Comparison of Features in Musical Instrument Identification Using Artificial Neural Networks

Róisín Loughran[1], Jacqueline Walker[1], Marion O'Farrell[1], Michael O'Neill[2]

[1] University of Limerick, Limerick, Ireland
{roisin.loughran, jacqueline.walker,
marion.ofarrell}@ul.ie
[2] University College Dublin, Belfield, Dublin 4, Ireland
{m.oneill}@ucd.ie

Abstract. This paper examines the use of a number of auditory features in identifying musical instruments. The Temporal Envelope, Centroid, Mel-frequency Cepstral Coefficients (MFCCs), Inharmonicity, Spectral Irregularity and Number of Spectral Peaks are all examined. By using these features to train a Multi-Layered Perceptron (MLP), it is determined that the MFCCs are the most efficient of these features in musical instrument identification. The Inharmonicity, Spectral Irregularity and Number of Spectral Peaks offered no benefit to the classifier. Of the instruments studied, the piano was most accurately classified and the violin was the least accurately classified instrument.

1 Introduction

The human ability to distinguish between musical instruments is a complex topic. A note played on a violin at the same pitch and loudness as a note played on a piano would sound different to a listener. It is this difference that helps the listener determine what instrument is playing. The quality of auditory sensation by which a listener can distinguish between two sounds of equal loudness, duration and pitch is defined as timbre [1]. Hence it can be said that aural recognition of an instrument is largely based on the timbre of that instrument. Unfortunately this definition really defines what timbre is not, as opposed to what it is. This lack of definition of timbre leads to further difficulties in measuring it – How can one measure what one cannot define? Many different features, such as the temporal envelope, spectral envelope or spectral centroid have been used to measure and explore the concept of timbre [2]. In this paper several features are examined and their effectiveness is compared in identifying musical instruments using a Multi-Layered Perceptron (MLP) as a classifier. Section 2 reviews some recent work in the area of instrument recognition. Section 3 outlines the proposal for the experiments and introduces the data used. Section 4 describes the results obtained and Section 5 contains some conclusions and proposes further work in the area.

2 Related Work

The field of musical sound analysis has produced numerous studies on timbre aimed at distinguishing between both musical families or groups and individual instruments. The multi-dimensionality of timbre and its lack of clear definition have contributed to the development of a 'timbre space'. Grey [3] asked human subjects to give 'similarity ratings' from pairs of notes. These similarity ratings were then used to create a timbre space by applying Multi-Dimensional Scaling (MDS). This idea of a timbre space has been explored in studies for distinguishing musical instruments. In [4], methods used in speech analysis were applied to musical sounds in order to construct a timbre space. The Mel-Cepstrum algorithm was applied to obtain parameters for the description of sounds and then Self-Organising Maps (SOM) and Principal Component Analysis (PCA) were applied to this data to produce a low-dimensional timbre space. This study concentrated on the spectral qualities of notes, specifically the Mel-frequency Cepstral Coefficients (MFCCs), and attempts to find a correlation between these and a physical timbre space. Although good spectral analysis was performed, no time varying qualities of sounds were examined in this study. Features were extracted from a wide range of musical instruments in [5]. Again, only spectral features were incorporated into this experiment. The features examined were analysed using a variety of different classification techniques. It was found that Quadratic Discriminant Analysis and Support Vector Machines showed comparable success rates in distinguishing between different instrument families.

Several studies have aimed to more distinctly identify specific musical instruments. Brown [6] distinguished between oboe and saxophone by calculating cepstral coefficients from training samples and applying a *k*-means algorithm to form clusters. Gaussian probability density functions were formed from the mean and variance of each of these clusters so that each sample from the test set could be classified using a Bayes decision rule. One significant difference in this study is that each sample contains numerous notes played on each instrument, rather than a single tone. Eronen and Klapuri [7] examined a wide range of temporal and spectral features from a large variety of orchestral instruments. The instruments were classified using a k-NN classifier and arranged into a hierarchical taxonomy. Martin and Kim [8] used features calculated from the log-lag correlogram rather than features based on the Short-Time Fourier Transform (STFT) to classify instruments hierarchically. This method is used to determine if it is better suited to represent inharmonic signals, as unlike traditional fourier methods, the correlogram does not assume that signals are periodic. Kaminsky and Materka [9] examined the short-term root mean square (RMS) energy envelope of a group of instruments and reduced this data using PCA. This data was then classified using an Artificial Neural Network and a Nearest Neighbour Classifier. Herrera et al [10], [11] give a more exhaustive account of various classification methods that have been used to distinguish between musical instruments.

3. Proposal

This paper aims to examine the effectiveness of some of the previously used features in distinguishing between musical instruments. Many of the studies mentioned in Section 2 classified a large number of instruments. It is debatable however, if enough samples were taken from these instruments to classify them accurately. From the number of samples quoted, it is unlikely that the instruments were sampled at different dynamic levels, or that different makes of each instrument were sampled. A large number of the studies mentioned in section 2, ([4], [5], [7] and [8]), use the MUMS (McGill University Master Samples) as training and test data. This database is used in the current study as test samples and it is known that although it contains many recordings from a wide range of instruments, there are no different dynamics or instrument makes within these samples. In [9], samples were recorded at different dynamic levels, although the range of each instrument was confined to just one specified octave. Preliminary experiments in classifying the data have shown that results are comparable whether the data is across the range of the instruments or confined to a smaller interval of just one octave on each instrument. The current study therefore examines samples across the physical pitch and dynamic range of each instrument. Furthermore, in the previous studies although numerous features were used, they were not compared against each other individually. It is the purpose of this study to discern which of the features examined are the most effective in instrument identification. This is achieved by repeating the experiment using different combinations of the features calculated. To ensure an exhaustive search of each instrument, the instrument selection was limited to the piano, violin and flute.

3.1 Data Sets

Samples were taken from the RWC Music Database (Music Instrument Sound) of the three selected instruments. Three makes of piano, Yamaha, Bosendorfer and Steinway were each sampled at dynamic levels *f, mf* and *p* across their range [12]. Violins manufactured by J.F Pressenda, Carcassi and Fiumebianca were sampled at these three dynamic levels with vibrato and at level *mf* without vibrato across their range [13]. Flutes manufactured by Louis Lot and Sankyo were sampled at the three levels both with and without vibrato [14]. In total this gave a training set of 2004 samples across the entire pitch range of the three instruments. In contrast, many of the previous studies already discussed use approximately 1000 samples ranging across 15 to 30 instruments.

The samples that make up the test dataset are from the MUMS database [15]. This smaller dataset consists of samples of the three instruments played at the same dynamic level. In total this dataset consists of 45 violin samples, 37 flute samples and 88 piano samples. A completely different dataset was used as a test set to realistically examine the generality of the trained classifier. It is hoped that the neural network described in this study, when trained properly, will be able to recognize a sample regardless of the source of the sound, as a human observer would. Using separate datasets, recorded under different conditions ensures that it is the tonal quality of the

instrument that is being categorized, and that superfluous qualities such as those arising from recording conditions or playing style, do not have an effect on the results.

3.2 Features Examined

This paper is focused on examining and comparing various features that have been used in distinguishing between musical timbres. The features examined are described below.

Temporal Envelope. The temporal envelope was found by calculating the root mean square (RMS) energy envelope of each sound, which was then filtered using a 3rd order low pass Butterworth filter with a cutoff frequency of 4410 Hz. This envelope was calculated over the length of each note and so includes temporal information on how the energy within the sound changes over time. Thus this envelope incorporates information regarding the attack time which has been shown to be of high importance to instrument classification [16], [17].

Evolution of Centroid. Physically the Centroid can be thought of as a measure of the power distribution of a sound, but perceptually it has been linked to the perceived quality of brightness [18]. While some of the previous experiments examined the average centroid, it is considered for this experiment that the evolution of the centroid over the duration of each note may be more informative. This shows how a specific spectral quality changes over the duration of the note. At sample k the centroid is calculated as:

$$\text{Centroid} = (\Sigma k f_k) / \Sigma f_k. \quad \textbf{(1)}$$

Mel-frequency Cepstral Coefficients. MFCCs are a way of representing the spectral information in a sound. They are commonly used in speech analysis, but have not been used extensively in music analysis as yet. Each coefficient has a value for each frame of the sound. This experiment examines how these coefficients change across the duration of the sound. MFCCs are a way of describing the sound using the Mel-scale. This Mel-scale is based on human perception of sound; it is a perceptual scale of pitches judged to be equidistant from one another. It can be approximated using a filterbank with triangular filters of constant bandwith up to 1 KHz and with a constant-Q above this frequency [4]. Obtaining the MFCCs involves analysing and processing the sound according to the following steps [19]:

- Divide the signal into frames
- Obtain the amplitude spectrum of each frame
- Take the log of these spectrums
- Convert to the Mel scale
- Apply the Discrete Cosine Transform (DCT)

This is implemented in Matlab using the melcepst function. These calculated coefficients change from frame to frame. These changes can be plotted as an envelope across the sound. The changes in these envelopes are distinctive to the instrument as illustrated below. Figure 1(a) shows the changes in the first MFCC for C5 on a piano whereas figure 1(b) shows the first MFCC for the same note on the flute.

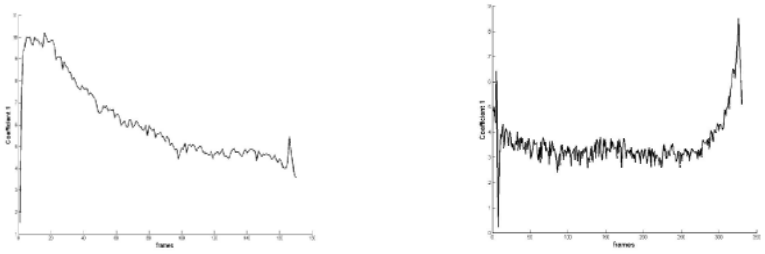


Fig. 1. Time-varying envelopes of the first MFFC for C5 on (a) a piano and (b) a flute

Inharmonicity. Inharmonicity is a measure of the deviation of upper partials from being perfect integer multiples of the fundamental frequency. Many instruments, such as the piano, rely on their upper partials being slightly 'detuned' to add warmth and character to their tone. As this quality is distinctive to each instrument, Inharmonicity may be used as a quality to identify instruments. The Inharmonicity of partial k is calculated from [2]:

$$I_k = (\Delta f_k/(k\Delta f_1)) - 1. \quad (2)$$

Spectral Irregularity. Spectral Irregularity is a measure of how much the actual spectral envelope varies in comparison to a smoothed version of itself. The smoothed version is calculated at each partial according to its two surrounding partials, and then the log of the variance of this smoothed version to the real envelope is calculated, [2]. Thus if at partial k the smoothed envelope Âk is calculated from:

$$\hat{A}k = (A_{k-1} + A_k + A_{k+1})/3. \quad (3)$$

Then the Spectral Irregularity may be calculated from the log of the standard deviation from each measured amplitude point to this smoothed envelope:

$$SIR = log[std(\hat{A}k - A_k)]. \quad (4)$$

Number of Peaks. In this experiment the Number of Peaks is defined as the number of spectral peaks that are at least one-tenth the strength of the strongest spectral peak (whether the strongest peak be the fundamental or not). This was calculated by performing a 512-point FFT on the sound, and examining the spectral peaks produced. Instruments with a rich timbre, such as the piano, contain many spectral peaks whereas those with a more pure tone, like the flute, tend to have fewer strong peaks. Hence the number of strong peaks is measured as an indicator of the richness of the sound produced.

3.3 Classification Methods

An MLP was used to classify the features described above. MLP use supervised learning so the number of possible categories is specified. A numerical result of how much a classified test sample fits into its given category can be obtained. Hence using MLPs can give a more quantifiable result than other clustering or self-organising neural networks such as Self-Organising Maps (SOMs). The features above are calculated from the training data (RWC samples) and used to train the MLP. Once this network is successfully trained, the corresponding features are calculated from the test samples. These features are then used to simulate the trained network so that the network can classify each test sample as a specific instrument. All features for each particular sample are presented to the network concurrently. This poses no problem for the Inharmonicity, Spectral Irregularity and Number of Peaks as they each contain only one data value per feature. The Temporal Envelope, Centroid and MFCCs, however, all consist of envelopes of data, giving multiple data values for each feature. Statistically much of this data is redundant and so a method must be employed to extract the most significant information from the data collected. This is achieved by applying PCA to the calculated features.

Principal Component Analysis. PCA is a standard technique commonly used in statistical pattern recognition and signal processing for performing dimensional reduction. PCA was implemented in Matlab for this experiment using the pca function in the Statistics Toolbox. Essentially the application of PCA transforms data orthonormally so that the variance of the data remains constant, but is concentrated in the lower dimensions. The matrix of data being transformed consists of one feature set (eg. Temporal Envelope) for each sample. The covariance matrix of this data matrix is calculated. The principal components for the data set can then be calculated from the eigenvectors of this covariance matrix [20]. This results in a set of principal components, ordered according to the percentage of explained variance from most to least. As such the most important data can be extracted, with minimum disruption to the original data collected. While this method may not result in particularly intuitive or meaningful data axes, it is an excellent method of reducing the dimensions of the calculated data.

Multi-layered Perceptron. MLPs are a specific type of Artificial Neural Network (ANN) that use supervised training to train multiple layers of interconnected perceptrons. MLPs contain at least one layer of hidden neurons – each of which includes a non-linear activation function, and they exhibit a high degree of connectivity [20]. These characteristics combine to make the theoretical analysis of an MLP difficult and as such the design of these systems is often, as in this case, unintuitive and based on trial and error. The network used in this experiment is trained using the backpropagation algorithm with two hidden layers of neurons. It was implemented in Matlab using the *newff* function from the Neural Network Toolbox. This was set up with a learning rate of 0.1 and a momentum constant of 0.95. It is batch trained with a Resilient Backpropagation Algorithm, *trainrp*, with a goal of 0.001 and a maximum number of epochs of 1000. This means that the network will keep training until it achieves an error value of 0.001 or below, or until it has tried to train the network 1000 times and fails. With this set up it was found that a network with 57 neurons in the first layer and two hidden layers containing 28 and 15 neurons respectively would be sufficient to train the data set. The details of this set up were found by trial and error and are not necessarily the only solution to designing a network such as this. It was found however that with this set-up the network could quickly process all of the data used in this experiment.

4. Results

The features discussed in Section 3 were all presented to the MLP in a variety of combinations to determine which of them were most useful in instrument identification. PCA was applied to some of the data sets to reduce the quantity of data values inputted to the MLP. The PCA algorithm implemented in Matlab results in a set of principal components as large as the original set of variables. As explained above, after the application of PCA most of the variance from these values will be concentrated in the first few components. To decide how many of these components to use, some preliminary experiments were run on the Temporal Envelope, Centroid and MFCC data sets.

4.1 Data Reduction on Temporal Envelope and Centroid

Once the principal components of each feature were calculated, the first three components were plotted to observe the separation between the instruments. This observed separation is an indication of how well the MLP will be able to categorise the samples. A plot of these 3 components for the Temporal Envelope and Centroid data can be seen in figures 2 (a) and (b) below. Clearly, the majority of points in these plots separate into 3 distinct regions, indicating that these features are indeed useful in identifying these instruments.

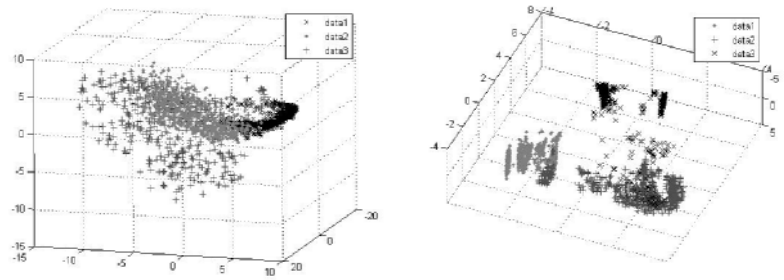


Fig. 2. Plot of the first 3 components of (a) Temporal Envelope and (b) Centroid Evolution

Although only 3 principal components can be plotted, the MLP is not limited to just 3 data values for each feature. The network was trained using 3, 4 and 5 principal components for each feature in order to determine the optimum method of incorporating these features into the classifier. For these experiments the MLP consisted of 57 neurons in the first layer with two hidden layers containing 22 and 8 neurons respectively. It was batch trained in Matlab with a Quasi-Newton Algorithm, *trainbfg*, with a goal of 0.001 and a maximum number of epochs of 1000. This is slightly different to the network described in Section 3.3. The Quasi-Newton Algorithm is a faster algorithm than the Resilient Backpropagation Algorithm but it requires more free memory to run. Hence it is not suitable for training and testing on all the features combined but it can be used in this instance where the data is largely reduced. Once the network is trained its accuracy is measured by simulating the network with the corresponding number of principal components of the test data, and noting the percentage of times the network classified the test instrument correctly. The results of these tests can be seen in table 1. The results indicate that 4 principal components give the optimum results for both sets of data. Inclusion of the 5^{th} component actually reduces this result, which may be due to the unintuitive manner in which PCA reduces data. It is not known what physical aspect, if any, each component relates to. It is possible that the 5^{th} component relates to a frequency or dynamic element of the sound that is not dependent on the type of instrument.

Table 1: Classification results of 3-5 principal components for the Temporal Envelope and Centroid Evolution data

# Principal Components	**Temporal Envelope (% correct)**	**Centroid Evolution (% correct)**
3	82.94	67.06
4	84.71	78.82
5	73.53	74.14

4.2 Number of MFCCs Used

As mentioned above, MFCCs have been used for some time in speech analysis. Implementing the Mel-Cepstrum algorithm gives a number of coefficients and it is not immediately obvious how many of these should be used. In speech analysis, it has been determined that 8-14 coefficients are sufficient to use and quite often 12 are chosen [21]. There has been no such recommendation for music analysis. In order to determine how many coefficients to include in this experiment, the network was trained and tested with different numbers of MFCCs to determine the optimum number to use. The accuracy of the trained network was judged as before – by determining the percentage of times the network trained on a specific number of MFCCs correctly identified the test instruments. Classification results from the first 3 principal components of the first 6-16 MFCCs are illustrated below in figure 3. These results indicate that results are consistently high when at least 10 MFCCs are included. The best result is obtained from using 15 MFCCs.

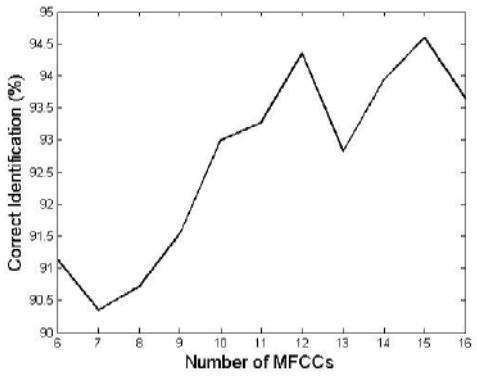


Fig. 3. Classification results for network trained and tested on 6-16 MFCCs

As before with the Envelope and Centroid data, these experiments were repeated to see how many principal components should be used. The bar chart in figure 4 indicates the classification accuracy on a network trained and tested on 3, 4 and 5 principal components from the first 11-16 MFCCs. Once again this shows that the best results are obtained each time from using 4 principal components.

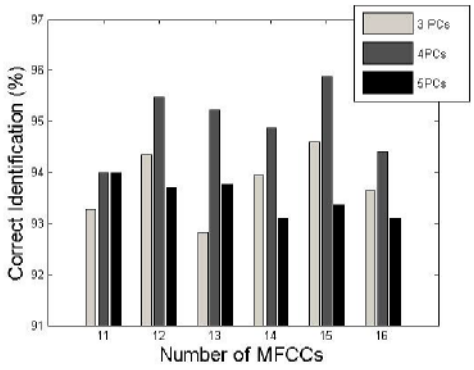


Fig. 4. Classification results of network trained on 3,4 and 5 principal components for 11-16 MFCCs.

4.3 Classification With All Features

As the optimum method of including the above features into the classifier has been established, the combination of these with the other features described in Section 3.2 can be examined. Table 2 displays the results from combinations of the 4 principal components from the Temporal Envelope, Evolution of the Centroid and the first 15 MFCCs. Again this table indicates the percentage of accurate classifications of test data from a network trained on these features. Each column of this table represents an experimental set-up – ie each 'X' indicates that this feature is used in this particular run of the experiment.

Table 2: Classification results from training on the Temporal Envelope, Centroid and MFCCs

Feature							
Envelope	X			X	X		X
Centroid		X		X		X	X
MFCC			X		X	X	X
% Correct:	**84.71**	**78.82**	**94.71**	**91.76**	**95.88**	**95.29**	**99.41**

These results clearly indicate that as individual features the MFCCs result in the most accurate classification, but also that more accurate results can be obtained by combining features. The highest result of 99.41% accuracy is obtained from using all three features. The rest of the features are now combined with this best result to see if it can be improved further. The classification results for these features are shown in table 3.

Table 3: Classification results from training on the Temporal Envelope, Centroid and MFCCs, combined with Inharmonicity, Spectral Irregularity and Number of Peaks

Feature							
Envelope	X	X	X	X	X	X	X
Centroid	X	X	X	X	X	X	X
MFCC	X	X	X	X	X	X	X
Inharm.	X			X	X		X
Spec. Ir		X		X		X	X
# Peaks			X		X	X	X
% Correct:	**97.06**	**98.82**	**97.06**	**95.29**	**97.65**	**93.53**	**94.12**

It can be seen from these results that rather than increasing the accuracy of the classifier, including these features actually decreases its performance. No matter which combination of these features is used with the three initial features, the accuracy is somewhat reduced. This is quite surprising, as these features have been used in numerous previous studies. It is possible that inclusion of these features 'overcomplicates' the input data, leading to misclassifications.

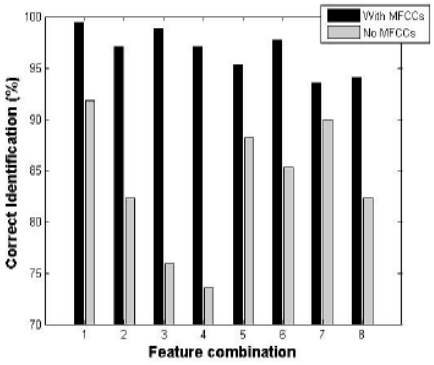
4.4 Training Without the MFCCs

One point to note about these features is that they are not all equal from a computational point of view. Inharmonicity, Spectral Irregularity and Number of Peaks all have only one data value each whereas the Temporal Envelope and the Centroid each have 4 values for each sound sample (one for each principal component). The MFCCs on the other hand have 4 principal component values for each of the first 15 coefficients. This gives 60 data values for each sound sample for this feature alone. Because of this the experiment was run again without the computational expense of the MFCCs, to discover how accurate the classifier could be without them. The results are shown in table 4.

It is clear from these results that the MFCCs are very important in instrument identification. None of the combinations come close to the accuracy of those achieved with the MFCCs present. A comparison of the results achieved from the various feature combinations both with and without the MFCCs present is illustrated in the bar chart in figure 5.

Table 4: Classification results from training on just the Temporal Envelope and Centroid, combined with Inharmonicity, Spectral Irregularity and Number of Peaks

Feature							
Envelope	X	X	X	X	X	X	X
Centroid	X	X	X	X	X	X	X
Inharm.	X			X	X		X
Spec. Ir		X		X		X	X
# Peaks			X		X	X	X
% Correct:	**82.35**	**75.88**	**73.53**	**88.24**	**85.29**	**90**	**82.35**

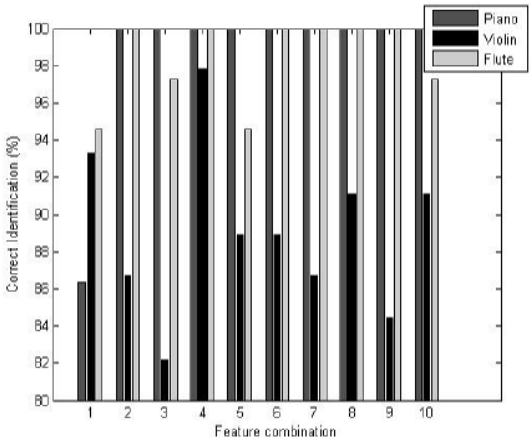


X-label	Features Included
1	Env, Cent
2	Env, Cent, Ih
3	Env, Cent, SIR
4	Env, Cent, #P
5	Env, Cent, Ih, SIR
6	Env, Cent, Ih, #P
7	Env, Cent, SIR, #P
8	Env, Cent, Ih, SIR, #P

Fig. 5. Comparison of classification accuracy of feature combinations with and without MFCCs

4.5 Classification of Specific Instruments

The above classification results are averaged across the 3 instrument sets. This section details the classification accuracy of the individual instruments. The training and test sets were run on all features as in Section 4.3 above. The classification results of the individual instruments are illustrated in the bar chart in figure 6. This indicates an interesting finding; regardless of which features are used, the violin is consistently the least accurately classified instrument. This lack of clarity in discerning the violin would indicate that the tone or timbre of the violin is somewhere 'in between' the timbres of the other two instruments. Clearly the piano is the most accurately classified instrument. It can be seen from figure 6 that the piano is recognised with 100% accuracy in every feature combination apart from the combination that does not include the MFCCs. The flute is the next most accurate with again achieving a 100% correct recognition rate for several of the feature combinations. The violin, however rarely achieves recognition accuracy of over 95%. This result encourages the inclusion of more instruments in further experiments to determine which other instruments are difficult to classify.



X-label	Features Included
1	Env, Cent
2	Cent, MFCC
3	Env, MFCC
4	Env, Cent, MFCC
5	Env, Cent, MFCC, Ih
6	Env, Cent, MFCC, Ih, SIR
7	Env, Cent, MFCC, Ih, #P
8	Env, Cent, MFCC, Ih, SIR, #P
9	Env, Cent, MFCC, #P
10	Env, Cent, MFCC, SIR, #P

Fig. 6. Comparison of Individual Instrument Classification using all features

5. Conclusion

The results described above have a number of implications for further work on such a classifier. It is evident from examining classification on the Temporal Envelope and the Evolution of the Centroid that these are important features and that they are best incorporated into such a classifier using 4 principal components. It was also discovered that the optimum number of MFCCs to use in such experiments is 15. More generally it can be said that of the features analysed here, the above results indicate that the most important feature in identifying musical instruments is the MFCCs. These combined with the Temporal Envelope and the Evolution of the Centroid give an accuracy of 99.41% when identifying a novel instrument. Surprisingly, the Inharmonicity, Spectral Irregularity and Number of Peaks were not found to be of significant benefit in a classifier such as this. Finally it was determined that of the three instruments examined, the piano had the highest rate of correct classification, the flute had the next highest and the violin was the least accurately classified instrument.

Future work in this topic would involve examining more features in this manner. It would also be advantageous to further generalise the classifier by training it on more instruments and to see how accurately these instruments are classified. As mentioned earlier in the paper, the design of an MLP is highly unintuitive and offers somewhat of a ‘black box’ solution to the problem. Using a more controllable network as a classifier may lead to greater insights into the manner in which the instruments are

classified. Hence it would be interesting to incorporate a more intuitive type of network such as an ARTMAP in this type of musical instrument identifier.

Acknowledgments

This study is funded by the Science Federation of Ireland (SFI), under the current National Development Plan and Strategy for Science Technology and Innovation (SSTI) 2006-2013.

References

1. ASA, Acoustical Terminology, New York: American Standards Association, New York, (1960)
2. Beauchamp, J. W.: Analysis, Synthesis and Perception of Musical Sounds, The Sound of Music. Springer Science & Business Media, LLC, New York (2007)
3. Grey, J. M. and Gordon, J. W.: Perceptual Effects of Spectral Modifications on Musical Timbres. J. Acoust. Soc. Am. 63**,** 1493--1500 (1978)
4. De Poli, G., Prandoni, P.: Sonological Models for Timbre Characterization. J. New Music Research, 26, 170--197 (1997)
5. Agostini, G., Longari, M., Pollastri, E.: Musical Instrument Timbres Classification with Spectral Features. EURASIP J. of Applied Sig. Proc., 1, 5--14 (2003)
6. Brown, J.: Computer Identification of Musical Instruments Using Pattern Recognition with Cepstral Coefficients as Features. J. Acoust. Soc. Am. 105, 1933--1941 (1998)
7. Eronen, A., Klapuri, A.: Musical Instrument Recognition Using Cepstral Coefficients and Temporal Features. In: IEEE International Conference on Acoustics, Speech and Signal Processing, pp. 753—756, Istanbul (2000)
8. Martin, K. D. and Kim, Y. E.: Musical Instrument Identification: A Pattern-Recognition Approach. In: 136th meeting of the Acoustical Society of America, Cambridge, MA 02139 (1998)
9. Kaminsky, I., Materka, A.: Automatic Source Identification of Monophonic Musical Instrument Sounds. In: IEEE Int. Conf. On Neural Networks, 1, 189--194 (1995)
10. Herrera-Boyer, P., Peeters, G., Dubnov, S.: Automatic Classification of Musical Instrument Sounds. J. New Music Research, **23**, 3--21 (2003)
11. Herrera, Pamatriain, X., Batlle, E., Serra, X.: Towards Instrument Segmentation for Music Content Description: A Critical View of Instrument Classification Techniques. In ISHMIR, (2000)
12. RWC Music Database: RWC-MDB-I-2001-W01, Instrument No.1: Pianoforte
13. RWC Music Database: RWC-MDB-I-2001-W05, Instrument No.15: Violin
14. RWC Music Database: RWC-MDB-I-2001-W09, Instrument No.33: Flute
15. McGill University Music Master Samples: http://www.mcgill.ca/music/events/samples/
16. McAdams, S., Winsberg, S., Donnadieu, S., De Soete, G., Krimphoff, J.: Perceptual Scaling of Synthsized Musical Timbres: Common Dimensions, Specificities, and Latent Subject Classes. Psychological Research, 58, 177--192 (1995)
17. Jensen, K.: Timbre Models of Musical Sounds. Unpublished Doctoral Dissertation, Department of Computer Science, University of Copenhagen, Copenhagen, Denmark (1999)

18.Logan, B.: Mel Frequency Cepstral Coefficients for Music Modelling. In ISMIR (2000)
20.Haykin, S.: Neural Networks A Comprehensive Foundation. Prentice Hall International (UK) Limited, London, (1999)
21.O'Shaughnessy, D.: Speech Communication Human and Machine, Addison-Wesley Series in Electrical Engineering, (1987)