Audio event analysis: detection, classification, and false positive reduction

Nonspeech audio event classification and detection (AEC/D) are important problems of computational auditory scene analysis. Since the events convey rich information about physical events taking place in the surroundings, they enables novel application areas such as safety, ambient assisted living, health monitoring to name a few. Compared to speech recognition, these problems are still in its infancy. However, the adaptation of automatic speech recognition framework to the problems has resulted in limited success. This talk will present our approaches covering different aspects of AEC/D. First, we propose a regression approach, based on random regression forest framework, to cope with the AED problem in continuous streams as well as study how the target events can be detected early in time even when their partial durations are observed without losing the overall detection performance. Second, we present two types of learned descriptors, namely data-specific and data-generic descriptors, for audio events to tackle the classification problem. The former, which is specific for a target dataset, is derived by utilizing the bank of class-specific regressors that are previously used for the detection purpose. The later, which is dataset-independent, is relied on similarities of a nonspeech audio event instance and different speech patterns. Finally, we revisit the AED problem and propose an improved detection pipeline in which the high-quality classifiers are used to verify and reject false-positive hypotheses outputted by the detection systems, leading to better overall performance.