

Wavelet ridges for musical instrument classification

M. Erdal Özbek · Nalan Özkurt · F. Acar Savacı

Received: 13 September 2010 / Revised: 25 January 2011 / Accepted: 25 January 2011 /
Published online: 8 February 2011
© Springer Science+Business Media, LLC 2011

Abstract The time-varying frequency structure of musical signals have been analyzed using wavelets by either extracting the instantaneous frequency of signals or building features from the energies of sub-band coefficients. We propose to benefit from a combination of these two approaches and use the time-frequency domain energy localization curves, called as wavelet ridges, in order to build features for classification of musical instrument sounds. We evaluated the representative capability of our feature in different musical instrument classification problems using support vector machine classifiers. The comparison with the features based on parameterizing the wavelet sub-band energies confirmed the effectiveness of the proposed feature.

Keywords Wavelet ridge · Musical instrument classification · Support vector machine

1 Introduction

The classification of musical instruments is one of the main issues in the automatic transcription of music. There have been many features offered to represent musical signals which also targeted to reveal the differences of musical instrument sounds

M. Erdal Özbek (✉) · F. Acar Savacı
Dept. of Electrical and Electronics Engineering,
İzmir Institute of Technology, Urla, 35430, İzmir, Turkey
e-mail: erdalozbek@iyte.edu.tr

N. Özkurt
Dept. of Electrical and Electronics Engineering,
Dokuz Eylül University, Buca, 35160, İzmir, Turkey

(Herrera-Boyer et al. 2003, 2006; Deng et al. 2008; Essid et al. 2006; Klapuri and Davy 2006). While the temporal evolution of the signal is characterized by the time-based features, the spectral features are extracted based on the short-time spectrum using time-frequency transformations such as short-time Fourier transform, constant-Q transform, and wavelets (Alm and Walker 2002; Pielemeier et al. 1996). Apparently, representing musical signals with time-frequency features are effective due to the nature of musical signals where the frequency varies with time.

The wavelet analysis is a very popular time-frequency analysis technique and it is widely used for the non-stationary signals in which the statistical properties vary with time (Mallat 2009). As musical signals are inherently non-stationary, wavelets are frequently used to analyze their properties and build representative features from them. The time-varying frequency structure of the signals is observed in a time-frequency plane with the continuous wavelet analysis, while the signal is divided into its sub-bands via the discrete wavelet transform (DWT) leading to wavelet packet analysis, which can also be interpreted with the multi-resolution concept (Vetterli and Kovačević 1995).

Accordingly, the use of wavelets in musical signals may be separated into two cases based on the following representations. In one representation, the emphasis is given to the extraction of instantaneous frequency (IF) (Boashash 1992) for analysis and re-synthesis purposes built upon sinusoidal modeling (Pielemeier et al. 1996; Beauchamp 2007; Goodwin and Vetterli 1996; Kronland-Martinet 1988). The IF of a signal can be extracted using wavelet analysis (Delprat et al. 1992) based on the time-frequency energy localization (see Sejdić et al. 2009; Shafi et al. 2009, for recent overviews). As the energy is concentrated in frequency about the IF (Boashash 1992), the required information can be extracted using the curves at the time-frequency plane along where the energy is locally maximum, namely the wavelet ridges (Mallat 2009; Delprat et al. 1992; Carmona et al. 1997; Todorovska 2001). These trajectories allow to characterize the frequency modulation law of the signals. Moreover, the wavelet transform coefficients along the ridges are called as the skeleton of the wavelet transform (Delprat et al. 1992), which can be used to de-noise and reconstruct a signal (Carmona et al. 1999; Lin 2006; Özkurt and Savacı 2006). The ridges can be computed using many techniques (Mallat 2009; Delprat et al. 1992; Carmona et al. 1997; Todorovska 2001; Guillemain and Kronland-Martinet 1996), where some examples for IF and ridge extraction can be found in Dai et al. (2008), Özkurt and Savacı (2005), Shuai and Yuan (2006) and Tse and Lai (2007), including musical rhythm detection (Smith and Honing 2008). Although the diagram of ridges associated with the musical instrument sounds has been presented in Delprat (1997), the features representing musical instrument sounds based on wavelet ridges have not been considered for the classification of instruments.

The other wavelet representation involves with the sub-band coefficients, their energy, and the distribution of the amplitude or energies of the coefficients. The first attempts for this case are focused on finding a suitable wavelet function to represent musical signals, such as harmonic and musical wavelets (Newland 1994), and log-Morlet wavelet function (Olmo et al. 1999). Later, the amplitude envelopes of the wavelet coefficients have been used rather than the raw wavelet coefficients or the sub-band energies (Hacıhabiboğlu and Canagarajah 2002). The employment

of DWT has been investigated in Tzanetakis et al. (2001), preceding to the usage of sub-band wavelet coefficients and wavelet packets for musical instrument sound classification (Pruysers et al. 2005). The sub-band power and pitch frequency features have been extracted for audio including musical instrument sounds in Lin et al. (2005). Furthermore, the musical instrument sounds have been classified by parameterizing the energy of each sub-band in the remarkable works of Kostek and Czyzewski (2001), Kostek (2004, 2005) and Wieczorkowska (2001, 2010). The observation stated in Kostek (2004) reveals that the distribution pattern of energy values within the sub-bands is similar for the wind instruments whereas the same parametric representation for the string instruments is different from the wind instruments. This is consistent with the assumption of Do and Vetterli (2002), declaring that the energy distribution in frequency domain identifies a texture, whereas the marginal densities of wavelet sub-band coefficients have been modeled with a Generalized Gaussian Density (GGD). This GGD modeling has been used for musical genre (Tzagkarakis et al. 2006) and then for musical instrument classification in Özbek and Savacı (2007). Following the same idea, histograms computed from the wavelet coefficients at different frequency sub-bands have been efficiently used for music information retrieval (Li and Ogihara 2006).

The presented paper proposes to benefit from a combination of these two representations. Using the sub-band energies is likely to result with an acceptable recognition rate, but certainly loses the temporal dimension of timbre which is essential in discriminating musical instruments. That explains the intention of offering a time-related parameter in Kostek (2004), in addition to the wavelet sub-band energies. As the energy is concentrated around the IF, wavelet ridges can be used to build features similar to the ones that has been built upon the spread of energies in wavelet sub-bands. Besides, wavelet ridges carry the time domain information which can be exhibited in the feature vector. By using the ridges, we offer to build a representation of musical signals reflecting their time-varying frequency components, mainly composed of the fundamental and harmonic frequencies, where the energy is localized. Although, the main purpose is to classify instruments, meanwhile, the note of the musical instrument may be determined based on the distribution of energy to the frequencies in those sub-bands.

The representative capability of our wavelet ridge based feature is evaluated in different musical instrument classification problems using support vector machine (SVM) classifiers (Cristianini and Shawe-Taylor 2000; Vapnik 1998) which have proven to demonstrate successful classification rates for audio and musical instrument classification (Deng et al. 2008; Essid et al. 2006; Lin et al. 2005; Wieczorkowska and Kubera 2009). As the study is focused on wavelet representations, the importance is given to the comparison results with different wavelet mother functions. Therefore, only the SVM classifiers are used for evaluation rather than comparing the results of different classifiers.

The organization of the paper is as follows: Section 2 reviews the wavelet transform and the method of obtaining the wavelet ridges of a signal. The experimental study performed for different musical instrument groups will be explained in Section 3. Finally, the corresponding performance results using SVMs and conclusions will be presented.

2 A review of wavelet transform and wavelet ridges

In continuous wavelet transform (CWT), the signal $s(t)$ is projected on a zero-mean family of functions called wavelets,

$$W_s(a, b; \psi) = \int_{-\infty}^{\infty} s(t)\psi_{a,b}^*(t)dt, \tag{1}$$

with translations and scaling of a mother wavelet function, where a and b are the scaling and translation coefficients, respectively. The $\psi(\cdot)$ is the mother wavelet function, the $*$ denotes the complex conjugate, the scaled and translated wavelet is given by

$$\psi_{a,b}(t) = \frac{1}{\sqrt{a}}\psi\left(\frac{t-b}{a}\right). \tag{2}$$

The local time-frequency energy density named as scalogram is defined in Delprat et al. (1992) and Mallat (2009) as

$$P_s(a, b; \psi) \triangleq |W_s(a, b; \psi)|^2. \tag{3}$$

For numerical computations, the DWT is obtained by regularly sampling CWT at discrete time and scale positions, hence defining the scaled and translated wavelets as

$$\psi_{m,n}(t) = a^{-m/2}\psi(a^{-m}t - bn) \quad m, n \in \mathbb{Z}, \tag{4}$$

then the scalogram matrix for discrete case can be defined as

$$P_s = |W_s(a_m, b_n; \psi_{m,n})|^2, \tag{5}$$

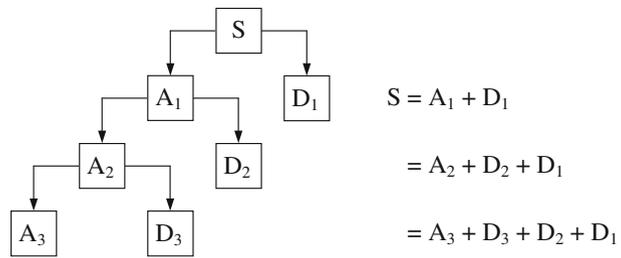
which are simply the local time-frequency domain energy densities evaluated at the discrete dilations a_m and the translations b_n .

A way of showing the time-frequency information is to define the ridges of the signal. There are several ridge detection methods including stationary phase method which calculates the ridges using stationary point theorem (Delprat et al. 1992; Todorovska 2001; Guillemain and Kronland-Martinet 1996), and the simple method which directly finds the local maxima of the scalogram (Mallat 2009; Carmona et al. 1997; Todorovska 2001).

The interpretation of the IF using the analytic function representation indicates that the derivative of the phase of a signal provides a measure of frequency (Boashash 1992; Delprat et al. 1992). Hence, the signals can be modeled using a frequency-modulated signal fitted to the change of the main frequency. For a multi-component signal $s(t)$ with L components, the instantaneous amplitudes $A_l(t)$ and the instantaneous phases $\phi_l(t)$ can be described by

$$s(t) = \sum_{l=1}^L A_l(t)e^{j\phi_l(t)}. \tag{6}$$

Fig. 1 Three level wavelet decomposition



Then the wavelet transform of the signal can be written as (Delprat et al. 1992; Carmona et al. 1997; Todorovska 2001)

$$W_s(a, b; \psi) = \frac{1}{2} \sum_{l=1}^L A_l(b) e^{j\phi_l(b)} \hat{\psi}^*(a\phi'_l(b)) + r(a, b), \tag{7}$$

with $r(a, b) \sim O(|A'_l|, |\phi''_l|)$ indicating higher orders where the primes denote the derivatives. Therefore, if the Fourier transform of the mother wavelet function $\hat{\psi}(\omega)$ is localized near a certain frequency $\omega = \omega_0$, the scalogram is localized around L curves

$$a^l = a^l(b) = \frac{\omega_0}{\phi'_l(b)}, \quad l = 1, \dots, L \tag{8}$$

which are named as the ridges of the wavelet transform or simply wavelet ridges (Delprat et al. 1992; Carmona et al. 1997; Todorovska 2001).

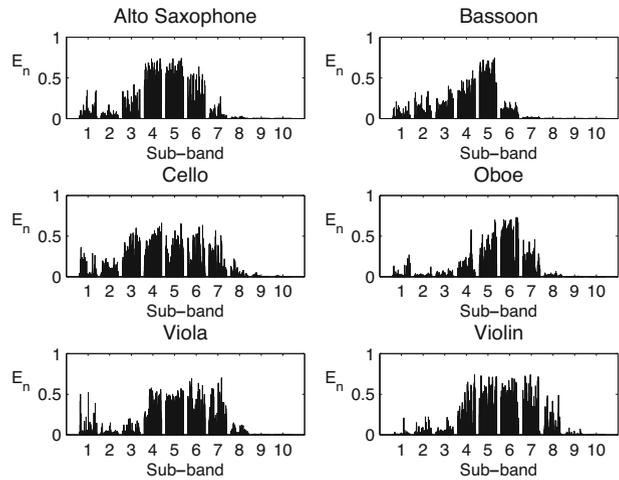
On the other hand, by selecting the variables $a = 2$ and $b = 1$ in (4), we obtain the dyadic grid which leads to the orthonormal bases with the special choices of the mother wavelet function ψ . Accordingly, the signal $s(t)$ can be written as a linear combination of the dyadic wavelet coefficients whereas reconstruction of the signal from those coefficients is possible. By interpreting this dyadic grid with very efficient and low complexity filter-bank structures based on the multi-resolution concept, which states that a given signal can be represented by the sum of a coarse approximation and details, the signal can be decomposed into many detail and approximation sub-band coefficients (Vetterli and Kovačević 1995). Figure 1 displays the decomposition where the signal S is decomposed to three level detail (D) and approximation (A) sub-band coefficients.

3 Experimental study

3.1 Instrument sound samples

In this study, we mainly used the University of Iowa Electronic Music Studios musical instrument samples (Fritts 1997). The samples were recorded in an anechoic chamber in mono channel, having 16 bit dynamic range and sampled at 44.1 kHz. The sound files with groups of note samples in the library have been separated note by note,

Fig. 2 Normalized energy values for 10 sub-bands of some musical instruments



and labeled accordingly in Özbek et al. (2007). Each of the samples is in one of the three dynamic ranges: fortissimo (ff), mezzo forte (mf), and pianissimo (pp). The wind instruments recordings include samples with and without vibrato whereas we excluded the string samples played by plucking (pizzicato).

It is accepted that the starting transients are the most important part of the signal for the recognition of musical sounds (Kostek 2005). Therefore, a frame length consisting of 2,048 samples were selected from the attack part of the musical instrument sounds.

3.2 Wavelet sub-band energy based feature

The energy related parameters for each of the instrument note samples were computed by decomposing the signals into 10 sub-bands as

$$E_n = \frac{E_i}{\sum_{i=1}^{10} E_i} \quad i = 1, \dots, 10 \quad (9)$$

where E_n denotes the partial energy parameter normalized with respect to the overall energies of the each sub-band energy parameterized by E_i . Figure 2 displays the energy-based feature (E_n) for some of the chosen musical instruments. For each bar group in the sub-bands, each micro-bar corresponds to the different note sample of that instrument. The ranges of the instruments can be seen from the distribution of the energies to the sub-bands.

Although the second order Daubechies wavelet (*db2*) was preferred as a mother wavelet function because of its lower computational load than the others as acknowledged in Kostek (2004), we also computed the energy parameters using the Daubechies *db8* wavelets as in Li and Ogihara (2006), as well as bi-orthogonal (*bior4.4*) wavelets in order to compare the efficiency. Furthermore, the first three moments, i.e., the mean, the variance, and the skewness values of the energy parameters for each sub-band were computed.

3.3 Wavelet ridge based feature

In order to determine the ridges, we applied the technique proposed in Özkurt and Savacı (2005) which uses singular value decomposition (SVD) to factorize the scalogram matrix given in (3). By selecting only the dominant components associated with the signal, an approximated scalogram matrix was obtained. Then, by selecting the localized energy points in the scalogram, the wavelet ridges were labeled. The complex Morlet wavelet

$$\psi(t) = \frac{1}{\sqrt{2\pi}} e^{j\omega_0 t} e^{-t^2/2}, \quad (10)$$

was selected as the mother wavelet function because the magnitude and phase of the wavelet coefficients can be easily separated due to the fact that it is described by an analytic function (Todorovska 2001). Moreover, the Gaussian shape of this mother wavelet function provides a smooth energy distribution, thus the resulting wavelet ridges effectively display this distribution over the time-frequency plane. Figure 3 shows an example of a Flute A4 note sample, its scalogram calculated using CWT, and the corresponding wavelet ridges marked by employing the SVD-based ridge determination procedure.

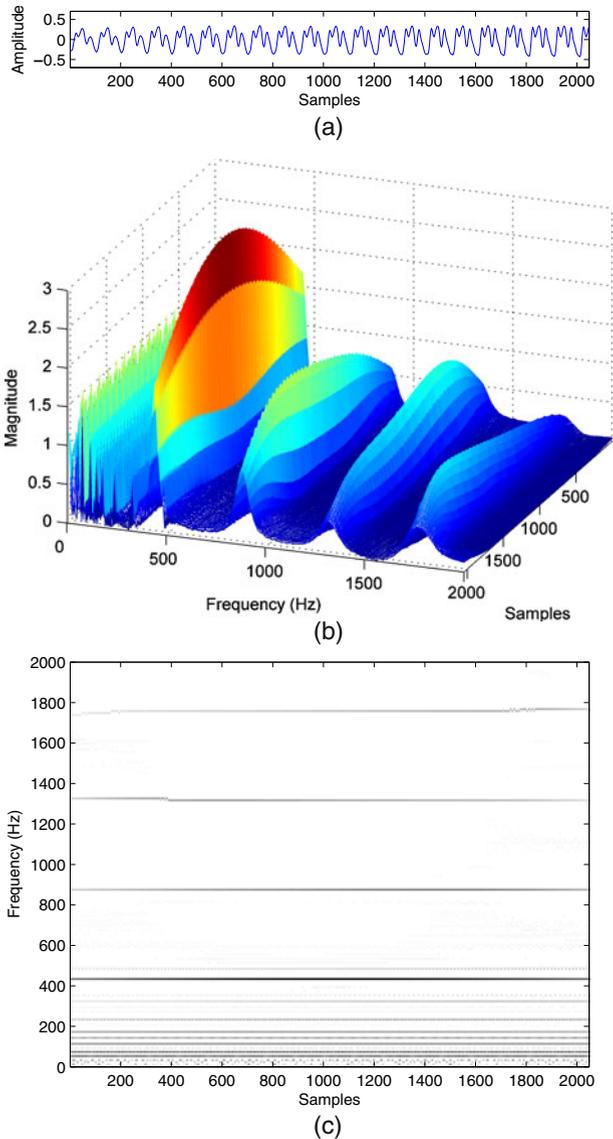
The values of the ridges are displayed using a gray level image in which darker points represent higher energy of the hills of the scalogram similar to the spectrogram. Apparently, the ridges denote the energies which are concentrated around the fundamental frequency and the corresponding harmonic frequencies of the sample signal as shown in the hills of the scalogram in (b) of Fig. 3. Due to the multi-resolution capability of wavelets, the scalogram can be referred as a high resolution spectrum. Another interpretation can be displayed based on the ridge energies for each frequency. Figure 4 shows the A4 note samples of Soprano Saxophone and Violin, their spectrograms, and the energy values of their ridges. While both of the instrument samples have the same note signal, the energies of the partials differ with respect to the instrument. Therefore, as in the case of the energy of the bins in the Fourier transform (shown in the spectrograms) or of the sub-bands of the wavelet transform, the energy of the ridges is likely to reveal the differences of the instruments.

Furthermore, while the wavelet ridges include the time evolution of the frequency, we eliminated the redundant information in time by downsampling. We selected the values of only the 8 time instants for a feasible feature vector length which directly effected the classification performance. Similarly, the frequency content was compacted in sub-bands. Thus, as in the energy based parameters of the sub-bands, we built features by summing the energy values of the ridges for 10 sub-bands, having 80 values. We included the mean, variance, and skewness values of the energy parameters and labeled the corresponding features with 'mvs'.

3.4 SVM classification

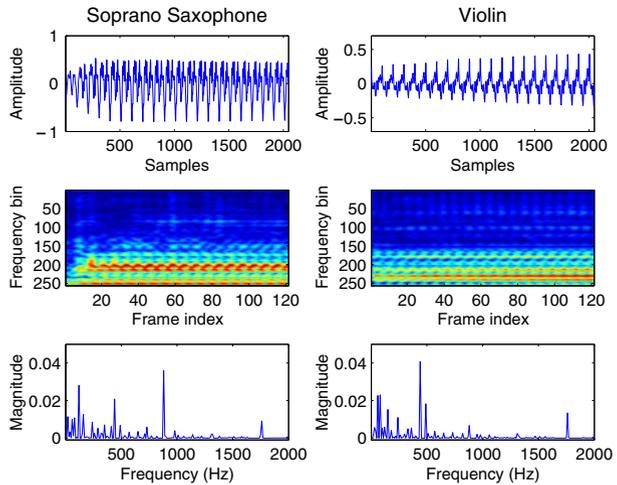
The SVM classifiers have been developed based on statistical learning theory (Vapnik 1998). They are widely used because of their generalization ability.

Fig. 3 **a** A Flute A4 note sample, **b** its scalogram, **c** and the corresponding wavelet ridges



Although they were originally designed for solving two-class classification problems, multi-class classifications can be performed using the two common methods: one-vs-all and one-vs-one. While the choice of the method depends on the problem, one-vs-all method often produces acceptable results (Schölkopf and Smola 2002). Therefore, we chose to implement SVM classifiers using one-vs-all method (Weston et al. 2003). Each classifier was implemented using the radial basis function (RBF) kernel. The feature vectors were normalized before introducing to the SVM classifiers. Half of the samples were used for training while the remaining half of the

Fig. 4 The Soprano Saxophone and Violin A4 note samples, their spectrograms, and corresponding ridge energy values



samples were left for testing. The presented results were obtained after a 10-fold stratified cross-validation scheme with kernel parameter $\sigma = 1$.

4 Results

We present the classification performance results in three different groups: string instruments, wind instruments, and both string and wind instruments. For the string instruments, we used the samples of Bass (289), Cello (352), Viola (271), and Violin (283), while for the wind instruments we selected Alto Saxophone (192), Flute (227), Oboe (104), and Trumpet (212) with the number samples given in parentheses. In each case, eight different feature vectors were evaluated, i.e., *R*, *Db2*, *Db8*, *Bior44*, *Rmus*, *Db2mus*, *Db8mus*, *Bior44mus*. Here, *R* denotes the feature of the wavelet ridges, where *Db2*, *Db8*, and *Bior44* denote the energy parameters calculated with DWT using the respective mother wavelet function. The labels including *mus* denote that the mean, variance, and skewness values of the features were included in the feature vector. The numbers printed as boldface in the tables denote the highest value of the corresponding result while all numbers are given in percentage.

For the string instruments, the confusion matrices obtained using the *Db2* and *R* features are displayed in the Table 1. An increase in the classification scores of Viola and Violin was observed with the proposed wavelet ridge-based feature. Bass had the highest performance which is mainly due to the frequency range of the instrument. Viola was found to have the lowest performance and it was confused mostly with Cello in both cases, while the confusion of Violin with Cello was decreased in the proposed feature case.

We observed similar results for the wind instruments, where the confusion matrix for the wavelet ridge based feature is given in the Table 2. As similar but lower rates were observed for other features, only the results obtained by the wavelet ridge based feature will be presented afterwards. The Trumpet was found to have the highest performance while Flute had the lowest score with similar misclassification

Table 1 Confusion matrices for the string instruments using Db2 and R features

Instrument	Db2				R			
	Classified as				Classified as			
	Bass	Cello	Viola	Violin	Bass	Cello	Viola	Violin
Bass	94.1	4.6	0.0	1.3	94.5	2.3	1.7	1.5
Cello	4.1	85.6	8.2	2.1	1.1	85.5	11.0	2.4
Viola	1.7	39.8	45.4	13.1	2.5	26.8	57.4	13.3
Violin	0.7	25.6	9.2	64.5	0.1	8.7	12.4	78.8

rates with the other three instruments. Afterwards, the string and wind instruments were combined and a new classification was performed for eight instruments. The confusion matrix for the wavelet ridge based feature is given in the Table 3.

Parallel to the observations from the string and wind instruments classification performance results, Bass and Trumpet had higher performance scores whereas Flute and Viola displayed lower classification rates. While the afore presented confusion matrices have denoted the performance of the wavelet ridge based feature for the classification of instruments, Table 4 displays the average accuracies for each of the string, wind, and all of the instruments computed using each of the feature set.

In all of the cases, the proposed wavelet ridge based feature outperforms the wavelet sub-band energy based feature sets. In most of the cases, adding the mean, variance, and the skewness values to the features resulted with better classification performance, however it was found to be the opposite when using the wavelet ridge based feature. This can be explained with the localization of energy around the ridges. By concentrating the highest energy in the frequencies of a sub-band, the variation of the energy within the sub-band mostly depends on the energy of the ridges. Therefore, adding the mean, variance, and skewness values to the feature may not supply extra information to be used in classification. Besides, by enlarging the length of the feature vector it may cause a decrease in the classification rates.

An advantage of the method is in the gathered information related to the frequency content. While the instruments are being classified, the extracted wavelet ridges may aid in identification of the notes of the instruments. For the notes in the C4-C5 octave range, based on the maximum energy value of the ridge, the notes of the isolated sample were correctly (within a semitone) identified with ratios of 95% for Alto Saxophone and 85% for Flute, whereas lower for the rest such as around 50% for Viola and Violin. The misclassifications were mainly due to the one frame analysis in the attack part of the signal. By using more frames in the wavelet analysis or by analyzing the sustain part of the notes where the frequency and correspondingly the spectral energy of the signal is stable, more precise identification of notes may be

Table 2 Confusion matrix for the wind instruments using R feature

Instrument	Classified as			
	Alto Saxophone	Flute	Oboe	Trumpet
Alto Saxophone	73.6	16.4	10.0	0.0
Flute	10.8	69.2	9.4	10.6
Oboe	4.6	7.9	86.5	1.0
Trumpet	0.7	7.6	0.4	91.3

Table 3 Confusion matrix for the string and wind instruments using R feature

Instrument	Classified as							
	Alto Saxophone	Bass	Cello	Flute	Oboe	Trumpet	Viola	Violin
Alto Saxophone	66.8	5.3	1.8	11.1	8.2	0.0	3.9	2.9
Bass	0.0	93.3	2.6	0.2	0.0	0.0	1.9	2.0
Cello	0.0	0.4	81.8	1.0	0.0	2.3	9.3	5.2
Flute	3.4	14.6	2.5	55.9	4.1	7.6	7.3	4.6
Oboe	7.1	1.0	0.0	10.5	75.0	4.9	0.4	1.1
Trumpet	0.1	0.0	0.9	5.6	1.8	89.7	1.3	0.6
Viola	1.7	3.7	26.5	2.4	0.0	2.8	49.3	13.6
Violin	2.9	0.7	7.7	0.1	0.0	0.4	13.7	74.5

performed. Besides, the frequency of the note may not always be the frequency with the highest energy as illustrated in the Fig. 4. Therefore, a proper identification of notes from tracking of harmonic frequency positions would be required as in Every and Szymanski (2006).

Although the proposed feature has better classification performance than the wavelet sub-band energy based feature sets, and some of the state-of-the-art features, it has not yet outperformed the mel frequency cepstral coefficients (MFCCs) which recently been shown as the best feature for many musical instruments (Wieczorkowska and Kubik-Komar 2010). In order to build a baseline for comparison purposes, the classification rates for the standard 13-length MFCC feature have been obtained as 91.8, 90.4 and 85.5% for the string, wind, and all of the instruments case, respectively. Similar with the MFCCs, the wavelet ridges display the spectral envelope of the signal. However, the MFCCs lack to display the temporal information which were included in the wavelet ridges representing the temporal evolution of the spectral variation. Therefore, we aimed to use this additional time domain information extracted from the attack part of the signals for the recognition of musical sounds. While the state-of-the-art features like MFCCs may result with better performance and therefore are widely used, some of the features may have better representations for some certain instruments. Moreover, improvements of MFCCs yield slightly better results which reflect that there is not much room to benefit further from MFCCs. As explained above, by using wavelet ridges, it may be possible to determine and follow the individual partials (notes) which may exist and possibly overlapping in a polyphonic music or in a mixture of musical instrument

Table 4 Performance results for the string, wind, and all cases for all feature sets

Feature	Performance (%)		
	String	Wind	All
R	79.7	79.9	73.9
Db2	69.9	66.2	64.6
Db8	73.1	61.7	62.7
Bior44	72.6	65.3	62.8
Rmvs	76.4	72.7	70.2
Db2mvs	76.1	71.2	67.1
Db8mvs	70.3	66.6	65.5
Bior44mvs	72.3	71.0	66.1

Table 5 Confusion matrix with the ridge based feature for the CD samples

Instrument	Classified as			
	Cello	Oboe	Piano	Violin
Cello	70.7	17.1	8.9	3.3
Oboe	0.3	64.7	9.0	26.0
Piano	16.6	14.3	43.3	25.8
Violin	18.9	22.6	1.6	56.9

sounds. This is similar to the use of the pitch frequency cepstral coefficients which intends to extract the frequency content with timbre information.

The benefit of using wavelet ridges is to reveal the information hidden in the time-frequency domain by following the energy paths. As the energy paths can be investigated in higher resolution, information for the discrimination of the instruments can be obtained accordingly. Although obtaining a unique representation for a musical instrument is still far, the skeleton information represented in the wavelet ridge based features is found efficient and can be used to build better representations.

We further evaluated our proposed feature using the solo excerpts of musical CDs obtained from personal collections and the library of Turkish Radio and Television Corporation. The collection was composed of 69 different tracks from four instruments, i.e., Cello, Oboe, Piano, and Violin. Only the left channel of the stereo tracks were used. The analyzed samples were extracted using an energy threshold in order to discard the silence. The same analysis frame length and SVM architecture were kept. The number of total samples for each instrument was not less than 1,400 which half of them were used for training and the rest was kept for testing. No test sample was selected from the recordings of the training sample. The confusion matrix with the wavelet ridge based feature is presented in Table 5.

The results display that the Cello was mostly confused with Oboe, while Oboe and Piano were mostly confused with Violin. Piano had the lowest performance due to the misclassifications with all three instruments. The performance results of the CD excerpts for all feature sets are displayed in Table 6.

Repeatedly, the wavelet ridge based feature set had the highest performance compared to the other feature sets. The lower performance results obtained for all of the instruments are mainly due to the selection of the analysis frame which may not be in accordance with the attack part as in the isolated note samples. While for the isolated note samples it was relatively easy to detect the attack part of the signal, there were no preprocessing for the frames excerpted from the CD recordings. Besides, all of the 10 sub-bands were used unlike Li and Ogihara (2006) where the

Table 6 Performance results of the CD excerpts for all feature sets

Feature	Performance (%)
R	59.9
Db2	48.3
Db8	52.4
Bior44	50.5
Rmvs	63.9
Db2mvs	43.7
Db8mvs	54.7
Bior44mvs	52.0

Table 7 Confusion matrices with the wavelet ridge based feature with trained/tested with isolated note samples/CD excerpts using R feature

Instrument	Trained with isolated note samples			Trained with CD excerpts		
	Classified as			Classified as		
	Cello	Oboe	Violin	Cello	Oboe	Violin
Cello	54.8	39.8	5.4	6.9	73.9	19.2
Oboe	48.4	26.7	24.9	0.0	100.0	0.0
Violin	14.9	31.8	53.3	0.7	37.5	61.8

features have been built by selecting the most efficient sub-bands. Obviously, a pre-selection of the sub-bands according to the instrument would be helpful in increasing the performance ratios.

The final experiments were made for the wavelet ridge based feature set using the isolated note samples as training data and the CD excerpts for testing data, and vice versa. Note that there are only three instruments in common, i.e., Cello, Oboe, and Violin, and Table 7 displays the confusion matrices obtained for both cases.

Because of the limited number of Oboe note samples, the performance was lower when trained with the isolated note samples. On the contrary, when trained with the CD excerpts, all Oboe samples were correctly classified. Therefore, for Oboe, the isolated note samples seem to lack a fully representation of the instrument. However, it is not the case for the Cello, which is mostly confused with Oboe when trained with the CD excerpts. This may be due to the CD recordings which possibly could not present all the note samples of the corresponding musical instrument. Results also demonstrated that the wavelet ridge based feature performs better for string instruments than the wind instruments, independent from the collection of sounds. This is in accordance with the previous results which demonstrates its ability to catch the variations in time and frequency of string instruments compared to the more stable sounds of wind instruments.

5 Conclusions

This paper offers wavelet ridges in order to build features for representing musical signals. The features based on parameterizing the wavelet sub-band energies have been found efficient but they lack the time domain information. By displaying the energy of the time-varying frequency content of the signal, the features based on wavelet ridges were found to have better classification rates than the wavelet sub-band energy based features, computed using different mother wavelet functions in a comparative study for some sample musical instrument classification problems.

Although the frequency content were summed in sub-bands in order to make fair comparisons with the sub-band based feature sets, the extraction of ridges can be performed at different scales or time resolutions according to a pre-defined precision. For example, making a finer analysis by adjusting the scales to the semitones of musical note frequencies will result with the determination of the exact note frequency while a classification of instruments is ongoing. Thus, it seems possible to fuse this information to solve the identification of note and instrument at the same time which will be helpful for musical information retrieval. However, our

observations justified that necessary caution should be taken when increasing the detail in time or frequency. Otherwise, increasing the length of the feature with redundant information is likely to result with a decrease in the classification accuracy.

The wavelet ridges offer a way of determining a unique identifier (skeleton) of each musical instrument by considering both time and frequency domain information. Therefore, similar/alternative features based on wavelet ridges can be used to exploit possible discrimination directions. However, as the ridge extraction method uses continuous wavelet transform, the necessity of high computation time is still cumbersome. An option may be to use the dual-tree complex wavelet transform (Selesnick et al. 2005) which also combines the advantages of both the continuous and discrete time wavelet analysis.

Acknowledgements The authors would like to thank anonymous reviewers for their comments and suggestions in improving the quality of the manuscript.

References

- Alm, J. F., & Walker, J. S. (2002). Time-frequency analysis of musical instruments. *SIAM Review*, 44(3), 457–476.
- Beauchamp, J. W. (ed) (2007). *Analysis, synthesis, and perception of musical sounds*. Springer.
- Boashash, B. (1992). Estimating and interpreting the instantaneous frequency of a signal - Part I: Fundamentals. *Proceedings of the IEEE*, 80(4), 520–538.
- Carmona, R. A., Hwang, W. L., & Torr sani, B. (1997). Characterization of signals by the ridges of their wavelet transforms. *IEEE Transactions on Signal Processing*, 45(10), 2586–2590.
- Carmona, R. A., Hwang, W. L., & Torr sani, B. (1999). Multiridge detection and time-frequency reconstruction. *IEEE Transactions on Signal Processing*, 47(2), 480–492.
- Cristianini, N., & Shawe-Taylor, J. (2000). *An introduction to support vector machines*. Cambridge University Press.
- Dai, Y., Ma, Q., & Tang, W. (2008). Efficient wavelet ridge extraction method for asymptotic signal analysis. *Review of Scientific Instruments*, 79, 124703. doi:10.1063/1.3036984.
- Delprat, N. (1997). Global frequency modulation laws extraction from the Gabor transform of a signal: A first study of the interacting components case. *IEEE Transactions on Speech and Audio Processing*, 5(1), 64–71.
- Delprat, N., Escud , B., Guillemain, P., Kronland-Martinet, R., Tchamitchian, P., & Torr sani, B. (1992). Asymptotic wavelet and Gabor analysis: Extraction of instantaneous frequencies. *IEEE Transactions on Information Theory*, 38(2), 644–664.
- Deng, J. D., Simmermacher, C., & Cranefield, S. (2008). A study on feature analysis for musical instrument classification. *IEEE Transactions on Systems, Man and Cybernetics - Part B: Cybernetics*, 38(2), 429–438.
- Do, M. N., & Vetterli, M. (2002). Wavelet-based texture retrieval using generalized Gaussian density and Kullback-Leibler distance. *IEEE Transactions on Image Processing*, 11, 146–158.
- Essid, S., Richard, G., & David, B. (2006). Instrument recognition in polyphonic music based on automatic taxonomies. *IEEE Transactions on Audio, Speech, and Language Processing*, 14(1), 68–80.
- Every, M. R., & Szymanski, J. E. (2006). Separation of synchronous pitched notes by special filtering of harmonics. *IEEE Transactions on Audio, Speech, and Language Processing*, 14(5), 1845–1856.
- Fritts, L. (1997). *The University of Iowa electronic music studios musical instrument samples*. <http://theremin.music.uiowa.edu>.
- Goodwin, M., & Vetterli, M. (1996). Time-frequency signal models for music analysis, transformation, and synthesis. In *Proc. of the IEEE-SP international symposium on time-frequency and time-scale analysis* (pp 133–136).
- Guillemain, P., & Kronland-Martinet, R. (1996). Characterization of acoustic signals through continuous linear time-frequency representations. *Proceedings of the IEEE*, 84(4), 561–585.
- Hachibaboğlu, H., & Canagarajah, C. N. (2002). Musical instrument recognition with wavelet envelopes. In *Proc. forum acusticum Sevilla*. Sevilla, Spain.

- Herrera-Boyer, P., Peeters, G., & Dubnov, S. (2003). Automatic classification of musical instrument sounds. *Journal of New Music Research*, 32(1), 3–21.
- Herrera-Boyer, P., Klapuri, A., & Davy, M. (2006). Automatic classification of pitched musical instrument sounds. In *Signal processing methods for music transcription* (pp. 163–200). Springer.
- Klapuri, A., & Davy, M. (eds) (2006). *Signal processing methods for music transcription*. Springer.
- Kostek, B. (2004). Musical instrument classification and duet analysis employing music information retrieval techniques. *Proceedings of the IEEE*, 92(4), 712–729.
- Kostek, B. (2005). *Perception-based data processing in acoustics*. Springer.
- Kostek, B., & Czyzewski, A. (2001). Representing musical instrument sounds for their automatic classification. *Journal of Audio Engineering Society*, 49(9), 768–785.
- Kronland-Martinet, R. (1988). The wavelet transform for analysis, synthesis, and processing of speech and music sounds. *Computer Music Journal*, 12(4), 11–20.
- Li, T., & Ogihara, M. (2006). Toward intelligent music information retrieval. *IEEE Transactions on Multimedia*, 8(3), 564–574.
- Lin, C. C., Chen, S. H., Truong, T. K., & Chang, Y. (2005). Audio classification and categorization based on wavelets and support vector machine. *IEEE Transactions on Speech and Audio Processing*, 13(5), 644–651.
- Lin, J. (2006). Ridges reconstruction based on inverse wavelet transform. *Journal of Sound and Vibration*, 294, 916–926.
- Mallat, S. G. (2009). *A wavelet tour of signal processing: The sparse way* (3rd edn). Academic Press.
- Newland, D. E. (1994). Harmonic and musical wavelets. *Proceedings: Mathematical and Physical Sciences*, 444(1922), 605–620.
- Olmo, G., Dovis, F., Benotto, P., Calosso, C., & Passaro, P. (1999). Instrument-independent analysis of music by means of the continuous wavelet transform. In *Proc. of the SPIE wavelet applications in signal and image processing VII* (Vol. 3813, pp. 716–726). Denver, USA.
- Özbek, M. E., & Savacı F. A. (2007). Music instrument classification using generalized Gaussian density modeling. In *IEEE 15th signal processing and communications applications conference*. Eskişehir, Turkey.
- Özbek, M. E., Delpha, C., & Duhamel, P. (2007). Musical note and instrument classification with likelihood-frequency-time analysis and support vector machines. In *European signal processing conference* (pp. 941–945). Poznań, Poland.
- Özkurt, N., & Savacı, F. A. (2005). Determination of wavelet ridges of nonstationary signals by singular value decomposition. *IEEE Transactions on Circuits and Systems-II: Express Briefs*, 52(8), 480–485.
- Özkurt, N., & Savacı, F. A. (2006). Reconstruction of nonstationary signals along the wavelet ridges. *International Journal of Bifurcation and Chaos*, 16(1), 191–198.
- Pielemeier, W. J., Wakefield, G. H., & Simoni, M. H. (1996). Time-frequency analysis of musical signals. *Proceedings of the IEEE*, 84(9), 1216–1230.
- Pruysers, C., Schnapp, J., & Kaminskyj, I. (2005). Wavelet analysis in musical instrument sound classification. In *Proc. of the 8th international symposium on signal processing and its applications* (Vol. 1, pp. 1–4).
- Schölkopf, B., & Smola, A. J. (2002). *Learning with Kernels*. MIT Press.
- Sejdić, E., Djurović, I., & Jiang, J. (2009). Time-frequency feature representation using energy concentration: An overview of recent advances. *Digital Signal Processing*, 19, 153–183.
- Selesnick, I. W., Baraniuk, R. G., & Kingsbury, N. G. (2005). The dual-tree complex wavelet transform. *IEEE Signal Processing Magazine*, 22(6), 123–151.
- Shafi, I., Ahmad, J., Shah, S. I., & Kashif, F. M. (2009). Techniques to obtain good resolution and concentrated time-frequency distributions: A review. *EURASIP Journal on Advances in Signal Processing*, 2009(673539), 43. doi:10.1155/2009/673539.
- Shuai, X. F., & Yuan, X. (2006). Instantaneous frequency extraction via wavelet ridge. In *IEEE international conference on communications, circuits and systems proceedings* (Vol. 1, pp. 253–257).
- Smith, L. M., & Honing, H. (2008). Time-frequency representation of musical rhythm by continuous wavelets. *Journal of Mathematics and Music*, 2(2), 81–97.
- Todorovska, M. I. (2001). *Estimation of instantaneous frequency of signals using the continuous wavelet transform*. Tech. Rep. CE 01-07. University of Southern California, Los Angeles, California.
- Tse, N. C. F., & Lai, L. L. (2007). Wavelet-based algorithm for signal analysis. *EURASIP Journal on Advances in Signal Processing*, 2007(38916). doi:10.1155/2007/38916.

- Tzagkarakis, C., Mouchtaris, A., & Tsakalides, P. (2006). Musical genre classification via generalized Gaussian and alpha-stable modeling. In *IEEE international conference on acoustics, speech, and signal processing* (Vol. 5, pp. 217–220). Toulouse, France.
- Tzanetakis, G., Essl, G., & Cook, P. (2001). Audio analysis using the discrete wavelet transform. In *WSES international conference on acoustics and music: Theory and applications*. Koukounaries, Greece.
- Vapnik, V. N. (1998). *Statistical learning theory*. John Wiley and Sons.
- Vetterli, M., & Kovačević, J. (1995). *Wavelets and subband coding*. Prentice Hall.
- Weston, J., Elisseeff, A., Bakır, G., & Sinz, F. (2003). Spider: Object-orientated machine learning library. <http://www.kyb.tuebingen.mpg.de/bs/people/spider>.
- Wieczorkowska, A. (2001). Musical sound classification based on wavelet analysis. *Fundamenta Informaticae*, 47, 175–188.
- Wieczorkowska, A. A. (2010). Towards musical data classification via wavelet analysis. *Lecture Notes in Computer Science*, 1932, 147–165.
- Wieczorkowska, A. A., & Kubera, E. (2009). Identification of a dominating instrument in poly-timbral same-pitch mixes using SVM classifiers with non-linear kernel. *Journal of Intelligent Information Systems*, doi:10.1007/s10844-009-0098-3.
- Wieczorkowska, A., & Kubik-Komar, A. (2010). Application of analysis of variance and post hoc comparisons to studying the discriminative power of sound parameters in distinguishing between musical instruments. *Journal of Intelligent Information Systems*. doi:10.1007/s10844-010-0140-5.