14th European Signal Processing Conference (EUSIPCO 2006), Florence, Italy, September 4-8, 2006, copyright by EURASIP

AN IDENTIFICATION SYSTEM OF MONOPHONIC MUSICAL TONES

Vincenzo Giovanni Di Salvo, Graziano Bertini

ISTI – CNR
Via G. Moruzzi, 1 – 56124 PISA, ITALY
phone: + 39.050.3153144, fax: + 39.050.3152810,
email: vincenzo.disalvo@isti.cnr.it, g.bertini@isti.cnr.it
web: www.isti.cnr.it

ABSTRACT

The present paper describes the problem of defining a method recognizing musical tones in the context of automatic identification systems. This study had arisen from the demand to have a tool to detect a musical tone among available alternatives of a library of previously recorded musical tones produced by instruments of the same class[1]. This work is based on a comparison criterion to measure the distance between the features of the actual tone and the reference database ones.
The algorithm is realized in two distinct steps. At first, digital processing techniques are used with the purpose to obtain features pattern vectors from the waveforms. These resulting patterns are, subsequently, elaborated using the Least Squares Optimal Filtering. The algorithm is relatively simple and may be implemented efficiently with low latency on DSP processors.

1. INTRODUCTION.

It is well known that sounds of a musical instrument are listened and identified on the basis of musical characteristic attributes, like loudness, pitch, timbre. The task to build a mathematical model, deciding which physical parameters are suitable to be identified by an automatic recogniser and how to combine them is not straightforward. It is important to estimate attributes that reflect unique characteristics of musical signal, that are more measurable and easy to identify by an automatic identification machine. This cognitive task is well performed by high-skilled human musicians.
The vast production of musical instruments shows many salient acoustic features of musical sound.
Identification of a musical tone requires accurate spectral estimation for the extraction of the fundamental frequency present in the song under analysis. In this article a solution regarding identification of monophonic stream from a database is proposed, alternative to several other solutions [1,2].

[1] *A preliminary investigation of the problem was developed in cooperation between the Norwegian University of Science and Technology of Trondheim and the ISTI-CNR of Pisa, in the framework of a stage at NTNU within the EU-RTN Project "Mosart" (2003).*

Some constraints on the music material, like the type of instruments or musical styles, are imposed.
The method consists of two main parts. First, extraction of energy-dependent features of the waveform corresponding to the musical tones is performed, using the Short-Time Fourier Transform (STFT).
Second, based on them, the identification of the actual musical tone follows.
Precisely, Amplitude Variations and Component Frequency Shifting of the tones has been take into account to define the model. This strategy is near to the fact that the human ear is sensitive to these parameters changes.

2. SYSTEM OVERWIEV

In this stage we present a brief summary of the proposed system, and it is shown in fig. 1.

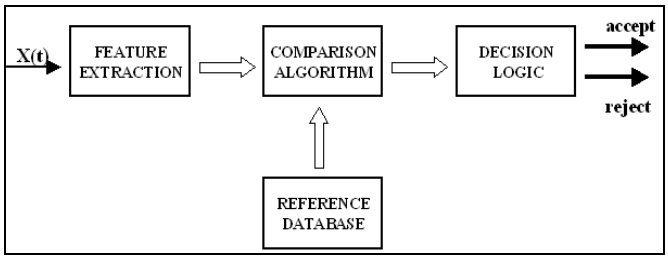


Figure 1 – Layout of the sound recognition system

First, the signal is pre-processed. This operation, not described here, preserves features relevant for tones identity. Subsequently, the vector of features extracted from the actual signal is compared with the available alternatives of the library. The comparison is made performing a measure of the distance, based on the LMS Algorithm, devoted to determine the best matching between actual tones and those of the reference database. The decision logic block, based on an established correlation coefficient treshold, gives (or not) the identification number of the corresponding database tone.
The solution proposed in this article has been developed considering the tones of the wind instruments. Figure 2 shows the sequence of three flute tones used [3]. In Appendix, figure (a) and (b) clarino and oboe single tones are shown.

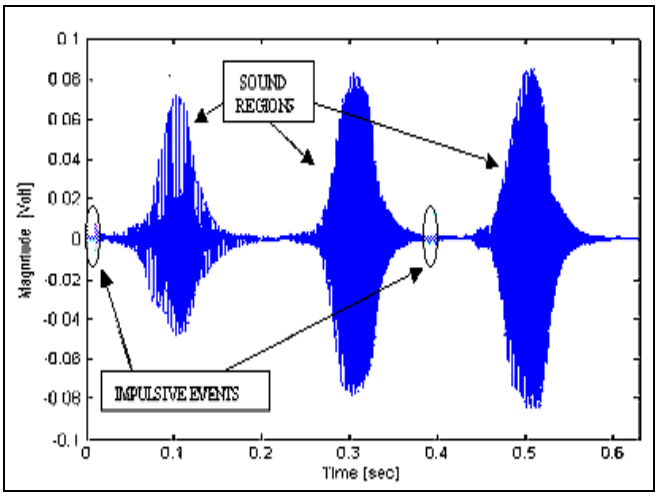


Figure 2 – Flute tones signal

Feature Extraction Block.
As well known musical signals s(n) are generally represented by a non stationary waveforms in the time domain. In the frequency domain the partials behaviour of these signals can be analyzed by the STFT, given by [4,5]

$$\mathbf{X}(f, t_i) = \sum_{n=-\infty}^{\infty} s(n, t_i)\, \exp^{-j2\pi fn}$$

where

$$s(n, t_i) = s(n)\, w(n, t_i)$$

$w(n, t_i)$ represents a weighting temporal window centred at $t = t_i$ and $s(n, t_i)$ the windowed sound segments, properly overlapped. Short window means good time resolution, longe one means good frequency resolution. Well defining them enable us to achieve the objectives in the best possible way.
A peculiarity of the STFT is that the Square Magnitude, given by

$$\boldsymbol{\Psi}(f, t_i) = |\mathbf{X}(f, t_i)|^2.$$

It can be thought as a two dimensional (2-D) energy-density. In some application (mechanical fields, vibrational fields, etc.) a three dimensional representation of the previous function is used, known as waterfall. In audio applications it is usually draws a two dimensional plot, known as spectrogram, specifying the energetic levels $\boldsymbol{\Psi}$ by a color palette.

Comparison Algorithm Block.
This block makes a comparison, expressed as least mean square error, between the features of the actual tone and, in turn, all the reference ones. To do this the optimal filtering theory, from a slightly different point of view, has been considered [4]. The problem is here regarded as purely deterministic. No prior knowledge of the statistical properties of the signals beforehand is assumed.
In general it is assumed that given an observation *x(n)* of an unknown sequence *s(n)*, a causal FIR filter of length P can be used to estimate *s(n)*.

Then

$$\hat{s}(n) = \sum_{k=0}^{P-1} h(k)\, x(n-k) \qquad (1)$$

where h(k), k = 0, …, P-1, are the filter coefficients to be calculated. P is the order of the filter and, the difference

$$\varepsilon(n) = s(n) - \hat{s}(n). \qquad (2)$$

defines the error between the signals.
The approach here is to design the filter to minimize the sum of squared error

$$\boldsymbol{S} = \sum_{[nI,\, nF]} |\varepsilon(n)|^2 \qquad (3)$$

where n_I and n_F are some initial and final values of *n* over which the minimization is performed.
The above approach is called Least Squares Criterion and could be regarded as purely deterministic [6].
Differentiating (3) one finds the taps of the FIR filter:

$$\mathbf{h} = \mathbf{Z}^{+}\mathbf{s} \qquad (4)$$

where

$$\mathbf{Z}^{+} = (\mathbf{Z}^{T}\mathbf{Z})^{-1}\mathbf{Z}^{T} \qquad (5)$$

is the observation matrix, known as the *Moore-Penrose Pseudoinverse* matrix.

Decision Logic Block
The decision logic block investigates the error signals ε(n) previously computed and extracts, from the database, the signal corresponding to the minimum distance.

3. METHOD DETAILS.

Features extractions
With respect the canonical features considered in the field of the recognition and identification of the audio signals, here the mathematical model adopted is based on two quantities: *Amplitude Variation* and *Component Frequency Shifting* (fig. 3) choosen to find attributes that reflect unique

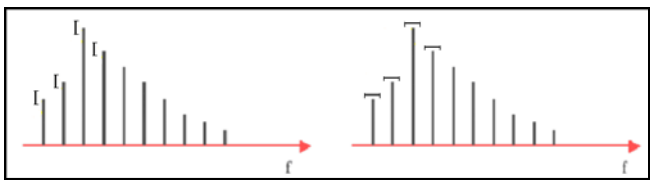


Figure 3 – *Amplitude Variation* and *Component Frequency Shifting*.

characteristics of the musical tones.
The *Amplitude Variation* refers to the energy variations of each component of the stochastic process obtained with the STFT.
The *Component Frequency Shifting* refers to the frequency deviation that occurs along the rows of the spectrogram matrix obtained with the STFT.
To measure the *Amplitude Variation* of each component, a good choice seemed to consider the energy values along the row's elements in the spectrographic matrix (fig. 4).

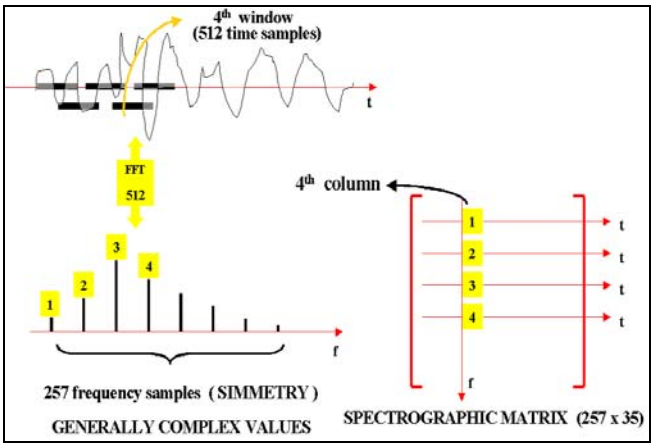


Figure 4 – Specgram matrix of Loudness Stability

The Component Frequency Shifting can be calculated using the phase differences between two consecutive column elements (fig. 5) with the formula:

$$\Delta f = \Delta\Theta / 2\pi\Delta n$$

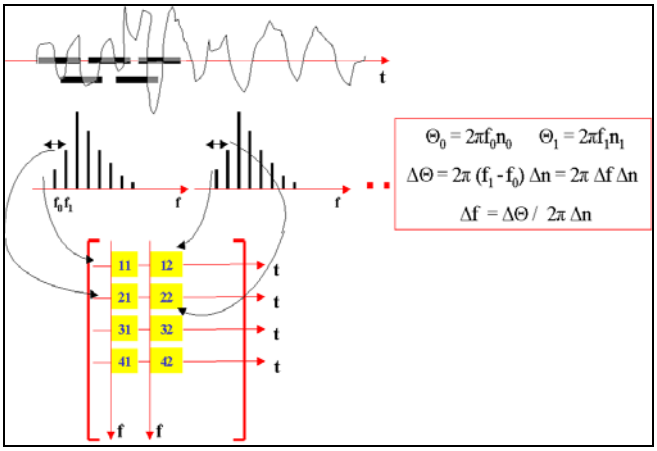


Figure 5 – Specgram matrix of Frequency Tone Stability

The above defined matrixes are well represented for flute instrument by the spectrogram shown in figure 6. In Appendix are shown spectrogram for clarino and oboe.

Identification System
Giving a look at the spectrograms one can observe how the signals are quite stationary. In the light of this the taps of FIR filter has been updated using the LMS algorithm in the feedback chain.

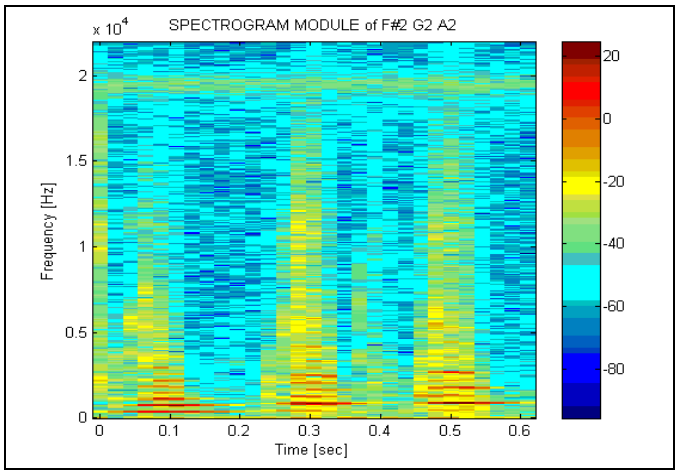


Figure 6 – Spectrogram of flute tones

This algorithm, in fact, assures a good convegence degree[2] if the signal is slowly changing in time, that is eigenvalues of the signal autocorrelation matrix are similar [7]. The core of the identification sistem is shown in fig.7.

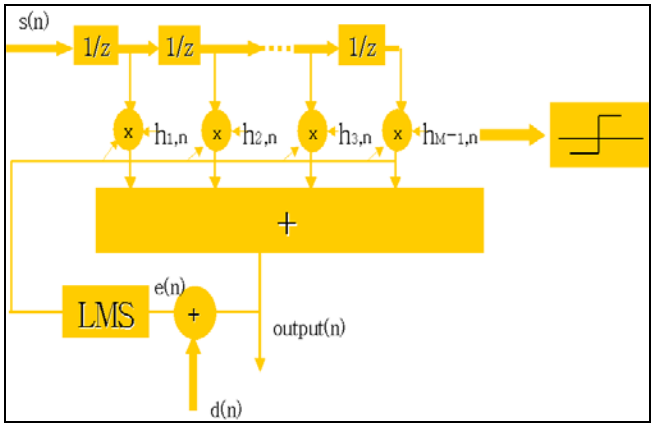


Figure 7 – Layout of the sound recognition system

As one can see the actual tone features s(n), average value of the database signals d(n), inputs to the FIR filter which gives the minimum error signal ε(n) vs database features realization d(n)

$$\varepsilon(n) = d(n) - \hat{s}(n)$$

The LMS algorithm block assures the minimization of the ε(n).
A subsequent decision block (fig. 8) implemented with a simple threshold discriminator makes a choice among the available alternatives [8].

[2] *The measure of the degree of convergence of the LMS algorithm consists in verifying that the eigenvalues of the autocorrelation matrix are similar.*

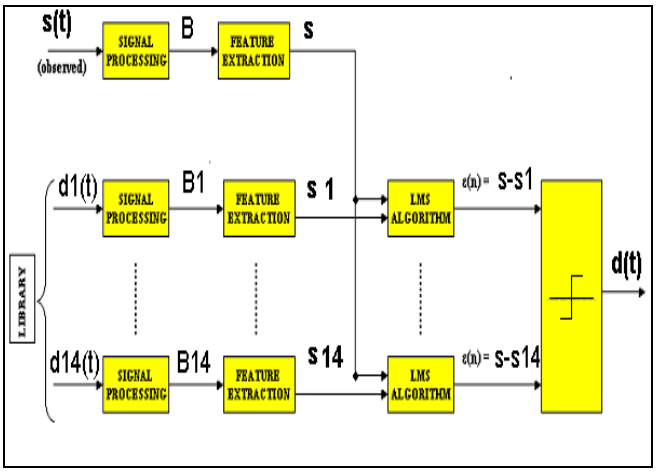


Figure 8 - Proposed sound recognition system

4. TEST AND FUTURE WORKS

The system described in this paper has been implemented in Matlab. FFTs of 1024 sample has been computed using Hamming window with an overlap of 80% .

For purpose of comparison the test has been conducted on single musical tones played by flute, clarino and oboe. The database containing 16 files single tone has been realized recording 16 single tone of instruments of the same class. Positive results has been reached with a correlation coefficient of 95% .

To improve the method developed in this work extensive experiments should be done with other classes of instruments.

A usefull practical application of such a system could be in the educational field of musical instruments training.

REFERENCES

[1] J. C. Brown "Feature dependence in the automatic identification of musical woodwind instruments" *J. Acoustical Soc. America (JASA)*, Vol. 109 (3), pp. 1064-1071, March 2001.

[2] R. P. Paiva, T. Mendes and A. Cardos "On the definition of musical notes from pitch tracks for melody detection in polyphonic recordings" in *Proc 8th Conference on Digital Audio Effects (DAFx-05)* Madrid, Spain, Sep 2005, pp. 208-213.

[3] Sølvi Ystad "Identification and Modeling of a Flute Source Signal", in *Proc 2nd COST G-6 Workshop on Digital Audio Effects (DAFx99)*, Trondheim, Dec 1999.

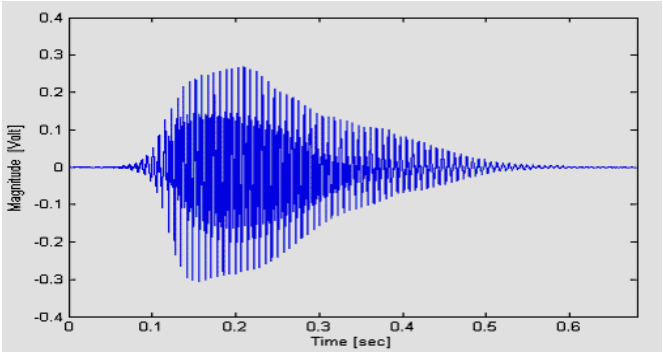
[4] Quatieri, *Discrete Time Speech Signal Processing*. Prentice-Hall, 2000

[5] Graziano Bertini, Massimo Magrini, "Spectral Data Management Tools for Additive Synthesis" in *Proc. of the COST G-6 Conference on Digital Audio Effects (DAFX-00),* pp. 195-200, Verona, Italy, Dec 7-9, 2000.

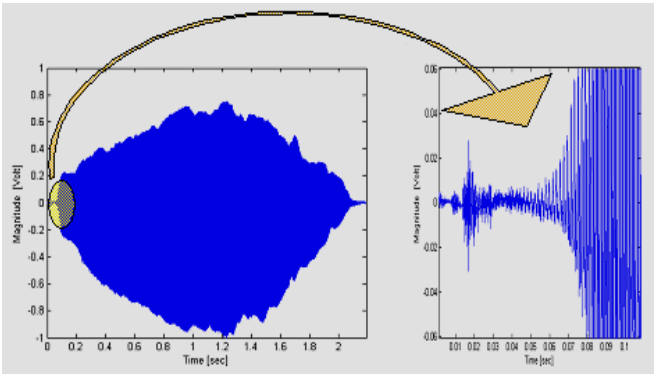
[6] C.W. Therrien, "The Lee-Wiener Legacy. A History of the Statistical Theory of Communication". *IEEE Signal Processing Magazine*. November 2002.

[7] L.R.Rabiner and R.W.Schafer, *Digital Processing of Speech Signal*, Prentice-Hall

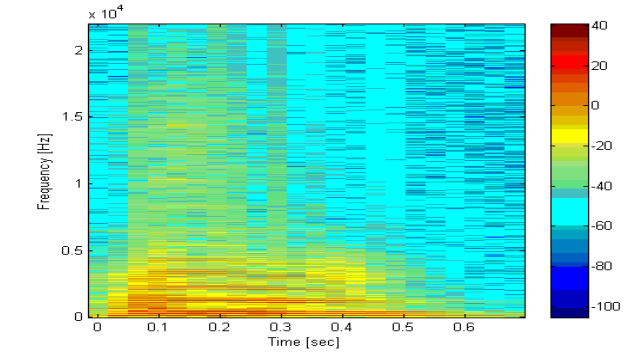
APPENDIX



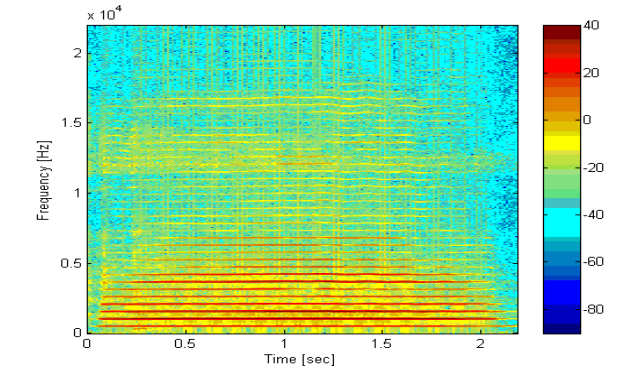
(a) – clarino tone



(b) – oboe tone



(a') – spectrogram of clarino



(b') – spectrogram of oboe