Within the framework of
ERC FACTORY

Informatics Research Institute of Toulouse

Thursday 21 February 2019
11h00 – 12h00
INP-ENSEEIHT, Salle des thèses

Thomas PELLEGRINI
Team SAMOVA - IRIT (France)

Weakly-supervised approaches for sound event detection

Abstract: Sound Event Detection (SED) systems aim to detect possibly overlapping audio events, and locate the events temporally in recordings, i.e. determining event onsets and offsets. In this talk, I will review weakly-supervised deep learning approaches for SED, a topic that is gaining momentum due to the availability of very large datasets such as Audioset, comprised of terabytes of audio extracts along with audio tags produced by YouTube users. There is still an issue, though, regarding annotated data availability for most audio event types and data-efficient/noisy label robust algorithms are needed. I will present state-of-the-art deep neural networks trained for this task when only “weak labels” are available for learning. Weak labels refer to audio tags at recording level with no information on temporal onsets and offsets of the annotated events. I will review two main research directions: i) the introduction of attention mechanisms in the network architecture, ii) the use of Multiple Instance Learning inspired objective functions. I will comment on their limitations and how these could be overcome.