



# REALISM IN ROBUST SPEECH AND LANGUAGE PROCESSING

## Guest Editors



**DAYANA RIBAS** [dayanaribasglez@gmail.com](mailto:dayanaribasglez@gmail.com)  
*Junior Researcher with Advanced Technologies Application  
Center (CENATAV), Cuba.*



**EMMANUEL VINCENT** [emmanuel.vincent@inria.fr](mailto:emmanuel.vincent@inria.fr)  
*Senior Research Scientist with Inria, France.*



**JOHN H.L. HANSEN** [john.hansen@utdallas.edu](mailto:john.hansen@utdallas.edu)  
*Associated Dean and Professor of Electrical Engineering and  
Director of the Center for Robust Speech Systems (CRSS) at  
the University of Texas at Dallas, USA.*



## Call for Papers

How can you be sure that your research has actual impact in **real-world applications**? This is one of the major challenges currently faced in many areas of speech processing, with the migration of laboratory solutions to real-world applications, which is what we address by the term "Realism". Real application scenarios involve several **acoustic, speaker and language variabilities** which challenge the robustness of systems. As early evaluations in practical targeted scenarios are hardly feasible, many developments are actually based on simulated data, which leaves concerns for the **viability of these solutions** in real-world environments.

**Simulated datasets** are not usually 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. Most also ignore Lombard effect which occurs as noise levels increase. 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.

Thus, a suitable dataset for a given task will be strongly depend on how well it matches the acoustic characteristics of the target application scenarios. **Found data** is easy to locate from different available online sources such as YouTube, among others. However the knowledge of the context, identity, acoustic variability and mismatch with specific system goals may cause such data to be less effective. Even, **well organized speech data** from data resource consortia such as LDC used for different purposes than they were originally collected for, may in fact be constructing irrelevant solutions. For example, the big amount of spontaneous telephone conversations included in NIST-SRE datasets, which have been considered as the standard corpora for speaker recognition, would be unrealistic in access control applications, considering the speech duration and modality. New advancements in machine learning technologies for speech and language, such as Deep Neural Networks (DNN), require exten-

sive datasets and researchers are increasingly moving towards found data with **less knowledge or understanding of the impact this has on final solutions**.

Information about **which conditions are required for a dataset to be realistic** and experimental evidence about **which ones are actually important for the evaluation of a certain task** is sparsely found in the literature. Motivated by the growing importance of robustness in commercial speech and language processing applications, this Special Issue aims to provide a venue for **research advancements, recommendations for best practices, and tutorial-like papers** about realism in robust speech and language processing.





Prospective authors are invited to submit original papers in areas related to the problem of realism in robust speech and language processing. Contributions may include, but are not limited to:

- Position papers from researchers or practitioners for best practice recommendations and advice regarding different kinds of real and simulated setups for a given task
- Objective experimental characterization of real scenarios in terms of acoustic conditions (reverberation, noise, sensor variability, source/sensor movement, environment change, etc)
- Objective experimental characterization of real scenarios in terms of speech characteristics (spontaneous speech, number of speakers, vocal effort, effect of age, non-neutral speech, etc)
- 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
- Analysis of advantages vs. weaknesses of simulated and/or real data, and techniques for addressing these weaknesses

Papers may address any area in speech and language processing including: speech enhancement, automatic speech, speaker and language recognition, language modeling, speech synthesis and perception, affective speech processing, paralinguistics, etc.

Papers written by practitioners and industry researchers are especially welcomed. If there is any doubt about the suitability of your paper for this special issue, please contact us before submission.

## Proposed time schedule

-  Deadline for first submission **MAY 31, 2017**
-  Completion of the first review round **SEPTEMBER 30, 2017**
-  Deadline for revised manuscripts **NOVEMBER 30, 2017**
-  Completion of the review and revision process **APRIL 30, 2018**