

# Instruments recognition using neural networks and spectral information

Ezzaidi Hassan

*Ermetis, Université du Québec à Chicoutimi, Chicoutimi, Québec, Canada, G7H 2B1*

hezzaidi@uqac.ca

**Abstract:** This paper addresses musical sounds recognition produced by different instrument. Various architectures of back-propagation and radial basis networks were applied as classifiers and their results were compared. The musical notes used in the test processing were not presented in the previous training session. The discrete Fourier transform vectors extracted from each segment were used as parameters, with the objective to evaluate only the magnitude of the spectral information conveyed to discriminate between musical sounds instrument. The Music Instrument Sample Database (UIOWA) was used for this experiment. The number of instrument was increased from 14 to 19 compared to previous research. With the proposed method, a perfect score is obtained where the recognition is achieved on the basis of the family level (String, Brass, Reed, and Flute). However, when the recognition was achieved which was based on the instrument level, the performance decreased. Spectral information 17 to 18 out of 19 instruments were identified and all the family was well recognized. The results show that temporal information is not important for family but it plays an important role for instrument identification.

**Key words:** audio feature, music instruments, neural networks, spectral information.

## INTRODUCTION

Since the early 1980s speech signal processing and analysis has been realized and has advanced progressively in various fields of practice. Over the years, many disciplines have emerged and concentrated research in areas such as speech recognition, speaker identification, speaker verification, speaker segmentation, segments clustering, language recognition and so forth.

Recently many studies have been oriented to musical signal analysis and processing in order to respond to the high demand of internet users and to countless multimedia applications. The demand includes audio indexing, automatic transcription, genre classification, singer identification and instrument recognition.

A musical signal can be viewed as random symbol (note) generated by a combination of an output of the hidden sources. When only one source (instrument) is present at any moment of time, we are addressed to the isolated note known as the monophonic recordings. Many previous studies have focused on the monophonic case. Brown et al. [Brown 01] in their studies compared various features in the automatic identification of woodwind musical instruments by using the speaker identification techniques. The examined features included cepstral coefficients,

constant-Q coefficients, spectral centroide, autocorrelation coefficients and moments of the time signal. The best accuracies results were identified between 79% to 84%. The higher-order statistics moments were examined by Dubnov et al. [Dubnov 97]. They concluded that these features are adequate to discriminate well between instruments of different families and are not robust to discriminate between instruments within the same family. Krishna and Sreenivas [Krishna 04] proposed Line Spectral Frequencies as features using a Gaussian Mixtures Models (GMM). They reported the best score of 95% at family level and 90% at 14 instrument level compared to MFCC and LPC features. Marques and Moreno [Marques 99] have proposed a GMM model and support vector machines using the FFT based cepstral coefficients, LPC and MFCC features set. A score of 70% was obtained with 9 instruments.

When more than one source can be presented at any moment in time, then the case is qualified as the polyphonic recordings. Research on the polyphonic case has been particularly challenging and complex with limited studies contributed to this field.

The following research investigates exclusively a combination of 2 or 3 simultaneous notes. Among other studies, Kashino and Murase [Kashino 99]

describe an approach of the sound source identification based on the template adaptation and music stream extraction. A missing features technique in the GMM classifier was introduced by Eggink and Brown [Eggink 03].

In this paper, we proposed the use of different neural networks structures for music instruments recognition. A neural network with magnitude spectral information without any temporal information may be considered to analyze indirectly how the spectral information contributes to characterize specifically the musical instrument. The hidden layer with a number of unit cells fewer than the input layer has also been used to explore the coding capacity and the dimension reduction of the input vector features. Comparatively to other experiments (studies) the database was augmented to 19 instruments originating from four families. All experiments in this study are focused on the monophonic recordings.

## 1. Models

Neural network is a formal model inspired from the organisation of biological cells to learn many tasks in a diver's field. Neural networks was applied with success in many application as pattern recognition, signal control and time series analysis. The main propriety is that we can learn only from input data (unsupervised learning) or from input-output data (supervised learning). In the following research, only supervised work is addressed.

### 1.1. Backpropagation neural network

A Multilayer Perceptron (MLP) neural network with back-propagation was used as classifier (see Fig. 1). The structure is composed from an input layer where the training data vector is injected. The hidden layer is supposed to realize a compression when the number of cells is lower than the input layer cells and an extension when the number of cells is greater than the input layer cells. The output is an encoding process that characterizes which categories are to be recognized. The network is an acyclique form fully connected. In the following study, the number of cells in hidden layer was taken from following numbers: 20, 40, 80 and 120.

The fastest and steepest descent with the momentum algorithm was used in the training process to update the networks weight and biases in the negative direction of the gradient.

The learning occurs according to the following parameters: the learning rate is fixed to 0.1, and the momentum constant is fixed to 0.9. The multilayer network uses the sigmoid transfer function that generates outputs from 0 to 1 at each unit cell. Effectively, the input goes between negative to positive infinity. In the case of the classification tasks, the output of the sigmoid function can be interpreted as the approximation of posteriors probabilities of classes.

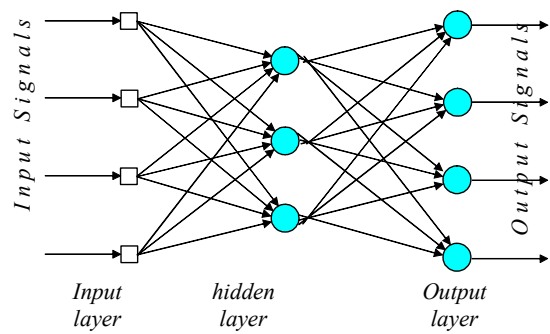


Figure 1: Multilayer Perceptron Neural Network Structure

### 1.2. Radial basis networks

The Radial Basis Functions (RBF) network is composed of two layers: the hidden radial basis layer and the output competitive layer. The first layer computes distance similarities between the input and each of the prototype vectors representing the classes. This can be interpreted as a projection over some mean or centre mass. The second layer based on the activity of the first layer produces a probability vector as output. The transfer function for the radial basis neuron is:  $\exp(-d^2)$ , where  $d$  is the similarity distance. All experiments are realized with the neural network toolbox of The MathWorks, Inc.[Matlab 03].

## 2. Methodology

### 2.1. Parameters estimation

The energy with a dynamic threshold is done as criterion to detect the start and the end of each note. Indeed the efficiency value of the threshold was verified and adjusted manually to minimize the error detection. The time duration for each note was divided on adjacent window off 1024 points to calculate the magnitude of the Fast Fourier Transform (FFT) vector. By symmetry, only 512 points were kept for the following process. Hence the input of all proposed networks is set to 512 cells. The number of the output cells was also set to 5 where only 19 combinations were used to characterize each instrument. The rest of the other combinations were considered as reject decisions.

### 2.2. Database

The collection of instruments used in this work was from the database of the University of Iowa,<sup>1</sup> Musical Instrument samples. The collection was composed of 19 instruments coming from four families (String, Brass, Reeds, and Flutes) as follows:

Violin, Sop Sax, Tenor Trombone, Oboe, French Horn, Flute, EbClarinet, Cello, BbClarinet, Bass Trombone, Bassoon, Bass Flute, Alto Sax, Alto Flute, BassClarinet, Bass, Trumpet, Tuba and Viola.

<sup>1</sup> University of Iowa's Music Instrument Samples, <http://theremin.music.uiowa.edu/MIS.html>.

A set of 248 isolated notes for all these instruments were used to train (first half of data) and to test (rest of data) the proposed systems.

### 2.3. Criterion

Two criteria namely Major Ratio (MAR) and Minor Ratio (MIR) were proposed to calculate the score performance.

The MAR criterion considers an instrument as recognized if the score performance is better than all the other instruments and higher than 50%. The MIR criterion express an instrument as recognized if the score performance is only better than all the other instruments and lower than 50%. The MIR seems to be important in discriminating better between instruments originating from the same family.

## 3. Results and Discussion

instruments	Number of unit cells for MLP			
	20	60	120	240
Violin	32%	72%	71%	70%
Sop Sax	29%	31%	36%	41%
Tenor Trombone	62%	56%	62%	57%
Oboe	32%	29%	37%	34%
French Horn	20%	27%	23%	19%
Flute	18%	19%	20%	22%
EbClarinet	20%	27%	31%	29%
Cello	62%	65%	68%	71%
BbClarinet	66%	68%	74%	73%
Bass Trombone	19%	23%	23%	23%
Bassoon	49%	49%	51%	49%
Bass Flute	26%	33%	36%	39%
Alto Sax	32%	43%	46%	47%
Alto Flute	32%	37%	38%	39%
BassClarinet	12%	14%	24%	24%
Bass	74%	82%	78%	79%
Trumpet	50%	44%	52%	52%
Tuba	59%	93%	95%	96%
Viola	79%	83%	85%	85%
<b>MIR</b>	<b>8/19</b>	<b>9/19</b>	<b>7/19</b>	<b>8/19</b>
<b>MAR</b>	<b>7/19</b>	<b>7/19</b>	<b>9/19</b>	<b>8/19</b>
<b>SR(19)</b>	<b>15/19</b>	<b>16/19</b>	<b>16/19</b>	<b>16/19</b>
<b>SR(14)</b>	<b>11/14</b>	<b>12/14</b>	<b>12/14</b>	<b>12/14</b>

**Table 1:** Score recognition for instruments identification with MLP networks: SR(19) is the score for 19 instruments and SR(14) is the score recognition for 14 instruments using the criterion MIR and MAR.

The following tables presented in this section illustrate the diagonal score recognition of the confusion matrix. The SR (19) line corresponds to the score performance with 19 instruments added to the score obtained respectively by the MAR and MIR criteria. The SR (14) line corresponds to the score performance with 14 instruments added to the score obtained respectively by the MAR and MIR criteria. The last score is proposed mainly to compare the

proposed system to previous studies that used different methods. The SR line also reports the score recognition for families as combined the MAR and the MIR scores.

Table 1 illustrates the score recognition of the instruments for various architectures of multilayer perceptron neural networks. With 19 instruments, the best performance was obtained with the architecture containing respectively 60, 120 and 240 unit cells in the hidden layer. With 120 unit cells, 7 instruments were recognized with MIR criterion and 9 instruments with the MAR criterion to produce a performance rate of 16/19. Some instruments were identified with a higher rate and the others were identified with a weaker score. This fact, can be interpreted as some instruments convey a specific spectral information that enhances the inter-variability between different instruments. The instruments then recognized with lower score are affected principally by the higher similarity related to instruments coming from the same family. One can remark that the individual performance rate increases from the neural structure with 20 nodes in the hidden layer to the structure with 60 nodes. Then the rate performance remains almost stable between the 120 and 240 nodes structures. We can conclude that optimal structure is there with 120 nodes in hidden layer in order to reduce the time calculation and the memory capacity.

	Number of unit cells for MLP			
	20	60	120	240
String	73%	83%	83%	84%
Brass	59%	64%	68%	64%
Reeds	51%	50%	53%	51%
Flutes	39%	45%	43%	45%
<b>MIR</b>	<b>1/4</b>	<b>1/4</b>	<b>1/4</b>	<b>1/4</b>
<b>MAR</b>	<b>3/4</b>	<b>3/4</b>	<b>3/4</b>	<b>3/4</b>
<b>SR</b>	<b>4/4</b>	<b>4/4</b>	<b>4/4</b>	<b>4/4</b>

**Table 2:** Score recognition for families instruments with MLP networks. SR is the score recognition using the criterion MIR and MAR

Table 2 illustrates the diagonal score recognition of the confusion matrix for various architectures of the multilayer perceptron networks with the family instruments accuracy. Only one family is recognized with MIR score and 3 out of 4 are identified with MAR. This shows that there is more similarity between instruments of the same family and therefore best performance is attained.

Table 3 and 4 illustrate the impact of modifying the training and testing data set. The collection of database was divided in two different ways in order to construct a new set of the data training and testing. Then, two neural networks models with 120 unit cells in hidden layer were used to learn from each data distribution. The performance score was increased for 19 instruments from a rate of 16/19 to 18/19 at the

instrument level and remained with the same performance at instrument level family. The results reveals for having an optimal performance, we should take into account the way of choosing the training data.

instruments	Number of unit cells for MLP	
	120	120
Violin	71%	77%
Sop Sax	36%	39%
Tenor Trombone	62%	50%
Oboe	37%	29%
French Horn	23%	16%
Flute	20%	19%
EbClarinet	31%	32%
Cello	68%	66%
BbClarinet	74%	73%
Bass Trombone	23%	41%
Bassoon	51%	43%
Bass Flute	36%	40%
Alto Sax	46%	40%
Alto Flute	38%	26%
BassClarinet	24%	17%
Bass	78%	77%
Trumpet	52%	53%
Tuba	95%	91%
Viola	85%	83%
<b>MIR</b>	<b>7/19</b>	<b>10/19</b>
<b>MAR</b>	<b>9/19</b>	<b>8/19</b>
<b>SR(19)</b>	<b>16/19</b>	<b>18/19</b>
<b>SR(14)</b>	<b>12/14</b>	<b>14/14</b>

**Table 3:** Impact of training data set. *Score recognition basis instruments with MLP networks: SR(19) is the score for 19 instruments, SR(14) is the score recognition for 14 instruments using the criterion MIR and MAR*

families	Number of unit cells for MLP	
	120	120
String	83%	85%
Brass	68%	67%
Reeds	53%	51%
Flutes	43%	41%
<b>MIR</b>	<b>1/4</b>	<b>1/4</b>
<b>MAR</b>	<b>3/4</b>	<b>3/4</b>
<b>SR</b>	<b>4/4</b>	<b>4/4</b>

**Table 4 :** Impact of the training data set. *Score recognition for families instruments with MLP networks. SR is the score recognition using the criterion MIR and MAR*

With 14 instruments the score increased for 19 from a rate of 12/14 to 14/14. Compared to a previous work [Krishna 04] using the same database with 14 instruments, the Line Spectral Frequencies (LSF) as parameters and 46 Mixtures Gaussians Models (GMM), approximately the same performance were

observed.

instruments	RBF
Violin	100%
Sop Sax	39%
Tenor Trombone	14%
Oboe	50%
French Horn	17%
Flute	24%
EbClarinet	24%
Cello	62%
BbClarinet	74%
Bass Trombone	21%
Bassoon	47%
Bass Flute	37%
Alto Sax	47%
Alto Flute	25%
<b>MIR</b>	<b>7/14</b>
<b>MAR</b>	<b>4/14</b>
<b>SR(14)</b>	<b>11/14</b>

**Table 5:** *Score recognition for instruments with RBF network. SR is the score recognition using the criterion MIR and MAR.*

Table 5 and 6 report the score recognition respectively for the 14 instruments and the 4 families with the radial basis functions. The total family score is the same as the multilayer perceptron neural network but the score elements of the diagonal confusion matrix decrease significantly. However with the instruments only 11/14 are registered as score performance with the RBF network.

Families	RBF
String	41%
Brass	28%
Reeds	46%
Flutes	44%
<b>MIR</b>	<b>4/4</b>
<b>MAR</b>	<b>0/4</b>
<b>SR</b>	<b>4/4</b>

**Table 6:** *Score recognition for families instruments with RBF network. SR is the score recognition using the criterion MIR and MAR*

#### 4. Conclusion

Music processing, recognition, identification, segmentation is now an immense challenge and practical reality. In this study, we are interested in the instrument identification task in the monophonic context. We used only the spectral information as parameters by applying a discrete Fourier transform. Different architectures and models of neural networks were experimented as classifiers. We found that the spectral information is sufficient to discriminate well the family instrument but it cannot only work well to recognize each instrument particularly those coming

from the same family. The best score is obtained with back-propagation neural network with 120 unit cells at the hidden layer. It was also noted that the choice of the training and testing data play a crucial role and must be considered as another factor for enhancing performance of the instruments recognition cases.

## REFERENCES

- Brown J. C., Houix O. and McAdams S., Feature dependence in the automatic identification of musical woodwind instruments, *J. Acoust. Soc. Am.*, Vol. No. 3, pp. 1064-1072, 2001.
- Dubnov S. , Tishby N. and Cohen D., Polyspectra as measures of sound texture and timbre, *J. New Music Res.* 26, 277-314, 1997.
- Eggink Jana and Brown Guy, A missing feature approach to instrument identification in polyphonic music, *Proc. of ICASSP*, pp. V-553-556, 2003.
- Kashino K. and Murase H., A sound source identification system for ensemble music stream extraction. *Speech Comm.* 27, pp. 337-349, 1999.
- Marques J. and Moreno P., A study of musical instrument classification using gaussian mixture models and support vector machines, *Cambridge Research Labs Technical Report Series CRL/4*, 1999.
- Matlab, the language of technical computing, version 6.0.0.88 Release 12. Copyright Mathworks, inc., 2003.
- Krishna A.G and T.V Sreenivas, Music instruments recognition: from isolated notes to solo phrases, *Proc of ICASSP*, pp. IV-265-268, 2004.
- University of Iowa's Music Instrument Samples, <http://theremin.music.uiowa.edu/MIS.html>