



A workshop on

**Consistent and reliable acoustic cues
for sound analysis**

September 2nd 2001
Aalborg, Denmark

Listeners are capable of sound processing tasks such as robust speech recognition in acoustic environments which defeat their automated counterparts. Part of this performance disparity may be due to different developmental conditions. Automated systems have typically attempted to ameliorate the effect of noise on representations and algorithms devised for clean conditions. Listeners, growing up outside the anechoic chamber, are immediately faced with the reality of perception in a world where competing sound sources are not simply there to be suppressed.

This workshop¹ focuses on approaches to robust sound processing which recognise the existence of such noisy realities. One theme of techniques developed in recent years has been to replace classification based on complete representations with a search for reliable acoustic cues followed by an identification process which accommodates sparse representations. The initial stage can be accomplished by techniques ranging from standard noise estimation algorithms to full-blown computational auditory scene analysis systems using pitch and location models. The subsequent identification process has been tackled using missing data and multistream processing.

Although the dominant application area is robust speech recognition, the workshop brings together research on music and non-speech audio, in addition to work on representations inspired by the auditory system. This diversity flows from the growing belief that progress in a wide range of application areas relies on confronting and solving the problem of general-purpose sound understanding faced by listeners.

CRAC Organizing Committee

Dan Ellis, Columbia University (co-chair)
Martin Cooke, Sheffield University (co-chair)

Frédéric Berthommier, Institut de la Communication Parlée, Grenoble
Andrzej Drygajlo, École Polytechnique Fédérale de Lausanne
Phil Green, Sheffield University
Andrew Morris, Institut Dalle Molle d'Intelligence Artificielle Perceptive
Hiroshi Gitchang Okuno, Kyoto University

1. The workshop grew out of the EU LTR Project RESPITE (28149) - Recognition of Speech by Partial Information Techniques.

Contents

PERCEPTION & AUDITORY MODELS

Enhancing Sound Sources by use of Binaural Spatial Cues

Johannes Nix & Volker Hohmann

Generalized Correlation Network model of auditory processing

Alain de Cheveigné

Sound resynthesis from Auditory Mellin Image using STRAIGHT

T. Irino, R. D. Patterson & H. Kawahara

On the various influences of envelope information on the perception of speech in adverse conditions: An analysis of between-channel envelope correlation

Olivier Crouzet & W.A. Ainsworth

Acoustic cues of voiced and voiceless plosives for determining place of articulation

Philip J.B. Jackson

Auditory Interpretation and Application of Warped Linear Prediction

Matti Karjalainen

Robust Phonetic Feature Extraction Under a Wide Range of Noise Backgrounds and Signal-to-Noise Ratios

Shuangyu Chang, Lokendra Shastri & Steven Greenberg

MUSIC & GENERAL AUDIO ANALYSIS

Automatic transcription of musical recording

Anssi Klapuri, Tuomas Virtanen, Antti Eronen & Jarno Seppänen

Reduced-Rank Spectra and Entropic Priors as Consistent and Reliable Cues for Generalized Sound Recognition

Michael A. Casey

Sound Classification in Hearing Instruments by means of Auditory Scene Analysis

Silvia Allegro, Michael Büchler & Stefan Launer

Equivalence between Frequency Domain Blind Source Separation and Frequency Domain Adaptive Beamformers

Shoko Araki, Shoji Makino, Ryo Mukai & Hiroshi Saruwatari

Fast Music Retrieval using Spectrum and Power Information

Tomoya Narita & Masahide Sugiyama

Optimization of Voice/Music Detection in Sound Data

Shin'ichi Takeuchi, Masaki Yamashita, Takayuki Uchida & Masahide Sugiyama

A Predominant-F0 Estimation Method for Real-world Musical Audio Signals: MAP Estimation for Incorporating Prior Knowledge about F0s and Tone Models

Masataka Goto

Detecting alarm sounds

Daniel P.W. Ellis

MISSING-DATA SPEECH RECOGNITION

Integrating bottom-up and top-down constraints to achieve robust ASR: The multisource decoder

Jon Barker, Martin Cooke & Dan Ellis

Detection of Reliable Features for Speech Recognition in Noisy Conditions Using a Statistical Criterion

Phillippe Renevey & Andrzej Drygajlo

A Binaural Model for Missing Data Speech Recognition in Noisy and Reverberant Conditions

Kalle J. Palomäki, Guy J. Brown & DeLiang Wang

On the Use of Missing Feature Theory with Cepstral Features

Juha Häkkinen & Hemmo Haverinen

Data Utility Modelling for Mismatch Reduction

Andrew C. Morris

Robust multi-stream speech recognition based on the combined reliabilities of the speech signal (voicing cue) and phonemes estimates using a bias prediction

Hervé Glotin

Multiband with contaminated training data

Stéphane Dupont & Christophe Ris

Robust Speech Recognition using Missing Features: the Case for Restoring Missing Input Features

Bhiksha Raj, Michael L. Seltzer & Richard M. Stern

APPROACHES TO HANDLING NOISY SPEECH

Speech enhancement and segregation based on the localisation cue for cocktail-party processing

Emmanuel Tessier & Frédéric Berthommier

Speech estimation biased by phonemic expectation in the presence of non-stationary and unpredictable noise

Ikuyo Masuda-Katsuse & Yoshimori Sugano

Analysis of Disturbed Acoustic Features in terms of Emission Cost

Laurens van de Werff, Johan de Veth, Bert Cranen & Louis Boves

A fundamental frequency estimation method for noisy speech based on instantaneous amplitude and frequency

Yuichi Ishimoto, Masashi Unoki & Masato Akagi

Evaluation of Robust Feature Extraction and Acoustic Modelling algorithms/systems by interfacing ASR systems

Joan Marí, José Manuel Ferrer Ruiz & Fritz Class

Effects of increasing modalities in understanding three simultaneous speeches with two microphones

Hiroshi G. Okuno, Kazuhiro Nakadai & Hiroaki Kitano

A recognition method using synthesis-based scoring that incorporates direct relations between static and dynamic feature vector time series

Yasuhiro Minami, Erik McDermott, Atsushi Nakamura & Shigeru Katagiri

Schedule

08:00	Registration
09:00	Session 1: Perception & Auditory Models (Phil Green)
10:30	Coffee break
11:00	Session 2: Music & General Audio Analysis (Dan Ellis)
12:30	Lunch
14:00	Session 3: Missing-Data Speech Recognition (Martin Cooke)
15:30	Coffee break
16:00	Session 4: Approaches to Handling Noisy Speech (Hiroshi G. Okuno)
17:30	Close

Per session:

Overview (session chair)	10 minutes
Lecture 1	15 minutes
Lecture 2	15 minutes
Posters	45 minutes