CLASSIFICATION OF PIZZICATO AND SUSTAINED ARTICULATIONS

G. E. Hall, H. Ezzaidi

Université du Québec à Chicoutimi,
Department of Applied Sciences,
555, boul. de l'Université,
Chicoutimi, Qc, Canada, G7H 2B1.
glennerichall@uqac.ca, hezzaidi@uqac.ca

+M. Bahoura, C. Volat

+Université du Québec à Rimouski,
Department of Engineering,
300, allée des Ursulines,
Rimouski, Qc, Canada, G5L 3A1.
Mohammed_Bahoura@uqar.ca,cvolat@uqac.ca

ABSTRACT

Musical instrument recognition has recently received growing attention from the research community and music industry. It plays a significant role in multimedia applications. Many approaches have been proposed to classify musical instruments. Particularly, the articulation refers to the style in which a song's note is played. In this paper, we propose a new avenue for musical instrument classification into two categories: Pizzicato and Sustain articulations. New features derived from chromagram contours are investigated by using the classical invariant moments. A comparison with a reference system using a feature vector constructed from 38 feature parameters and using *k*-NN classifier is provided. The standard RWC database is used for all experiments.

1. INTRODUCTION

The frequency range of signals resulting from the vibration of material (air) is located between the infrasound and ultrasound. Specifically, sounds audible to the human ear is between 20 Hz and 20000 Hz. The sound waves upon arriving at the ear are analyzed by the auditory cortex to extract and generate a characteristic auditory sensation such as speech, noise, wind, whistling, musical notes, etc. Music is the sound composition class that produces an auditory sensation similar to what a painter produces with his brush as sensation to human eye. Music can be defined as a state of sound art to express joy, sadness, melancholy, humor, anger, affection, in short our state of being. Unlike speech, music signal contains a wide variety of descriptors characterizing the wealth of information contained in the audio signal. Some of this information is often: pitch, harmony, beat, rhythm, melody, onset, offset, attack, melody, chorus, meter, timbre, artist identity, genre, etc. The music composition and analysis are fundamentally built on 4 basic elements that are often interrelated: melody, harmony, rhythm and arrangement. The timbre (also called color) is another characteristic element of the musical instrument that allows us to distinguish between two instruments playing the same melody with the same sound's pitch, and loudness. In fact, each note from musical instrument may have an even greater variety of frequencies according to the type of instrument. The number and energy associated to the fundamental, the all harmonics and their relationship to each other, create the different musical timbre. Therefore, there is a direct relationship between timbre and musical instrument identification. The challenge is to determine which attributes characterize best the multidimensional perceptual timbre. Psycho-acousticians sketches timbre as a geometric construction built from similarity ratings. Multidimensional scaling is generally used to find sound attributes that correlate best with the perceptual dimensions (brightness, smoothness, compactness, etc.) [1, 2]. From the same idea, research in musical instrument identification began with the construction of a vector space describing the timbre or commonly named the space timbre. The main idea is to reduce the dimension of the feature vectors while preserving the natural topology of the instrument timbre; a practical interpretation should emerge. Recently, many works are based on the hierarchical natural taxonomy to achieve the task of musical instrument recognition. Natural taxonomy separates in the first step, the pizzicato instruments, whose attack is abrupt and the sustained instruments, where the holding time is constant. Pizzicato instruments have particularity that the excitation source is given by a pulse and the holding time depends on the intensity of the pulse. Sustained instruments have the particularity that the excitation source is applied consistently until you release the note. At the second level, the instruments are grouped by family and mode of production. In the subclass of pizzicato instruments, only one family is present (stringed instruments). For the subclass of sustained instruments, four groups of instruments are present: brass, flutes/piccolo, reed and stringed instruments. In this paper, we are interested to propose new parameters for the instruments classification into two categories: pizzicato and sustained. The proposed parameters give a new description based only on chromagram contours allowing the identification of both tonal content and the timbre instrument (identity). A comparison with a reference system using most common components in domain is presented.

The proposed sub-system allows recognition systems to proceed with a first hierarchical classification.

2. RELATED WORK

All features proposed in the last years attempt to describe the multidimensional vector representing the perceptive human sensation into the timbre space. Since several decades, various parameters derived from time attack, time release, spectral centroid, harmonic partials, onset, and cutoff frequency exhibit relevant information to characterize quality attributes of timbre instruments as orchestral instruments, bowed string, brightness, harmonic and inharmonic structure, etc [1, 2, 3]. Recently, many features related to characterization of the excitation sound source and the resonant instrument structure extracted from transformed correlogram were suggested in [4]. All the 31 features extracted from each tone based on statistical measures are related to pitch, harmonic structure, attack, tremolo and vibrato proprieties. They are assumed to capture a partial information of tone color (timbre). In addition, assuming that the human auditory perception system is organized to recognize sounds in a hierarchical manner, a similar classification scheme was suggested and compared in the same work [4]. Results show a score improvement about of 6% for individual instrument and 8% for instrument family recognitions. Instead, Eronen [5] exploited the psychoacoustic knowledge to determine feature parameters describing the music timbre. Essentially, statistical measures based pitch, onset, amplitude modulation, Mel Frequency Cepstral Coefficients (MFCC), Linear Predictive Coding (LPC) and their derivatives are investigated as parameters. Results show that the MFCC and derivatives extracted from the onset and steady state segments give mainly the best performance, comparatively to others aggregated features. Performance comparison between direct and hierarchical classification techniques was examined in [4, 6] showing a particular interest with the last technique. Particularly, Hall et al. [6] used 6698 notes with the hierarchical classification proposed in [5] and constructed a system where the feature vector is dynamic and changes depending on each level and each node of the hierarchical tree. The feature vector was thus optimized and determined with the Sequential Forward Selection (SFS) algorithm. Using the Real World Computing (RWC) music database [7], the results showed a score gain in musical instrument recognition performance [6]. Kitahara et al. [8, 9] used pitch-dependent algorithms as an F0-dependent multivariate normal distribution, where each element of the mean vector is represented by a function of F0.

Table 1. Database Description.

Instrument	Notes	Instrument	Notes
Accordion	282	Acoustic Guitar	463
Alto Sax	198	Banjo	208
Baritone Sax	198	Bassoon (Fagotto)	240
Cello	377	Clarinet	240
Cornet	62	Electric Bass	676
Electric Guitar	468	English Horn	60
Flute	148	French Horn	218
Harmonica	168	Mandolin	283
Oboe	132	Pan Flute	74
Piccolo	200	Pipe Organ	56
Recorder	150	Soprano Sax	198
Tenor Sax	196	Trombone	194
Trumpet	141	Tuba	180
Ukulele	144	Viola	360
Violin	384		
Total : 6698			

3. METHODOLOGY

3.1. Database

Database "RWC Music Database for Musical Instrument Sound " [7] was chosen in this work. In this database, each audio file contains the signal of a single instrument played with isolated notes. This database provides multiple records for each instrument: different manufacturers for the same instrument and different musicians took part to generate records and provide a range of several instrumental signatures. In principle, it contains three variations for each instrument: three manufacturers, three musicians and three different dynamics. For each instrument, the musician is playing each note individually at an interval of a semitone over the entire possible range of the instrument. In terms of string instruments, the full range for each chord is played. Dynamics also varied with intensities strong, mezzo and piano. Table 1 gives specification over the type and number of musical instruments used in current experiments.

3.2. Reference system

The reference system proposed in [6, 10] uses a feature vector built from 38 feature parameters: 13 MFCC, 14 LPC, spectral centroid, spectral spread, spectral kurtosis, spectral skewness, zero crossing rate (ZCR), onset time, envelope slope, envelope centroid, envelope spread, envelope kurtosis and envelope skewness. The popular k-Nearest Neighbor algorithm (k-NN) is used to the classification and decision task. The metric used for the k-NN classifier is the euclidian distance and the number of neighbors was set to 4, value determined by empirical testing. Also, the components of the feature vector are optimized by employing the Sequential Forward Selection (SFS) algorithm. Therefore, the feature vector is reduced to keep only the best discriminating factors for the instrument identification task. Grossly speaking, the SFS method selects the best attributes from the score obtained by an objective function. Unlike the PCA, it allows thinning down the set of features available by keeping only the most significant ones. This has the advantage to bring out the features most likely to have an impact on the system and therefore by provide a better understanding of the phenomenon.

3.3. Proposed system

3.3.1. Chromagram estimation

Chromagram is defined as the whole spectral audio information mapped into one octave. Each octave is divided into 12 bins representing each one semitone. The same strategy based on instantaneous frequency (IF), presented in [11], is adopted in this work to compute the features chroma. The audio signal, with sampling frequency of 11025 Hz, is split up into frames (1024 points) interlaced over 512 points. Motivation behind the IF is to track only real harmonics. Figure 1 illustrates the chroma obtained for two instrument playing two different notes. We note that the tonal information of each note is captured, however, no information on the type of instrument is taken into account.

3.3.2. Chromatimbre estimation

Each two-dimensional chroma matrix is associated with time axis and bin frequency axis (semitone note). We utilize the *contour* function of MATLAB, which determines 10 contours levels by using a linear interpolation. Each contour tracking represents the intensity variation with respect to a fixed threshold for yielding a segmentation of chromagram representation (image) producing several regions. Hence, contours delimiting the frontiers give some description equivalent to the acoustical scene auditory activity. To deal with variability level, all contours are set to the same intensity. This is similar to transforming a color image to black and white. This binary encoding approximation is used just to accelerate and facilitate the continuation of this exploratory study [12]. Fig. 2 and fig. 3 illustrate chromatimbre representation with only one contour for flute, piano, violin and trumpet instruments playing different tones. According to geometrical shape contours, it is clear that chroma shows a great energy concentrated at small interval centred at bin number 4 and 11 for different note modes (C4, C6, G5, and G6). The contours representation in Fig. 2 and fig. 3 with the same instrument, exhibits rather than the tonal content, a particular pattern shape assumed to characterize the timbre information. Illustrations beside pattern shapes seem to keep and conserve the same geometrical propriety when an instrument played different notes. Pizzicato is especially clearly visible and there is no ambiguity to distinguish sustained instruments from pizzicato instruments. The chromatimbre of the accordion is especially easy to recognize because of its unique signature. However, the shapes of chromatimbre are not trivial and it would be difficult to enumerate all the characteristics that can have each instrument. In fact, the chromatimbre can be exploited to extract many information as the envelope, amplitude modulation, frequency modulation, attack time, sustain and release of the note. As a first investigation, we assume the representation of chroma as a 2D image, where the points corresponding to the contour are set to 1 and other points to 0. The problem is therefore how to characterize each instrument by describing the geometric contours of its chromatimbre. This can be achieved by applying the moment invariants discussed in the next sub-section.

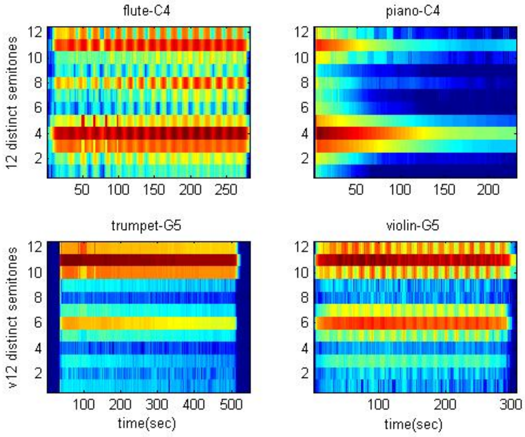

Fig. 1. Description by Chromagram: Only tonal information is preserved

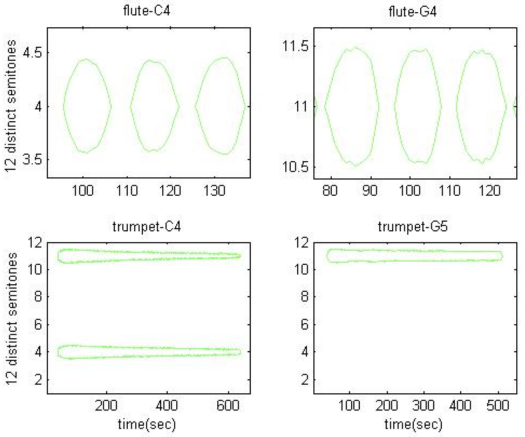

Fig. 2. Description by the proposed chromatimbre: flute and trumpet instruments playing C4, G4 and G6 notes.

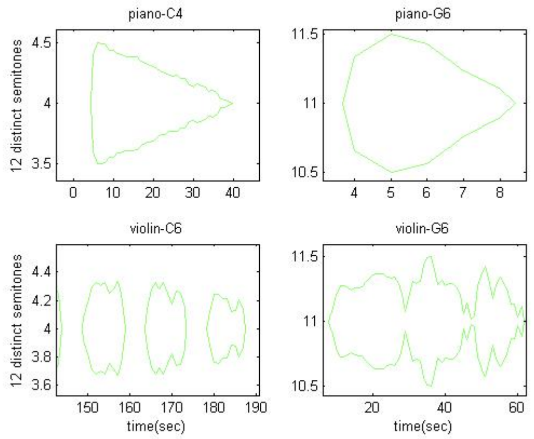

Fig. 3. Description by the proposed chromatimbre: piano and violin instruments playing C4 and G6 notes.

3.3.3. *Invariant moments*

Invariant moments are recognized as a classical technique for pattern recognition during the last years. They were introduced by [13], who derived seven moments invariant to translation, rotation and scale of 2D objects:

$$I_1 = \eta_{20} + \eta_{02} \tag{1}$$

$$I_2 = (\eta_{20} - \eta_{02})^2 + (2\eta_{11})^2 \tag{2}$$

$$I_3 = (\eta_{30} - 3\eta_{12})^2 + (3\eta_{12} - \eta_{03})^2 \tag{3}$$

$$I_4 = (\eta_{30} + \eta_{12})^2 + (\eta_{12} + \eta_{03})^2 \tag{4}$$

$$\begin{aligned} I_5 =& (\eta_{30} - 3\eta_{12})(\eta_{30} + \eta_{12})((\eta_{30} + \eta_{12})^2 - \\ & 3(\eta_{12} + \eta_{03})^2) + (3\eta_{21} - \eta_{03})((\eta_{21} + \eta_{03}) \\ & (3(\eta_{30} + \eta_{12})^2 - (\eta_{12} + \eta_{03})^2) \end{aligned} \tag{5}$$

$$\begin{aligned} I_6 =& (\eta_{20} - \eta_{02})((\eta_{30} + \eta_{12})^2 - (\eta_{20} + \eta_{03})^2 + \\ & 4\eta_{11}(\eta_{30} + \eta_{12})(\eta_{21} + \eta_{30})) \end{aligned} \tag{6}$$

$$\begin{aligned} I_7 =& (3\eta_{21} - \eta_{03})(\eta_{30} + \eta_{12})((\eta_{30} + \eta_{12})^2 - \\ & 3(\eta_{12} + \eta_{03}) - (\eta_{30} - 3\eta_{12})(\eta_{21} + \eta_{03}) \\ & (3(\eta_{30} + \eta_{12})^2 - (\eta_{12} + \eta_{03})^2)) \end{aligned} \tag{7}$$

where

$$\eta_{ij} = \frac{\mu_{ij}}{\mu_{00}^{(1+\frac{i+j}{2})}} \tag{8}$$

is the normalized moment with $i + j \geq 2$, and:

$$\mu_{pq} = \sum_x \sum_y (x - \bar{x})^p (y - \bar{y})^q I(x, y) \tag{9}$$

is the central moment of order (p, q) and $I(x, y)$ is the pixel intensity level at (x, y) coordinates.

4. RESULTS AND DISCUSSION

The best approach uses 10 normalized contours of a 24 bins chromagram giving a recognition rate of 86.85% (see Table 2). The reference system gives a score of 97.50% (see Table 3). As the chromatimbre technique utilizes only 7 feature parameters for the feature vector, the reference system deals with 38 feature parameters which is more than 5 times the number of parameters in the chromatimbre system. A feature selection algorithm has also been applied in the reference system, forcing more complex calculations to reduce the feature vector dimensions.

Separation of pizzicato instruments from sustained instruments is trivial using only the shape of chromatimbre. Visual discrimination of the two classes yields near perfect results as the human eye can easily distinguish between the two characteristic shapes of pizzicato and sustained chromatimbres. On the other hand, extracting feature parameters from the chromatimbre is more difficult and needs more sophisticated techniques than the invariant moments. Image processing techniques could be used in conjunction with the chromatimbre of the instruments in order to extract features from the audio signals.

5. CONCLUSION

A new proposal has been examined and evaluated using the chromatimbre with invariant moments as feature parameters. The importance of chromatimbre is that it can encode instrument timbre information in conjunction with its tonal content. Comparison with a reference system revealed that this approach gives results almost similar without having to optimize the parameter vector and consider the ideal classification strategy. Chromatimbre has promising avenues in a hierarchical classification system where the articulation defines a tree level. New parameterization techniques could also be constructed from the chromatimbre representation that wields better results with other hierarchical levels.

Table 2. Best confusion matrix obtained from chromagram parameters.

86,85%	pizzicato	sustained
pizzicato	1861	381
sustained	500	3956

Table 3. Confusion matrix of reference system.

97,50%	pizzicato	sustained
pizzicato	2233	9
sustained	8	4448

6. REFERENCES

[1] J. W. Beauchamp, “Time-variant spectra of violin tones,” *Journal of the Acoustical Society of America*, vol. 56, no. 3, pp. 995–1004, 1974.

[2] M. D. Freedman, “Multidimensional perceptual scaling of musical timbres,” *Journal of the Acoustical Society of America*, vol. 41, no. 4A, pp. 793–806, 1967.

[3] J. M. Grey, “Multidimensional perceptual scaling of musical timbres,” *Journal of the Acoustical Society of America*, vol. 61, no. 5, pp. 1270–1277, 1977.

[4] K. D. Martin and Y. E. Kim, “Musical instrument identification: A pattern-recognition approach,” in *Presented at the 136th meeting of the Acoustical Society of America*, 1998. [Online]. Available: http://sound.media.mit.edu/Papers/kdm-asa98.pdf

[5] A. Eronen, “Automatic Musical Instrument Recognition,” Master’s thesis, Department of Information Technology, Tampere University of Technology, Tampere, Finland, 2001.

[6] G.-E. Hall, H. Ezzaidi, and M. Bahoura, “Hierarchical parametrization and classification for instrument recognition,” in *the 11th International Conference on Information Science, Signal Processing and their Applications (ISSPA)*, Montreal, Canada, 2-5 July 2012, pp. 1066–1071.

[7] M. Goto, H. Hashiguchi, T. Nishimura, and R. Oka, “RWC Music Database: Music Genre Database and Musical Instrument Sound Database,” in *the 4th International Conference on Music Information Retrieval (ISMIR 2003)*, Baltimore, Maryland, 26-30 October 2003, pp. 229–230.

[8] T. Kitahara, M. Goto, and H. G. Okuno, “Pitch-Dependent Identification of Musical Instrument Sounds,” *Applied Intelligence*, vol. 23, pp. 267–275, 2005.

[9] ——, “Musical instrument identification based on F0-dependent multivariate normal distribution,” in *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP’03)*, vol. 5, 6-10 April 2003, pp. 421–424.

[10] G.-E. Hall, H. Ezzaidi, and M. Bahoura, “Study of feature categories for musical instrument recognition,” in *International Conference on Advanced Machine Learning Technologies and Applications (AMLTA12)*, Cairo, Egypt, 8-10 December 2012.

[11] D. Ellis, “Classifying Music Audio with Timbral and Chroma Features,” in *the 8th International Conference on Music Information Retrieval (ISMIR 2007)*, Vienna, Austria, 23-30 September 2007, pp. 339–340.

[12] E. Hassan, M. Bahoura, and G.-E. Hall, “Towards characterization of music timbre based contour chroma,” *International Conference on Advanced Machine Learning Technologies and Applications (AMLTA12)*, 2012.

[13] M. K. Hu, “Visual pattern recognition by moment invariants,” *IRE Transactions on Infomation Theory*, vol. 8, p. 179187, 1962.