

REALISM IN ROBUST SPEECH PROCESSING

Special Session



2016
SAN FRANCISCO
Hyatt Regency | September 8-12



ORGANIZERS



Dayana Ribas (dribas@cenatav.co.cu) is a Young Researcher with CENATAV, Cuba. She has authored more than 15 papers on robust speaker recognition, noise compensation methods and speech quality measures, currently she is finishing the PhD. She is member of ISCA and IAPR.



Emmanuel Vincent (emmanuel.vincent@inria.fr) is an Experienced Research Scientist with Inria, France. He has authored more than 150 papers on speech and audio source separation, uncertainty propagation, and robust automatic speech recognition, and given tutorials and keynotes in conferences such as Interspeech 2012 and WASPAA 2015. He is a founder of the SiSEC and CHiME series of evaluation campaigns on audio source separation and robust automatic speech recognition and the chair of ISCA's Robust Speech Processing (RoSP) SIG.



John H.L. Hansen (john.hansen@utdallas.edu) is Assoc. Dean and Professor of Electrical Engineering, and Director of the Center for Robust Speech Systems (CRSS) at the Univ. of Texas at Dallas (USA). His research interests span the areas of analysis and modeling of speech and speaker traits, feature estimation in noise, robust speech recognition, speaker/language/dialect ID in diverse environmental conditions, and general problems in man-machine interaction. He has published over 580 papers in the field. He is an IEEE Fellow and an ISCA Fellow.

MOTIVATION

One of the challenges currently faced in many areas of speech processing is the migration of laboratory results to real applications. Real applications raise a bulk of robustness problems due to acoustic variability (reverberation, background noise, sensor noise, source/sensor movement...), speaker variability (stressed/Lombard speech, speaker's voice, speaking rate, accent...), and language variability (dialects, topic changes...).

Ultimately, the best way to ensure that a scientific result has practical value is to evaluate it in the targeted scenarios of use. This is rarely feasible in the early stages of research, though, and researchers have resorted to recording or simulating data in controlled scenarios. It is usual that simulated datasets are not acoustically realistic. For example, many popular datasets obtained by mixing speech and noise

at fixed signal-to-noise ratios include some levels and types of distortion that never happen in real life. Even real datasets are often not ecologically realistic because they were not collected in the real conditions of use. This can result in satisfactory performance of the tested methods on scenarios that will never happen in practice, while their performance may be much worse in real scenarios. Furthermore, complex and expensive methods might be obtained that are actually not required.

The information about which conditions are required for a dataset to be realistic and which ones are actually important for the evaluation of a certain task is sparsely found in the literature and in regular conference sessions. Motivated by the growing importance of robustness in commercial speech processing applications, we consider that it is time to bring this issue to the table.

This session aims to provide a forum for the cross-fertilization of expertise and experimental evidence about “realism” across different areas of robust speech processing. Through the study of the state of the art and the exchange of specialized experiences, we aim to characterize real scenarios by measuring the ranges and combinations of different parameters, and to establish “good practices” regarding which parameter violations are acceptable or not given the task to be solved and the limitations of today’s data collection and simulation tools.

SESSION FORMAT

The session will consist of short oral presentations followed by a 30-minute panel discussion. The duration of the oral presentations will be adapted to the number of accepted papers. The panelists will be chosen among the authors of the best rated and most controversial papers. The discussion will be moderated by the organizers and the outcomes will be summarized in a paper to be submitted to Speech Communication or Computer Speech and Language

CALL FOR PAPERS

Prospective authors are invited to submit original papers in areas related to the problem of realism in robust speech processing, including but not limited to

- Objective experimental characterization of real scenarios in terms of acoustic conditions (reverberation, noise, sensor variability, source/sensor movement, environment change...).
- Objective experimental characterization of real scenarios in terms of speech characteristics (spontaneous speech, number of speakers, vocal effort, effect of age, non-neutral speech...).
- Objective experimental characterization of real scenarios in terms of language variability.
- Real data collection protocols,
- Data simulation algorithms,
- New datasets suitable for research on robust speech processing,
- Performance comparison on real vs. simulated datasets for a given task and a range of methods.

Contributions may cover all areas of robust speech processing, including but not limited to speech enhancement, automatic speech, speaker and language recognition, paralinguistics...

