# Toward real-time recognition of acoustic musical instruments

Angela Fraser and Ichiro Fujinaga

Peabody Conservatory of Music
Johns Hopkins University
Baltimore, MD USA 21202
ich@peabody.jhu.edu

A real-time timbre recognition system based of Miller Puckette's fiddle program was tested using the attack portions of acoustic musical instruments. The dynamically changing spectra are quantified by the velocities of the integral, the centroid, the standard deviation, and the skewness of the spectra and the velocity of the estimated pitches. The mean and the standard deviation of the five parameters were also calculated. These features were stored in the database for an exemplar-based learning system, which is based on a k-nearest neighbor classifier. The system is enhanced by a genetic algorithm, which finds the optimal set of feature weights to improve the recognition rate.

Compared to a previous experiment using only the steady-state portion of the sounds, the current system achieves 10-20% increase in the recognition rates. For example, the recognition rate of the trumpet, the clarinet, and the violin using the dynamic spectra was 98% compared to 80% using the steady-state spectrum and the recognition rate involving 39 timbre improved from 50% to 64%.

## 1. Introduction

In many musical situations, it is often useful to know not only the pitch of the sound but the instrument that is producing the sound. A real-time timbre recognition system based of Miller Puckette's fiddle program (Puckette et al. 1998) was tested using the attack portions of acoustic musical instruments (Puckette's accompanying program bonk is designed to recognizes timbre of percussion instruments). The dynamically changing spectra are quantified by the movment of centroid and other moments of the spectra. These and other features were stored in the database for an exemplar-based learning system, which is based on a k-nearest neighbor classifier. The system is enhanced by a genetic algorithm, which finds the optimal set of feature weights to improve the recognition rate.

## 2. Features involving moments

The method of moments is a versatile tool for decomposing arbitrary shape into a finite set of character features. In general, moments describe numeric quantities at some distance from a reference point or axis. Moments are commonly used in statistics to characterize the random variable distribution and in mechanics to characterize bodies by spatial distribution mass. Here, the spectral shape is considered to be a density distribution function.

For each spectral function in an analysis window (2048 points), the following were calculated: the mass or the integral of the curve (zeroth-order moment), the centroid (first-order moment), the standard deviation (square root of the second-order central moment), the skewness (third-order central moment), and the estimated pitch. In order to capture the dynamic evolution of the spectra, the total distance traveled over multiple analysis windows by each of the five principle features was calculated:

$$\sum_{i=1}^{N-1} \left| x_i - x_{i-1} \right|$$

where x is a principle feature and N is the number of analysis windows (varied from 3 to 20). The mean and the standard deviation of the distances were also used as features. A snapshot of the spectrum of the last analysis window is also included using the five principle features of the last window. If an attack has not been detected by the fiddle program at the last window, the snapshot is taken at the window when the attack is reported.

Thus the velocity (distance), the mean, and the standard deviation of each of the five principle features and the five features of the spectrum at the last analysis window, for a total of 20 features, were stored for each pitch of each instrument.

## 3. Exemplar-based classifier

An exemplar-based classifier is implemented by the k-NN classifier (Cover and Hart 1967), which is a classification scheme to determine the class of a given sample by its feature vector. Distances between feature vectors of an unclassified sample and previously classified samples are calculated and the class represented by the majority of k-nearest neighbors is then assigned to the unclassified sample.

This type of classifier is based on the idea that objects are categorized by their similarity to one or more stored examples. It differs both from rule-based or prototype-based models of concept formation in that it assumes no abstraction or generalizations of concepts (Nosofsky 1984, 1986). The flexibility of the classifier is apt here because many different types of musical timbre and performance styles exist.

The particular implementation of k-NN classifier was enhanced by modifying the feature space, or equivalently, changing the weights in the distance measure (Kelly and Davis 1991). A commonly-used weighted-Euclidean metric between two vectors **X** and **Y** in an N-dimensional feature space is defined as:

$$d = \left( \sum_{i=0}^{N} \omega_i \left( x_i - y_i \right)^2 \right)^{1/2}$$

By changing the weights $\omega_i$ the shape of the feature space can be changed. A genetic algorithm (Holland 1975) was used to determine the values of the weights in order to maximize the recognition rate (Wettschereck, Aha, and Mohri 1997). The standard leave-one-out procedure was used to calculate the recognition rate.

## 4. Data

The data used in this experiment was taken from the orchestral instrument tones from the McGill University Master Samples, which are digital recordings of live musical performers. As in the previous experiment (Fujinaga 1998), 39 different timbre from 23 orchestral instruments, some with different articulations played at different pitches were used. These were: violin, Bb clarinet, C trumpet, oboe, violin pizzicato, flute, French horn, trombone, cello, bassoon, bass, viola, tuba, bass clarinet, cello pizzicato, bass flute, alto flute, piccolo, contrabassoon, English horn, conrabass clarinet, bass clarinet, Eb clarinet, French horn mute, bass trombone, alto trombone, trombone mute, Bach trumpet, Bass martele, bass mute, cello martele, cello mute, viola martele, viola mute, viola pizzicato, violin martele, violin mute, and violin ensemble.

Each sound file in the McGill CDs contains one instrument played at different pitches with a short pause between each note. To simulate real live environment, these sound files were used as the direct input to the fiddle program without any editing.

## 5. Program

The fiddle program (Puckette et al. 1998) is a robust and efficient real-time pitch detection software available for various platforms. The Pd version of the program was slightly modified to process data from sound files and to generate spectral features for the analysis. The list of spectral peaks (frequencies and amplitudes) generated to estimate the pitch is used to calculate the spectral centroid and other features for identifying instruments.

The FFT window size in the fiddle object was set to 2048 samples with a hop size of 1024 samples. For each sound, up to 20 windows from the beginning of the sound were analyzed.

Although a considerable amount of time is needed for the genetic algorithm to determine the set of weights, the calculation time of the actual k-NN classifier is insignificant and can be performed in real-time.

## 6. Results

Table I shows the recognition rates for various numbers of instruments and two different analysis lengths. The number of windows 3 and 20 corresponds to approximately 93 ms and 490 ms, respectively.

|  | Number of instruments |  |  |  |  |  |
|---|---|---|---|---|---|---|
| Number of windows | 3 | 5 | 8 | 10 | 15 | 39 |
| 3 | 98.3 | 92.9 | 88.7 | 81.5 | 74.4 | 54.3 |
| 20 | 93.9 | 92.6 | 88.0 | 86.5 | 81.6 | 63.6 |

Table I. The recognition rate (%) for a various number of instruments and two analysis lengths.

Compared to the previous experiment using only the steady-state portion of the sounds, the current system achieves 10-20% increase in the recognition rate. For example, the recognition rate of the trumpet, the clarinet, and the violin using the dynamic spectra is 98% compared to 80% using the steady-state spectrum and the recognition rate involving 39 timbre improved from 50% to 64%. These results are comparable to those reported by Martin and Kim (1998) who also used the McGill CDs. Note that these results are considerably better than reported human performances (Kendall 1986, Saldanha and Corso 1964).

In a typical ensemble situations, one is usually dealing with handful of instruments. Results here show that a reliable real-time system for recognizing instruments within the attack portion (less than 100 ms) is possible.

## 7. Conclusions

Using exemplar-based learning model, the computer was able to identify orchestral instruments quite accurately and quite quickly, especially when a small number of instruments were involved. These results indicate the feasibility of using timbre recognition tasks in real-time and in real-life applications. The obvious next step in this research is to experiment with other sources of timbre.

## Bibliography

Cover, T., and P. Hart. 1967. Nearest neighbor pattern classification. *IEEE Transactions on Information Theory* 13 (1): 21–7.

Fujinaga, I. 1998. Machine recognition of timbre using steady-state tone of acoustic musical instruments. *Proceedings of the International Computer Music Conference*. ??-?.

Holland, J. H. 1975. *Adaptation in natural and artificial systems*. Ann Arbor: U. of Michigan Press.

Kendall, R. A. 1986. The role of acoustic signal partitions in listener categorization of musical phrases. Music Perception. 4 (2): 185–214.

Martin, K. D., and Y. E. Kim. 1998. Musical instrument identification: A pattern-recognition approach. Paper read at the 136th meeting of the Acoustical Society of America.

Nosofsky, R. M. 1986. Attention, similarity, and the identification categorization relationship. *Journal of Experimental Psychology: General* 115 (1): 39–57.

Nosofsky, R. M. 1984. Choice, similarity, and the context theory of classification. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 10 (1): 104–14.

Puckette, M. S., T. Apel, and D. D. Zicarelli. 1998. Real-time audio analysis tools for Pd and MSP. *Proceedings of the International Computer Music Conference*. ??-?.

Saldanha, E. L., and J. F. Corso. 1964. Timbre cues and the identification of musical instruments. *Journal of the Acoustical Society of America*. 36(11): 2021–6.

Wettschereck, D., D. W. Aha, and T. Mohri. 1997. A review and empirical evaluation of feature weighting methods for a class of lazy learning algorithms. *Artificial Intelligence Review* 11 (1–5): 272–314.