

ACOUSTICS2008/2617
**Hierarchical approach to voice activity detection in realistic
environmental noise**

Jörg-Hendrik Bach and Jörn Anemüller
Universität Oldenburg, Medizinische Physik, Fakultät V, 26111 Oldenburg, Germany

The automatic identification of acoustic objects, particularly speech, in adverse backgrounds can pose a difficult problem. In this work, the problem of speech detection in different backgrounds is tackled using a hierarchical decision tree: The root node provides a coarse classification of the noise background, while the next nodes are optimised to detect speech in one particular background each. The results of this approach are compared to the ITU G729B standard VAD and to a binary classifier trained on speech vs a mixture of the backgrounds. The features employed are amplitude modulation spectra of the signals (Kollmeier & Koch, 1993) that represent the modulation content of a signal as a two-dimensional (modulation-frequency vs centre frequency) spectral pattern. The best features for classification at each node are selected independently (node-specific) using a sequential floating forward search algorithm. The statistical learner is an open source support vector machine (Chang & Lin, 2001).