ICT PHD

Research project for a PhD curriculum in Industrial and Environmental Engineering “Enzo Ferrari” – Industry 4.0 (E4E school)

**Tutor**: Prof. Sonia Bergamaschi

**(\*) Italian [Industrial] Co-tutor:** Leonardo Badino (Istituto Italiano di Tecnologia)

**(\*\*) Foreign Co-tutor:**

**Proposed Title of the research:**

Deep Learning for Audio-Visual Speech Enhancement and Separation

**Keywords: (5)**

Deep Learning, Spoken Language Processing, Speech Enhancement, Cocktail Party, Audio-Visual integration

**Research objectives: --(max 10 rows)**

In the context of speech perception, the cocktail party effect is the ability of the brain to recognize speech in complex and adverse listening conditions where the attended speech is mixed with competing sounds/speech. This is an ill-posed problem in that many different hypotheses about what the target speaker says are consistent with the mixture signal. Yet, it can be solved by exploiting some additional information associated to the speaker of interest and/or by leveraging some prior knowledge about speech signal properties. The goal of this research is to explore the use of video of a talking face (e.g. lip movements) as additional information in a deep learning-based architecture. This choice is motivated by speech perception studies that have shown that watching speaker's face movements could dramatically improve our ability at recognizing the speech of a target speaker in a multi-talker environment.

**Proposed research activity -- (max 10 rows)**

Expected activities (not limited to):

* The PhD student will study the state-of-the-art audio-only and audio-visual speech enhancement and separation models and systems, performing benchmarking on real-world datasets.
* The PhD student will work with machine learning and deep learning framework (scikit-learn, TensorFlow, PyTorch) to implement new speech enhancement and separation systems.

Ideal candidate should have one of the following degrees:

* Computer Engineering/Science
* Statistics
* Mathematics/Physics

**Supporting research projects (and Department)**

CINECA Big Data Research Agreement

**Possible connections with research groups, companies, universities.**

Istituto Italiano di Tecnologia – Dr. Leonardo Badino (Speech and Communication Team - Center for Translational Neurophysiology, Ferrara)

(\*) optional

(\*\*) optional/to be completed on the second year