Classification of audio signals using statistical features on time and wavelet transform domains

Conference Paper *in* Acoustics, Speech, and Signal Processing, 1988. ICASSP-88., 1988 International Conference on · May 1998

DOI: 10.1109/ICASSP.1998.679665 · Source: IEEE Xplore

5 authors, including:

Tryphon Lambrou
University of Lincoln

Panos Kudumakis
Queen Mary, University of London

Mark Brian Sandler
Queen Mary, University of London

Alf Linney
University College London

CLASSIFICATION OF AUDIO SIGNALS USING STATISTICAL FEATURES ON TIME AND WAVELET TRANSFORM DOMAINS.

T. Lambrou[1], P. Kudumakis[2], R. Speller[1], M. Sandler[2], A. Linney[1]

[1]Dept. of Medical Physics, University College London, 11-20 Capper Str., London WC1E 6JA, UK

[2]Dept. of Electrical Electronic Engineering, King's College London, Strand, London WC2R 2LS, UK

ABSTRACT

This paper presents a study on musical signal classification, using wavelet transform analysis in conjunction with statistical pattern recognition techniques. A comparative evaluation between different wavelet analysis architectures in terms of their classification ability, as well as between different classifiers is carried out. We seek to establish which statistical measures clearly distinguish between the three different musical styles of rock, piano, and jazz. Our preliminary results suggest that the features collected by the adaptive splitting wavelet transform technique performed better compared to the other wavelet based techniques, achieving overall classification accuracy of 91.67%, using either the Minimum Distance Classifier or the Least Squares Minimum Distance Classifier. Such a system can play a useful part in multimedia applications which require content based search, classification, and retrieval of audio signals, as defined in MPEG-7.

1. INTRODUCTION

The widespread usage of computer networks has led to the need for systems capable for searching, classifying and retrieving audio files. Today, audio files are commonly treated as text files and are classified by name, file format, sampling rate etc. Various researchers have implement algorithms capable of extracting audio structure from a sound [1]. These algorithms were tuned to specific musical patterns and were not appropriate for all sounds. Other researchers have focused on indexing audio databases using neural networks [2], with some success. E. Wold *et.al.* [3], proposed a system which collects perceptual and acoustical features and then is using the Minimum Distance Classifier, for classification, search and retrieval of audio signals.

This paper attempts an investigation on the usage of statistical features collected from the time and wavelet transform domains, using several different classifiers, for applications on audio classification and retrieval, under the "Multimedia Content Description Interface" protocol, (i.e. MPEG-7) [4].

2. PATTERN RECOGNITION

Our statistical pattern recognition approach uses the classical steps of feature extraction, classification and feature selection, which are further described below.

2.1 Feature Extraction

The first step of our pattern recognition approach is the feature extraction step, which is the transformation of patterns into features that are regarded as a compacted representation. Overall eight statistical signal features were collected from each signal, given by category as: First Order Statistics [5], i.e. Mean, Variance, Skewness, and Kurtosis. Second Order Statistics [6], i.e. Angular Second Moment, Correlation, and Entropy. Finally, the number of zero crossings was evaluated since it indicates the noise behavior of the signal.

2.2 Classifiers.

Four statistical classifiers were constructed and employed in this study. The classifiers used are: 1) the Minimum Distance Classifier (MDC) [7], which employs as classification criterion the minimum Euclidean distance between the unknown entry and the mean values of each of the other classes, 2) the k-Nearest Neighbor Distance Classifier (k-NNC) [7], where the classification criterion is the minimum Euclidean distance between the unknown entry and the k-Nearest Neighbor elements of any other class, 3) the Least Squares Minimum Distance Classifier (LSMDC) [8][9], where the classification rule is again the minimum Euclidean distance between the unknown entry and the mean values of each of the other classes, using a linear equation within the least squares technique in order to minimize the errors, and 4) the Quadrature Classifier (QC) [8][9], which employs the same decision rule as the previous classifier, but using a quadrature equation within the least squares method.

2.3 Feature Selection

The performance of the classifiers was evaluated by using the Leave-One-Out method. This involves the re-classification of all the signals (one at the time) to their *a priori* known categories (or classes). In addition, for each set of features all possible combinations were tested up to three-dimensional decision space. Those features which achieve the best classification rate are selected, in order to be used in the pattern recognition process. This phase is called feature selection, and aims to reduce the features set to a subset which consists only of meaningful information (i.e. features which characterize best) about the signals we want to classify.

3. EXPERIMENTS AND DISCUSION

In this study we used twelve (12) musical signals (4 Rock, 4 Piano, and 4 Jazz), for the training stage of the classification procedure. We seek to establish which statistical measures clearly distinguish between these three different musical styles. The signals were analyzed using different wavelet transform architectures such as: the logarithmic splitting (LOG) [10], uniform splitting, or wavelet packet (UNIF) [11] and adaptive splitting (ADAP) [12]. The Daubechies 4-TAP wavelet filter was used, in all the wavelet architectures. As a criterion for the adaptive wavelet transform algorithm the zero crossing measurement for each frame of the signal was applied. From all of the signals and their wavelet transform coefficients their first and second order statistical values, as well as the zero crossing measurements were collected. Using the Leave-One-Out method we selected the following combinations of features, which achieved classification accuracy better than 90.0%, in our case those features capable of correctly characterizing eleven signals out of the total twelve. The measurements which are stated below are given in the fashion of x vs. y (or x vs. y. vs. z) coordinates, as they appear into the decision space of the classifiers. So, for the time domain, the feature combinations which achieved this classification accuracy threshold, are: Skewness vs. Entropy, Kurtosis vs. Entropy, and Correlation vs. Entropy. For the features collected from the adaptive splitting wavelet transform the feature combinations are: Entropy vs. Zero crossing, Kurtosis vs. Entropy, and Skewness vs. Kurtosis vs. Entropy. For the features collected after the uniform splitting wavelet transform the feature combination of Skewness vs. Entropy was best. The best classification rates for all the classifiers which exceed the arbitrary threshold of 65.0% in accuracy, for the unprocessed (i.e. original) signals and their wavelet transform coefficients using different wavelet analysis architectures, for each classifier are presented in Table 1. As we can observe, the features selected from the wavelet transform coefficients using the adaptive splitting wavelet method achieved classification accuracy of $\approx 92.0\%$, under the classification rules of the Minimum Distance Classifier (MDC) and that of the Least Squares Minimum Distance Classifier (LSMDC). It is obvious from Table 1, that the features selected from the adaptive splitting wavelet method, outperformed, in terms of classification accuracy, those of the logarithmic and the uniform splitting methods. The classification rates achieved by the features collected from the original signals are almost equal to those of the adaptive wavelet method. Finally, from Table 1, we can draw conclusions about the performance of all the different classifiers used in this study. The Least Squares Minimum Distance Classifier (LSMDC) performed better than the others, achieving the highest classification accuracy, using features which were collected by the original signals, and the wavelet coefficients of the adaptive, logarithmic and uniform splitting methods. The k-Nearest Neighbor Classifier (k-NNC) performed surprising poorly, and the classification accuracy it achieved was below the arbitrary threshold of 65.0%, so it is not included in Table 1.

Figure 1, illustrates the decision space of the Least Squares Minimum Distance Classifier (LSMDC), using the features of Kurtosis vs. Entropy of the original signals, which achieved classification accuracy equal to 91.67%. From Figure 1, we can observe that one signal of the Jazz category has been misclassified, but the rest of the signals have been correctly classified into their categories.

Figure 2, illustrates the decision space of the LSMDC classifier, using the features of Kurtosis vs. Entropy, of the wavelet transform coefficients obtained by the adaptive splitting wavelet technique of the signals, which gave classification accuracy of 91.67%, with the misclassification of one Rock signal.

Figure 3, presents the decision space of the LSMDC classifier, using the features of Kurtosis vs. Entropy of the wavelet transform coefficients obtained by the uniform splitting wavelet technique of the signals. The classification accuracy of this feature combination is equal to 75.0%, with the misclassification of two Rock, and one Piano signals.

	TIME	ADAP.	LOG.	UNIF.
MDC	83.33%	91.67%	66.67%	75.0%
k-NNC	---	---	---	---
LSMDC	91.67%	91.67%	83.33%	91.67%
QC	83.33%	83.33%	75.0%	75.0%

Table 1. The best classification scores for each classifier, using features selected from the Time Domain (TIME), the Adaptive Splitting Wavelet Transform (ADAP), the Logarithmic Splitting Wavelet Transform (LOG), and the Uniform Splitting Wavelet Transform (UNIF).

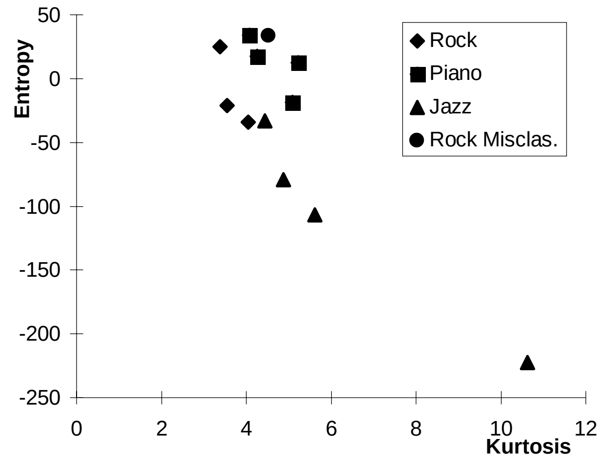

Figure 1. Decision space of the LSMDC, using the features of Kurtosis vs. Entropy, from the time domain.

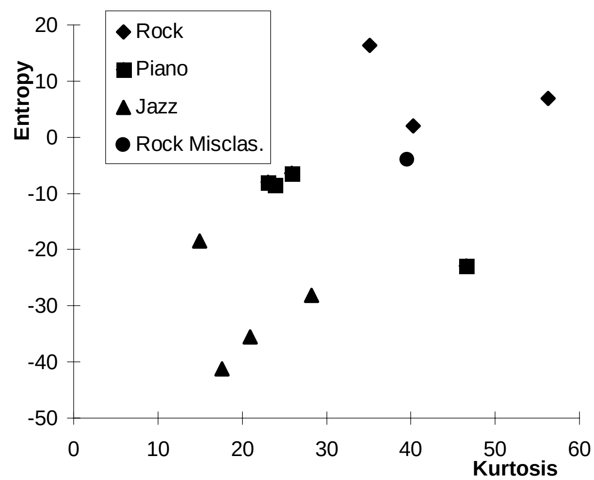

Figure 2. Decision space of the LSMDC, using the features of Kurtosis vs. Entropy, from the wavelet transform coefficients of the adaptive splitting algorithm.

4. CONCLUSIONS.

In this paper an investigation on musical signal classification, using different wavelet analysis techniques in conjunction with statistical pattern recognition methods was presented. In our preliminary study twelve musical signals (4 Rock, 4 Piano, and 4 Jazz) were used, in order to establish which features distinguish better these three different musical signals. Eight statistical measurements were collected, from the original signals as well as from their different wavelet transform coefficients. We found using the Leave-One-Out method that the combination of the features: Skewness vs. Entropy, Kurtosis vs. Entropy, Correlation vs. Entropy, Entropy vs. Zerocrossing, and Skewness vs. Kurtosis vs. Entropy, achieved classification accuracy more than 90.0%. Overall the features selected by the adaptive analysis wavelet transform coefficients performed better than the other techniques, under the classification rule of either the Minimum Distance Classifier (MDC) or the Least Squares Minimum Distance Classifier (LSMDC). Another advantage of using the wavelet transform coefficients, instead of the time domain signal, is that the processing delay/cost needed in the feature extraction stage is a lot less due to the compacted representation of the wavelet transform. In addition we demonstrated that high classification accuracy can be reached using only compacted data, which is in line with the MPEG-7 protocol. Possible applications of systems like the one presented in this paper are on content based classification, search and retrieval of audio signals, detection of infringement of the copyright law etc.

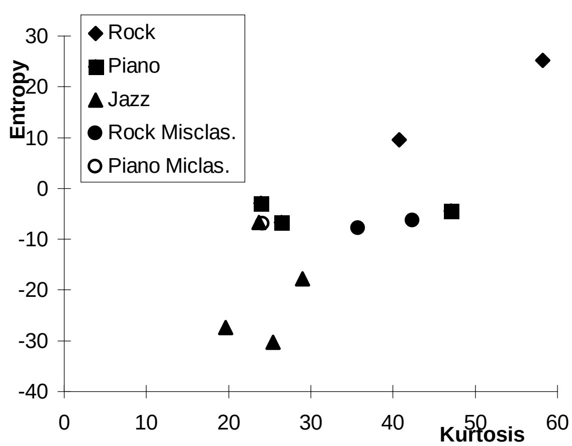

Figure 3. Decision space of the LSMDC, using the features of Kurtosis vs. Entropy, from the wavelet transform coefficients of the uniform splitting algorithm.

5. REFERENCES

[1] S. Foster, W. Schloss, A.J. Rockmore, “Towards an Intelligent Editor of Digital Audio: Signal Processing Methods”, Computer Music Journal, 6, 1, 42-51, 1982.

[2] B. Feiten, S. Gunzel, “Automatic Indexing of a Sound Database Using Self-Organizing Neural Net”, Computer Music Journal, 18, 3, 53-65, 1994.

[3] E. Wold, T. Blum, D. Keislar, J. Wheaton, “Content-Based Classification, Search, and Retrieval of Audio”, IEEE Multimedia Magazine, 27-36, Fall 1996.

[4] MPEG Requirements Group, "MPEG-7: Context and Objectives", Document edited by R. Koenen, ISO/MPEG N1733, Stockholm MPEG Meeting, 1-8, July 1997.

[5] P.C. Chen, T. Pavlidis, “Segmentation by Texture Using a Co-occurrence Matrix and a Split-and-Merge Algorithm”, Computer Graphics and Image Processing, 10, 172-182, 1979.

[6] R.M. Haralick, K. Shanmugam, I. Dinstein, “Textural Features for Image Classification”, IEEE Transactions on Systems, Man, and Cybernetics, SMC-3, 6, 610-621, 1973.

[7] R.A.G. Dyer, S.A Dyer, P.K. Bhagat, “Myocardial Tissue Characterisation Using Pattern Recognition Procedures on Backscattered Ultrasonic Signals”, Ultrasonic Imaging, 8, 181- 195, 1986.

[8] J.R. Ullmann, “Pattern Recognition Techniques”, Butterworth & Co, London, 1973.

[9] J.M. Mendel, K.S. Fu, “Adaptive Learning and Pattern Recognition Systems, Theory and Applications”, Academic Press, New York, 1970.

[10] P. Kudumakis, “Sunthesis and Coding of Audio Signals Using Wavelet Transforms for Multimedia Applications”, PhD Thesis, King’s College University of London, 1997.

[11]M. Vetterli, J. Kovacevic, “Wavelets and Subband Coding”, Prentice Hall, Englewood Cliffs NJ., 1995.

[12]M.V. Wickerhauser, “Adapted Wavelet Analysis from Theory to Software”, A.K. Peters, Wellesley Massachusetts, 1994.