Automatic timbre classification of ethnomusicological audio recordings

D. Fourer, J-L. Rouas, P.Hanna and M. Robine
LaBRI - CNRS UMR 5800
{fourer, rouas, hanna, robine}@labri.fr

Abstract

Automatic timbre characterization of audio signals can help to measure similarities between sounds and is of interest for automatic or semi-automatic databases indexing. The most effective methods use machine learning approaches which require qualitative and diversified training databases to obtain accurate results. In this paper, we introduce a diversified database composed of worldwide non-western instruments audio recordings on which is evaluated an effective timbre classification method. A comparative evaluation based on the well studied Iowa musical instruments database shows results comparable with those of state-of-the-art methods. Thus, the proposed method offers a practical solution for automatic ethnomusicological indexing of a database composed of diversified sounds with various quality. The relevance of audio features for the timbre characterization is also discussed in the context of non-western instruments analysis.

Purpose of this work

Automatic timbre classification of ethnomusicological audio recordings: instrumental familly and playing style.

Motivation:

- automatic audio database indexing
- retrieval of instrument samples with a timbre similarity measure for musicological analysis
- instruments acoustic properties analysis

Context:

- ground recording conditions (noise, interferences, poor quality medium)
- diversified database with uncommon non-western instruments (e.g. struck bamboo)
- generalized musicological instrument taxonomy [5]

Corpus description

The CREM database [1] is a research database composed of diversified sound samples directly recorded by ethnomusicologists in various conditions (i.e. no recording studio) and from diversified places all around the world.

- more than 7000 hours of audio recordings since 1932,
- digitized sounds from various audio media (e.g. magnetic tapes, vynil, etc.)
- ground recording conditions
- presence of uncommon non-western instruments (e.g. lute or the Ngbaka harp, struck bamboo, etc.)

Class name	Duration (s)		#	
aerophones-blowed	1,383		146	
cordophones-struck	357	1,229	37	128
cordophones-plucked	715		75	
cordophones-bowed	157		16	
idiophones-struck	522	753	58	82
idiophones-plucked	137		14	
idiophones-clinked	94		10	
membranophones-struck	170		19	
Total	3,535		375	

The Iowa database [2] is a well known database on which existing methods were successfully evaluated.

- Composed of western pitched instruments
- studio recording condition

T1 class name	T2 equivalence	Duration (s)	#
aero-blowed	reed/flute and brass	5,951	668
cordo-struck	struck strings	5,564	646
cordo-plucked	plucked strings	5,229	583
cordo-bowed	bowed strings	7,853	838
Total		24,597	2,735

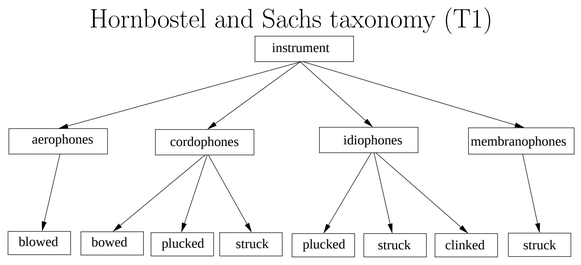
Timbre quantization and classification

The **Timbre quantization** is based on 164 audio acoustic features as proposed by Peeters *et al.* [4] which can be organized as follows:

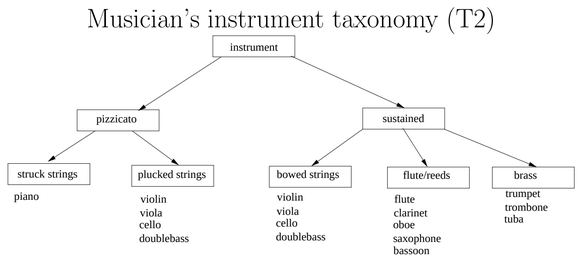
- *Temporal descriptors* (e.g. log attack time, temporal increase, zero-crossing rate, etc.).
- *Harmonic descriptors* (e.g. noisiness, inharmonicity, etc.).
- *Spectral descriptors* (e.g. spectral centroid, spectral decrease, etc.)
- *Perceptual descriptors* are computed from auditory-filtered bandwidth versions of signals which aim at approximating the human perception of sounds.

Classification taxonomies

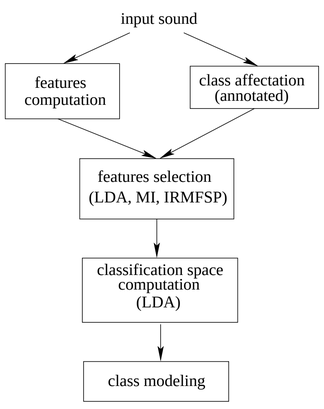
Hornbostel and Sachs taxonomy (T1)



Musician's instrument taxonomy (T2)



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 (IRMFSP) [3].

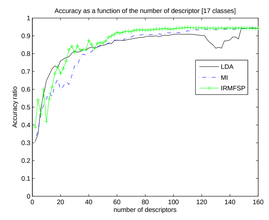
Class modeling and automatic classification
Each class k is modeled into the projected classification space of descriptors d, by a probability density function (pdf) $p(k|d) = p(k)p(d|k)/p(d)$. The classification decision which affect a class $\hat{k}$ to an input sound represented by a projected vector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

$$\hat{k} = \arg\max_{k} p(k|x) \;\; \forall k \in [1, K]. \qquad (1)$$

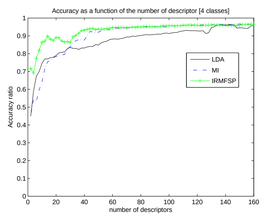
Results

Classification accuracy

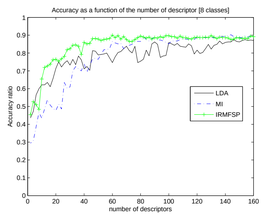
Iowa database using T2



Iowa database using T1



CREM database using T1



Confusion matrices

CREM database using 20 descriptors (selected by IRMFSP):

	aero	c-struc	c-pluc	c-bowed	i-pluc	i-struc	i-clink	membr
aero	**71**	3	9	5		7		5
c-struc	5	**92**		3				
c-pluc	5	8	**74**	4		8		1
c-bowed			13	**80**	7			
i-pluc					**79**	14		7
i-struc	8	2	5		2	**79**		4
i-clink							**100**	
membr			11			17		**72**

CREM+Iowa databases merged using 20 descriptors:

	aero	c-struc	c-pluc	c-bowed	i-pluc	i-struc	i-clink	membr
aero	**75**	14	5	3	2	1		
c-struc	12	**70**	10	5	1			2
c-pluc	1	7	**58**	29	1	2		2
c-bowed	3	6	33	**54**	1	3		
i-pluc		7		14	**79**			
i-struc	2	2	4	11	2	**51**		28
i-clink	11						**89**	
membr				6		16		**78**

Conclusion and future work

- a novel diversified ethnomusicological database with non-western instruments was introduced
- an efficient solution (about 80% of accuracy) was adapted and successfully evaluated for automatic timbre classification

Future work will consist in:

- a further analysis of the most relevant features
- application to instrument detection in polyphonic recordings

References

[1] Center for Research in Ethnomusicology (CREM). Sound archives of the cnrs - musée de l'homme, freely available online at: http://archives.crem-cnrs.fr/.

[2] L. Fritts. Musical instrument samples. Univ. Iowa Electronic Music Studios, 1997. [Online]. Available: http://theremin.music.uiowa.edu/MIS.html.

[3] G. Peeters. Automatic classification of large musical instrument databases using hierarchical classifiers with intertia ratio maximization. In *115th convention of AES*, New York, USA, Oct. 2003.

[4] G. Peeters, B. Giordano, P. Susini, N. Misdariis, and S. McAdams. The timbre toolbox: Audio descriptors of musical signals. *Journal of Acoustic Society of America (JASA)*, 5(130):2902–2916, Nov. 2011.

[5] E. v. Hornbostel and C. Sachs. The classification of musical instruments. *Galpin Society Journal*, 3(25):3–29, 1961.