

## Abstract

Advances in time-frequency distributions and spectral analysis techniques (i.e., for the estimation of amplitude and/or frequency modulations) allow a better representation of non-stationary signals like speech, highlighting their fine structure and dynamics. Although such representations are very useful for analysis purposes, they complicate the classification tasks due to the large number of parameters extracted from the signal (“curse of dimensionality”). For such tasks, a significant dimensionality reduction is required.

In this thesis, the problem of dimensionality reduction of these time/frequency-frequency representations is studied; selection criteria of the optimal parameters are suggested, based on their relevance to a given classification task. Relevance is defined based on mutual information. First, using tools from multilinear algebra, such as High Order SVD, the initial dimensions and the noise components of the representation are reduced. Then, feature selection proceeds based on maximum relevance criterion. It is shown that the suggested process is equivalent to the maximum dependency criterion for feature selection, without, however, the need of the multivariate probability densities estimation.

The feature selection approach suggested in the thesis is applied on a number of audio classification tasks, including speech detection in broadcast news and voice pathology detection and discrimination from vowel recordings. The complementarity of the modulation spectral features to the state-of-the-art Mel frequency cepstral coefficients is shown for the above classification tasks. A system for the automatic discrimination of pathological heart murmurs using a high resolution time-frequency analysis of the phonocardiogram (PCG) is also presented. The classification accuracy of the system is comparable to the diagnostic accuracy of experienced paedo-cardiologists on the same PCG dataset.