NEWSLETTER

Centre for Acoustic Signal Processing Research (CASPR)

January 2021

The Centre for Acoustic Signal Processing Research (CASPR) is a research centre at the Section for Artificial Intelligence & Sound, Department of Electronic Systems, Aalborg University, Denmark. CASPR is supported by the Demant Foundation, Oticon A/S, and Aalborg University.

Contents

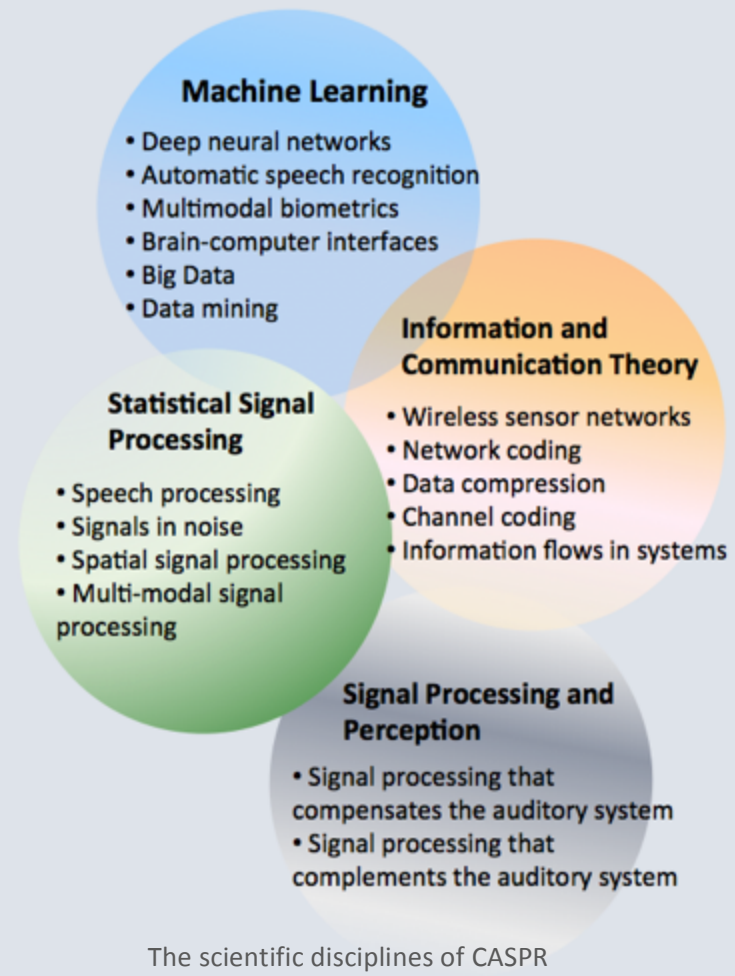
Department of Electronic Systems
Section on Artificial Intelligence & Sound
Aalborg University

Research in CASPR

CASPR is conducting research related to advanced statistical signal processing solutions for assistive listening. The research finds direct use in communication devices such as hearing aids, helmets, headphones, cochlear implants, ear monitor, ear pieces, diagnostic equipment, etc. However, the envisioned research also finds use in related areas such as robust speech/speaker recognition, brain computer interfaces, acoustic event detection, etc.

The scientific scope of CASPR encompasses

- statistical signal processing.
- machine learning.
- information and communication theory with applications to wireless exchange of information between listening devices and other external devices.
- pattern recognition.
- data mining in body worn sensor data.
- perception-based statistical signal processing.



The scientific disciplines of CASPR

CASPR will navigate in a rapidly changing technological landscape: we envision a near future, where the technological landscape allows very different, and better, hearing assistive devices than are known today. Specifically, we envision that near-future hearing assistive devices will:

- increase the wireless exchange of information with each other, with other body-worn devices and with devices outside the body.
- make use of additional microphones on or outside the body, and will employ other types of body-worn or outside-the-body sensors.
- work in a much closer symbiosis with the user.

Teaching in CASPR

Current courses related to CASPR

CASPR is heavily involved in teaching and education at B.Sc., M.Sc., PhD., and Postdoc level in disciplines that are relevant to the scientific scope of CASPR:

- Optimization Methods (Master Course), Autumn 2020.
- Information Theory (Master course), Autumn 2020.
- Machine Learning (Master course), Autumn 2020.
- Platforms and Methods for Multi-Modal System Architectures (Master course), Autumn 2020.
- Research in Vision, Graphics and Interactive Systems (Master course), Autumn 2020.

CASPR is currently involved in 7 student projects:

Project 1
Self-Supervised Deep Learning of Representations with Application to Speech Classification.
Signal Processing and Acoustics, long Master thesis project with RTX A/S.
Bjørn Uttrup Dideriksen and Kristoffer Calundan Derosche

Project 2
Predictability-Based Objective Evaluation of Sound.
Mathematical Engineering, long thesis project with Oticon A/S.
Thor Pilgaard Knudsen

Project 3
Acoustic Transfer Function Estimation for Adaptive Beamformers in Hearing Aids Using Deep Learning".
Mathematical Engineering, 7th semester thesis project with Oticon A/S.
Dennis Grøndahl, Andersen Kristian Søgaard, Mikkel Fjord Olsen, Simone Birk Bols Thomsen.

Project 4
Deep Neural Networks on Constrained Devices".
Mathematical Engineering, 7th semester thesis project with RTX A/S.
Andreas Anton Andersen, Martin Voigt Vejling, Morten Stig Kaaber, Daniel Bernard van Diepen.

Project 5
Audio Codec Noise Modelling.
Mathematical Engineering, 7th semester thesis project with RTX A/S.
Lars Vedel Friis, Lukas Menholt og Nicolai André Weinreich.

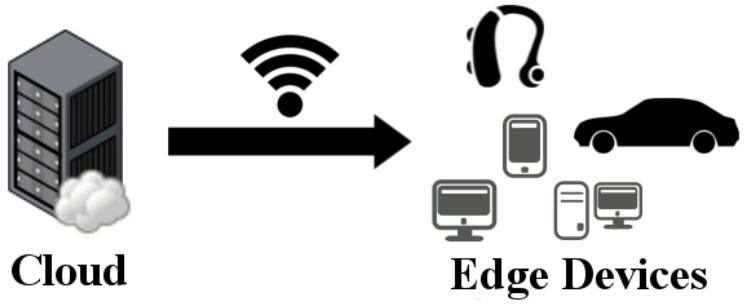
Project 6
Analysis of Scale-Invariance in EEG Microstates due to Acoustic Stimuli.
Mathematical Engineering, long thesis project.
Rasmus Lykke Vestergaard.

Project 7
Dual-mic Voice Activity Detection.
Signal Processing and Acoustics, 7th semester project with RTX A/S.
Nicolai Almskou Rasmussen, Max Væhrens, Victor Mølbach Nissen, Joachim Roland Hejslet

PhD Project: Training Methods for DNNs Under Computational Resource Constraints
Start date: September 1st, 2019.

Morten Ø. Nielsen

Deep learning has shown superior performance over the last couple of decades in various fields and applications such as speech, image, and object recognition and enhancement. As more advanced and powerful computers have become available, the size and computational complexity of state-of-the-art deep neural networks (DNNs) have increased accordingly. Large-scale DNNs are prohibited from operating on edge devices, e.g. microcontrollers, smartphones, and hearing assistive devices, due to the requirements of limited power consumptions and computational resources. Instead, the DNNs could be stored and operating on cloud servers, which can be accessed by the edge devices. This introduces issues in terms of data privacy and latency. For example, real-time sound processing, e.g., for hearing assistance, requires minimal latency in order to present a sound signal for the user on time. However, accessing DNNs on a cloud-server could introduce long communications delays.
A branch of deep learning is focusing on different methods to compress DNNs with minor reduction in the overall performance compared to the original model.



In this project, the latest trends in deep learning algorithms and compression techniques will be studied. The aim is to develop a training procedure for learning sparse parameters of a DNN under computational resource constraints, e.g. memory consumption and floating-point operations. In order to avoid a reduction in the overall performance of the DNNs, recent information theoretical results in deep learning will be employed. Our current findings have shown that exploiting a variational information-bottleneck approach for training the DNNs, can lead to significant reductions in the size of the DNNs at the price of only insignificant performance degradations.

NEWS

A new demo on *audio-visual speech enhancement*
based on the work of PhD researcher Daniel Michelsanti has been released.

All demos are available on the CASPR website:
http://caspr.es.aau.dk/demos

Daniel Michelsanti successfully defended his PhD thesis entitled "Audio-Visual Speech Enhancement Based on Deep Learning" on Friday 18th, December, 2020.

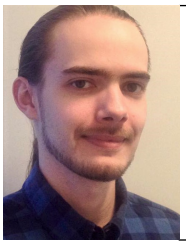
The defense was fully online and the opponents were:
John H. L. Hansen, University of Texas
Xavier Alameda-Pineda, University of Grenoble
Flemming Christensen, Aalborg University

Jan Østergaard has been appointed Full Professor in Information Theory and Signal Processing at the Department of Electronic Systems, AAU.
He gave his inaugural lecture entitled: "Information in Biological and Physical Systems" online on the 13th of January, 2021.

JAN ØSTERGAARD: NEW FULL PROFESSOR IN INFORMATION THEORY AND SIGNAL PROCESSING

A new industrial PhD project was initiated with Eriksholm Research Center, Oticon A/S. The project is supported by Innovation Fund Denmark and Eriksholm Research Center, Oticon A/S.

- Title***: Hearing loss compensation using computational models of hearing and deep learning***
- Company: Eriksholm Research Center, Oticon.
- PhD student: Peter Asbjørn Leer Bysted



Mathias Pedersen and Daniel Michelsanti presented their work at INTERSPEECH 2020

- Vocoder-Based Speech Synthesis from Silent Videos. D. Michelsanti, O. Slizovskaia, G. Haro, E. Gómez, Z.-H. Tan and J. Jensen. *Proc. Interspeech,* 2020.
- A Neural Network for Monaural Intrusive Speech Intelligibility Prediction. M. B. Pedersen, A. H. Andersen, S. H. Jensen and J. Jensen, *Proc. Interspeech,* 2020.

Jesper Jensen was invited to participate in Danish national radio program called "Kortsluttet" (Hardwired) to talk about digitally enhanced senses in the context of current and future hearing aid systems. The program (in Danish) can be found here: https://www.dr.dk/radio/p1/kortsluttet/kortsluttet-42

Recent CASPR Related Research

Journal papers

1. Minimum Processing Beamforming. A. Zahedi, M.S. Pedersen, J. Østergaard, T.U. Christiansen, L. Bramsløw, J. Jensen. Accepted for publication in IEEE Trans. Audio, Speech, Language Process. 2021.

2. Speech Intelligibility Prediction Using Spectro-Temporal Modulation Analysis. A. Edraki, W.-Y. Chan, J. Jensen, and D. Fogerty. IEEE Trans. Audio, Speech, Language Process., Vol. 29, pp. 210-225, 2021.

3. Online Multichannel Speech Enhancement Based on Recursive EM and DNN-Based Speech Presence Estimation. J. M. Martin-Donas, J. Jensen, Z.-H. Tan, A. M. Gomez, A. M. Peinado. IEEE Trans. Audio, Speech, Language Process., Vol. 28, pp. 3080-3094, 2020.

4. Spatially Correct Rate-Constrained Noise Reduction For Binaural Hearing Aids in Wireless Acoustic Sensor Networks. J. Amini, R. C. Hendriks, R. Heusdens, M. Guo, and J. Jensen. IEEE Trans. Audio, Speech, Language Process, Vol. 28, pp. 2731-2743, 2020.

5. Bounds on the Sum-Rate of MIMO Causal Source Coding Systems with Memory under Spatio-Temporal Distortion Constraints. P.A. Stavroe, J. Østergaard, M. Skoglund. Entropy 2020, 22(8), 842.

6. Deep InterBoost Networks for Small-sample Image Classification. X. Li, D. Chang, Z. Ma, Z.-H. Tan, J.-H. Xue, J. Cao and J. Guo. Accepted by Neurocomputing, 2020.

7. OSLNet: Deep Small-Sample Classification with an Orthogonal Softmax Layer. X. Li, D. Chang, Z. Ma, Z.-H. Tan, J.-H. Xue, J. Cao, J. Yu and J. Guo. IEEE Transactions on Image Processing, vol. 29, pp. 6482-6495, May 2020.

8. On the Comparisons of Decorrelation Approaches for non-Gaussian Neutral Vector Variables. Z. Ma, X. Lu, J. Xie, Z. Yang, J.-H. Xue, Z.-H. Tan, B. Xiao and J. Guo. Accepted by IEEE Transactions on Neural Networks and Learning Systems, 2020.

Conference Papers

1. Low Delay Robust Audio Coding by Noise Shaping, Fractional Sampling, and Source Prediction. J. Østergaard. Accepted for IEEE Data Compression Conference, March 2021.

2. UIAI System for Short-Duration Speaker Verification Challenge 2020. M. Sahidullah, A.K. Sarkar, V. Vestman, X. Liu, R. Serizel, T. Kinnunen, Z.-H. Tan, E. Vincent, Proc. of the 8th IEEE Spoken Language Technology Workshop (SLT 2021).

3. CC-loss: Channel Correlation Loss for Image Classificaiton. Z. Song, D. Chang, Z. Ma, X. Li and Z.-H. Tan, Proc. of the 25th International Conference on Pattern Recognition (ICPR 2020).

4. End-to-end Speech Intelligibility Prediction Using Time-Domain Fully Convolutional Neural Networks. M. B. Pedersen , M. Kolbæk , A. H. Andersen , S. H. Jensen , J. Jensen. Proc. Interspeech 2020.

5. Shouted Speech Compensation for Speaker Verification Robust to Vocal Effort Conditions. S. Prieto, A. Ortega, I. López-Espejo, E. Lleida, Proc. Interspeech 2020.

6. Vocoder-Based Speech Synthesis from Silent Videos. D. Michelsanti, O. Slizovskaia, G. Haro, E. Gómez, Z.-H. Tan, J. Jensen, Proc. Interspeech 2020.

7. Exploring Filterbank Learning for Keyword Spotting. I. López-Espejo, Z.-H. Tan, J. Jensen, Proc. of European Signal Processing Conference (EUSIPCO) 2020.

8. Estimation of Directed Dependencies in Time Series Using Conditional Mutual Information and Non-linear Prediction. P.S. Baboukani, C. Graversen, and J. Østergaard, Proc. Eusipco 2020.

CASPR Summer School – May 2021

Building on the success of our previous CASPR Winter School in 2017, we are happy to announce that there will be a CASPR Summer School in May 2021. The Summer School is also open for Industry.

Hearing assistive devices (HADs) are ubiquitous. They include, for example, devices such as headsets for speech communication in noisy environments (air plane crews, emergency/rescue teams, combat soldiers, police forces, etc.), headsets for office use, gaming, etc., and hearing care systems, e.g. hearing aids and cochlear implants.

The Summer school consists of lectures and hands-on exercises, which allow the participants to understand in-depth the technical problems related to HADs and their potential solutions. The school has three main parts. The first part is an introductory part, which lays the foundation for the rest of the course, covering fundamental topics such as auditory perception (normal and impaired hearing) and a discussion of the basic principles of HADs. The second part provides an overview of fundamental signal processing problems encountered in HADs, and an in-depth treatment of state-of-the art solutions. These include methods for beamforming and noise reduction, direction-of-arrival estimation, feedback control, hearing loss compensation, etc. Furthermore, an overview is given of methodologies for evaluating HADs with a particular focus on methods for intelligibility assessment and estimation.
The third part of the Summer school presents emerging technologies for hearing assistive devices, including machine learning techniques for processing of speech in noise, audio-visual signal processing, methods for own voice processing, etc.
While the course focuses on the HAD application, many of the discussed techniques are very general and find use in the much broader field of general sound processing. The course is multi-disciplinary, including topics such as basic auditory perception, statistical signal processing, deep learning, practical do's and don'ts.

The course also bridges the gap between theoretical background and practical/robust application. The course is a one-week concentrated course to be held in the Summer, 2021. The course involves course preparation (approximately 1 ECTS), course presence (1 ECTS), assignment finalization and hand-in (1 ECTS).

Date of Summer School: 17. – 21. May, 2021.
Place: Aalborg University, Fredrik Bajers Vej 7b, 9220 Aalborg, Denmark.
The Summer School will be held physically in Aalborg, and it will not be possible to participate online. Due to the Corona situation, we restrict the number of participants to at most 30. Should the Corona situation prevent us from meeting physically, an online version of the Summer School will be established.

For more information about the Summer School, feel free to contact the organizers:
Prof. Jesper Jensen (jje@es.aau.dk), Prof. Zheng-Hua Tan (zt@es.aau.dk), and Prof. Jan Østergaard (jo@es.aau.dk).

To register for the Summer School please use the following link, which takes you to the page of courses:
https://phd.moodle.aau.dk/course/index.php?categoryid=222

Course Fee: Free for PhD students and postdocs and 7500 DKK (1000 EUR) for industrial participants.