Proceedings of the 5th WSEAS International Conference on Signal Processing, Istanbul, Turkey, May 27-29, 2006 (pp151-154)

Musical Instrument Classification Using Neural Networks

MUSTAFA SARIMOLLAOGLU[1] and COSKUN BAYRAK[2]
[1]Dept. of Applied Science, [2] Dept. of Computer Science
University of Arkansas at Little Rock
2801 S. University Avenue Little Rock, Arkansas 72204
USA

Abstract: - In this paper, a system for automatic classification of musical instrument sounds is introduced. As features mel-frequency cepstral coefficients and as classifiers probabilistic neural networks are used. The experimental dataset included 4548 solo tones from 19 instruments of MIS database (The University of Iowa Musical Instrument Samples). Experiments for different system structures (hierarchical and direct classification) were carried out and compared. The best performance in direct classification was 92% for individual instruments and 97% for families; and 89% for individual instruments when hierarchical approach is used.

Key-Words: - Musical instrument classification, probabilistic neural networks, PNN

1 Introduction

Music Information Retrieval (MIR) has gained increasing research attention over the recent years. Apart from their academic merits, robust MIR systems will have important commercial and social implications. They will add significant value to the existing music libraries by making them easily accessible; enabling automatic classification, organization, indexing, and searching. Musical instrument classification, where the idea is to recognize the instruments playing in a musical sound, is one of the signal analysis problems in MIR.

Musical instrument recognition is a difficult task and is far from being solved and applicable to real-world musical signals [1]. The problem is rather easy for monophonic sounds compared to polyphonic ones, where multiple instruments played together. Assuming a preliminary source separation has been performed, classification research has been concentrated on monophonic sounds.

The recognition of audio signals consists of two basic steps; defining and extracting the features that distinguish the sources, and design of a system (classifier) to recognize the sources using those features. Many features (cepstral, spectral, temporal) are introduced in the literature and a comparison of them with regard to recognition performance can be found in [2]. K-Nearest Neighbors (k-NN), Hidden Markov Models, Gaussian Mixture Models (GMM), Naive Bayesian classifiers, Support Vector Machines, and Artificial Neural Networks (ANN) are some of the techniques used for instrument classification.

Eronen and Klapuri's system used a wide set of features and tested on 30 instruments [3]. They utilized a hierarchical framework for classification and used Gaussian or k-NN classifier at each node but direct classification performed better. Best recognition accuracy was 94% for instrument family and 80% for individual instruments. Krishna and Sreenivas proposed line spectral frequencies as features and obtained 95% and 90% accuracy for family and instruments respectively [4]. They classified 14 instruments using GMMs and K-NN classifiers. Bolat compared the performances of three statistical neural networks, on four reed instruments using linear prediction coefficients as features [5]. PNN achieved highest accuracy of 93% compared to GRNN (90%) and RBF (47%) networks.

In this paper we focus on the performance of probabilistic neural networks on classification of a large number of instruments, using only the cepstral features.

2 Feature Extraction

Mel-Frequency Cepstral Coefficients (MFCCs) are the features used in our system to model the tones. MFCCs have been proven to be useful in a broad range of classification applications, such as speech classification [6], speaker identification [7], musical genre classification [8], etc. In [2], a large set of features are compared in terms of recognition performance in an instrument recognition system. Among the others, MFCCs, their standard

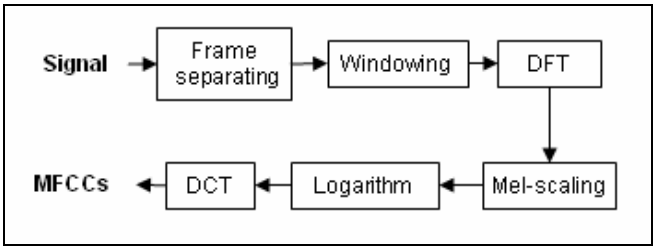

Figure 1. Computation of MFCCs

deviations, and deltas seemed to be the most successful ones [2].

MFCCs provide a compact representation of the spectral envelope and are extracted as given in Figure 1. Mel-scaling emphasizes the perceptually meaningful frequencies by mapping the spectrum coefficients into a non-linear manner. Discrete cosine transform (DCT) is used to reduce the dimension of power spectrum.

In this work, the input signal was processed in 256 point frames overlapped by 50 %. First 12 MFCCs (excluding 0^{th}) were calculated. Resulting 12-dimensional vectors were used as the base feature vectors. Keeping the same file structure as in MIS, we had several feature files for each instrument.

Prior to the classification stage, vector quantization was applied to further reduce the amount of data and complexity. LBG algorithm, which is an efficient variant of k-means algorithm, was used in quantization [9].

Three sets of quantized sample vectors were produced to be used in appropriate experiments. In the first step, feature vectors in every feature file were individually clustered to 100 vectors (Set 1). In the second step, for each instrument, 70% of the feature files from Set 1 were selected randomly and they were merged and clustered into 300 vectors to represent each instrument (Set 2). In the last step, samples in Set 2 were combined within the instrument families and clustered into 300 vectors (Set 3). Cluster size was chosen to be 300 as a trade-off between computational complexity and a good representation of the class. Remaining 30% of the Set 1 samples were left for testing purpose while Set 2 and 3 were used for training. Hence, testing and training samples were completely different.

3 Classifier

Artificial Neural Networks (ANNs) are processing structures that consist of interconnected neurons. Connectivity pattern and weights between neurons represents the mappings between input and output vectors. Some ANNs have the ability of approximating any function, but in general it takes very long to train the network and adjust the parameters [1].

Probabilistic Neural Network (PNN) is a type of statistical neural networks. PNN learns to approximate the probability density function of the training examples. The only parameter that needs to be selected for training is the spread, which is the deviation of the Gaussian functions. Spread is chosen experimentally to find the best results. For more information about PNN, please refer to [7].

PNNs' advantages over the other methods are flexibility and the straightforward design. Training time is much faster than the other types of ANNs. They enable incremental training, where new training examples can be incorporated without difficulties. And they are robust to noise. However, the major disadvantage of the PNN is that it is slower to operate, because it performs more computations then other ANN models [10].

In our work, PNNs were used for classification. In the non-hierarchic case, one network was used. The hierarchic scheme composed of five networks in two stages; one for family classification and four for instrument classification within the family.

Decision Mechanism

PNN provides best matching instrument for each input vector. However, in order to increase the accuracy, multiple input vectors from each sample are needed. At the decision level, a predefined number of outputs from the PNN are buffered and summed together. This way, each class in the system gets a score on the period of sample applied to the system. Then the class with the highest score is chosen to be the source of the sample [7].

4 Experiments and Results

Sample database used in experiments consist of 4548 tones from 19 instruments of MIS database, as detailed in Table 1. All samples are in mono, 16 bit and 44.1 kHz. These samples include several different articulation styles; all strings include pizzicato and some instruments include vibrato. All instruments have samples of three dynamic levels (ff,mf,pp).

Two systems were implemented; one with direct classification and one with hierarchical classification. Direct classification system has one PNN, having 19 classes, which are as many as instruments. It is trained with features from Set 2 (detailed in section 2). The hierarchical system on the other hand, has a structure consisting of two

Table 1. Sample Database

Family	Instrument	Number of tones
Strings	Bass	589
	Cello	731
	Viola	552
	Violin	597
Brass	Horn	96
	Bass Trombone	131
	Tenor Trombone	99
	Alto Sax	192
	Soprano Sax	192
	Tuba	111
	Trumpet	212
Reeds	Bassoon	123
	Oboe	105
	Bb Clarinet	139
	Eb Clarinet	119
	Bass Clarinet	138
Flute	Flute	221
	Alto Flute	99
	Bass Flute	102

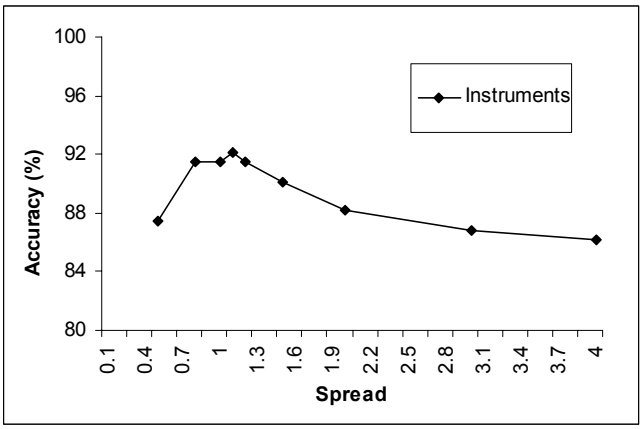

Figure 2. Results for direct classification system

stages. In the first stage there is one PNN for instrument family recognition. This network is trained with features from Set 3. In the second stage there are four PNNs; one for each family. These networks are trained with Set 2 features. According to the decision about family, input vector is then directed to appropriate branch for final classification.

In the first experiment, direct classification approach was used. Features from Set 2 were used to create the PNN. Accuracy is highly dependent on the spread value selected for PNN. Results for different spread values are given in Figure 2. The accuracy of 92% was achieved in the best case. It was also observed that 90% of the errors were made within the family (Table 2); i.e. instrument was classified as another instrument of the same family.

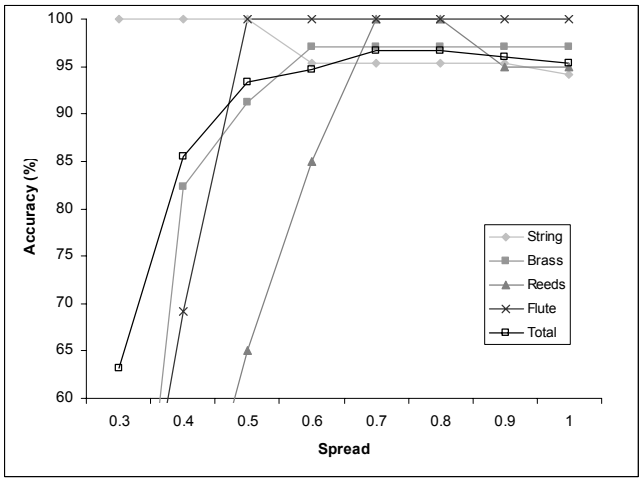

Figure 3. Results for family recognition

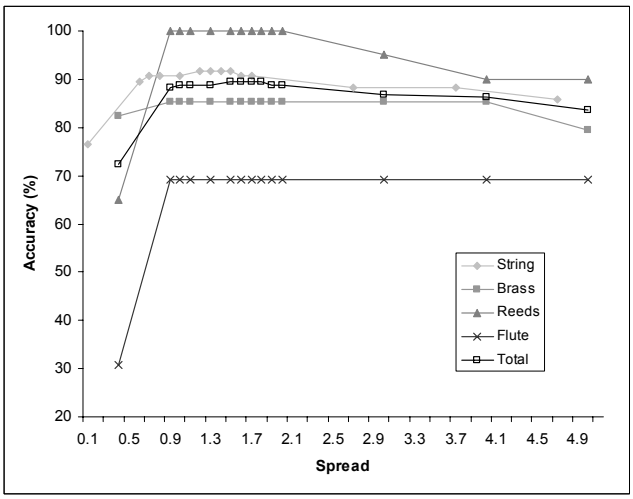

Figure 4. Results for hierarchical classification system

In the second experiment first stage of the hierarchical system was considered. The accuracy of 97% was achieved for instrument family recognition. Total performance along with the family performances is plotted in Figure 3. As spread increases, accuracy of strings tends to decrease, while the accuracies of others increase.

In the last experiment both stages of the hierarchical system were used. First stage was set to the best case of the previous experiment. Accuracy of the whole system was investigated (Figure 4), and found to be 89% in the best case. Since the results of second stage totally depend on the results of first stage, it was not expected to get more accuracy then the second experiment. But hierarchical structure is expected to give better results than the direct classification since each network has a smaller number of classes to choose one from. This result may be due to the fact that 90% of the errors were made within the family in direct classification.

Table 2. Results for best cases

	No Hierarchy				Hierarchy			
	s	c/t	%	fe/ewf	S	c/t	%	fe/ewf
Families	0.7	147/152	96.7	5/0	0.7	147/152	96.7	5/11
Strings	N/A	81/85	95.3	4/0	1.6	78/85	91.8	4/3
Brass	N/A	33/34	97.1	1/0	1.6	29/34	85.3	1/4
Reeds	N/A	20/20	100	0/0	1.6	20/20	100	0/0
Flute	N/A	13/13	100	0/0	1.6	9/13	69.2	0/4
Instruments	1.1	140/152	92.1	2/10	N/A	136/152	89.5	5/11

s: spread for PNN
c/t: correctly recognized samples/total test samples
fe: family errors
ewf: errors within family

Our system seems to perform better than the studies [3,4,5] mentioned in Section 1. However, it should be noted that these results can be misleading in terms of comparison between classifiers' accuracy, since it is greatly affected by the factors such as number of instruments, number of samples, features, and testing scheme.

5 Conclusions

Systems for automatic classification of musical instrument sounds were presented. Probabilistic neural networks were used as classifiers and MFCCs were used as features. Multi-level quantization was applied to the features prior to classification. Experiments for different system structures were carried out. Hierarchical and direct classification structures were compared. In separate trials, 97% and 92% accuracies obtained for family and instrument classification respectively. But the combination of them, the hierarchical approach, did not perform better than the direct instrument classification (89%), at least, in our setup.

Using only one type of feature and applying it to 19 instruments, results suggest that PNNs are suitable for musical instrument classification. Addition of other types of features may be considered for a better performance.

References:

[1] P. Herrera-Boyer, G. Peeters and S. Dubnov, "Automatic Classification of Musical Instrument Sounds" Journal of New Music Research, 2003, Vol. 32, No.1, pp. 3-21.

[2] A. Eronen, "Comparison of features for musical instrument recognition" IEEE Workshop on Applications of Signal Processing to Audio and Acoustics, 2001.

[3] A. Eronen, and A. Klapuri, "Musical instrument recognition using cepstral coefficients and temporal features" IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP '00), 2000, Vol. 2, pp. II753-II756.

[4] A. Eronen, and A. Klapuri, "Music instrument recognition: from isolated notes to solo phrases" IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP '04), 2004, Vol. 4, pp. IV265-IV268.

[5] B. Bolat, "Recognition of the Reed Instrument Sounds by Using Statistical Neural Networks", Sigma Journal of Engineering and Natural Sciences, 2005/2, pp. 36-42.

[6] S. B. Davis and P. Mermelstein, "Comparison of parametric representations for monosyllabic word recognition in continuously spoken sentences", IEEE Transactions on Acoustics, Speech and Signal Processing, 1980.

[7] M. Sarimollaoglu, S. Dagtas, K. Iqbal and C. Bayrak "A Text-Independent Speaker Identification System Using Probabilistic Neural Networks" in International Conference on Computing, Communication and Control Technologies (CCCT), Austin, USA 2004, pp. 407-411.

[8] G. Tzanetakis and P. Cook, "Musical genre classification of audio signals" IEEE Transactions on Speech and Audio Processing, 2002.

[9] Linde, Y., A. Buzo & R.M. Gray, "An algorithm for vector quantizer design," IEEE Transactions on Communication COM-28, 1980, pp. 84-95.

[10] H. Demuth, M. Beale, Neural Network Toolbox User's Guide, Version 4, MathWorks Inc, 2003, pp. 7.12-7.13