Musical Instrument Identification Using Principal Component Analysis and Multi-Layered Perceptrons

Róisín Loughran
University of Limerick, Limerick, Ireland
roisin.loughran@ul.ie

Jacqueline Walker
University of Limerick, Limerick, Ireland
jacqueline.walker@ul.ie

Michael O'Neill
University College Dublin, Dublin, Ireland
m.oneill@ucd.ie

Marion O'Farrell
University of Limerick, Limerick, Ireland
marion.ofarrell@ul.ie

Abstract

This study aims to create an automatic musical instrument classifier by extracting audio features from real sample sounds. These features are reduced using Principal Component Analysis and the resultant data is used to train a Multi-Layered Perceptron. We found that the RMS temporal envelope and the evolution of the centroid gave the most interesting results of the features studied. These results were found to be competitive whether the scope of the data was across one octave or across the range of each instrument.

1. Introduction

Musical sound analysis and source identification has been a subject of investigation over the past number of years. Most people possess the ability to distinguish between familiar musical instruments. Although specific a priori knowledge of the instrument may lead to a very certain distinction e.g. a double bass is known to have a much lower pitch range to a violin, in general instruments are identifiable even when played at the same pitch and loudness. As defined in [1] that quality of auditory sensation by which a listener can distinguish between two sounds of equal loudness, duration and pitch is known as timbre. Unfortunately, unlike pitch and loudness, timbre is a quality that has proven to be somewhat difficult to measure or quantify.

In this paper, Section 2 reviews some of the more relevant automatic classifiers that have been developed in recent years. An introduction of the data used, the features extracted and the methods used for classification is given in Section 3. Section 4 outlines the results obtained and finally Section 5 discusses conclusions that can be drawn from the results.

2. Related Work

Research into timbre and instrument classification has become more popular in recent years. In [2], 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 provides good spectral analysis, but no temporal measures were incorporated in the analysis. Features were extracted from a wide range of musical instruments in [3]. These were analysed using a variety of different classification techniques. It was found that Quadratic Discriminant Analysis performed best in distinguishing between instrument families.

Experiments to distinctly identify specific musical instruments have also been reported in recent years. Brown [4] distinguished between oboe and saxophone by calculating cepstral coefficients and applying a *k*-means algorithm to form clusters. Eronen and Klapuri [5] examined a wide range of temporal and spectral features from a large variety of orchestral instruments. Martin and Kim [6] used features calculated from the log-lag correlogram rather than features based on the Short-Time Fourier Transform (STFT) to classify instruments hierarchically. Kaminsky and Materka [7] examined the RMS 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 [8], give a more exhaustive account of various classification methods that have been used to distinguish between musical instruments.

3. Proposal

This study proposes to create and automatic musical instrument classifier by extracting and examining relevant features. These features are used as representations of the timbre of the instrument. The effectiveness of each of these features is examined on a number of instruments as explained in this section.

3.1 Training and Test Datasets

In classification studies, such as this one, the range and specifications of the samples used and the manner in which they are analysed are imperative to the accuracy and consistency of the result. Many of the studies mentioned in Section 2 classified a large number of instruments. From the number of samples quoted, it is unlikely that multiple samples for each instrument were included. It was decided for this study to exhaustively search just three instruments – the piano, violin and flute. Samples were taken from the RWC Music Database (Music Instrument Sound) of these 3 instruments. Three makes of piano, Yamaha, Bosendorfer and Steinway were each sampled at dynamic levels *f*, *mf* and *p* across their range [9]. Violins manufactured by J.F Pressenda, Carcassi and Fiumebianca were sampled at these three loudness levels with vibrato and at level *mf* without vibrato across their range [10]. Flutes manufactured by Louis Lot and Sankyo were sampled at the three levels both with and without vibrato [11]. In total this gave 2004 samples across the entire pitch range of the three instruments.

The samples that make up the test dataset are from the MUMS (McGill University Master Samples) database [12]. This smaller database 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. Each instrument was sampled and recorded across their entire range. A completely different dataset from the training set was used, as this should test the generality of this classifier.

3.2 Features Examined

It is evident from the literature reviewed that both temporal and spectral features are necessary in order to give and accurate description of timbre. The features first examined in this study comprised of the temporal envelope, spectral envelope, temporal residual envelope, spectral residual envelope and the evolution of the centroid.

3.2.1. Temporal and Spectral Envelopes. The temporal envelope was found by calculating the RMS energy envelope of each sound, which was then filtered using a 3rd order low pass Butterworth filter. 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 [13]. The temporal envelope was then subtracted from the original sound to find the residual. The temporal residual envelope was calculated from the RMS of this residual.

The spectral envelope was calculated from the envelope of the FFT of the sound. The FFT of a sound contains a measure of the spectral content of a sound. Taking the envelope of this measure will give some indication to the number and strength of the partials present. The spectral residual envelope was found by taking the FFT of the temporal residual calculated above.

3.2.2. Evolution of the Centroid. Physically the centroid can be thought of as a measure of the power distribution, but perceptually it has been linked to the perceived quality of brightness [14]. 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 gives an indication of how a specific spectral quality changes over the duration of the note. The centroid is calculated as:

$$\text{Centroid} = (\Sigma_k f_k) / \Sigma f_k .$$

Where f_k is the frequency at sample k.

3.3 Classification Methods

It was decided to use a MLP to classify the features described above. These features are calculated from the training data (RWC samples) and used to train an MLP. The features described above, however, have too many points per feature to be useful to the MLP and as such this data needs to be reduced. This is achieved by applying PCA to the calculated features. Essentially it transforms data orthonormally so that the variance of the data remains constant, but is concentrated in the lower dimensions. This results in a set of principal components, the first of which comprises the maximum variance of the data, the second the next highest variance and so on, [15].

MLPs are a specific type of Artificial Neural Network (ANN) that use supervised training on multiple layers of interconnected perceptrons. MLPs contain at least one layer of hidden neurons – each of

which includes a non-linear activation function exhibiting a high degree of connectivity [16]. 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 is worth mentioning that the above method is computationally quite expensive. The current experiment is implemented in Matlab, and so the run-time is largely dependent on the processor speed of the machine on which it is compiled. Both PCA and the training of the MLP involve a large number of calculations. Implementing this study in real-time is not considered here but if it was to be in the future, some complexity analysis on these calculations would need to be undertaken.

4. Results

4.1 PCA Results

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.

4.2.1. Results Over One Octave. A 3-dimensional plot of the first three principal components from the temporal envelope across the range C5 to C6 can be seen below in *figure 1*. This plot shows both the training and the test data sets on the same plot. This is encouraging as the three instruments can clearly be seen to segregate from each other. The piano samples have very clearly segregated themselves into a distinct group. This is not surprising as the strong attack in the envelope of the piano is very distinct from the other two more sustained instruments. The violin and flute samples also segregate, but there is some overlap between the two. Hence another feature is needed to distinguish these instruments distinctly.

A similar plot of the principal components extracted from the Centroid Evolution data is shown in *figure 2*. Again this shows quite good separation between the instruments. In particular the flute samples are distinctly segregated from the rest of the samples. This clear distinct separation between samples is a good indication that these principal components from these measures would be useful input to the MLP.

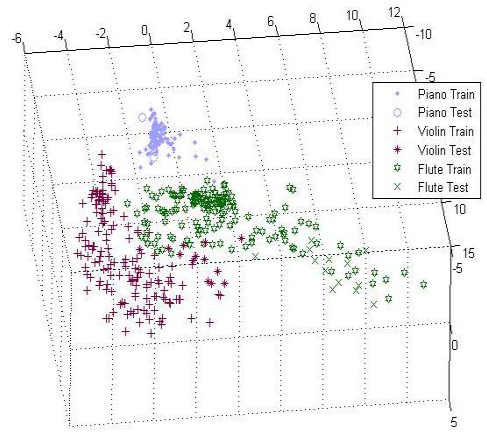

Figure 1. Plot of the first 3 principal components of Envelope data across one octave of each instrument

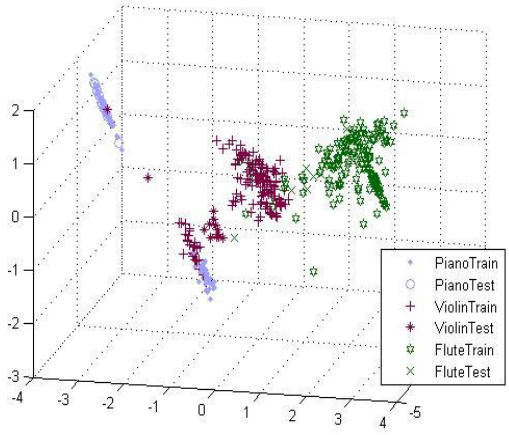

Figure 2. Plot of the first 3 principal components of Envelope data across one octave of each instrument

4.2.2. Results Over Range of Instrument. The plot obtained from PCA on the Temporal Envelope data of the entire training set of data is shown in *figure 3*. Again this shows good separation between the instruments, again particularly with the piano. The Centroid Evolution shown in *figure 4* also displays good separation between the instruments when the whole range of each instrument is examined. Again as with the envelope data there is much more overlap between the instruments and as such the boundaries between the instruments are not always clear.

The spectral and both the temporal residual and spectral residual envelopes, however, did not provide such a useful separation between instruments. The plots obtained from their 3 principal components did not separate out clearly. As the spectral envelope is a frequency measure, it is possible that it would be more

useful to determine between the pitches of the notes. The large amount of pitches used (88 separate pitch for each piano set) may have proved too difficult for the PCA to reduce between instruments.

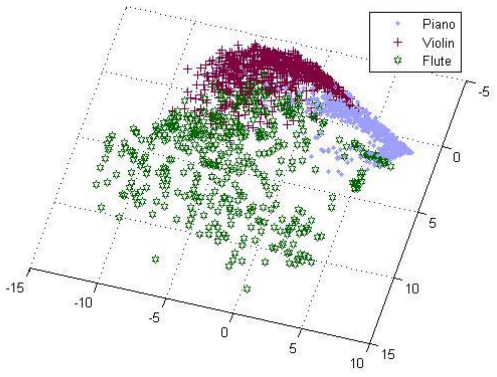

Figure 3. Plot of the first 3 principal components of the Envelope data across the range of the instruments

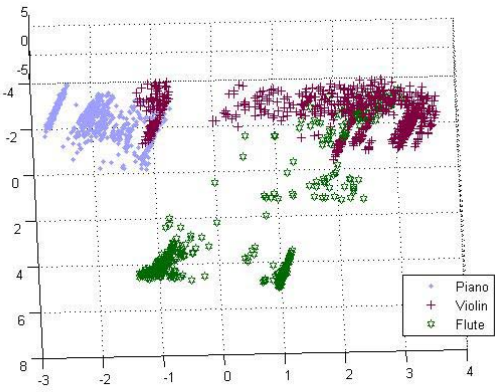

Figure 4. First 3 principal components for the centroid data across the range of the instruments

4.2 Multi-Layered Perceptron Classification

Once the data had been reduced and the principal values extracted, these values were used to train a MLP. Our MLP 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 Quasi-Newton Algorithm, *trainbfg*, with a goal of 0.001 and trained up to maximum epochs of 1000. With this set up it was found that a network with 57 neurons in the first layer and two hidden layers containing 22 and 8 neurons respectively would be sufficient to train the larger data set. A smaller network would most likely train the smaller one-octave set, but it was decided to use this set-up for both data sets for ease of comparison.

4.2.1. Classification Over One Octave. The classification results over one octave of each instrument, are shown in *table 1* below. This indicates the percentage of times the network trained on the training samples correctly identified a new test sample. The experiment was repeated for the first 3, 4 and 5 principal components to see if the inclusion of more data was worthwhile. As can be seen, choosing 4 principal components from the temporal envelope data produces the most accurate results. On the other hand varying the number of principal components for the centroid data does not seem to have much effect – the results are consistently high. These results may seem somewhat unusual – that increasing the amount of principal components can increase accuracy in one instance yet not in another. The manner in which PCA reduces data is quite unintuitive however. It is not known what physical aspect each component relates to – or indeed if it does relate to one. This lack of intuitiveness is a drawback of PCA, however its ability to reduce data so efficiently encourages us to overlook this drawback. The consistent results in the centroid data are most likely due to the small data set tested. The next section gives the results of the larger dataset.

4.2.2. Classification Over Range of Instrument. A network of similar structure to that used above was also used to examine the larger data set. This network was trained with the RWC samples across the range of each instrument and then tested using the MUMS samples across the same range. The results can be seen below in *table 2*. Although the accuracy of classification using the centroid data has diminished, it can be seen that increasing the number of principal components used may increase the accuracy. As before, this does not work for the envelope data however and may decrease results. It is evident from *table 1* and *table 2* that the overall best performance was obtained from the centroid data across the one-octave range. This performance clearly decayed across the range of the instrument but still gave encouraging results considering the increase in the search space was from one octave to over seven octaves in the case of the piano.

Table 1. Classification Results for samples ranged across one octave

# PCs	Temporal Envelope (% correct)	Centroid Evolution (% correct)
3	69.23	92.31
4	76.92	92.31
5	74.36	92.31

Table 2. Classification Results for samples across the range of the instruments

# PCs	Temporal Envelope (% correct)	Centroid Evolution (% correct)
3	82.94	67.06
4	81.76	78.82
5	73.53	74.14

5. Conclusion and Further Work

From the PCA plots obtained, it can be concluded that the features found to be most useful for separating these sounds were the temporal envelope and the evolution of the centroid across the sound. This agrees with previous literature that has found these features to be perceptually very important [13]. Other features examined – the residual envelope and spectral envelope did not produce such good results. It is planned to continue this method of investigation by looking at other features such as spectral irregularity, inharmonicity and Mel-Frequency Cepstral Coefficients among others. As discussed before, the MLPs offer somewhat of a 'black box' solution to this problem and so other types of ANN, that offere more control over the system, such as an ARTMAP, may be investigated, to compare and confirm the results.

These results show that the best classification was seen in the centroid data across one octave. Increasing the range decreased the accuracy in classification but still gave encouraging results for pursuing classification across the physical range of instruments. An interesting point about these results is that rather than automatically reducing the accuracy of the classifier by increasing the range of notes examined, in the case of the temporal envelope data the accuracy of the classifier actually increased. This is particularly interesting, as most preceding studies on this topic have purposely constricted the range of notes so that only a common pitch range is studied across each instrument. These results show that in fact widening the search space to a more realistic range can in some cases be beneficial to the system. Hence, future studies can confidently continue with developing an automatic instrument classifier across the natural range of instruments.

6. Acknowledgments

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

7. References

1. ASA, Acoustical Terminology, New York: American Standards Association, New York (1960)
2. De Poli, G., Prandoni, P.: Sonological Models for Timbre Characterization. J. New Music Research, **26**, 170-197, (1997)
3. Agostini, G., Longari, M., Pollastri, E.: Musical Instrument Timbres Classification with Spectral Features. EURASIP J. of Applied Sig. Proc., **1**, 5-14 (2003)
4. Brown, J.: Computer Identification of Musical Instruments Using Pattern Recognition with Cepstral Coefficients as Features. J. Acoust. Soc. Am. **105**, 1933-1941 (1998)
5. 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 (2000)
6. 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)
7. Kaminsky, I., Materka, A.: Automatic Source Identification of Monophonic Musical Instrument Sounds. In: IEEE Int. Conf. On Neural Networks, **1**, 189-194 (1995)
8. Herrera, Pamatriain, X., Batlle, E., Serra, X.: Towards Instrument Segmentation for Music Content Description: A Critical View of Instrument Classification Techniques. In ISHMIR, (2000)
9. RWC Music Database: RWC-MDB-I-2001-W01, Instrument No.1: Pianoforte
10. RWC Music Database: RWC-MDB-I-2001-W05, Instrument No.15: Violin
11. RWC Music Database: RWC-MDB-I-2001-W09, Instrument No.33: Flute
12. http://www.music.mcgill.ca/resources/mums/html/mums.html
13. 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)
14. Jensen, K.: Timbre Models of Musical Sounds. In: Department of Computer Science, University of Copenhagen (1999)
15. http://www1.cs.columbia.edu/~jebara/htmlpapers/UTHESIS/node64.html
16. Haykin, S.: Neural Networks A Comprehensive Foundation. Prentice Hall International (UK) Limited, London, (1999)