HIERARCHICAL GAUSSIAN TREE WITH INERTIA RATIO MAXIMIZATION FOR THE CLASSIFICATION OF LARGE MUSICAL INSTRUMENT DATABASES

Geoffroy Peeters

Ircam – CUIDADO I.S.T.
1, pl. Igor Stravinsky - 75004 Paris - France
`peeters@ircam.fr`

Xavier Rodet

Ircam
1, pl. Igor Stravinsky - 75004 Paris - France
`rod@ircam.fr`

1. ABSTRACT

This paper addresses the problem of classifying large databases of musical instrument sounds. We propose an efficient algorithm for selecting the most appropriate features for a given classification task. This algorithm, called IRMFSP, is based on the maximization of the ratio of the between-class inertia to the total inertia combined with a step-wise feature space orthogonalization. The IRMFSP algorithm is then compared successfully to the widely used feature selection algorithm CFS. We then show the limits of usual flat (all classes considered on a same level) classifiers for large database classification and propose the use of hierarchical classifiers. Finally, we show the applicability of our system for large database classification. Especially considered is the application when our classification system is trained on a given database and used for the classification of another database possibly recorded in completely different conditions.

2. INTRODUCTION

During the last decades, sound classification has been the subject of many research efforts [1] [2] [3] [4]. However, few of them address the problem of generalization of the sound source recognition system i.e. applicability to several instances of the same source possibly recorded in different conditions, with various instrument manufacturers and players. In this context, Martin [5] reports only 39% recognition rate for individual instrument (76% for instrument family), using the output of a log-lag correlogram for 14 different instruments. Eronen [6] reports 35% (77%) recognition rate using mainly MFCCs and some other features for 16 different instruments.

Sound classification systems rely on the extraction of a set of signal features (such as energy, spectral centroid…) from the signal. This set is then used to perform classification according to a given taxonomy. This taxonomy is defined by a set of textual attributes defining the properties of the sound such as its source (speaker genre, music genre, sound effects class, instrument name...) or its perception (bright, dark...).

The choice of the features depends on the targeted application (speech/music/noise discrimination, speaker identification, sound effects recognition, musical instruments recognition). The most appropriate set of features can be selected a priori - having a prior knowledge of the feature discriminative power for the given task -, or a posteriori by including in the system an algorithm for automatic feature selection. In our system, in order to allow the coverage of a large set of potential taxonomies, we have implemented a large set of features. This set of features is then filtered automatically by a feature selection algorithm.

Because sound is a phenomenon which changes over time, features are computed over time (frame by frame analysis). The set of temporal features can be used directly for classification [2]; or the temporal evolution of the features can be modeled. Modeling can be done before the modeling of the classes (using mean, std, derivative values, modulation or polynomial representation [4]) or during the modeling of the classes (using for example a Hidden Markov Model [7]). In our system, temporal modeling is done before that of the classes.

The last major difference between classification systems concerns the choice of the model to represent the classes of the taxonomy (multi-dimensional gaussian, gaussian mixture, KNN, NN, decision tree, SVM...).

The system performance is generally evaluated, after training on a subset of a database, on the rest of the database. While such an evaluation gives an insight on the quality of the classification system it does not prove any applicability of the system for the classification of sounds which do not belong to the database. In particular, the system may fail to recognize sounds recorded in completely different conditions. In this paper we evaluate such performances.

3. FEATURE EXTRACTION

Many different types of signal features have been proposed for the task of sound recognition coming from the speech recognition community, previous studies on musical instrument sounds classification [1] [2, 3] [4] [8] and results of psycho-acoustical studies [9] [10]. In order to allow the coverage of a large set of potential taxonomies, a large set of features has been implemented, including - features related to the *temporal shape* of the signal (attack-time, temporal increase/decrease, effective duration), - *harmonic features* (harmonic/noise ratio, odd to even and tristimuls harmonic energy ratio, harmonic deviation), - *spectral shape features* (centroid, spread, skewness, kurtosis, slope, roll-off frequency, variation), - *perceptual features* (relative specific loudness, sharpness, spread, roughness, fluctuation strength), - Mel-Frequency Cepstral Coefficients (plus Delta and DeltaDelta coefficients), auto-correlation coefficients, zero-crossing rate, as well as some MPEG-7 Low Level Audio Descriptors (spectral flatness and crest factors [11]). See [12] for a review.

4. FEATURE SELECTION

Using a high number of features for classification can cause several problems: 1) bad classification results because some features are irrelevant for the given task; 2) over fitting of the model to the training set (this is especially true when using, without care, data reduction techniques such as Linear Discriminant Analysis), 3) the models are difficult to interpret by human.

For this reason, feature selection algorithms attempt to detect the minimal set of
1) informative features with respect to the classes
2) features that provide non redundant information.

4.1. Inertia Ratio Maximization using Feature Space Projection (IRMFSP)

Feature selection algorithms (FSA) can take three main forms (see [13]): -embedded (the FSA is part of the classifier) – filter (the FSA is distinct from the classifier and used before the classifier – wrapper (the FSA makes use of the classification results). The FSA we propose is part of the Filter techniques.

Considering a gaussian classifier, the *first criterion* for FSA can be expressed in the following way: "feature values for sounds belonging to a specific class should be separated from the values for all the other classes". If it is not the case then the gaussian pdfs will overlap, and class confusion will increase. In a mathematical way this can be expressed by looking at features for which the ratio r of the Between-class inertia B to the Total class inertia T is maximum. For a specific feature f_i, r is defined as

$$r = \frac{B}{T} = \frac{\sum_{k=1}^{K} \frac{N_k}{N}(m_{i,k} - m_i)(m_{i,k} - m_i)'}{\frac{1}{N}\sum_{n=1}^{N}(f_{i,n} - m_i)(f_{i,n} - m_i)'}$$

where N is the total number of data, N_k is the number of data belonging to class k, m_i is the center of gravity of the feature f_i over all the data set, and $m_{i,k}$ is the center of gravity of the feature f_i for data belonging to class k. A feature f_i with a high value of r is therefore a feature for which the classes are well separated with respect to their within spread.

The *second criterion* should allow taking into account the fact that a feature with a high value of r could bring the same information as an already selected feature and is therefore redundant. While other FSAs, like the CFS one [14][1], use a weight based on the correlation between the candidate feature and already selected features, in the IRMFSP algorithm, an orthogonalization process is applied after the selection of each new feature f_i. If we note $\underline{\underline{F}}$ the feature space (space where

---

[1] In the *CFS* algorithm (Correlation-based Feature Selection), the information brought by one specific feature is computed using symmetrical uncertainty (normalized mutual information) between discretized features and classes. The second criterion (features independence) is taken into account by selecting a new feature only if its cumulated correlation with already selected features is not too large.

each axis represents a feature), $\underline{f_i}$ the last selected feature and $\underline{g_i}$ its normalized form ($\underline{g_i} = \underline{f_i}/\|f_i\|$), we project $\underline{\underline{F}}$ on $\underline{g_i}$ and keep $\underline{f_j}'$:

$$\underline{f_j}' = \underline{f_j} - (\underline{f_j} \cdot \underline{g_i}) \quad \underline{g_i} \quad \forall j \in \underline{\underline{F}}$$

This process (ratio maximization followed by space projection) is repeated until the gain of adding a new feature $\underline{f_i}$ is too small. This gain is measured by the ratio r_l obtained at the lth first iteration to the one at the first iteration. A stopping criterion of $t = \frac{r_l}{r_1} < 0.01$ has been chosen. In Part 7.2.1, the CFS and IRMFSP algorithm are compared.
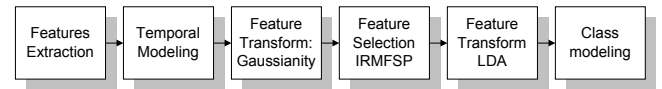
5. FEATURE TRANSFORMATION

Figure 1 *Classification system flowchart*

In the following, two feature transform algorithms (FTA) are considered.

The *first FTA* is the Linear Discriminant Analysis (LDA) which was proposed by [3] in the context of musical instrument sound classification and evaluated successfully in our previous classifier [15]. LDA allows finding a linear combination among features in order to maximize discrimination between classes. From the initial feature space $\underline{\underline{F}}$ (or a selected feature space $\underline{\underline{F}}'$), a new feature space $\underline{\underline{G}}$ of dimension smaller than $\underline{\underline{F}}$ is obtained.

Classification models based on gaussian distribution makes the underlying assumption that modeled data (in our case signal features) follow a gaussian probability density function (pdf). However, this is rarely verified by features extracted from the signal. Therefore a *second FTA*, a non-linear transformation, can be applied to each feature individually in order to make its pdf fit as much as possible a gaussian pdf. The set of considered non-linear functions depending on the parameters λ is defined as

$$f_\lambda(x) = \frac{x^\lambda - 1}{\lambda} \text{ if } \lambda \neq 0 \text{ and } f_\lambda(x) = \log(x) \text{ if } \lambda = 0$$

For a specific value of λ, the gaussianity of $f_\lambda(x)$ is measured by the correlation factor between the percent point function ppf (inverse of the cumulative distribution) of $f_\lambda(x)$ and the theoretical ppf of a gaussian function. For each feature x, we find the best non-linear function (best value of λ) defined as the one with the largest gaussianity.

6. CLASS MODELING

Among the various existing classifiers (multi-dimensional gaussian, gaussian mixture, KNN, NN, decision-tree, SVM...)

(see [12] for a review), only the gaussian classifier and its hierarchical formulation have been considered.

6.1. Flat gaussian classifier (F-GC)

A flat gaussian classifier models each class k by a multi-dimensional gaussian pdf. The parameters of the pdf (mean $\underline{m}_k$ and covariance matrix $\underline{\underline{\Sigma}}_k$) are estimated by maximum-likelihood given the selected features for sounds belonging to class k. The term "flat" is used here since all classes are considered on a same level. In order to evaluate the probability that a new sound belongs to a class k, Bayes formula is used.

The training and evaluation process of a flat gaussian classifier system is illustrated in Figure 2 [Left part].

6.2. Hierarchical gaussian classifier (H-GC)

A hierarchical gaussian classifier is a tree of flat gaussian classifiers, i.e. each node of the tree is a flat gaussian classifier with its own feature selection (IRMFSP), its own LDA, its own gaussian pdfs. Hierarchical classifiers have been used by *[3]* for the classification of 14 instruments (derived from the McGill Sound Library) using a hierarchical KNN-classifier and Fisher multiple discriminant analysis combined with a gaussian classifier. During the *training*, only the subset of sounds belonging to the classes of the current node (example: the bowed-string node is trained using only bowed-string sounds, the brass node is trained using only brass sounds) is used. During the *evaluation,* the maximum local probability at each node (probability $p(k|\underline{f})$) decides which branch of the tree to follow. The process is then pursued until reaching a leaf of the tree. Contrary to binary trees, the construction of the tree structure of a H-GC is supervised and requires a previous knowledge of class organization (oboe belongs to double-reeds family which belongs to sustained sounds).

Advantages of Hierarchical Gaussian Classifiers (H-GC) over Flat Gaussian Classifiers (F-GC).

Learning facilities: Learning a H-GC (feature selection and gaussian pdf model parameter estimation) is easier since it is easier to characterize the difference in a small subset of classes (learning the difference between brass instruments only is easier than between the whole set of classes).

Reduced class confusion: In a F-GC, all classes are represented on the same level and are thus neighbors in the same multi-dimensional feature space. Therefore, annoying class confusions, as for example confusing an "oboe" sound with an "harp" sound, are likely to occur. In a H-GC, because of the hierarchy and the high recognition rate at the higher levels of the tree (such as non sustained /sustained sounds node), this kind of confusion is unlikely to occur.

The training and evaluation process of a hierarchical gaussian classifier system is illustrated in Figure 2 [Right part]. The gray/white box connected to each node of the tree is the same as the one of Figure 2 [Left part].
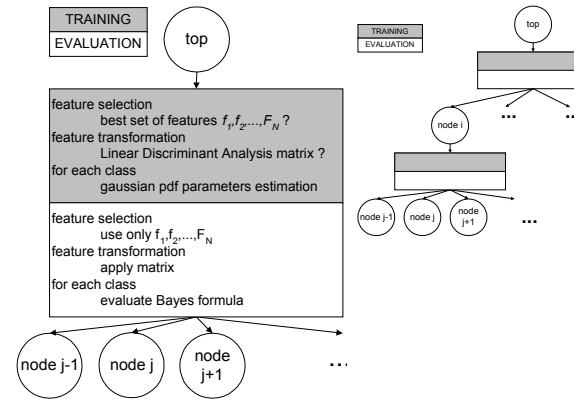

Figure 2 *[Left] Flat gaussian classifier learning and evaluation [Right] Hierarchical gaussian classifier*

7. EVALUATION

7.1. Methodology

7.1.1. Evaluation process

For the evaluation of the models, three methods have been used.

The *first evaluation method* used is the random 66%/33% partition of the database where 66% of the sounds of each class of a database are randomly selected in order to train the system. The evaluation is then performed on the remaining 33%. In this case, the result is given as the mean value over 50 random sets.

The second and third evaluation methods were proposed by Livshin [16] for the evaluation of large database classification, especially for testing the applicability of a system trained on a given database when used for the recognition of another database

The *second evaluation method*, called O2O (One to One), uses in turns each database for training the system and measure the recognition rate on each of the remaining ones. If we note A, B and C the various databases, the training is performed on A, and used for the evaluation of B and C; then the training is performed on B, and used for the evaluation of A and C, …

The *third evaluation method*, called the LODO (Leave One Database Out), uses all databases for the training except one which is used for the evaluation. All possible left out databases are chosen in turns. The training is performed on A+B, and used for the evaluation of C; then the training is performed on A+C, and used for the evaluation of B; …

7.1.2. Taxonomy used

The instrument taxonomy used during the experiment is represented in Figure 3. In the following experiments we consider taxonomies at three different levels:
1. a 2 classes taxonomy: sustained/ non-sustained sounds. We call it *T1* in the following.
2. a 7 classes taxonomy corresponding to the instrument families. We call it *T2* in the following.

3. a 27 classes taxonomy corresponding to the instrument names. We call it *T3* in the following.

This taxonomy is of course subject to discussions, especially - the piano, which is supposed to belong here to the non-sustained family – the inclusion of all saxophone instruments in the same family as the oboe.
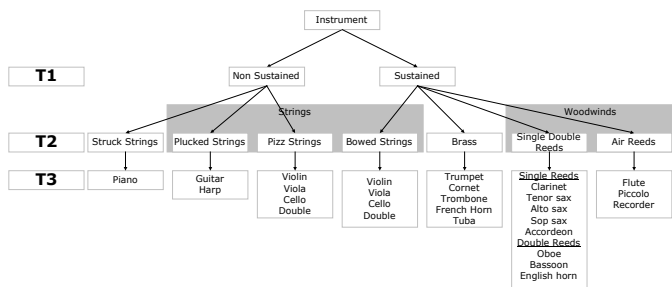

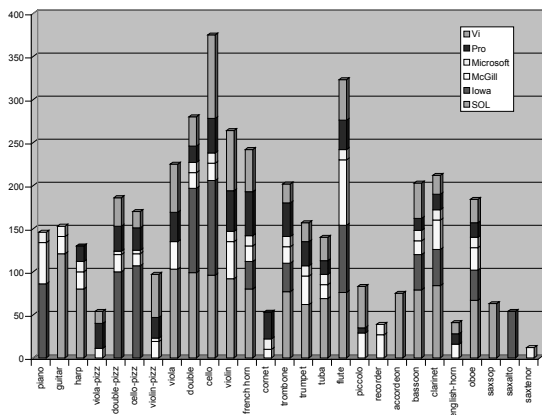Figure 3 *Instrument Taxonomy used for the experiment*

Figure 4 *Instrument distribution of the six databases*

7.1.3. Test set

Six different databases were used for the evaluation of the models:

- the Ircam Studio OnLine [17] (1323 sounds, 16 instruments),
- the Iowa University database [18] (816 sounds, 12 instruments),
- the McGill University database [19] (585 sounds, 23 instruments),
- sounds extracted from the Microsoft "Musical Instruments" CD-ROM [20] (216 sounds, 20 instruments),
- two commercial databases the Pro (532 sounds, 20 instruments) and the Vi databases (691 sounds, 18 instruments), for a total of 4163 sounds.

It is important to note that a large pitch range has been considered for each instruments (4 octaves on average). In the opposite, not all the sounds from each database have been considered. In order to limit the number of classes, the muted sounds, the martele/ staccato sounds and some more specific type of playing have not been considered. The instrument distribution of each database is depicted in Figure 4.

7.2. Results

7.2.1. Comparison of feature selection algorithms

In Table 1, we compare the result of our previous classification system [15] (which was based on Linear Discriminant Analysis applied to the whole set of features combined with a flat gaussian classifier) with the results obtained with the flat gaussian classifier applied directly (without feature transform) to the output of the two feature selection algorithms CFS (weka [21] implementation) and IRMFSP. The result is given for the Studio OnLine database for taxonomies T1, T2 and T3. Evaluation is performed using the 66%/33% paradigm with 50 random sets.

Discussion: Comparing the result obtained with our previous classifiers (LDA) and the result obtained with the IRMFSP algorithm, we see that using a good feature selection algorithm not only allows to reduce the number of features but also increases the recognition rate. Comparing the results obtained using the CFS and IRMFSP algorithms, we see that for T3 IRMFSP performs better than CFS. Since the number of classes is larger at T3, the number of required features is also larger and features redundancy is more likely to occur. CFS fails at T3, perhaps because of a potentially high feature redundancy.

7.2.2. Comparison of classification algorithms for large database classification

In Table 2, we compare the recognition rate obtained using flat gaussian (F-GC) and hierarchical gaussian (H-GC) classifiers. The results are indicated using the O2O evaluation method for the six databases. The results are indicated as mean values over the 30 (6*5) O2O experiments. Feature transform algorithms (LDA and Gaussianity) are not used here considering that the number of data inside each database is too small for a correct estimation of FTA parameters. Features have been selected using the IRMFSP algorithm with a stopping criterion of t<0.01 and a maximum of 10 features per node.

Discussion: Compared to the results of Table 1, we see that good results with flat gaussian classifier using 66%/33% paradigm on a single database does not prove any applicability of the system for the recognition of another database (30% using F-GC at T3 level). This is partly explained by the fact that each database contains a single instance of an instrument (same instrument played by the same player in the same recording conditions). Therefore the system mainly learns the instance of the instrument instead of the instrument itself and is unable to recognize another instance of it. Results obtained using H-GC are higher than with H-GC (38% at T3 level). This can be partly explained by the fact that, in a H-GC, lower levels of the tree benefit from the classification results of higher levels. Since the number of instances used for the training at the higher level is larger (at the T2 level, each family is composed of several instruments, thus several instances of the family) the training of higher level is able to be generalized and the lower level benefits from this. Not indicated here are the various recognition rates of each individual O2O experiment. These results show that when the training is performed on either Vi, McGill or Pro database, the model is applicable for the recognition of most other databases. On the other hand, when training is performed on Iowa database, the model is poorly applicable to other databases.

In order to increase the number of instances of each instrument, several databases can be combined as in the LODO evaluation method. The results of the LODO experiment are indicated in Table 3 as mean values over the 6 Left Out databases. Features have been selected using the IRMFSP algorithm with a stopping criterion t<0.01 and a maximum of 40 features per node.

Discussion: As expected, the recognition rate increases with the number of instances of each instrument used for the training (F-GC O2O 30% to F-GC LODO 53%, H-GC O2O 38% to H-GC LODO 57%). The best results are again obtained with the hierarchical gaussian classifiers. In Table 3, the effect of applying feature transform algorithm (gaussianity and LDA) for both F-GC and H-GC is observed. In the case of H-GC, it increases the recognition rate from 57% to 64%.

	T1	T2	T3
LDA	96	89	86
CFS weka	99.0 (0.5)	93.2 (0.8)	60.8 (12.9)
IRMFSP (t=0.01, nbdescmax=20)	99.2 (0.4)	95.8 (1.2)	95.1 (1.2)

Table 1 *Feature selection algorithms: comparison in terms of recognition rate, mean (standard deviation)*

	T1	T2	T3
F-GC	89	57	30
H-GC	93	63	38

Table 2 *Comparison of flat and hierarchical gaussian classifier using O2O methodology*

	T1	T2	T3
F-GC	98	78	53
F-GC (G+LDA)	99	81	52
H-GC	98	80	57
H-GC (G+LDA)	99	85	64

Table 3 *Comparison of flat and hierarchical gaussian classifier using LODO methodology*

8. CONCLUSION

In this paper we investigated the classification of large musical instrument databases. We proposed a new feature selection algorithm based on the maximization of the ratio of the between-class inertia to the total inertia, and compared it successfully with the widely used CFS algorithm. We studied the use of hierarchical gaussian classifiers for large database classification. The recognition rate obtained with our system (64% for 23 instruments, 85% for instrument families) must be compared to the results reported by previous studies: Martin (respectively Eronen), 39% for 14 instruments, 76% for instrument families (respectively 35% for 16 instruments, 77% for instrument families). The increased recognition rates obtained here can be mainly explained by the use of new signal features.

APPENDIX

In Figure 5, we present the main selected features by the IRMFSP algorithm for each node of the H-GC tree.

In Figure 6, we represent the *mean confusion matrix* (expressed in percent of the sounds of the original class) for the 6 experiments of the LODO evaluation method. The lower row of the figure represents the total number of sounds used for each instrument class. Clearly visible in the matrix, is the low confusion between sustained and non-sustained sounds. The largest confusions occur inside each instrument family (viola recognized at 37% as a cello, violin at 14% as a viola and 16% as a cello, French-horn at 23% as a tuba, cornet at 47% as a trumpet, English-horn at 49% as a oboe, oboe at 20% as a clarinet). Note that the classes with the smallest recognition rate (cornet at 30% and English-horn at 12%) are also the classes for which the training set was the smallest (53 cornet sounds and 41 English-horn sounds). More surprising are the confusions inside the non-sustained sounds (piano recognized as guitar or harp, guitar recognized as cello-pizz). Cross-family confusions as the trombone recognized at 12% as a bassoon, recorder recognized at 10% as a clarinet or clarinet recognized at 23% as a flute can be explained perceptually (we have considered a large pitch range for each instrument, therefore the timbre of a single instrument can drastically change).

ACKNOWLEDGEMENT

sust./non-susts	among non-sust.	among sust.	among bowed-string	among brass	among air reeds	among sing/dbl reeds
temporal increase *temporal decrease* *temporal log-attack*	*temporal decrease* *temporal centroid*	*temporal decrease*				*temporal decrease*
spectral centroid spectral spread	spectral centroid spectral spread spectral skewness	spectral spread spectral skewness spectral kurtosis + std spectral variation spectral decrease std	spectral centroid spectral spread sharpness spectrall skewness std spectral kurtosis	spectral centroid spectral skewness spectrall kurtosis std	spectral skewness spectral kurtosis + std spectral slope spectral variation std	spectral centroid spectral spread spectral skewness
harmonic deviation		*harmonic deviation*	*tristimulus*	*noisiness*	*harmonic deviation* *tristimuls std*	*tristimulus* *harmonic deviation*
mfcc2,6 std	various mfcc		mfcc3,4,6	xcorr 3, 6, 8	xcorr3	xcorr3

Figure 5 *Main selected features by the IRMFSP algorithm for each node of the H-GC tree*

recognized \ original	piano	guitar	harp	viola-pizz	bass-pizz	cello-pizz	violin-pizz	viola	bass	cello	violin	french-horn	cornet	trombone	trumpet	tuba	flute	piccolo	recorder	bassoon	clarinet	english-horn	oboe
piano	36	3	4	4	2	2	1					1			2								
guitar	29	48	12		1	8																	
harp	24	22	68	2	3	5	2																
viola-pizz	1		6	85		1	9											4		2			
bass-pizz	1	3			76	12																	
cello-pizz	2	20	2	4	18	71				1					1								
violin-pizz	3	1		6		1	88										2	4		3			
viola								44		5	14												
bass			2						93	4	6					1	2				1	3	
cello	1							37	5	68	16	1					2				1	1	5
violin								14		3	55						1					10	1
french-horn	1										1	50		13	1	15		4	5	2			
cornet									2			1	30	3	13		2	1		2			4
trombone												15	15	49	7		1	1	3		2		2
trumpet											1		47	10	61								
tuba	2		2									23		7		79							
flute								1	2			5	2	3	4	1	77	10	10	1	23	2	4
piccolo															1		4	71	5		5		8
recorder									1								2	4	59				
bassoon	1	4	5							1		2	2	12					3	81		12	1
clarinet	1				1			2	1	7	2				4		5	1	10		46	10	20
english-horn										1	1	1							3	3	1	12	4
oboe										4	1			4		9	3	1	3	1	14	49	58
number of sounds	146	159	130	54	186	170	97	225	280	356	264	242	53	202	157	140	323	83	39	203	212	41	184

Figure 6 *Overall confusion matrix ((expressed in percent of the sounds of the original class) for the LODO evaluation method. Thin lines separate the instrument families while thick lines separate the sustained/non-sustained sounds.*

9. REFERENCES

[1] Scheirer, E. and M. Slaney. *Construction and evaluation of a robust multifeature speech/music discriminator.* in *ICASSP*. 1997. Munich, Germany.

[2] Brown, J., *Computer identification of musical instruments using pattern recognition with cepstral coefficients as features.* JASA, 1999. **105**(3): p. 1933-1941.

[3] Martin, K. and Y. Kim. *2pMU9. Instrument identification: a pattern-recognition approach.* in *136th Meet. Ac. Soc. of America.* 1998.

[4] Wold, E., et al., *Classification, search and retrieval of audio*, in *CRC Handbook of Multimedia Computing*, B. Furth, Editor. 1999, CRC Press: Boca Raton, FLA. p. 207-226.

[5] Martin, K., *Sound source recognition: a theory and computational model.* 1999, MIT: Cambridge.

[6] Eronen, A. *Comparison of features for musical instrument recognition.* in *WASPAA (IEEE Workshop on Applications of Signal Processing to Audio and Acoustics).* 2001. New York, USA.

[7] Zhang, T. and J. Kuo. *Hierarchical classification of audio data for archiving and retrieving.* in *IEEE ICASSP.* 1999. Phoenix, AR.

[8] Jensen, K. and K. Arnspang. *Binary Decision Tree Classification of Musical Sounds.* in *ICMC.* 1999. Bejing, China.

[9] Krimphoff, J., S. McAdams, and S. Windsberg, *Caractérisation du timbre des sons complexes. II: Analyses acoustiques et quantification psychophysique.* Journal de physique, 1994. **4**: p. 625-628.

[10] Peeters, G., S. McAdams, and P. Herrera. *Instrument sound description in the context of MPEG-7.* in *ICMC.* 2000. Berlin, Germany.

[11] MPEG-7, *Information Technology - Multimedia Content Description Interface - Part 4: Audio*, in *ISO/IEC JTC 1/SC 29*. 2002: ISO/IEC FDIS 15938-4:2002.

[12] Herrera, P., G. Peeters, and S. Dubnov, *Automatic Classification of Musical Instrument Sounds.* Journal of New Musical Research, 2003.

[13] Molina, L., L. Belanche, and A. Nebot. *Feature Selection Algorithms: A Survey and Experimental Evaluation.* in *International Conference on Data Mining.* Dec. 2002. Maebashi City, Japan.

[14] Hall, M., *Feature Selection for Discrete and Numeric Class Machine Learning.* 1999.

[15] Peeters, G. and X. Rodet. *Automatically selecting signal descriptors for Sound Classification.* in *ICMC.* 2002. Goteborg, Sweden.

[16] Livshin, A., G. Peeters, and X. Rodet. *Studies and Improvements in Automatic Classification of Musical Sound Samples.* in *submitted to ICMC.* 2003. Singapore.

[17] Ballet, G., *Studio OnLine.* 1998.

[18] Fritts, L., *University of Iowa Musical Instrument Samples.* 1997.

[19] Opolko, F. and J. Wapnick, *McGill University Master Samples CD-ROM for SampleCell VOLUME 1.* 1991.

[20] Microsoft, *Musical Instruments CD-ROM.*, Microsoft.

[21] E. Frank, et al., *Weka: Waikato Environment for Knowledge Analysis.* 1999-2000: Waikato.

[22] Vinet, H., P. Herrera, and F. Pachet. *The CUIDADO Project.* in *ISMIR.* 2002. Paris, France.