

# AUTOMATIC LOW-DIMENSIONAL ANALYSIS OF AUDIO DATABASES

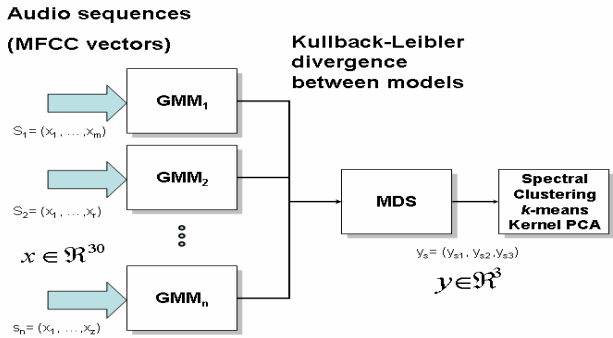
José Anibal Arias, Régine André-Obrecht, Jérôme Farinas  
 Université Paul Sabatier – IRIT, Toulouse, France  
 {arias, obrecht, jfarinas}@irit.fr

## Motivations

- Map variable size audio sequences into fixed-length vectors.
- Explore the contents of audio databases (music, speakers and languages).

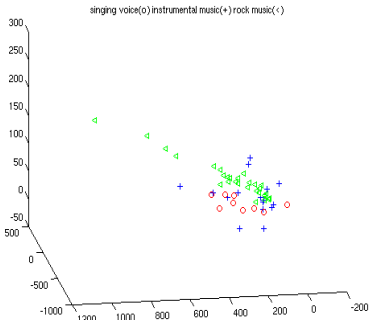
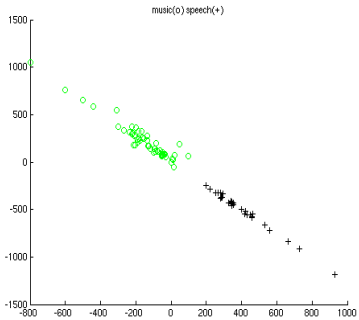
## General system

- MFCC parameterization of each sequence  $s_i$  in the database
- $GMM_i$  training (initialization step -VQ- followed by an optimization step -EM-).
- Symmetric Kullback-Leibler divergence between  $GMM_i$  stored in a squared matrix.
- Multidimensional scaling of “dissimilarities” matrix. MDS finds a low-dimensional Euclidian space for the sequences.
- Cluster and Kernel methods can be applied afterwards.
- Several variants can be tested: different parameterizations, GMM creation (through UBM adaptation), statistical dissimilarities.



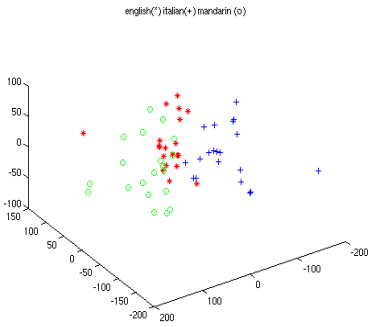
## Speech-music discrimination

- If we mix singing voice files, rock/jazz sequences, instrumental music and speech files from OGI database, we obtain two clusters.
- Within music cluster, some structure is revealed.



## Language identification

- OGI database (multilanguage spontaneous telephonic speech).
- Silence elimination and shifted delta cepstra (config. 7-1-3-3) parameterization is used.



## Speaker identification

- ANITA database (high quality studio speech, phonetically balanced sequences).
- Automatic clustering of 180 sequences.

