Nonlinear Analyses and Algorithms for Speech Processing

International Conference on Non-Linear Speech Processing, NOLISP 2005
Barcelona, Spain, April 19-22, 2005
Revised Selected Papers
Preface

We present in this volume the collection of finally accepted papers of NOLISP 2005 conference. It has been the third event in a series of events related to Nonlinear speech processing, in the framework of the European COST action 277 “Nonlinear speech processing”.

Many specifics of the speech signal are not well addressed by conventional models currently used in the field of speech processing. The purpose of NOLISP is to present and discuss novel ideas, work and results related to alternative techniques for speech processing, which depart from mainstream approaches.

With this intention in mind, we provide an open forum for discussion. Alternate approaches are appreciated, although the results achieved at present may not clearly surpass results based on state-of-the-art methods.

The call for papers was launched at the beginning of 2005, addressing the following domains:

1. Non-Linear Approximation and Estimation
2. Non-Linear Oscillators and Predictors
3. Higher-Order Statistics
4. Independent Component Analysis
5. Nearest Neighbors
6. Neural Networks
7. Decision Trees
8. Non-Parametric Models
9. Dynamics of Non-Linear Systems
10. Fractal Methods
11. Chaos Modeling
12. Non-Linear Differential Equations
13. Others

All the main fields of speech processing are targeted by the workshop, namely:

1. Speech Coding: The bit rate available for speech signals must be strictly limited in order to accommodate the constraints of the channel resource. For example, new low-rate speech coding algorithms are needed for interactive multimedia services on packet-switched networks such as the evolving mobile radio networks or the Internet, and nonlinear speech processing offers a good alternative to conventional techniques. Voice transmission will have to compete with other services such as data/image/video transmission for the limited bandwidth resources allocated to an ever growing, mobile network user base, and very low bit rate coding at consumer quality will see increasing demand in future systems.
2. Speech Synthesis: New telecommunication services should include the capability to produce speech in a “natural way”; to this end, much research is required for improving the voice quality of text-to-speech and concept-to-speech systems. Enriched output signals of self-excited nonlinear feedback oscillators are expected to permit matching synthetic voices better to human voices. In this area, the COST Action has built on results obtained in signal generation by COST Action 258 “The Naturalness of Synthetic Speech”.

3. Speaker Identification and Verification: Security in transactions, information access, etc. is another important question to be addressed in the future, and speaker identification/verification is perhaps one of the most important biometric systems, because of its feasibility for remote (telephonic) recognition without additional hardware requirements. This line of work has built on results from COST Action 250 “Speaker Recognition in Telephony”.

4. Speech Recognition: Speech recognition plays an increasingly important role in modern society. Nonlinear techniques allow us to merge feature extraction and classification problem and to include the dynamics of the speech signal in the model. This is likely to lead to significant improvements over current methods which are inherently static.

In addition, other topics have been discussed in detail, such as Voice Analysis, where the quality of the human voice is analyzed (including clinical phonetics applications) and where techniques for the manipulation of the voice character of an utterance are developed, and Speech Enhancement, for the improvement of signal quality prior to further transmission and/or processing by man or machine.

After a careful review process, 33 papers were accepted for publication, including the contribution of two invited speakers. A total of 15 sessions containing 43 papers were accepted for presentation, covering specific aspects like speaker recognition, speech analysis, voice pathologies, speech recognition, speech enhancement and applications.

NOLISP 2005 was organized by the Escola Universitaria Politècnica de Mataró, for the European COST action 277, “nonlinear speech processing”. Sponsorship was obtained from the Spanish Ministerio de Educación y Ciencia (TEC2004-20959-E), and Fundació Catalana per a la Recerca i la Innovació. We also want to acknowledge the support of ISCA, EURASIP and IEEE.

We would like to express our gratitude to the members of the NOLISP organizing committee, and to all the people who participated in the event (delegates, invited speakers, scientific committee). The editors would like to address a special mention to the people who helped in the review process as special or additional reviewers.
Finally, we would like to thank Springer, and specially Alfred Hofmann, for publishing this post-conference proceedings.

September 2005

Marcos Faundez-Zanuy, EUPMt
Léonard Janer, EUPMt
Anna Esposito, IASS
Antonio Satue-Villar, EUPMt
Josep Roure-Alcobé, EUPMt
Virginia Espinosa-Duró, EUPMt
Organization

Scientific Committee:

Frédéric Bimbot       IRISA, Rennes, France
Gérard Chollet        ENST, Paris, France
Marcos Faundez-Zanuy  EUPMt, Barcelona, Spain
Hynek Hermansky       OGI, Portland, USA
Eric Keller           University of Lausanne, Switzerland
Bastiaan Kleijn       KTH, Stockholm, Sweden
Gernot Kubin          TUG, Graz, Austria
Jorge Lucero          University of Brasilia, Brasil
Petros Maragos        Nat. Tech. Univ. of Athens, Greece
Stephen Mclaughlin    University of Edinburgh, UK
Kuldip Paliwal        University of Brisbane, Australia
Bojan Petek           University of Ljubljana, Slovenia
Jean Schoentgen       Univ. Libre Bruxelles, Belgium
Jordi Sole-Casals     Universitat de VIC, Barcelona, Spain
Isabel Trancoso       INESC, Portugal

Organising Committee:

Marcos Faundez-Zanuy  EUPMt
Jordi Sole-Casals     UVIC
Léonard Janer         EUPMt
Josep Roure           EUPMt
Antonio Satue-Villar  EUPMt
Virginia Espinosa-Duró EUPMt

Scientific Reviewers:

Amir Hussain
Andrew Morris
Anna Esposito
Atonio Satue-Villar
Bastiaan Kleijn
Bojan Petek
Dijana Petrovska
Enric Monte-Moreno
Erhard Rank
X

Organization

Fernando Diaz-de-María
Gernot Kubin
Jean Rouat
Jean Schoentgen
Jordi Solé-Casals
Léonard Janer
Marcos Faundez-Zanuy
Miguel-Ángel Ferrer-Ballester
Mohamed Chetouani
Najim Dehak
Pedro Gómez-Vilda
Peter Murphy
Robert van Kommer
Stephen McLaughlin
Yannis Stylianou

The sponsorship and support of:

- ESCOLA UNIVERSITÀRIA POLITÈCNICA DE MATARÓ (EUPMt), Spain
- EUROPEAN COMMISSION COST ACTION 277: Nonlinear Speech Processing
- MINISTERIO DE CIENCIA Y TECNOLOGÍA (TEC2004-20959-E), Spain
- FUNDACIÓ CATALANA PER A LA RECERCA I LA INNOVACIÓ
- INTERNATIONAL SPEECH COMMUNICATION ASSOCIATION (ISCA)
- EUROPEAN ASSOCIATION FOR SIGNAL, SPEECH AND IMAGE PROCESSING (EURASIP)
- IEEE INDUSTRIAL APPLICATIONS SOCIETY

are gratefully acknowledged
# Table of Contents

The COST-277 European Action: An Overview  

Neuro-fuzzy Logic in Signal Processing for Communications: From Bits to Protocols  
_Anna Pérez-Neira, Miguel A. Lagunas, Antoni Morell, Joan Bas_  ...... 10

Connected Operators for Signal and Image Processing  
_Philippe Salembier_ ................................. 37

## Speaker Recognition

Exploiting High-Level Information Provided by ALISP in Speaker Recognition  
_Aasma El Hannani, Dijana Petrovska-Delacrétaz_ ................. 66

MLP Internal Representation as Discriminative Features for Improved Speaker Recognition  
_Dalei Wu, Andrew Morris, Jacques Koreman_ ................................. 72

Weighting Scores to Improve Speaker-Dependent Threshold Estimation in Text-Dependent Speaker Verification  
_Javier R. Saeta, Javier Hernando_ ................................. 81

Parameter Optimization in a Text-Dependent Cryptographic-Speech-Key Generation Task  
_L. Paola García-Perera, Juan A. Nolazco-Flores, Carlos Mex-Perera_ ................................. 92

The COST-277 Speech Database  
_Marcos Faundez-Zanuy, Martin Hagmüller, Gernot Kubin, W. Bastiaan Kleijn_ ................................. 100

## Speech Analysis

Children’s Organization of Discourse Structure Through Pausing Means  
_Anna Esposito_ ................................. 108
<table>
<thead>
<tr>
<th>Title</th>
<th>Author(s)</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>F0 and Intensity Distributions of Marsec Speakers: Types of Speaker Prosody</td>
<td>Brigitte Zellner Keller</td>
<td>116</td>
</tr>
<tr>
<td>A Two-Level Drive – Response Model of Non-stationary Speech Signals</td>
<td>Friedhelm R. Drepper</td>
<td>125</td>
</tr>
<tr>
<td>Advanced Methods for Glottal Wave Extraction</td>
<td>Jacqueline Walker, Peter Murphy</td>
<td>139</td>
</tr>
<tr>
<td>Cepstrum-Based Estimation of the Harmonics-to-Noise Ratio for Synthesized and Human Voice Signals</td>
<td>Peter J. Murphy, Olatunji O. Akande</td>
<td>150</td>
</tr>
<tr>
<td>Pseudo Cepstral Analysis of Czech Vowels</td>
<td>Robert Vich</td>
<td>161</td>
</tr>
<tr>
<td>Voice Pathologies</td>
<td></td>
<td></td>
</tr>
<tr>
<td>On the Acoustic-to-Electropalatographic Mapping</td>
<td>Asterios Toutios, Konstantinos Margaritis</td>
<td>186</td>
</tr>
<tr>
<td>Synthesis of Disordered Voices</td>
<td>Julien Hanquinet, Francis Grenez, Jean Schoentgen</td>
<td>231</td>
</tr>
<tr>
<td>Title</td>
<td>Authors</td>
<td>Page</td>
</tr>
<tr>
<td>----------------------------------------------------------------------</td>
<td>-------------------------------------------------------------------------</td>
<td>------</td>
</tr>
<tr>
<td>Voice Pathology Detection by Vocal Cord Biomechanical Parameter</td>
<td>Pedro Gómez, Rafael Martínez, Francisco Díaz, Carlos Lázaro, Agustín Álvarez, Victoria Rodellar, Víctor Nieto</td>
<td>242</td>
</tr>
<tr>
<td>Estimation</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Speech Recognition</td>
<td></td>
<td></td>
</tr>
<tr>
<td>A Speech Recognizer Based on Multiclass SVMs with HMM-Guided</td>
<td>D. Martín-Iglesias, J. Bernal-Chaves, C. Peláez-Moreno, A. Gallardo-Antolín, F. Díaz-de-María</td>
<td>257</td>
</tr>
<tr>
<td>Segmentation</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Segment Boundaries in Low Latency Phonetic Recognition</td>
<td>Giampiero Salvi</td>
<td>267</td>
</tr>
<tr>
<td>Third-Order Moments of Filtered Speech Signals for Robust Speech</td>
<td>Kevin M. Indrebo, Richard J. Povinelli, Michael T. Johnson</td>
<td>277</td>
</tr>
<tr>
<td>Recognition</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Models for Speech Feature Extraction</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Noise Robust Automatic Speech Recognition with Adaptive Quantile</td>
<td>Casper Stork Bonde, Carina Graversen, Andreas Gregers Gregersen, Kim Hoang Ngo, Kim Nørmark, Mikkel Purup, Thomas Thorsen, Børge Lindberg</td>
<td>291</td>
</tr>
<tr>
<td>Based Noise Estimation and Speech Band Emphasizing Filter Bank</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Spotting Multilingual Consonant-Vowel Units of Speech Using Neural</td>
<td>Suryakanth V. Gangashetty, C. Chandra Sekhar, B. Yegnanarayana</td>
<td>303</td>
</tr>
<tr>
<td>Network Models</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Speech Enhancement</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Novel Sub-band Adaptive Systems Incorporating Wiener Filtering</td>
<td>Amir Hussain, Stefano Squartini, Francesco Piazza</td>
<td>318</td>
</tr>
<tr>
<td>for Binaural Speech Enhancement</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Tran Huy Dat, Kazuya Takeda, Fumitada Itakura ........................................ 328

Applications

Modeling Fluctuations of Voiced Excitation for Speech Generation Based on Recursive Volterra Systems

Karl Schnell, Arild Lacroix ............................................................................. 338

A Simple, Quasi-linear, Discrete Model of Vocal Fold Dynamics

Max Little, Patrick McSharry, Irene Moroz, Stephen Roberts .............. 348

Blind Channel Deconvolution of Real World Signals Using Source Separation Techniques

Jordi Solé-Casals, Enric Monte-Moreno .................................................... 357

Method for Real-Time Signal Processing Via Wavelet Transform

Pavel Rajmic ........................................................................................................ 368

Author Index .................................................................................................... 379