

## Thomas Pellegrini



**Bio.** Since 2013, Thomas Pellegrini is an Associate Professor in Computer Science at Université Paul Sabatier in Toulouse, associated to the IRIT lab. He holds an engineering degree in Physics from Ecole Supérieure de Physique et Chimie Industriels de Paris (ESPCI, 2004), a M.S. in Computer Science with a specialization in audio processing (Master ATIAM, 2004), and a PhD from Université Paris-Sud at LIMSI-CNRS (2008) on lexicon modeling in speech recognition for less-represented languages.

From 2008 to 2013, he worked as a post-doc at the Spoken Language Systems Lab (L2F) of INESC-ID in Lisbon on speech recognition for the elderly, on linguistic data sharing (METANET), and on audio event detection in authentic videos (VIDI-VIDEO). Since his arrival at IRIT in 2013, he contributes to the group research lines related to speech and audio processing, with a strong interest these last years for deep learning applied to audio signal processing. In 2018, he was awarded a Jeune Chercheur project fellowship on lightly-supervised and unsupervised discovery of audio units using deep learning.

## **Deep learning with weakly-annotated data: a sound event detection use case (and hate speech detection here and there)**

Thomas Pellegrini  
IRIT & Université de Toulouse, France  
thomas.pellegrini@irit.f

**Abstract.** Weakly-annotated data correspond to data manually annotated with "weak" labels. Weak labels refer to global tags, at document level, with no information about the precise location (in time or space) of the events of interest. Deep neural networks can be trained with these data as predictors of the tags of interest. We would like to design methods to go further by trying to use these networks to also predict where the events of interest are localized within the input data. Weakly-supervised deep learning approaches will be described, with sound event detection and hate speech detection as a use cases. 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.

**Keywords.** weakly-annotated data, lightly-supervised deep learning, sound event detection