International Journal of Applied Information Systems (IJAIS) – ISSN : 2249-0868
Foundation of Computer Science FCS, New York, USA
Volume 1– No.3, February 2012 – www.ijais.org

Musical Instrument Recognition using Zero Crossing Rate and Short-time Energy

Sumit Kumar Banchhor
Assistant Professor,
Department of Electronics and Telecommunication,
GD Rungta College of Engineering and Technology,
Chhattisgarh Swami Vivekanand Technical University, Bhilai,
Chattisgarh, India

Arif Khan
Assistant Professor,
Department of Electronics and Telecommunication,
GD Rungta College of Engineering and Technology,
Chhattisgarh Swami Vivekanand Technical University, Bhilai,
Chattisgarh, India

ABSTRACT

Traditionally, musical instrument recognition is mainly based on frequency domain analysis (sinusoidal analysis, cepstral coefficients) and shape analysis to extract a set of various features. Instruments are usually classified using k-NN classifiers, HMM, Kohonen SOM and Neural Networks. Recognition of musical instruments in multi-instrumental, polyphonic music is a difficult challenge which is yet far from being solved. Successful instrument recognition techniques in solos (monophonic or polyphonic recordings of single instruments) can help to deal with this task.

We introduce an instrument recognition process in solo recordings of a set of instruments (flute, guitar and harmonium), which yields a high recognition rate. A large solo database is used in order to encompass the different sound possibilities of each instrument and evaluate the generalization abilities of the classification process. The basic characteristics are computed in 1sec interval and result shows that the estimation of zero crossing rate and short time energy reflects more effectively the difference in musical instruments.

Keywords

Musical Instrument classification, generalization, zero crossing rate, short time energy.

1. INTRODUCTION

Recognizing objects in the environment from the sounds they produce is arguably the primary function of the auditory system. An organism that can sense a threat at a distance has a competitive advantage (in the evolutionary sense) over one that cannot. Recognition is possible, in part, because acoustic features of sounds often betray physical properties of their sources. As a simple example, large objects tend to produce sound energy at frequencies lower than those produced by small objects. If an organism's goal is to recognize sounds as arising from particular source classes, recognition should be based on those acoustic features that are invariant across the sounds within each class yet distinguish between the sounds of different classes. For many classes of sound sources, acoustic characteristics that correlate with physical or behavioral properties are examples of such highly discriminatory features. Successful automatic classification of musical sounds is useful in many applications –classification of audio files scattered on the Internet, automatic scoring of recorded music, automatic indexing of recordings, multimedia labeling and many others. Computational auditory scene analysis (CASA), automatic music transcription frameworks and content-based search systems, all find such a capability to be extremely helpful. However, musical instrument recognition has not received as much research interest as, for instance, speech and speaker recognition, even though both the amateur music lover and the professional musician would benefit from such systems. The challenge of automatic classification of musical sounds poses many questions: Accuracy - is it possible to distinguish among virtually identical sounds coming from different instruments, for example certain sounds of Viola and Violin? Taxonomy - what should be the classes? Should sounds recorded in different environments using different instruments and playing techniques, classified in the same class? e.g. when classifying into musical instruments, should recordings of a string ensemble in a noisy environment and a pizzicato sound of a single violin recorded in an anechoic chamber considered the same class? Which instruments should be classified in the same classes when categorizing samples into instrument families? Generality - which are the common qualities of sounds of a specific class (e.g. the sounds of a classical guitar) which separate them from other classes, regardless of the sound database being used and the recording conditions? Validity of data - are the sound databases consistent? Do they contain "bad" or misclassified samples? One of the broad goals of computational auditory scene analysis research is to create computer systems that can learn to recognize the sound sources in a complex auditory environment.

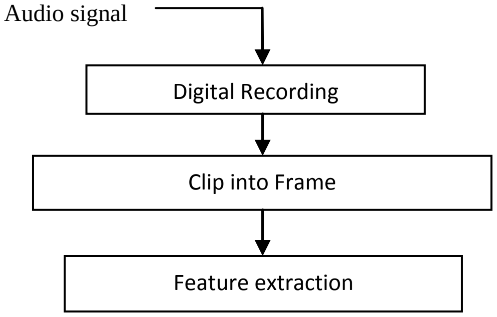

Fig 1: Basic processing flow of audio content analysis.

Fig 1 shows the basic processing flow which discriminates between speech and music signal. After feature extraction, the input digital audio stream is classified into speech, non speech and music.

2. PREVIOUS WORK

Many attempts in music instrument recognition have taken place in the last thirty years. Most of them have focused on single, isolated notes (either synthesized or natural) and tones taken from professional sound data-bases [1]. Recent works have operated on real-world recordings, polyphonic or monophonic, multi-instrumental or solo [2]. However, the issue is yet far from being solved. The work on recognition from separate notes still remains crucial, since it can lead to further optimization of the methods used and to insights on the recognition of multi instrumental, commercial recordings.

The majority of the recognition systems used so far concentrate on the timbral-spectral characteristics of the notes. Discrimination is based on features such as pitch, spectral centroid, energy ratios, spectral envelopes and mel frequency cepstral coefficients [3, 4]. Temporal features, other than attack, duration and tremolo, are seldom taken into account. Classification is done using k-NN classifiers, HMM, Kohonen SOM and Neural Networks [5, 6]. A limitation of such methods is that in real instruments the spectral features of the sound are never constant. Even when the same note is being played, the spectral components change. One has to take into consideration many timbral components and the way they can vary, which is often rather random, in order to develop a robust recognition system.

3. METHODOLOGY

The target sample was manually segmented using GOLDWAVE software and stored with .wav extension.

4. EXPERIMENT AND RESULT

4.1 Result using Zero Crossing Rate

The zero-crossing rate (ZCR) indicates the frequency of signal amplitude sign change. To some extent, it indicates the average signal frequency as:

$$ZCR = \frac{\sum_{n=1}^{N} \left|\operatorname{sgn} x(n) - \operatorname{sgn} x(n-1)\right|}{2N} \quad \text{...... (1)}$$

Where *sgn[]* is a signum function and *x(m)* is the discrete audio signal.

In mathematical terms, a "zero-crossing" is a point where the sign of a function changes (e.g. from positive to negative), represented by a crossing of the axis (zero value) in the graph of the function. The zero-crossing is important for systems which send digital data over AC circuits, such as modems, X10 home automation control systems, and Digital Command Control type systems for Lionel and other AC model trains. Counting zero-crossings is also a method used in speech processing to estimate the fundamental frequency of speech.

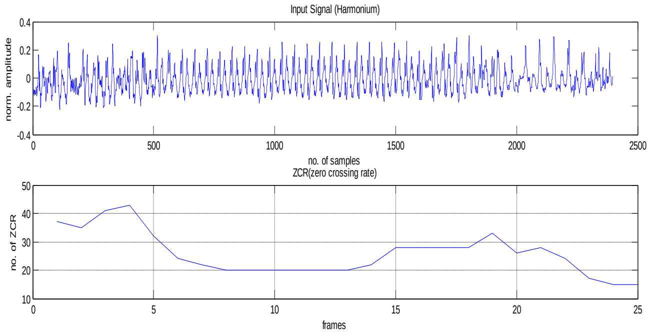

Fig 2: ZCR of harmonium.

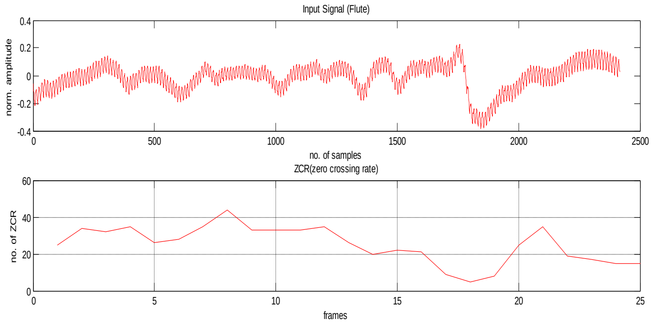

Fig 3: ZCR of flute.

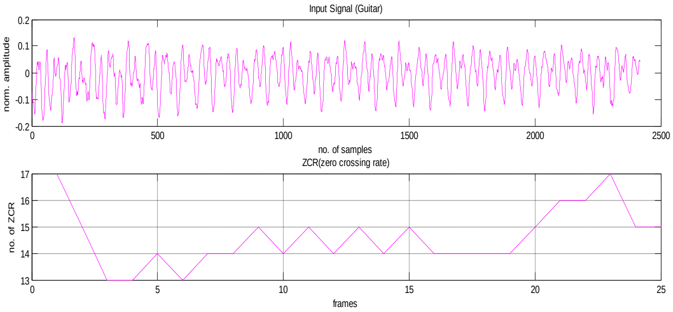

Fig 4: ZCR of guitar.

Table 1. ZCR of Harmonium, Flute and Guitar.

Frames Instruments	0	5	10	15	20
Harmonium	38	32	33	28	26
Flute	24	26	20	22	25
Guitar	18	14	14	15	15

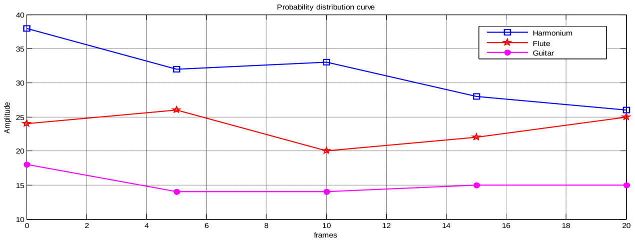

Fig 5: ZCR of Harmonium, Flute and Guitar.

Fig 5 displays the probability distribution curve of zero crossing rates of harmonium, flute and guitar.

4.2 Result using Short Time Energy

The short-time energy (STE) measurement of a speech signal can be used to determine voiced vs. unvoiced speech. Short time energy can also be used to detect the transition from unvoiced to voiced speech and vice versa. The energy of voiced speech is much greater than the energy of unvoiced speech.

$$E_n = \left(\sum_{m=-\infty}^{\infty} x^2(m)h(n-m)\right) \quad .... (1)$$

Eq. (1) defines the short time energy for a sampled signal where h(n-m) is a windowing function. For simplicity a rectangular windowing function is used as defined in eq. (2).

$$H(n) = \begin{matrix} 1 & 0 \le n \le N-1 \\ 0 & otherwise \end{matrix} \quad .... (2)$$

N in eq. (2) is the length of the window in samples. The selection of the window size is a compromise since a high pitched female or child's voice may have a pitch period as small as 16 samples (at an 8 kHz sampling rate) up to 200 samples for a low pitched male voice. A window size of 160 samples or about 20 msec. is a good compromise.

We record the input signal at fs=8 KHz. Now using Hamming window with the following specifications: Window size=256 samples, Window step=100 samples, Window overlap=156 samples and number of frames = (length of input – window size)/(window step), we calculate the STE for each frame using the following formula.

$$E = \sum_{m=0}^{N-1} \left| x(n)^2 / (N) \right|$$

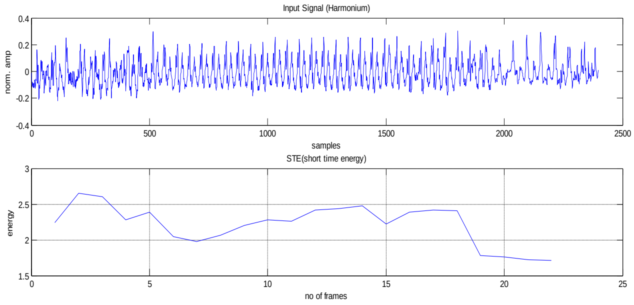

Fig 6: STE of harmonium.

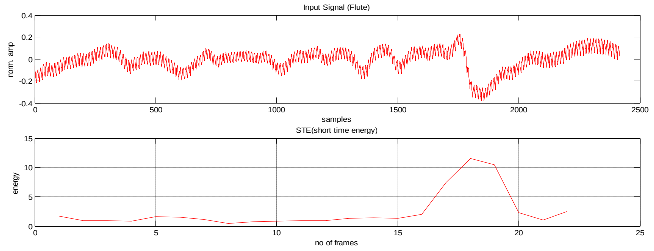

Fig 7: STE of flute.

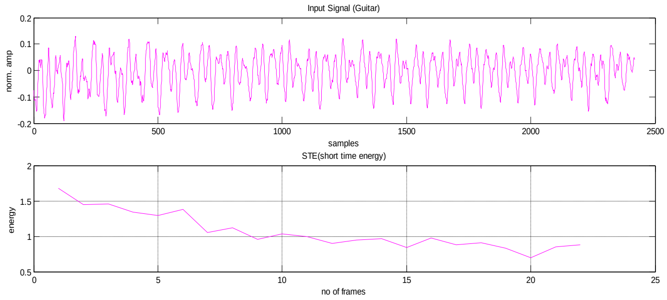

Fig 8: ZCR of guitar.

Table 2. STE of Harmonium, Flute and Guitar.

Frames / Instruments	0	5	10	15	20
Harmonium	2.2	2.4	2.3	2.2	2.3
Flute	2	1.6	1	1.3	1.8
Guitar	1.7	1.3	0.8	0.8	0.7

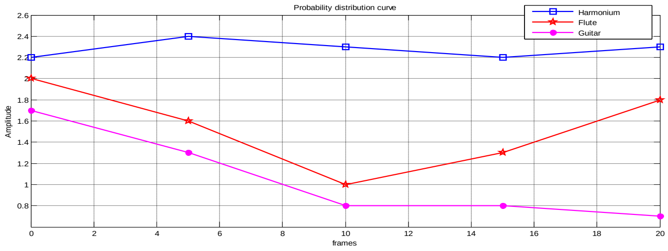

Fig 9: STE of Harmonium, Flute and Guitar.

Fig 9 displays the probability distribution curve of short time energy of harmonium, flute and guitar.

5. DISCUSSION AND CONCLUSION

Please In this paper, we dealt with recognition of sound samples and presented several methods to improve classification results. Tones are extracted from a large database of three musical instruments (harmonium, flute and guitar).

We use two different parameters in the analysis. From the experiments, we could observe evident results for zero crossing rate and short time energy. Zero crossing rates for the entire instrument's tones are always greater than short time energy. Zero crossing rate and short time energy is highest for harmonium then for flute and least for guitar.

Result shows that the estimation of zero crossing rate and short time energy reflects more effectively the difference in musical instrument.

6. REFERENCES

[1] K.D. Martin: Sound-Source Recognition: A Theory and Computational Model, Ph.D. thesis, MIT, 1999

[2] A. Livshin, X. Rodet: Musical Instrument Identification in Continuous Recordings, Proc. of the 7th Int. Conference on Digital Audio Effects (DAFX-04), Naples, Italy, October 5-8, 2004

[3] A. Eronen, A. Klapuri: Musical Instrument Recognition Using Cepstral Coefficients and Temporal Features, Proc. of the IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP 2000, pp. 753-756

[4] T. Kitahara, M. Goto, H. Okuno: Musical Instrument Identification Based on F0-Dependent Multivariate Normal Distribution, Proc. of the 2003 IEEE Int'l Conf. on Acoustic, Speech and Signal Processing (ICASSP '03), Vol.V, pp.421-424, Apr. 2003

[5] A. Eronen: Musical instrument recognition using ICA-based transform of features and discriminatively trained HMMs, Proc. of the Seventh International Symposium on Signal Processing and its Applications, ISSPA 2003, Paris, France, 1-4 July 2003, pp. 133-136

[6] G. De Poli, P. Prandoni: Sonological Models for Timbre Characterization, Journal of New Music Research, Vol 26 (1997), pp. 170-197, 1997