

# An Efficient Audio Classification Approach Based on Support Vector Machines

Lhoucine Bahatti

Electrical Engineering Department  
ENSET, Hassan II University  
Mohammedia, Morocco

Omar Bouattane

Electrical Engineering Department  
ENSET, Hassan II University  
Mohammedia, Morocco

My Elhoussine Echhibat

Mechanical Engineering Department  
ENSET, Hassan II University  
Mohammedia, Morocco

Mohamed Hicham Zaggaf

Electrical Engineering Department  
ENSET, Hassan II University  
Mohammedia, Morocco

**Abstract**—In order to achieve an audio classification aimed to identify the composer, the use of adequate and relevant features is important to improve performance especially when the classification algorithm is based on support vector machines.

As opposed to conventional approaches that often use timbral features based on a time-frequency representation of the musical signal using constant window, this paper deals with a new audio classification method which improves the features extraction according the Constant Q Transform (CQT) approach and includes original audio features related to the musical context in which the notes appear.

The enhancement done by this work is also lay on the proposal of an optimal features selection procedure which combines filter and wrapper strategies.

Experimental results show the accuracy and efficiency of the adopted approach in the binary classification as well as in the multi-class classification.

**Keywords**—Classification; features; selection; timbre; SVM; IRMFSP; RFE-SVM; CQT

## I. INTRODUCTION

Music Information Retrieval (MIR) is a growing field that benefits signal processing development and communication media tools. It uses the pattern recognition techniques to solve problems of music digital transcription including classification process. This classification is aimed to identify the artist in a given musical track.

In order to improve the performance, many researches are focused on the choice of an efficient classification algorithm, and on the use of relevant features which are able to answer questions in query. To do so, the conventional solution is to choose a set of elements that reflect the musical timbre and reduce their dimensionality by using filters method, wrappers method or embedded strategy [1][2][3]. The purpose of this work is firstly to introduce and operate a new family of so-called transition features which characterize the musical context in which the notes appear and optimize the classification algorithm. Secondly, we improve the

performances by combining features selections approaches.

The rest of this paper is organized as follows: Section 2 reviews the previous works as state of the art on the automatic audio classification algorithms and the features selection. The used classifier is presented in Section 3. Section 4 is devoted to the proposed approach which consists of extracting features and describes into details the reduction dimensionality technique. The parameters classifier and approach practical implementation results are presented in Section 5. Finally, we conclude the paper and suggest future related work in Section 6.

## II. STATE OF THE ART

To achieve an audio recognition task, different classification algorithms have been designed and tested [4][5][6][7][8]. Among several approaches, it seems that a consensus has been developed around the use of Support Vector Machines (SVM) [9][10][11][7][12][13] because of their flexibility, computational efficiency, capacity to handle high dimensional data and their profits of the feature selection [11].

The features selection method has been a central topic in a variety of fields such as pattern recognition and machine learning. The objective is to find an optimal subset of relevant and not redundant features in order to guarantee classification accuracy, computational efficiency and learning convergence [13][14][15]. In filter strategy, the features selection is performed independently of the learning classifier and the Inertia Ratio Maximization using Feature Space Projection (IRMSFP) algorithm [6][13] is the most popular due to its simplicity, rapidity and efficiency. For the wrappers approach, the well-known feature selection method combined with Support Vector Machines is the Recursive Feature Elimination (RFE-SVM) algorithm [14]. It generates the ranking of features by using backward data elimination until the highest classification accuracy is obtained. However, the RFE-SVM is a greedy method that only hopes to find the best possible combination for classification [16] and may be biased when the features are highly correlated [17].

### III. SVM CLASSIFIER

In this section, we first introduce the basic theory of the SVM for two-class classification problem, second the used strategy to solve multiclass problem and third the features selection method which is combined with Support Vector Machines.

#### A. Basic Theory of SVMs

Given a training set of instance-label pairs  $(x_i, y_i)_{i=1, \dots, m}$  where  $x_i \in \mathbb{R}^n$  and  $y_i \in \{1, -1\}^m$ . The support vector machines (SVMs)[18] are used to find a hyperplane  $W \cdot x + b = 0$  to separate the data with the maximum margin. This requires the solution of the following optimization problem:

$$\text{minimize } M = \frac{1}{2} W^T W$$

Subject to

$$y_i (W^T \phi(x_i) + b) \geq 1 \quad (1)$$

Using a soft-margin instead of a hard-margin, the primal problem for SVMs is obtained:

$$\text{minimize } \frac{1}{2} W^T W + C \sum_{i=1}^n \xi_i \quad (2)$$

Subject to

$$y_i (W^T \phi(x_i) + b) \geq 1 - \xi_i; \quad \xi_i \geq 0 \quad (3)$$

Here:

- $\{\xi_i\}$  are slack variables which allow for penalized constraint violation through the penalty function  $F(\xi)$  which defined by (4):

$$F(\xi) = \sum_{i=1}^n \xi_i \quad (4)$$

- $C$  is the parameter controlling the trade-off between a large margin and less constrained violation

- $\phi(\cdot)$  represents the mapping from the input space to the features space. However researchers prefer to use a kernel function  $K(\cdot, \cdot)$  given by the following expression:  $K(x_i, x_j) = \phi(x_i)^T \phi(x_j)$ . Practically, the most commonly used kernel functions are:

- Linear:  $K(x_i, x_j) = x_i^T x_j \quad (5)$

- Polynomial:  $K(x_i, x_j) = (\gamma x_i^T x_j + r)^d, \gamma > 0 \quad (6)$

- Radial basis function (RBF):

$$K(x_i, x_j) = \exp(-\gamma \|x_i - x_j\|^2), \gamma > 0 \quad (7)$$

- Sigmoid:  $K(x_i, x_j) = \tanh(\gamma x_i^T x_j + r) \quad (8)$

Here,  $\gamma$ ,  $r$  and  $d$  are kernel parameters, furthermore a practical use and implementation of the SVM classifier is presented in [19].

#### B. Multiclass Classification

As investigated in [20] the multiclass classification based

on SVMs is commonly performed by one of the two methods “one-vs-one” or “one-vs-all”. Both consider the multi-class problem as a collection of “two-class classification” problems. For  $k$ -class classification, “one-vs-all” method constructs  $k$  classifiers where each classifier constructs a hyperplane between one class and the rest  $(k - 1)$  classes. A majority vote or some other measures are applied over the all possible pairs for decision. For the “one-vs-one” approach,  $\frac{k(k-1)}{2}$  classifications are realized between each possible class pairs and similarly a voting scheme is applied for decision.

For more speed and reliability, Direct Acyclic Graph SVM (DAGSVM) [20][21] is often adopted. In this approach, the testing phase uses a rooted binary directed acyclic graph which has  $\frac{k(k-1)}{2}$  internal nodes and  $k$  leaves. Each node is a binary SVM of  $i^{\text{th}}$  and  $j^{\text{th}}$  classes.

Due to that hierarchical classification based on an acyclic tree structure, the DAG paradigm allows both of:

- Bringing a multiclass classification in a set of two-class classifications.
- Computing the classification process of  $k$  classes with  $(k - 1)$  comparisons instead of  $\frac{k(k-1)}{2}$  ones required for the basic “one-vs-one” approach.

#### C. Recursive Feature Elimination-SVM

The well-studied RFE-SVM is a feature selection algorithm for supervised classification which forms part of the wrapper method. It integrates filtering in the SVM learning process not only to evaluate each subset using SVM classifier but also to have information on each feature contribution in the separating hyperplane construction. The RFE-SVM is based on ranking all the features according to some score function and eliminating recursively one or more features with the lowest score.

According to [14], the RFE-SVM algorithm can be decomposed into four steps:

- 1) Train an SVM on the training set;
- 2) Order features using the weights of the resulting classifier;
- 3) Eliminate features with the smallest weight;
- 4) Repeat the process with the training set restricted to the remaining features.

### IV. PROPOSED APPROACH

As already mentioned, the classification approach is based on three steps: using new features and timbral ones, features selection by combining filter and wrapper methods and optimizing the SVM classifier. The related block diagram is illustrated in Figure 1.

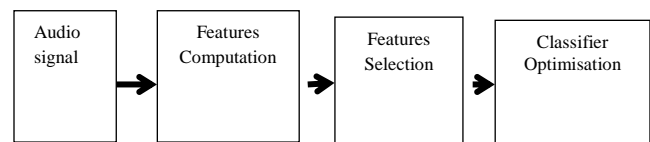


Fig. 1. Bloc diagram of our classification process

A. Features extraction

The basis of our approach focuses on the features of musical signal according to the block diagram scheme as in Figure 1. The obtained features are divided into the following two families.

- A classical family features that reflects the musical timbre and well known by the Music Information Retrieval “MIR” community [13][22][6][23][24].

- An additional and interesting family of features that reflects the transition segments between successive notes in the musical signal.

1) Features representing the musical signal timbre

The features part of this family are designed to represent the most important perceptual properties of a sound. They constitute a set of scalar parameters related to the spectral description of the musical signal. They were the subject of an extensive literature; further studies of their extraction are presented in [23]. The choice of these features, said low level, are generally depending on the desired application and on extraction duration. According to their modality of calculation as detailed in [24], they are organized as follow:

- Temporal features: convey information about the signal time evolution.

- Energetic features: features referring to various energy of the signal.

- Spectral features: those features are computed from signal time frequency representation without prior waveform model.

- Cepstral features: represent the shape of the spectrum with few coefficients using Mel-bands instead of the Fourier spectrum.

- Harmonic features: those features are computed from the detected pitch events associated with a fundamental frequency (F<sub>0</sub>).

- Perceptual features: are computed from auditory filtered bandwidth versions of signals which aim at approximating the human perception of sounds.

The presentation of these features is illustrated by the following table No. 1

TABLE I. LIST OF FEATURES RELATED TO THE MUSICAL TIMBRE

Features list	number
<b>Temporal Features</b>	
Log Attack Time	1
Temporal Increase	1
Temporal Decrease	1
Temporal Centroid	1
Effective Duration	1
Signal Auto-correlation function	12
Zero-crossing rate	1
<b>Energy Features</b>	
Total energy	1
Total energy Modulation (frequency, amplitude)	2
Total harmonic energy	1
Total noise energy	1
<b>Spectral Features</b>	
Spectral centroid	6
Spectral spread	6
Spectral skewness	6

Spectral kurtosis	6
Spectral slope	6
Spectral decrease	1
Spectral rolloff	1
Spectral v ariation	3
Spectral flatness	4
Spectral crest	4
<b>Cepstral Features</b>	
MFCC	12
Delta MFCC	12
Delta Delta MFCC	12
<b>Harmonic Features</b>	
Fundamental frequency	1
Fundamental fr. Modulation (frequency, amplitude)	2
Noisiness	1
Inharmonicity	1
Harmonic Spectral Deviation	3
Odd to Even Harmonic Ratio	3
Harmonic Tristimulus	9
HarmonicSpectral centroid	6
HarmonicSpectral spread	6
HarmonicSpectral skewness	6
HarmonicSpectral kurtosis	6
HarmonicSpectral slope	6
HarmonicSpectral decrease	1
HarmonicSpectral rolloff	1
HarmonicSpectral v ariation	3
<b>Perceptual Features</b>	
Loudness	1
RelativeSpecific Loudness	24
Sharpness	1
Spread	1
Perceptual Spectral centroid	6
Perceptual Spectral spread	6
Perceptual Spectral skewness	6
Perceptual Spectral kurtosis	6
Perceptual Spectral Slope	6
Perceptual Spectral Decrease	1
Perceptual Spectral Rolloff	1
Perceptual Spectral Variation	3
Odd to Even Band Ratio	3
Band Spectral Deviation	3
Band Tristimulus	9
<b>Total Number of Features</b>	<b>166</b>

In this work, we will extract the parameters of a signal related to the Oriental music known by its richness in melody [25] and generated by a lute. That signal is therefore relatively short, non-stationary and assumed to contain an almost percussive sound.

Spectral features, Cepstral features, Harmonic features, and Perceptual features are computed based on a Short Time Fourier Transform which is expressed according to (9):

$$S(t, f_K) = \int_{-\infty}^{+\infty} s(\tau - t)w(\tau)exp(-j2\pi \cdot f_K \cdot \tau) d\tau \quad (9)$$

For best time-frequency localization, using a window w(t) with a variable length is more efficient [26]. Features extraction is then based on splitting the audio signal s(t) into successive frames where each frame s<sub>k</sub>(t) of index k and duration T<sub>k</sub> is calculated by (10):

$$s_K(t) = s(t) \cdot w_K(t) \quad (10)$$

w<sub>K</sub>(t) is the Hann window expressed by (11) :

$$W_K(t) = \begin{cases} 0.5 + 0.5 \cos\left(2\pi \frac{t}{T_K}\right); & \text{if } \frac{T_K}{2} \leq t \leq \frac{T_K}{2} \\ 0 & \text{elsewhere} \end{cases} \quad (11)$$

According to the Constant Q Transform (CQT) approach applied in the oriental music context [26], the coefficient  $Q=37$  and the duration  $T_K$  is given by (12)

$$T_K = \frac{Q}{f_K} \quad (12)$$

2) Transition features between notes

The majority of classical timbre features is restricted only to the spectral characteristics of short duration without modeling a music temporal aspect. Thus, in order to describe better a melody that is made for monophonic recordings, we propose, in this section, to use additional most representative features of musical timbre, taking into account the pattern musical note and the musical context in which the note appears.

So, when examining the envelopes of lute's notes, we see that they are closest to a well-known pattern modeled by the classical "ADSR" model [25] (Figure 2).

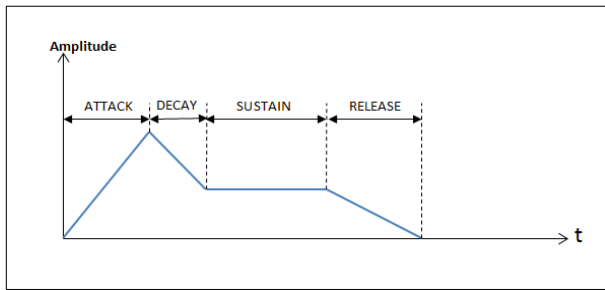


Fig. 2. Model ADSR of the note's envelope

The melody of a musical signal can only be achieved if the notes follow each other according to the sequence required by the musical score. In such a succession, energy  $E_{XX}$  of the resulting signal can be represented according to the ADSR model of Figure 3 as follows.

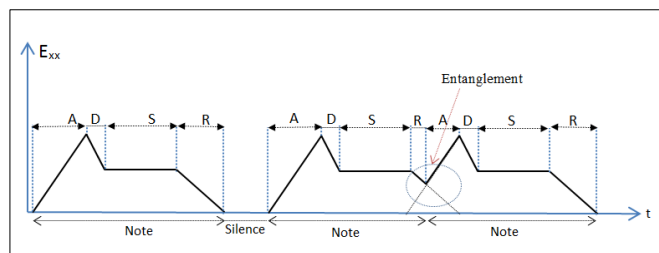


Fig. 3. Schematic view of the intra-note segmentation energy

Certainly, when two consecutive notes are separated by a "silence", the problem of intra-note correlation hardly arises, unlike the majority of cases where they have a temporal entanglement.

To characterize this phenomenon, we introduce a segment called "Transition" which starts at the beginning of the first note's Release phase and finishes at the end of Attack of the following one. We represent this articulation in Figure 4 while showing the evolution of the energy and fundamental frequency ( $F_0$ ) during this new segment. We denote:

- $T_c$ : Instant related to the energy envelope minimum.

- $T_{init}$ : Instant of first note's Release phase beginning.
  - $E_{LT}$ : Energy of the first note at the instant  $T_{init}$
  - $T_{end}$ : Instant of the following note Attack end
  - $E_{RT}$ : Energy of the following note at the instant  $T_{end}$
- Features associated with the transition between these types of notes are:

- The duration:  $T_D$  (13),
- The energy change during the interval  $T_D$  (14)

$$T_D = T_{end} - T_{init} \quad (13)$$

$$L_t = \frac{E_{RT} - E_{LT}}{T_D} \quad (14)$$

We also use two features introduced and investigated in [27] namely:

- The ratio between the instant  $T_c$  and the transition duration  $T_D$  (15). This parameter has proven useful when reconstructing the amplitude envelope's during transitions

$$E_{TPOS} = \frac{T_c}{T_D} \quad (15)$$

- A feature representing the legato which indicates how to link music notes together. This descriptor, whose relevance was assessed in [28], is described in (16).

$$LEG = \frac{A_1}{A_1 + A_2} = \frac{\int_{t_{init}}^{t_{end}} (L_t(t) - E_{XX}(t)) dt}{\int_{t_{init}}^{t_{end}} L_t(t) dt} \quad (16)$$

To compute LEG, we use the schematic view of figure 4. First, we join start and end points on the energy envelope contour using a line  $L_t$  which represents smoothest case of detachment. Then we calculate both the area  $A_2$  below envelope energy and the area  $A_1$  between the energy envelope and the joining line  $L_t$ . Our legato feature is finally defined as determined by (16).

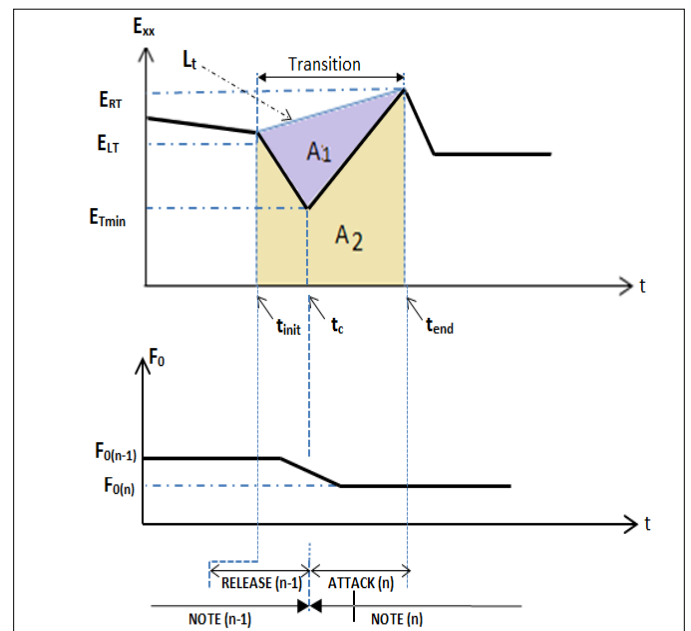


Fig. 4. Schematic view of the characterization of the "Transition" segment

### B. Features selection

Since all calculated features are not relevant for the classification task, they must be previously processed by reducing dimensionality in order to keep only relevant candidates and therefore to facilitate the classification task. When the filter is adopted, the solution is very simple and fast, but features are selected based only on their intrinsic characteristics and regardless of the used classifier. In the wrapper method, the system is accurate but without any guarantee of rapid learning and features redundancy. Taking into account this fact, the adopted approach consists of combining the simplicity and the accuracy by associating both filter and wrapper methods as mentioned in the Figure 5.

In this topology, original features are firstly filtered in order to eliminate the redundant ones and the outliers. Due to this operation which selects variables as a pre-processing step, the obtained features are less correlated and a significant reduction is already performed. This will allow a better understanding of the contained information in this subset of selected features. The computing duration will be therefore short for the wrapper which selects only features that improve the prediction accuracy and optimize the classification performance.

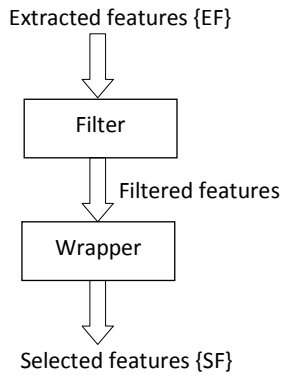


Fig. 5. Schematic view of features selection approach

To perform the filter, we choose a structure that simultaneously meets two fundamental criteria: Criterion A: Choosing a subset of informative features within each class. Criterion B: Selecting non-redundant and uncorrelated features.

For this, we adopt an algorithm based on linear discrimination analysis (LDA) strategy, called Inertia Ratio Maximization with Feature Space Projection (IRMFSP) [6][13]. This simple and efficient filter whose relevance was assessed in [13], selects features to satisfy iteratively criterion A (Inertia Ratio Maximization) and criterion B (Feature Space Projection). The implementation of IRMFSP is composed of two steps. The first one selects at iteration 1, the non-previously selected feature which maximizes the ratio between inter-class inertia and the total inertia expressed as follow (17):

$$\hat{d}^{(l)} = \arg \max_d \frac{\sum_{k=1}^K n_k (\mu_{d,k} - \mu_d)(\mu_{d,k} - \mu_d)^T}{\sum_{i=1}^n (f_{d,i}^{(l)} - \mu_d)(f_{d,i}^{(l)} - \mu_d)^T} \quad (17)$$

Where n is the total number of features,  $n_k$  is the number of features belonging to class k,  $f_{d,i}^{(l)}$  denotes the value of feature of

index d affected to the vector i.  $\mu_{d,k}$  and  $\mu$  respectively denote the average value of feature d into the class k and for the total dataset. The second step of this algorithm aims at orthogonalizing the remaining feature for the next iteration as follows (18):

$$f_{d,i}^{(l+1)} = f_{d,i}^{(l)} - (f_{d,i}^{(l)} \cdot g_{\hat{d}}) g_{\hat{d}} \quad \forall d \neq \hat{d}^{(l)} \quad (18)$$

Where  $f_{\hat{d}}^{(l)}$  is the vector of the previously selected feature  $\hat{d}^{(l)}$  and  $g_{\hat{d}} = \frac{f_{\hat{d}}^{(l)}}{\|f_{\hat{d}}^{(l)}\|}$  is its normalized form.

Due to that filtering operation, the RFE-SVM wrapper will be able to measure rapidly and accurately the impact of each selected feature on the classification procedure.

## V. PROCESS IMPLEMENTATION

### A. Materials

In this section, we are interested in the real tracks resulting from oriental music. Various artists and associated musical signals are shown in the following Table 2.

TABLE II. DATASET OF ARTISTS AND ASSOCIATED MUSICAL SIGNALS

Artist: Code	Learning database	Test database
<b>Badr El Ouarzazi : BO</b>	2 Tracks: BO1 : Duration : 4mn 42 s BO2 : Duration : 4 mn 52 s	8 Tracks : BO3, BO4,...BO10 : Duration: 5 mn
<b>Farid El Attrach : FE</b>	2 Tracks: FE1 : Duration: 5mn 30s FE2 : Duration: 4mn 48 s	8 Tracks : FE3, FE4,..... FE10 : Duration: 6 mn
<b>Nasser Chamma : NC</b>	2Tracks: NC1 : Duration:10 mn 30s NC2 : Duration: 7 mn 24 s	8 Tracks: NC3, NC4,...NC10 : Duration:10 mn
<b>Nouamane Lahlou : NL</b>	2Tracks: NL1 : Duration: 3mn 39 s NL2 : Duration :2mn 45s	2 Tracks: NL3: Duration : 4mn 31 s NL4: Duration : 3mn
<b>Said Chraïbi : SC</b>	2 Tracks: SC1 : Duration: 5mn 10s SC2 : Duration: 5mn 24 s	8 Tracks : SC3, SC4,..... SC10 : Duration: 7 mn

As justified in [29], the most appropriate criteria function to evaluate the efficiency and accuracy of such classification process is to use the F-Measure indicator (FMSR), which is the harmonic mean of the recall (RCL) and precision (PRC). FMSR is given by (19):

$$FMSR = \frac{2 \cdot RCL \cdot PRC}{RCL + PRC} \quad (19)$$

### B. Classifier Optimisation:

This section aims to set the classifier parameters based on binary classification. Beyond the fundamental principle of parsimony research, the SVM approach leaves, in practice, a number of options and settings to the user such as: the choice of the regularization parameter, and the choice of kernel type.

#### 1) The kernel function:

The adopted kernel is the Exponential Radial Basis Function (ERBF) as it represents the best way to follow the non-linear decision surfaces.

For greatest robustness, instead of (7), we use a structure which takes into account the number of learning elements[18].The kernel function expression is given by (20).

$$k(x, y) = \exp\left(-\frac{\|x - y\|^2}{m\sigma^2}\right) \quad (20)$$

- $m$  is the dimension of the observation vectors:
- $\sigma$  represents the width of the Gaussian function. It is a main parameter that affects the complexity of decision surfaces. Interesting choices are situated in the interval  $[0, 1]$ . The default value (before refining parameter) is  $\sigma = 1$ . Optimize the classifier involves determination of that parameter in order to maximize the FMSR performance on the training dataset. The obtained result is shown in Figure 6.

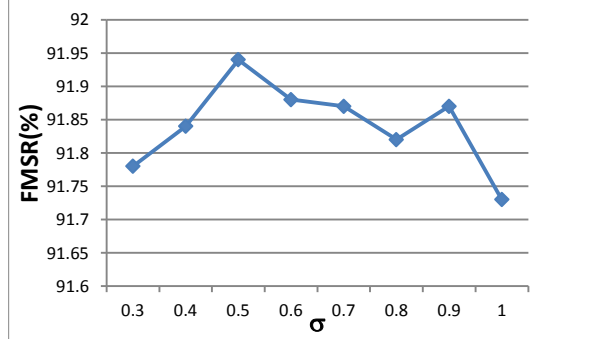


Fig. 6. Setting of SVM classifier

It can be seen that the overall performance shows a peak for  $\sigma=0.5$  ( $F=91.94\%$ ). Nevertheless, the performance variations are very weak (about 0.24%) this justifies the robustness of the SVM classifier.

### 2) Controlling parameter

As mentioned in section III, this is a factor that controls the tradeoff between maximizing the margin of class's separation and minimizing of classification errors on the training set. It is a balancing parameter to set a priori, in order to make floppy the margin's SVM. Certainly, there is no strict relation to the exact calculation of  $C$  value. However the best practical results are usually obtained using an adaptive value " $C_{dat}$ " of that penalty parameter based on the number of " $m$ " learning elements [18]. Thus,  $C_{dat}$  is obtained according to (21) wherein the kernel function  $k(x_i, x_i)$  is defined by (20).

$$C_{dat} = \frac{1}{\frac{1}{m} \sum_{i=1}^m k(x_i, x_i)} \quad (21)$$

### 3) Performance of features selection approaches:

After the SVM optimization, we evaluate the efficiency of used features selection algorithms by comparing them according to the desired learning elements size.

The first comparison is about the complexity criterion. In this context, we set 50 optimal features and measure the processing duration for each features selection algorithm. The obtained result is presented in Table 3 below and shown clearly that the IRMFSP filter is faster than RFE-SVM wrapper. When

two algorithms are used, the IRMFSP ensures well its role of preprocessing step which reduces the learning duration.

TABLE III. COMPLEXITY OF FEATURES SELECTION ALGORITHMS

Selection Features algorithm	Learning duration
IRMFSP	1,21 s
RFE-SVM	6,18 s
IRMFSP + RFE-SVM	2,34 s

The second comparison is made by F-measure criterion. Figure 7 compares that classification performance obtained as a function of the used features number.

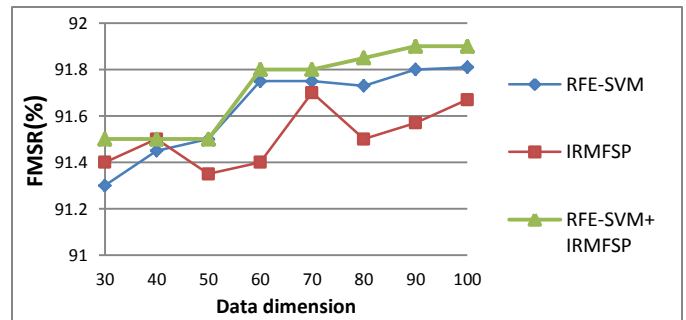


Fig. 7. Performance of the classification based on the features number

The performance of the IRMFSP filter is almost constant. The efficiency and robustness of the RFE-SVM wrapper alone or combination (RFE-SVM+ IRMFSP) are noticeable especially when the number of features is relatively high (>50).

Overall, both algorithms are less sensitive to the selected features number reduction and they perform a more reliable ranking of the most useful features by positioning those most effective at the forefront.

According to the results from Tables 6 and 7, a direct comparison between those two selection approaches (RFE-SVM vs IRMFSP) proves that obtained results are better when the two algorithms are combined regardless of filter simplicity and its efficiency.

### 4) Multiclass classification:

To evaluate this classification process in the multiclass context, we aim to highlight the effectiveness of the used features and their handling. The obtained results are presented as confusion matrix. (Table 4)

TABLE IV. CONFUSION MATRIX (EXPRESSED IN PERCENT) OF MULTI-CLASS CLASSIFICATION

SVM classifier : « ERBF Kernel », $\sigma = 0.5$ and $C = C_{dat}$		real classes				
		Artist 1	Artist 2	Artist 3	Artist 4	Artist 5
estimated classes	Artist1 (BO)	76.25	4.6	16.9	0.87	1.39
	Artist2 (FE)	3.45	68.70	6.32	9,11	12.40
	Artist3 (NC)	17.58	5.12	74.94	2.02	0.43
	Artist4 (NL)	8.22	2.43	3.65	78.62	7.15
	Artist5 (SC)	1.22	3.28	4.81	5.20	85.50
		Overall average rate of recognition = 76.88%				

The resulting confusion matrix of dataset using 50 audio descriptors is presented in Table 4 and shows an average classification accuracy of 76,88 % where each artist is well classified with a minimal accuracy of 68,7% for the Artist N°2. These results are good and somewhat better than those described in literature [6][13] which uses only the timbral features without optimizing the SVM classifier.

## VI. CONCLUSION AND FUTURE WORKS

In this paper, we have proposed a reinforcing audio recognition method by improving the extraction of well-known timbre features and taking into account features that reflect the transition between musical notes. We have also developed a practical method of the SVM classifier optimization as well as a features selection method that benefits from both of filters and wrappers advantages. In this approach, the filter has achieved a simple and straightforward features selection in order to get a subset of relevant factors most suitable to interpret and to deal with the Wrapper. Although the obtained results are interesting and encouraging, some aspects may be developed in future works. So, as perspective, we intend to investigate the use of other features kind and explore other features selection algorithms in such classification process.

### REFERENCES

- [1] H. Liu and L. Yu, "Toward integrating feature selection algorithms for classification and clustering", *IEEE Transactions on Knowledge and Data Engineering*, vol. 17, no. 4, (2005), pp. 491–502. doi>10.1109/TKDE.2005.66
- [2] Verónica Bolón-Canedo · Noelia Sánchez-Marroño · Amparo Alonso-Betanzos. A review of feature selection methods on synthetic data. *Knowledge and Information Systems*; Spring, March 2013, Volume 34, Issue 3, pp 483-519; ; DOI 10.1007/s10115-012-0487-8
- [3] Tushar Ratanpara, Narendra Patel. Singer identification using perceptual features and cepstral coefficients of an audio signal from Indian video songs. *EURASIP Journal on Audio, Speech, and Music Processing* 2015, 2015:16. doi:10.1186/s13636-015-0062-9
- [4] Perfecto Herrera, Geoffroy Peeters, and Shlomo Dubnov. Automatic classification of musical sound. *Journal of New Music Research*, 32:1 pages 2–21, 2003. <http://dx.doi.org/10.1076/jnmr.32.1.3.16798>
- [5] Kaminsky, I., and A. Materka. "Automatic source identification of monophonic musical instrument sounds." *Neural Networks*, 1995. *Proceedings, IEEE International Conference on*. Vol. 1. IEEE, 1995. DOI: 10.1109/ICNN.1995.488091
- [6] G. Peeters. Automatic classification of large musical instrument databases using hierarchical classifiers with inertia ratio maximization. In *Proc. of AES 115th Convention*, New York, USA, 2003. [http://recherche.ircam.fr/anasyn/peeters/ARTICLES/Peeters\\_2003\\_AES\\_SoundClassification.pdf](http://recherche.ircam.fr/anasyn/peeters/ARTICLES/Peeters_2003_AES_SoundClassification.pdf)
- [7] Dhanalakshmi, P. et al., Classification of audio signals using SVM and RBFNN, *Expert Systems with Applications* (2008); doi:10.1016/j.eswa.2008.06.126
- [8] Guodong Guo and Stan Z. Li. Content-Based Audio Classification and Retrieval by Support Vector Machines. *IEEE transactions on neural networks*, vol. 14, no. 1, January 2003. 209
- [9] Guodong Guo and Stan z. Li. content-based audio classification and retrieval by support vector machines *IEEE transactions on neural networks*, vol. 14, no. 1, January 2003 209. <http://www.ee.columbia.edu/~sfchang/course/spr/papers/guo-li-svm-audio00.pdf>
- [10] Qi, Zhiquan, Yingjie Tian, and Yong Shi. "Robust twin support vector machine for pattern classification." *Pattern Recognition* 46.1 (2013): 305-316. doi:10.1016/j.patcog.2012.06.019
- [11] Moraes, R., et al. Document-level sentiment classification: An empirical comparison between SVM and ANN. *Expert Systems with Applications* (2012), <http://dx.doi.org/10.1016/j.eswa.2012.07.059>
- [12] Rocamora, M., & Herrera, P. (2007, September). Comparing audio descriptors for singing voice detection in music audio files. In *Brazilian Symposium on Computer Music*, 11<sup>th</sup>. San Pablo, Brazil (Vol. 26, p. 27).
- [13] Dominique Fourer, Jean-Luc Rouas, Pierre Hanna, Matthias Robine. automatic timbre classification of ethnomusicological audio recordings. *International Society for Music Information Retrieval Conference (ISMIR 2014)*, Oct 2014, Taipei, Taiwan. 2014.
- [14] Pember A. Mundra and J. C. Rajapakse, "SVM-RFE with relevancy and redundancy criteria for gene selection," in *PRIB*, J. C. Rajapakse, B. Schmidt, and L. G. Volkert, Eds., vol. 4774, Springer, (2007), pp. 242–252.
- [15] Y. Tang, Y.-Q. Zhang and Z. Huang, "Development of two-stage SVM-RFE gene selection strategy for microarray expression data analysis", *IEEE/ACM Trans. Comput. Biology Bioinform*, vol. 4, no. 3, (2007), pp. 365–381.
- [16] Mouhamadou Lamine Samb ; Fodé Camara ; Samba Ndiaye; Yahya Slimani and Mohamed Amir Esseghir; A Novel RFE-SVM-based Feature Selection Approach for Classification ; *International Journal of Advanced Science and Technology*; Vol. 43, June, 2012.
- [17] Ke Yan; David Zhang. Feature selection and analysis on correlated gas sensor data with recursive feature elimination; Elsevier. *Sensors and Actuators B* 212(2015); 353–363; DOI: 10.1016/j.snb.2015.02.025
- [18] T.Joachims. Svm light support vector machine. <http://svmlight.joachims.org>
- [19] Chang, Chih-Chung, and Chih-Jen Lin. "LIBSVM: A library for support vector machines." *ACM Transactions on Intelligent Systems and Technology (TIST)* 2.3 (2011); 27. doi>10.1145/1961189.1961199
- [20] Chih-Wei Hsu and Chih-Jen Lin. A Comparison of Methods for Multiclass Support Vector Machines. *IEEE TRANSACTIONS ON NEURAL NETWORKS*, VOL. 13, NO. 2, MARCH 2002
- [21] Mostafa Sabzekar, Mohammad GhasemiGol, Mahmoud Naghibzadeh, Hadi Sadoghi Yazdi . "Improved DAG SVM: A New Method for Multi-Class SVM Classification". *International Conference on Artificial Intelligence ; ICAI'09*. July 13-16, 2009, Las Vegas Nevada, USA.
- [22] Guodong Guo and Stan Z. Li. Content-Based Audio Classification and Retrieval by Support Vector Machines. *IEEE transactions on neural networks*, vol. 14, no. 1, January 2003. 209
- [23] Richard, G.; Sundaram, S.; Narayanan, S., "An Overview on Perceptually Motivated Audio Indexing and Classification," *Proceedings of the IEEE*, vol.101, no.9, pp.1939, 1954, Sept. 2013. DOI: 10.1109/JPROC.2013.2251591
- [24] G. Peeters A large set of audio features for sound description (similarity and classification) in the CUIDADO project. Technical Report 2004. [http://recherche.ircam.fr/anasyn/peeters/ARTICLES/Peeters\\_2003\\_cuidadoaudiofeatures.pdf](http://recherche.ircam.fr/anasyn/peeters/ARTICLES/Peeters_2003_cuidadoaudiofeatures.pdf)
- [25] Lhoucine Bahatti, Omar Bouattane, Mimoun Zazoui and Ahmed Rebbani: "Fast Algorithm for In situ transcription of musical signals: Case of lute music". *IJCSI International Journal of Computer Science Issues*, Vol.10, Issue 1, No1, January 2013. Page 444. <http://ijcsi.org/papers/IJCSI-10-1-1-444-452.pdf>
- [26] L. Bahatti, M. Zazoui, O. Bouattane, and A. Rebbani: "Short-Term Sinusoidal Modeling of an Oriental Music Signal by Using CQT Transform". *Journal of Signal and Information Processing*, 2013, 4, 51-56. doi:10.4236/jsip.2013.41006
- [27] Ramirez Rafael, Esteban Maestre, and Xavier Serra. "Automatic performer identification in commercial monophonic jazz performances." *Pattern Recognition Letters* 31.12 (2010): 1514-1523. doi:10.1016/j.patrec.2009.12.032
- [28] Maestre, E., Gomez, E., 2005. Automatic characterization of dynamics and articulation of monophonic expressive recordings. In: *Proc. 118th AES Convention*, Barcelona, Spain.
- [29] Juan José Burred and Geoffroy Peeters. An Adaptive System for Music classification and Tagging. 3rd International Workshop on Learning Semantics of Audio Signals. *Proceeding* pages 3-16. December 2, 2009. SAMT Conference - Graz, Austria.