# Spectro-temporal matching for the classification of acoustic events

Project proposal

Acoustic event detection and classification is usually based on techniques from automatic speech recognition. At INCAS3 the approach is different. Using a model of the mechanical part of human hearing, we create a representation of the sound called a cochleogram[[1]](#footnote-1). The cochleogram gives the energy in the target sound as a function of time and frequency. From this cochleogram we determine features such as the loudness, duration and pitch of sounds. These features are then used for classification of the target by comparing them to values representing a class.

One of the features that is used to differentiate classes is called spectral matching. A representative spectrum of the target O(f) is compared to a reference spectrum of the class M(f).

The higher the value of α, the better the match.

For several sounds it is difficult to find a representative spectrum O(f). Extending the method to incorporate the time direction, creating spectro-temporal matching, would allow us to classify sounds with much higher accuracy. For the temporal direction, however, the situation can be more complicated than a simple shift in time. There are many sound sources that show variation in timing. In Dynamic Time Warping (DTW) a mapping is found between the time axis of O and the time axis of M which optimizes the match between O and M. Since it uses spectral matches, the procedures described for calculating α can be used directly.

Cochleograms can be provided for a variety of databases of environmental and office sounds. Matlab code for DTW for audio signals is available for synchronizing two utterances of the same word from <http://labrosa.ee.columbia.edu/matlab/dtw/>. This should provide an excellent starting point to calculate matching values based on DTW.

1. The name cochleogram refers to the cochlea, the human inner ear. [↑](#footnote-ref-1)