Department of Architecture, Design
and Media Technology – Audio Analysis Lab
Rendsurggade 14
9000 Aalborg

Invitation to PhD Defense

Alfredo Esquivel Jaramillo will defend his PhD Dissertation

Thursday, November 18, 2021, 13.00, presented via Zoom

Title:

"Pre-processing of Speech Signals for Robust Parameter Estimation"

Sign Up & Questions:

The defense will be presented via Zoom. To sign up or to pose questions regarding the PhD defense, please contact secretary Kristina Wagner Rojen.

Phone: +4599409926
Email: kwro@create.aau.dk

Assessment Committee:

Professor Maria Sandsten
Department of Mathematical Statistics, Lund University, Sweden

Associate Professor Christian Fischer Pedersen
Department of Electrical and Computer Engineering, Aarhus University, Denmark

Associate Professor Markus Löchtefeld
Department of Architecture, Design & Media Technology, Aalborg University, Denmark

Supervisors:

Professor Mads Græsbøll Christensen
Department of Architecture, Design & Media Technology, Aalborg University, Denmark

Associate Professor Jesper Kjær Nielsen
Siemens Gamesa, Denmark

Agenda:

13:00-13:05	Moderator Markus Löchtefeld welcomes the guests
13:05-13:50	Presentation by PhD student Alfredo Esquivel
13:50-14:05	Break

AALBORG UNIVERSITY

Department of Architecture, Design
and Media Technology – Audio Analysis Lab
Rendsurggade 14
9000 Aalborg

14:05-16:00	Questions
16:00-16:30	Assessment and announcement from committee

Abstract:

The topic of this thesis is methods of pre-processing speech signals for robust estimation of model parameters in models of these signals. Here, there is a special focus on the situation where the desired signal is contaminated by colored noise. In order to estimate the speech signal, or its voiced and unvoiced components, from a noisy observation, it is important to have robust estimators that can handle colored and non-stationary noise.

Two important aspects are investigated. The first one is a robust estimation of the speech signal parameters, such as the fundamental frequency, which is required in many contexts. For this purpose, fast estimation methods based on a simple white Gaussian noise (WGN) assumption are often used. To keep using those methods, the noisy signal can be pre-processed using a filter. If the colored noise is modelled as an autoregressive (AR) process, whose parameters are estimated from the noisy signal, it is possible to render the noise component closer to white with a simple pre-processing filter (pre-whitener). This makes it possible to estimate the fundamental frequency using the aforementioned assumption of white Gaussian noise. In non-stationary noise scenarios, it is possible to obtain better estimates of the noise spectral envelope as well as a higher degree of spectral flatness by using an adaptive pre-whitening filter based on supervised noise statistics estimates, than one based on unsupervised noise statistics. A pre-whitening filter also improves the accuracy of a source localization method. The problem of joint estimation of the parameters of the voiced speech and the stochastic signal parts (i.e., unvoiced and additive noise) is solved first by the cascade of a pre-whitening filter and the nonlinear least squares (NLS) fundamental frequency estimator, followed by an iterative estimation of the pre-whitening filter, based on the modelled residual, and a re-estimation of the fundamental frequency. This will further reduce the number of gross errors of fundamental frequnecy estimates and the voicing detection errors.

The second aspect is as follows: after a more accurate estimation of the parameters is obtained, the extraction of individual speech components (i.e., voiced and unvoiced speech) from a noisy speech signal, is investigated through linear filtering based on the statistics of the individual components. A Wiener filtering approach allows for a better recovery of both components when compared to the state-of-the-art decomposition methods, which assume that the additive noise is small and insignificant. Instead of using a fixed segment length for the extraction, we also propose to use time-varying segment lengths that are adapted to the signal. The optimal segmentation is obtained once the parameter estimates of a hybrid speech model have been found for all possible candidate models and segment lengths.