ISCA ITRW
SPEECH ANALYSIS AND PROCESSING
FOR KNOWLEDGE DISCOVERY

JUNE 4-6, 2008, AALBORG UNIVERSITY
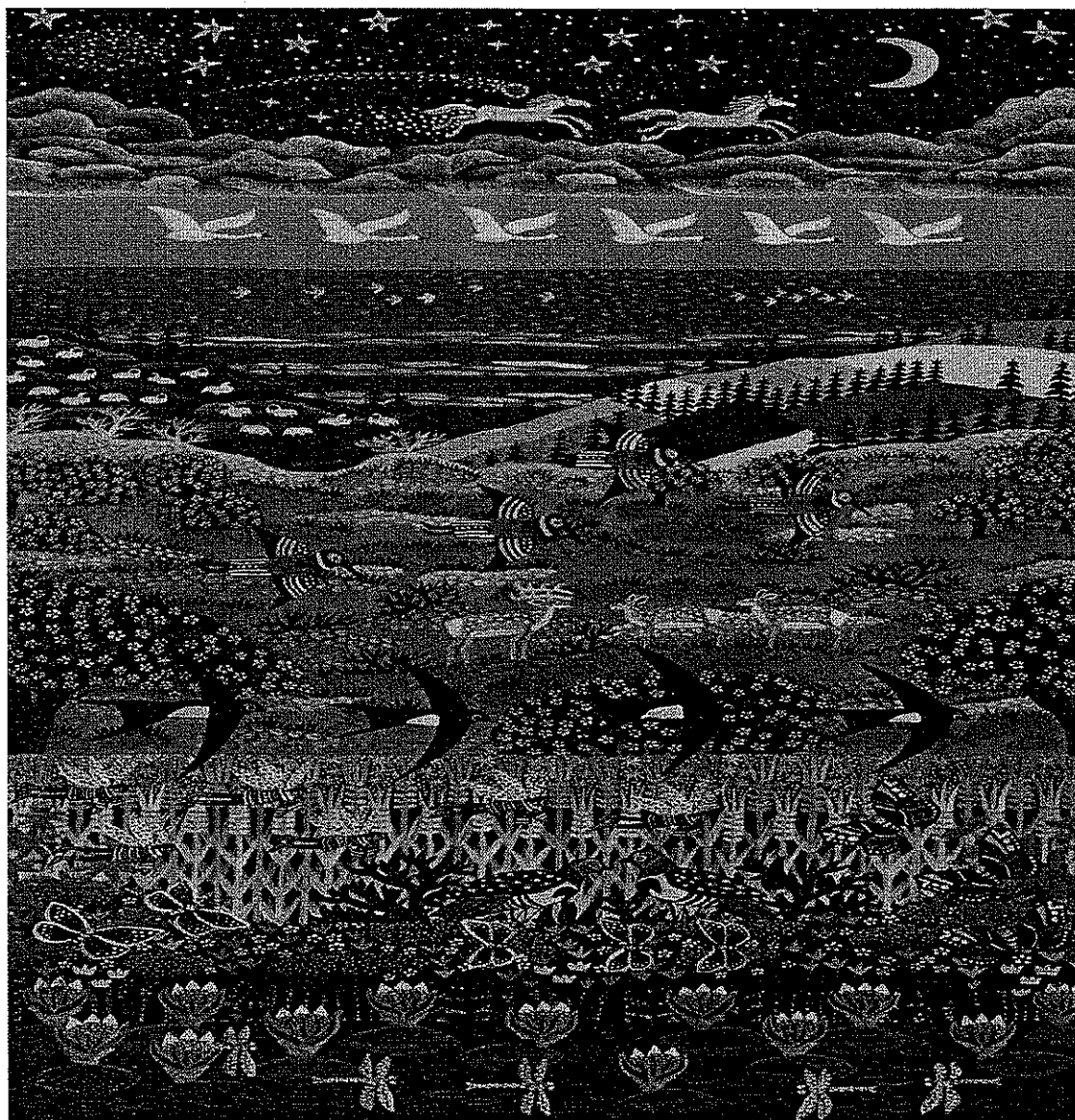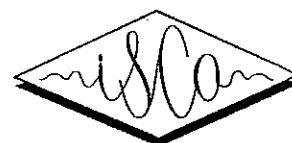## DENMARK

Proceedings

AALBORG UNIVERSITY

ISCA ITRW
SPEECH ANALYSIS AND PROCESSING
FOR KNOWLEDGE DISCOVERY

JUNE 4-6, 2008, AALBORG UNIVERSITY
DENMARK

Proceedings

Edited by
Paul Dalsgaard, Christian Fischer Pedersen and Ove Andersen
Section for Multimedia Information and Signal Processing
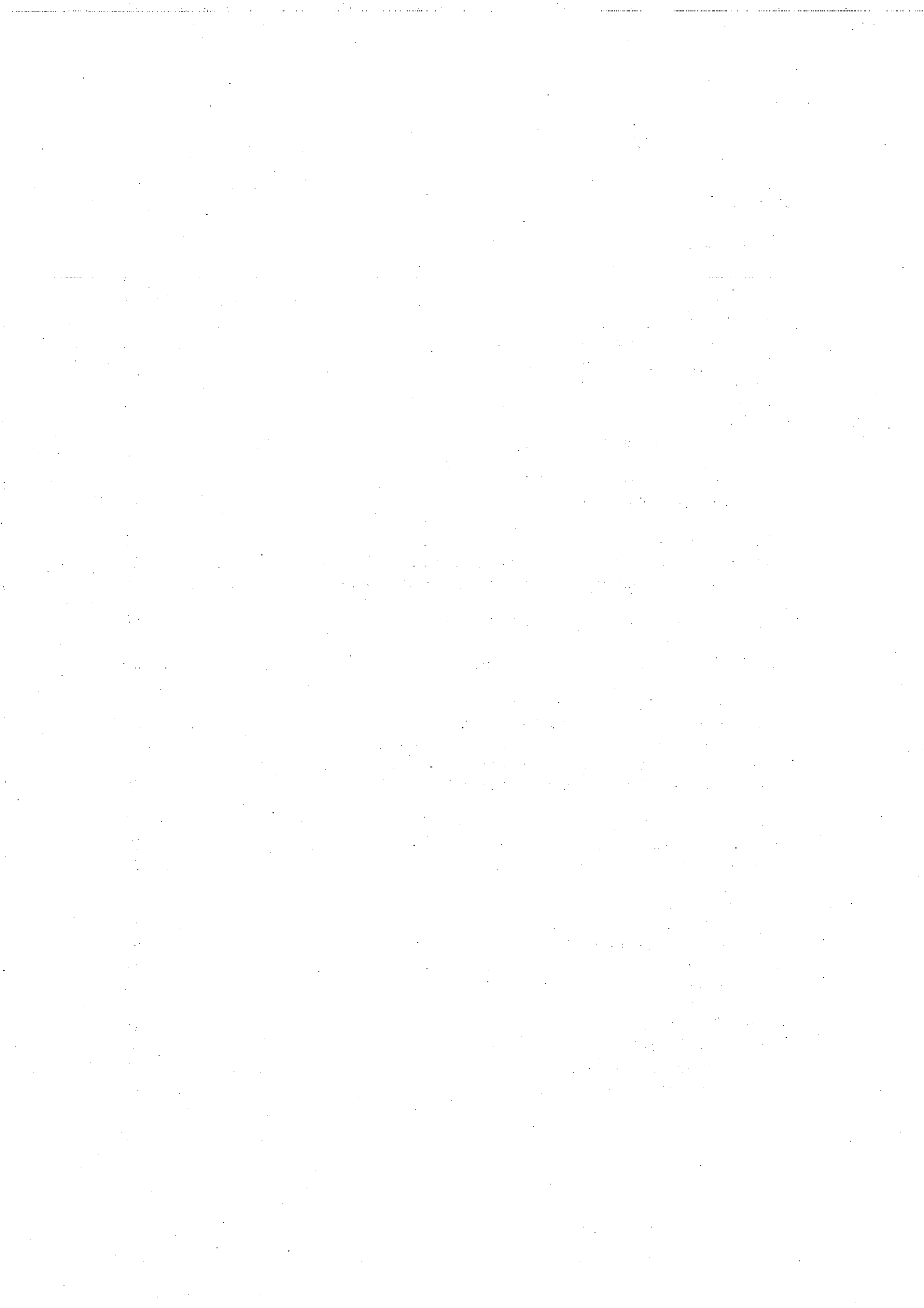Department of Electronic Systems
Aalborg University, Denmark

Contents

1 Welcome

It is a pleasure to welcome you to this ISCA Tutorial and Research Workshop on "Speech Analysis and Processing for Knowledge Discovery".
The first part of the title "Speech Analysis and Processing" is generally very well known to us who are doing research and development in speech technology.
However, the issue is very broad in the sense that it covers many different ways of conducting pre-analysis and post-processing. The aim of this workshop is to focus on methods that are expected to bring the state-of-the art in speech technology forward by doing re-search on existing analysis and processing methodologies and offer new methodologies that together will make it possible to estimate speech features that until this time has been ignored or not proven to be important for gaining an overall higher general performance of speech technology applications. Therefore, the second part of the title "Knowledge Discovery" as seen in the context of the first part is meant to narrow down the overall focus of the workshop to research that search for the unknown details by discovering where this information is hidden.

The focus of the workshop is thus to present and discuss new ways or combinations of traditional analysis and processing and newly developed methods, and to demonstrate how this can be exploited in, e.g. post-processing modules.

To underline this, the workshop is initiated by two invited lectures. The first by Professor Sarah Hawkins, Cambridge University, UK, who will give us basic insight in important acoustic-phonetic details, the second by Professor Christophe d'Alessandro, CNRS-LIMSI, France, who will focus on alternative ways of conducting speech analysis. Sarah Hawkin's background is in phonetics, Christophe d'Alessandro's is in mathematics and engineering.
We have no doubt that the two lectures will lay the ground for numerous discussions throughout the workshop.

The technical programme is divided into a series of six oral sessions and one two-slot long poster session. Each of the oral sessions addresses a relative narrow topic. By forming homogeneous sessions comprising papers of related topics, we hope to have created the right atmosphere for fruitful and lively discussions among all the workshop participants.
The technical programme is:

1. Estimating Speech Production Parameters

2. Attribute Detection and Knowledge Discovery

3. Features for Speaker Recognition

4. Acoustic Event Detection

5. Speech Analysis and Modelling for Production and Recognition

1

6. Speech Attributes and Knowledge Discovery

7. Speech Recognition and Classification

The final session is organised as a "Round Table Discussion" with the title "What have we learned and future initiatives". The discussion will be opened by comments from the two invited speakers which hopefully will lead to discussions on future research activities with the focus on new speech analysis and processing methods. Paul Dalsgaard will act as moderator.

To round off this welcome, the Organising Committee would like to thank the invited speakers for their willingness to contribute to this workshop and all authors for their invaluable contributions and presentations. We would also like to thank all members of the scientific committee for devoting their time and efforts in carefully reviewing all submission, as well as session chairs for preparing and running their session.

Welcome to Aalborg and to what we hope will be a scientifically fruitful ISCA event.

May 19th, 2008

Paul Dalsgaard, Ove Andersen and Christian Fischer Pedersen
Aalborg University
Department of Electronic Systems
Denmark

2 Committees

2.1 Organising committee

Paul Dalsgaard	Aalborg University	Denmark
Yegnanarayana Bayya	International Inst. of Information Tech.	India
Chin-Hui Lee	Georgia Institute of Technology	USA
Paavo Alku	Helsinki University of Technology	Finland
Rolf Carlson	The Royal Institute of Tech., Stockholm	Sweden
Torbjørn Svendsen	Norwegian Uni of Sci. and Tech., Trondheim	Norway

2.2 Scientific committee

Chin-Hui Lee	Georgia Institute of Technology	USA
Hynek Hermansky	IDIAP Research Institute	Switzerland
Isabel Trancoso	INESC,	Portugal
James Glass	MIT	USA
Kuldip Paliwal	Griffith University	Australia
Lin-Shan Lee	National Taiwan University	Taiwan
Lou Boves	Radboud University	The Netherlands
Louis C. W. Pols	Univeristy of Amsterdam	The Netherlands
Louis ten Bosch	Radboud University	The Netherlands
Magne H. Johnsen	NTNU, Trondheim	Norway
Mahadeva Prasanna	Indian Institue of Technology	India
Mark Hasegawa Johnson	University of Illinois	USA
Mats Blomberg	KTH, Stockholm	Sweden
Ove Andersen	Aalborg University	Denmark
Paavo Alku	HUT, Helsinki	Finland
Peter Murphy	University of Limerick	Ireland
Piero Cosi	Inst. of Cogn. Sciences and Technologies	Italy
Roger K. Moore	University of Sheffield	UK
Rolf Carlson	KTH, Stockholm	Sweden
Sadaoki Furui	Tokyo Institute of Technology	Japan
Steven Greenberg	Silicon Speech & Tech. Uni. of Denmark	USA & Denmark
Thiery Dutoit	Faculté Polytechnique de Mons	Belgium
Tom Bäckström	HUT, Helsinki	Finland
Torbjørn Svendsen	NTNU, Trondheim	Norway
Yannis Stylianou	University of Creete	Grece
Yegnanarayana Bayya	International Inst. of Information Tech.	India

2.3 Workshop secretariat

Charlotte Skindbjerg Pedersen, Aalborg University, Department of Electronic Systems, Section for Multimedia Information and Signal Processing, Denmark.

2.4 Web design and webmaster

Christian Fischer Pedersen, Aalborg University, Department of Electronic Systems, Section for Multimedia Information and Signal Processing, Denmark.

3 Schedule

Tuesday June 3	Wednesday June 4	Thursday June 5	Friday June 6
ASIP-NET seminar on Audio Content Analysis	Registration – 08.30 – 09.30		
The Audio Signal Processing Network in Denmark	Opening – 09.30 – 09.50	Oral session III - 09.00 - 10.30 Features for speaker recognition	Oral session V - 09.00 - 10.30 Speech attributes and knowledge discovery
11.00 – 11.05: Welcome 11.05 – 12.05: Music content analysis, Reader Dr. Mark Plumbley, Dept. of Elec. Engineering, Queen Mary University, London, United Kingdom	Tutorial lectures – 10.00 – 12.30: Sarah Hawkins & Christophe d'Alessandro	Coffee/tea break	Coffee/tea break
		Oral session IV - 11.00 - 12.30 Acoustic event detection	Oral session VI - 11.00 - 12.30 Speech recognition and classification
Lunch 12.05 - 13.00	Lunch 12.30 - 13.30	Lunch 12.30 - 13.30	Lunch 12.30 - 13.30
13.00-13.30: The ISP toolbox and a tempo-insensitive distance measure for cover song identification based on chroma features, Ph.D. student, Jesper Højvang Jensen, AAU 13.30-14.00: Discovering music strcture via similaritiy fusion, Ph.D., Anders Meng, Oticon A/S & DTU 14.00-14.30: Coffee break 14.30-15.00: Content-based search in broadcast news, Ph.D. student Lasse L. Mølgaard, DTU 15.00-16.00: TBA 16.00-16.15: Wrap up	Oral session I - 13.30 - 15.00 Estimating speech production parameters	POSTER session I - 13.30 - 15.00 Speech analysis and modelling for production and recognition	Round table discussion - 13.30 - 15.00 "What is learned and future initiatives?"
	Coffee/tea break	Coffee/tea break	Rounding off - 15.00-15.15 by Paul Dalsgaard
	Oral session II - 15.30 - 17.00 Attribute detection and knowledge discovery	POSTER session II- 15.30 - 17.00 Speech analysis and modelling for production and recognition	
	Welcome reception – 17.15 – 18.30	Ansgar church - 18.30 - 19.30 Organ concert by Christophe d'Alessandro Workshop dinner - 20.00 -	

4 Technical programme

4.1 Program - Wednesday June 4th

08.30-09.30	Registration
09.30-09.50	Workshop opening by Paul Dalsgaard Link to opening speech.
10.00-11.00	Invited lecture I Professor Sarah Hawkins Cambridge University, UK Chair: Rolf Carlson Abstract: Section 5.1 on page 15 Link to slides (ppt). Link to music file for slide no. 37.
11.00-11.30	Coffee / tea break
11.30-12.30	Invited lecture II Professor Christophe d'Alessandro LIMSI-CNRS, France Chair: Paavo Alku Abstract: Section 5.2 on page 16 Link to slides (pdf). Without sound clips. Link to slides (pps). With sound clips.
12.30-13.30	Lunch
13.30-15.00	Oral session I Estimating speech production parameters Chair: Torbjörn Svendsen
	A unified approach for F0 extraction and aperiodicity estimation based on a temporally stable power spectral representation Hideki Kawahara[1], Masanori Morise[1], Toru Takahashi[2], Ryuichi Nisimura[1], Hideki Banno[3], Toshio Irino[1] [1] Faculty of Systems Engineering, Wakayama University, Wakayama, Japan - [2] Graduate School of Informatics, Kyoto University, Kyoto, Japan - [3] Faculty of Science and Technology, Meijo University, Nagoya, Japan See abstract 6.1.1 on page 19 Link to slides.Link to movie.
	On the Estimation of the Speech Harmonic Model Yannis Pantazis[1], Olivier Rosec[2], Yannis Stylianou[1] [1]Institute of Computer Science, CSD, UoC, Greece - [2] Orange Labs, Lannion, France See abstract 6.1.2 on page 19 Link to slides.
	Analysis of Stop Consonants in Indian Languages Using Excitation Source Information in Speech Signal Yegnanarayana B[1], Sri Rama Murty K[2], Rajendran S[1] [1]International Institute of Information Technology - [2]Indian Institute of Technology Madras See abstract 6.1.3 on page 20 Link to slides.
15.00-15.30	Coffee / tea break

15.30-17.00	Oral session II Attribute detection and knowledge discovery Chair: Jim Baker
	Integration of Asynchronous Knowledge Sources in a Novel Speech Recognition Framework Hugo Van hamme[1] [1]K.U.Leuven, dept. ESAT See abstract 6.2.1 on page 20 Link to slides
	Incorporating Suprasegmental Knowledge For Phone Recognition With Conditional Random Fields Prateeti Mohapatra[1], Eric Fosler-Lussier[1] [1]The Ohio State University See abstract 6.2.2 on page 21 Link to slides.
	An Experimental Study on Continuous Phone Recognition with Little or No Language-Specific Training Data Dau-Cheng Lyu[1], Sabato Marco Siniscalchi[2], Chin-Hui Lee[3] [1]Department of Electrical Engineering, Chang Gung University - [2]Department of Electronics and Telecommunications, NTNU, Norway - [3]School of ECE, Georgia Institute of Technology See abstract 6.2.3 on page 21 Link to slides.
17.15-18.30	Welcome reception

4.2 Program - Thursday June 5th

09.00-10.30	**Oral session III** **Features for speaker recognition** Chair: Yegnanarayana Bayya
	On the Relative Importance of the Short-Time Magnitude and Phase Spectra Towards Speaker Dependent Information Kamil Wojcicki[1], Kuldip Paliwal[1] [1]Signal Processing Laboratory, Griffith University, Nathan QLD 4111, Australia See abstract 7.1.1 on page 23 Link to slides. Link to sound clips.
	Spectral Slope Measurements in Emotionally Expressive Speech Lucas Tamarit[1], Martijn Goudbeek[2], Klaus Scherer[1,2] [1]Swiss Center for Affective Sciences, Geneva, Switzerland - [2]Geneva Emotion Research Group, University of Geneva, Geneva, Switzerland See abstract 7.1.2 on page 23 Link to slides
	Automated Speaker Recognition Using Compressed Temporal-Spectral Dynamics Information of Password Spectrograms Amitava Das[1], Gokul Chittaranjan[1] [1]MSR-India See abstract 7.1.3 on page 24 Link to slides.
10.30-11.00	Coffee / tea break
11.00-12.30	**Oral session IV** **Acoustic event detection** Chair: Ove Andersen
	Innovative acoustic probes to test predictions of wider utterance context Dave Davies[1], Bruce Millar[2] [1]University of Canberra - [2]Australian National University See abstract 7.2.1 on page 24 Link to slides.
	Time-Varying Cepstral Coefficients Trond Skogstad[1], Torbjørn Svendsen[1] [1]NTNU See abstract 7.2.2 on page 24 Link to slides.
	Effective Segmentation based on Vocal Effort Change Point Detection Chi Zhang[1], John Hansen[1] [1] Erik Jonsson School of Engineering & Computer Science University of Texas at Dallas, Richardson, Texas 75083, USA See abstract 7.2.3 on page 25 Link to slides.Link to sound clips.
12.30-13.30	Lunch
13.30-15.00	**Poster session I** **Speech analysis and modelling for production and recognition** Chair: Chin-Hui Lee
	Fitting Mass-Spring Models to Glottal Flow Estimates Tom Bäckström[1] [1]TKK (Helsinki University of Technology) See abstract 7.3.1 on page 25 Link to poster.

	Investigating Explicit Model Transformations for Speaker Normalization Mats Blomberg[1], Daniel Elenius[1] [1]Dept Speech, Music and Hearing, CSC/KTH, Stockholm, Sweden See abstract 7.3.2 on page 26 Link to poster
	Noise robust digit recognition using sparse representations Jort Gemmeke[1], Bert Cranen[1] [1]Centre for Language and Speech Technology (CLST), Radboud University, P.O. Box 9103, NL-6500 HD Nijmegen, The Netherlands See abstract 7.3.3 on page 26 Link to poster.
	The Hartley Phase Spectrum as a noise-robust feature in speech analysis Ioannis Paraskevas[1], Maria Rangoussi[1] [1]Department of Electronics, Technological Education Institute of Piraeus, Athens, Greece See abstract 7.3.4 on page 27 Link to poster.
	Complex Wavelet Modulation Sub-Bands and Speech Jean-Marc Luneau[1], Jérôme Lebrun[2], Søren Holdt Jensen[1] [1]ES-MISP, Aalborg University - [2]I3S-CNRS, Sophia Antipolis, France. See abstract 7.3.5 on page 27 Link to poster
15.00-15.30	**Coffee / tea break**
15.30-17.00	**Poster session II** **Speech analysis and modelling for production and recognition** Chair: Chin-Hui Lee
	Comparing Human and Machine Recognition Performance on a VCV Corpus Odette Scharenborg[1], Martin Cooke[2] [1]Centre for Language and Speech Technology, Radboud University Nijmegen, The Netherlands - [2]Speech and Hearing Research Group, Department of Computer Science, University of Sheffield, UK See abstract 7.4.1 on page 28 Link to poster.
	Acoustic profiles in emotion - the GEMEP corpus Martijn Goudbeek[1], Klaus. R Scherer[2] [1]University of Geneva - [2]Swiss Center for Affective Sciences See abstract 7.4.2 on page 28 Link to poster.
	Speech Analysis by Time-Varying Lattice Filters Karl Schnell[1] [1]Institute of Applied Physics, Goethe-University Frankfurt, Germany See abstract 7.4.3 on page 29 Link to poster.
	Using Zeros of the z-transform in the Analysis of Speech Signals Paul Dalsgaard[1], Christian F. Pedersen[1], Ove Andersen[1], Yegnanarayana Bayya[2] [1]Aalborg University - [2]IIIT, Hyderabad See abstract 7.4.4 on page 29 Link to poster.
17.00-18.30	**Break**

18.30-19.30	Organ concert by Christophe d'Alessandro Ansgar Church
20.00-	Workshop dinner Hotel Hvide Hus Restaurant upper floor Right in front of Church Ansgar

4.3 Program - Friday June 6th

09.00-10.30	Oral session V Speech attributes and knowledge discovery Chair: Rolf Carlson
	An acoustic investigation of the [ATR] feature effect on vowel-to-vowel coarticulation Christina Orphanidou[1], Greg Kochanski[1], John Coleman[1] [1]Oxford University Phonetics Laboratory See abstract 8.1.1 on page 31 Link to slides.
	Joint Optimization of Event Detectors and Evidence Merger for Continuous Phone Recognition Sabato Marco Siniscalchi[1], Øystein Birkenes[1], Magne H. Johnsen[1], Torbjørn Svendsen[1] [1]NTNU See abstract 8.1.2 on page 31 Link to slides.
	Unsupervised detection of words - questioning the relevance of segmentation Louis ten Bosch[1], Hugo Van hamme[2], Lou Boves[1] [1]Radboud University Nijmegen - [2]ESAT, KUL, Leuven See abstract 8.1.3 on page 32 Link to slides.
10.30-11.00	Coffee / tea break
11.00-12.30	Oral session VI Speech recognition and classification Chair: Paavo Alku
	Enhancing Noise Robustness in Automatic Speech Recognition Using Stabilized Weighted Linear Prediction (SWLP) Jouni Pohjalainen[1], Carlo Magi[1], Paavo Alku[1] [1]Helsinki University of Technology See abstract 8.2.1 on page 32 Link to slides.
	Feature selection algorithms for the creation of multistream speech recognizers Yotaro Kubo[1], Shigeki Okawa[2], Akira Kurematsu[1], Katsuhiko Shirai[1] [1]Waseda University, Tokyo, Japan. - [2]Chiba Institute of Technology, Narashino, Japan. See abstract 8.2.2 on page 33 Link to slides.
	Discrimination of Speech from Nonspeeech in Broadcast News Based on Modulation Frequency Features Maria Markaki[1], Yannis Stylianou[1,2] [1] Computer Science Department, University of Crete, Greece. [2] Institute of Computer Science, FORTH, Greece See abstract 8.2.3 on page 33 Link to slides.
12.30-13.30	Lunch
13.30-15.00	Round table discussion What have we learned and future initiatives? Sarah Hawkins, Christophe d'Alessandro, Chin-Hui Lee, Hideki Kawahara, Yegnanarayana Bayya, Paavo Alku, Torbjörn Svendsen, Ove Andersen Moderator: Paul Dalsgaard

15.00-15.15	Rounding off Paul Dalsgaard