2005.

Vol. 3359: G. Grieser, Y. Tanaka (Eds.), Intuitive Human Interfaces for Organizing and Accessing Intellectual Assets. XIV, 257 pages. 2005.

Vol. 3346: R.H. Bordini, M. Dastani, J. Dix, A.E.F. Seghrouchni (Eds.), Programming Multi-Agent Systems. XIV, 249 pages. 2005.

Vol. 3345: Y. Cai (Ed.), Ambient Intelligence for Scientific Discovery. XII, 311 pages. 2005.

Vol. 3343: C. Freksa, M. Knauff, B. Krieg-Brückner, B. Nebel, T. Barkowsky (Eds.), Spatial Cognition IV. XIII, 519 pages. 2005.

Vol. 3339: G.I. Webb, X. Yu (Eds.), AI 2004: Advances in Artificial Intelligence. XXII, 1272 pages. 2004.

Vol. 3336: D. Karagiannis, U. Reimer (Eds.), Practical Aspects of Knowledge Management. X, 523 pages. 2004.

Vol. 3327: Y. Shi, W. Xu, Z. Chen (Eds.), Data Mining and Knowledge Management. XIII, 263 pages. 2005.

Vol. 3315: C. Lemaître, C.A. Reyes, J.A. González (Eds.), Advances in Artificial Intelligence – IBERAMIA 2004. XX, 987 pages. 2004.

Vol. 3303: J.A. López, E. Benfenati, W. Dubitzky (Eds.), Knowledge Exploration in Life Science Informatics. X, 249 pages. 2004.

Vol. 3301: G. Kern-Isberner, W. Rödder, F. Kulmann (Eds.), Conditionals, Information, and Inference. XII, 219 pages. 2005.

Vol. 3276: D. Nardi, M. Riedmiller, C. Sammut, J. Santos-Victor (Eds.), RoboCup 2004: Robot Soccer World Cup VIII. XVIII, 678 pages. 2005.

Vol. 3275: P. Perner (Ed.), Advances in Data Mining. VIII, 173 pages. 2004.

Vol. 3265: R.E. Frederking, K.B. Taylor (Eds.), Machine Translation: From Real Users to Research. XI, 392 pages. 2004.

Vol. 3264: G. Paliouras, Y. Sakakibara (Eds.), Grammatical Inference: Algorithms and Applications. XI, 291 pages. 2004.

Vol. 3259: J. Dix, J. Leite (Eds.), Computational Logic in Multi-Agent Systems. XII, 251 pages. 2004.

Vol. 3257: E. Motta, N.R. Shadbolt, A. Stutt, N. Gibbins (Eds.), Engineering Knowledge in the Age of the Semantic Web. XVII, 517 pages. 2004.

Vol. 3249: B. Buchberger, J.A. Campbell (Eds.), Artificial Intelligence and Symbolic Computation. X, 285 pages. 2004.

Vol. 3248: K.-Y. Su, J. Tsujii, J.-H. Lee, O.Y. Kwong (Eds.), Natural Language Processing – IJCNLP 2004. XVIII, 817 pages. 2005.

Vol. 3245: E. Suzuki, S. Arikawa (Eds.), Discovery Science. XIV, 430 pages. 2004.

Vol. 3244: S. Ben-David, J. Case, A. Maruoka (Eds.), Algorithmic Learning Theory. XIV, 505 pages. 2004.

Vol. 3238: S. Biundo, T. Frühwirth, G. Palm (Eds.), KI 2004: Advances in Artificial Intelligence. XI, 467 pages. 2004.

Vol. 3230: J.L. Vicedo, P. Martínez-Barco, R. Muñoz, M. Saiz Noeda (Eds.), Advances in Natural Language Processing. XII, 488 pages. 2004.

Vol. 3229: J.J. Alferes, J. Leite (Eds.), Logics in Artificial Intelligence. XIV, 744 pages. 2004.

Vol. 3228: M.G. Hinchey, J.L. Rash, W.F. Truszkowski, C.A. Rouff (Eds.), Formal Approaches to Agent-Based Systems. VIII, 290 pages. 2004.

Vol. 3215: M.G.. Negoita, R.J. Howlett, L.C. Jain (Eds.), Knowledge-Based Intelligent Information and Engineering Systems, Part III. LVII, 906 pages. 2004.

Vol. 3214: M.G.. Negoita, R.J. Howlett, L.C. Jain (Eds.), Knowledge-Based Intelligent Information and Engineering Systems, Part II. LVIII, 1302 pages. 2004.

Vol. 3213: M.G.. Negoita, R.J. Howlett, L.C. Jain (Eds.), Knowledge-Based Intelligent Information and Engineering Systems, Part I. LVIII, 1280 pages. 2004.

Vol. 3209: B. Berendt, A. Hotho, D. Mladenic, M. van Someren, M. Spiliopoulou, G. Stumme (Eds.), Web Mining: From Web to Semantic Web. IX, 201 pages. 2004.

Vol. 3206: P. Sojka, I. Kopecek, K. Pala (Eds.), Text, Speech and Dialogue. XIII, 667 pages. 2004.

Vol. 3202: J.-F. Boulicaut, F. Esposito, F. Giannotti, D. Pedreschi (Eds.), Knowledge Discovery in Databases: PKDD 2004. XIX, 560 pages. 2004.

Vol. 3201: J.-F. Boulicaut, F. Esposito, F. Giannotti, D. Pedreschi (Eds.), Machine Learning: ECML 2004. XVIII, 580 pages. 2004.

Vol. 3194: R. Camacho, R. King, A. Srinivasan (Eds.), Inductive Logic Programming. XI, 361 pages. 2004.

Vol. 3192: C. Bussler, D. Fensel (Eds.), Artificial Intelligence: Methodology, Systems, and Applications. XIII, 522 pages. 2004.

Vol. 3191: M. Klusch, S. Ossowski, V. Kashyap, R. Unland (Eds.), Cooperative Information Agents VIII. XI, 303 pages. 2004.

Vol. 3187: G. Lindemann, J. Denzinger, I.J. Timm, R. Unland (Eds.), Multiagent System Technologies. XIII, 341 pages. 2004.

Vol. 3176: O. Bousquet, U. von Luxburg, G. Rätsch (Eds.), Advanced Lectures on Machine Learning. IX, 241 pages. 2004.

Vol. 3171: A.L.C. Bazzan, S. Labidi (Eds.), Advances in Artificial Intelligence – SBIA 2004. XVII, 548 pages. 2004.

Vol. 3159: U. Visser, Intelligent Information Integration for the Semantic Web. XIV, 150 pages. 2004.

Vol. 3157: C. Zhang, H. W. Guesgen, W.K. Yeap (Eds.), PRICAI 2004: Trends in Artificial Intelligence. XX, 1023 pages. 2004.

Vol. 3155: P. Funk, P.A. González Calero (Eds.), Advances in Case-Based Reasoning. XIII, 822 pages. 2004.

Vol. 3139: F. Iida, R. Pfeifer, L. Steels, Y. Kuniyoshi (Eds.), Embodied Artificial Intelligence. IX, 331 pages. 2004.

Vol. 3131: V. Torra, Y. Narukawa (Eds.), Modeling Decisions for Artificial Intelligence. XI, 327 pages. 2004.

Vol. 3127: K.E. Wolff, H.D. Pfeiffer, H.S. Delugach (Eds.), Conceptual Structures at Work. XI, 403 pages. 2004.

Vol. 3123: A. Belz, R. Evans, P. Piwek (Eds.), Natural Language Generation. X, 219 pages. 2004.

Vol. 3120: J. Shawe-Taylor, Y. Singer (Eds.), Learning Theory. X, 648 pages. 2004.

Vol. 3097: D. Basin, M. Rusinowitch (Eds.), Automated Reasoning. XII, 493 pages. 2004.

Vol. 3071: A. Omicini, P. Petta, J. Pitt (Eds.), Engineering Societies in the Agents World. XIII, 409 pages. 2004.

Vol. 3070: L. Rutkowski, J. Siekmann, R. Tadeusiewicz, L.A. Zadeh (Eds.), Artificial Intelligence and Soft Computing - ICAISC 2004. XXV, 1208 pages. 2004.

Vol. 3068: E. André, L. Dybkjær, W. Minker, P. Heisterkamp (Eds.), Affective Dialogue Systems. XII, 324 pages. 2004.

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)

Table of Contents

Task Oriented Speech Applications

Some Notes on Nonlinearities of Speech

Anna Esposito[1,2] and Maria Marinaro[2,3]

[1] Seconda Università di Napoli, Dipartimento di Psicologia, Via Vivaldi 43, Caserta, Italy
anna.esposito@unina2.it, iiass.annaesp@tin.it
[2] IIASS, Via Pellegrino 19, 84019, Vietri sul Mare, Italy, INFM Salerno, Italy
[3] Università di Salerno, Via S. Allende, Baronissi, Salerno, Italy
marinaro@sa.infn.it

Abstract. Speech is exceedingly nonlinear. Efforts to propose non-linear models of its dynamics are worth to be made but difficult to implement since nonlinearity is not easily handled from an engineering and mathematical point of view. This paper is an attempt to make accessible to untrained people the notion of nonlinearity in speech, revising several nonlinear speech phenomena and the engineering endeavour for modeling them.

1 Introduction

Understanding speech is basic for facing a very broad class of challenging problems including language acquisition, speech disorders, and speech communication technologies. Verbal communication appears to be the most common, or at the least, the easiest method for conveying information and meanings among humans. Successful verbal communication is possible only if the addresser and the addressee use the same language but more importantly are provided with the same language building blocks which are constituted by the phonemes (or speech sounds), the lexicon (words), and the syntax (the rules for linking words together).

The implementation of the communication process requires a set of steps that could be summarized as follow:

1. A communicative intention of the addresser;
2. A code that assembles the communicative intention into words (the language);
3. A motor program that controls the movements of speech articulators and allows the transformation of air that emerges from the lungs into speech sounds (the code) through the appropriate configuration assumed by the vocal tract;
4. A physical channel (the air medium that conveys the produced sound to the addressee);
5. A transducer (the auditory apparatus) that converts the produced sound into the firing of the auditory neurons;
6. The addressee's understanding of the message.

The communication process could be exemplified through the ***Stimulus-Response*** model of behavioural psychology [44]. In this model there is a *sender* that encodes the communicative intention into a *message.* The message is sent out through a channel. A receiver is supposed to receive, decode, and provide feedback to the transmitted message. The communication will be considered successful if it results in a *transfer of meaning.*

G. Chollet et al. (Eds.): Nonlinear Speech Modeling, LNAI 3445, pp. 1–14, 2005.

The most influential schema of such a model, was proposed by Shannon and Weaver [75] and is reported in Figure 1.

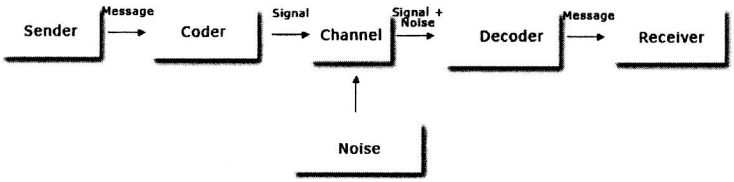


Fig. 1. An exemplification of the transmission model proposed by Shannon and Weaver [75]

The above model has several limitations, among those the most significant are: 1) it does not account for the effectiveness of the interaction between the sender and the receiver, since it assumes that communication is implemented by just carefully packaging the message to be transmitted; 2) feedback is not taken into account since it is problematic to implement it. However, some blocks in the above schema and in particular the *sender*, the *channel* and the *receiver* could be interpreted as an oversimplified description of the speech production system, the speech signal, and the speech perception system respectively. Each block has several nonlinear features and in the following sections we shall highlight some of them, in the attempt to clarify the notion of speech nonlinearities.

2 Nonlinearities in Speech Production

Speaking is a motor ability that consists of controlled and coordinated movements, performed primarily by the organs of the vocal tract (glottis, velum, tongue, lips) acting on the air in the respiratory passages (trachea, larynx, pharynx, mouth, nose) to produce speech sounds. The vocal organs generate a local disturbance of the air molecule at several positions in the vocal tract creating the sources for speech sound generation. The most common speech sources are: 1) the quasi-periodic vibration of the vocal cords – *voiced source*; 2) the turbulent noise generated by the passage the air through a quasi-narrow constriction (generally shaped by the tongue) in the oral cavity – *turbulent source*; 3) the plosive noise that follows the release of air compressed behind a complete obstruction of the oral cavity – *transient source*. However, the complex structure of speech sounds is not only due to the source generation features, but primarily to the response characteristics of the vocal tract that depends on the vocal tract configuration. The vocal tract configuration changes according to the motion of the vocal organs which modify its length, cross-sectional areas and response characteristics. The structure of speech sounds is generated by the combined effect of sound sources and vocal tract characteristics.

In absence of sounds, the vocal tract could be modeled as a single tube and the air molecules in it can be thought as a linear oscillator which responds to a disturbance with small displacements from the rest position. The conditions are extremely more complex when speech sounds are produced, since the motion of the vocal organs changes the vocal tract shape. A coarse model of the vocal tract in these conditions is a set of overlapping tubes of different lengths and cross-sectional areas. Due to its length and section area, each tube is subject to a different air pressure, which in turn,

generates different forces acting on the air molecules, causing their very complicated motion. In this case, a linear description fails to describe this complex dynamics and a non-linear approach should be used. However, nonlinearities introduce uncertainty and multiple solutions to several speech problems. As an example, let us report the problem known as the "*acoustic-to-articulatory mapping*". It consists in identifying, for a chosen vocal tract model constituted by a set of overlapping tubes, a set of parameters so that the resonances of the model correspond to the formants observed in a given produced speech sound. As is has been observed in [69], distinct vocal tract shapes can produce the same set of formant frequencies and therefore, a given set of formant values cannot univocally identify the vocal tract shape that has generated them (inverse problem). There are infinite solutions for the inverse problem. Even when functional constraints are imposed (such as minimal deformation, or minimal deformation rate of the vocal tract about a reference shape, or minimal deformation jerk) the inverse mapping does not fix the model, i.e., the real vocal tract shape that has produced that sound. To highlight the problems involved in the implementation of the acoustic-to-articulatory mapping, Schoentgen [69] reports a series of experiments where formant frequencies measured from sustained American English sounds are used to identify the corresponding vocal tract shapes. This is done taking three aspects into account: a) the accuracy of the vocal tract shapes estimate via formant-to-area mapping in comparison to the real vocal shapes; b) the underlying vocal tract models used; c) the numerical stability. Results shown that a good approximation is guaranteed only for speech sounds that are produced with a simple vocal tract configuration "*single cavity, single constriction, convex tongue, as well as constrained in the laryngo-pharynx, and, possibly, at the lips*" [69]; models based on a small number of conical tubelets with continuously varying cross area sections are preferred to exponential tubelets and cylindrical tubelets; the convergence to the desired format frequency values could be obtained with a precision greater than 1 Hz, even though the estimated vocal tract shapes could quantitatively and qualitatively differ from those built via the observed formant frequencies.

Another open problem is in the approximation of the glottal cycle waveform, i.e. the shape of the airflow produced at the glottis, before the signal is modified by the effects of the vocal tract configuration. The glottal waveform plays an important role in determining voice quality, which is defined as: "*the characteristic auditory coloring of an individual's voice, derived from a variety of laryngeal and supralaryngeal features and running continuously through the individual's speech*" [82]. An effective speech synthesizer should be able to adequately control the voice quality such that the synthesized speech sounds have a natural and distinctive tone. Several attempts have been made to model the glottal cycle waveform with the aim to identify the features of the glottal wave in accord to varying voice quality, and a "*conceptual framework*" of this research field is discussed in depth in this volume by Keller [40]. The most direct and automatic method, that does not involves invasive measurements is the inverse filtering technique [60], [61], that requires the recording (through a mask) of the glottal airflow at the mouth and its processing with a filtering system that separates the vocal tract characteristics from those of the glottal source. However, due to of the nonlinearities inside the transfer function of the vocal tract, this separation is not straightforward and the resulting glottal waveform models cannot account for several voice quality features. Nonlinearities are due to several factors, among these