Automatic timbre classification of ethnomusicalological audio recordings:
instrumental family and playing style.

Motivation:
- automatic audio database indexing
- retrieval of instrument samples with a timbre similarity measure
- diversified database with musicalological analysis: uncommon non-western instruments (e.g., struck strings)
- diversified database with acoustic properties analysis

Context:
- ground recording conditions (noise, interference, poor quality medium)
- diversified database with acoustic properties analysis: uncommon non-western instruments (e.g., struck strings)
- diversified database with musicalological instrument taxonomy [5]

Method overview

Linear Discriminant Analysis (LDA)
Goal: find the best projection or linear combination of all descriptors which maximizes the average distance between classes (inter-class distance) while minimizing distance between individuals from the same class (intra-class distance).

Features selection method
Goal: select the most relevant features for automatic timbre classification. 3 methods were compared:
- Analysis of eigenvectors resulting from LDA
- Maximizing the Mutual Information (MI) between features and classes
- Inertia Ratio: Maximization using Features Space Projection (RFSP) [5]

Class modeling and automatic classification
Each class k is modeled into the projected classification space of descriptors d by a probability density function (pdf) p(d|k) = p_k(d)/p(k). The classification decision which affects a class k to an input sound represented by a projected vector of features x, maximizes the resulting pdf.

k = arg max p_k(x) ∀k ∈ [1, K]  
(1)

Results

Classification accuracy
Iowa database using T2: Iowa database using T1

Timbre quantization and classification

The Timbre quantization is based on 364 audio acoustic features as proposed by Poeters et al. [4] which can be organized as follows:
- Temporal descriptors (e.g., log attack time, temporal increase, zero-crossing rate, etc.)
- Harmonic descriptors (e.g., linearity, inharmonicity, etc.)
- Spectral descriptors (e.g., spectral centroid, spectral decrease, etc.)

Perceptional descriptors are computed from auditory-filtered binaural listening signals, which aim at approximating the human perception of sounds.

Classification taxonomies

Conclusion and future work

- Timbre quantization and classification
- Classification accuracy
- Confusion matrices
  - CREM database using T1, 20 descriptors
  - CREM database using T2, 20 descriptors

References