Turk J Elec Eng & Comp Sci, Vol.20, No.6, 2012, © TÜBİTAK
doi:10.3906/elk-1102-1031

Skewed alpha-stable distributions for modeling and classification of musical instruments

Mehmet Erdal ÖZBEK*, **Mehmet Emre ÇEK**, **Ferit Acar SAVACI**
Department of Electrical and Electronics Engineering, İzmir Institute of Technology, Urla, 35430, İzmir-TURKEY
e-mail: erdalozbek@iyte.edu.tr

Received: 09.02.2011

Abstract

Music information retrieval and particularly musical instrument classification has become a very popular research area for the last few decades. Although in the literature many feature sets have been proposed to represent the musical instrument sounds, there is still need to find a superior feature set to achieve better classification performance. In this paper, we propose to use the parameters of skewed alpha-stable distribution of sub-band wavelet coefficients of musical sounds as features and show the effectiveness of this new feature set for musical instrument classification. We compare the classification performance with the features constructed from the parameters of generalized Gaussian density and some of the state-of-the-art features using support vector machine classifiers.

Key Words: *Musical instrument classification, skewed alpha-stable distribution, generalized Gaussian density, support vector machine.*

1. Introduction

The classification of musical instruments plays an important part in music information retrieval (MIR) and specifically for transcription of music [1]. In order to transcribe the note symbols of each instrument, each instrument should be classified based on some discriminative properties known as features. There have been many features proposed for musical signals representing either temporal or spectral properties of the signals [1, 2, 3, 4]. Time based features like zero-crossing rate, autocorrelation coefficients, or rise time characterize the temporal evolution of the signal. On the other hand, the spectral features were extracted using time-frequency transformations such as short-time Fourier transform, constant-Q transform, and wavelets [5]. Many features characterize the short-time spectrum of the signal, e.g., spectral centroid, flux, flatness, and mel frequency cepstral coefficients (MFCCs). The selection of features is an important issue with many models, methods, and selection criteria [6], although relevant features were generally selected by some ranking and dimension reduction

*Corresponding author: Department of Electrical and Electronics Engineering, İzmir Institute of Technology, Urla, 35430, İzmir-TURKEY

techniques [2, 7, 8]. Based on the definition of the general classification procedure [4], feature extraction/selection was followed by various classification algorithms including neural networks (NN) and support vector machines (SVM) which demonstrated successful classification rates [2, 3, 9, 10, 11, 12]. An evaluation of the different features were further investigated with different classifiers, cross-validation schemes, and variance analysis [7, 13].

Despite the existing ones, new methods and features are being investigated aiming to achieve higher classification rates. Depending on the effectiveness of time-frequency features, wavelets were frequently used. Musical sounds were classified by dividing the spectrum into frequency bands and then by parameterizing the energy of each sub-band [3, 14]. For each of the musical instruments, the spectrum was divided into 10 wavelet sub-bands where the energy values of each band were presented. In [3], it was observed that the distribution pattern of energy values within the sub-bands was similar for the wind instruments, whereas the same parametric representation for the string instruments was different from the wind instruments. This observation is consistent with the assumption of [15], stating that the energy distribution in frequency domain identifies a texture, which led to model a texture by the marginal densities of wavelet sub-band coefficients. Following the same idea, histograms computed from the wavelet coefficients at different frequency sub-bands were shown to be efficient in music information retrieval [16].

The feature extraction and similarity measurement method given in [15] for image retrieval was based on modeling the distribution of the sub-band wavelet coefficients with Generalized Gaussian Density (GGD). The similarity measurement between two wavelet sub-bands was calculated with Kullback-Leibler (KL) distance between their distributions by using only the model parameters. This GGD modeling was used for musical genre classification [17], and further for classification of shallow-water acoustic signals including an extension of the feature extraction model to the symmetric alpha-stable distributions [18]. The same GGD modeling was also used for musical instrument classification with various mother wavelet functions [19].

However, most of the real world signals have distributions with heavier tails than of the Gaussian density, moreover, they may have unsymmetrical (skewed) distributions rather than symmetric distributions. Therefore, the class of stable distribution provides a better model which includes Gaussian distribution as a special case [20]. The superiority was shown empirically for the audio signals considering only symmetric alpha-stable distributions [21]. Although the symmetric alpha-stable distributions solve the problem of representing data with heavy tailed distributions, the assumption of symmetrically distributed data poses an important restriction. For example in [22], the skewed alpha-stable distributions were shown to model the textures in images sufficiently. Recently, it was shown that the alpha-stable distributions satisfactorily model the gene expression distributions [23].

Musical signals often have significant spectral content in frequency ranges for which sub-band coefficients are not necessarily distributed in a symmetrical or Gaussian distribution. Therefore, in order to represent musical instrument signals in terms of features, we argue that the musical instrument signals might also be modeled with the skewed alpha-stable distribution. Based on this motivation, the features representing information associated with different musical instruments can be obtained by determining the alpha-stable distribution parameters of the sub-band wavelet coefficients of the musical instrument note samples. Thus, in this paper, we propose to build novel features by modeling the distribution of the sub-band wavelet coefficients of musical instrument samples by the skewed alpha-stable distribution. We calculated the parameters of the alpha-stable distribution from the sub-band wavelet coefficients of isolated note samples of different musical instruments. In order to show that our proposed alpha-stable model represents the musical instrument samples efficiently, we also formed features using the parameters of GGD model and computed some of the state-of the-art features (i.e., MFCCs, MIR, and MPEG-7 features) [1, 2, 7]. Then the classification performance of each feature set was evaluated using SVM classifiers.

The paper is organized as follows: Section 2 summarizes the parameter estimation methods of the GGD and the skewed alpha-stable distribution models. Section 3 describes the experimental work including the composition of the alpha-stable and the GGD features for representing musical instrument sounds. Section 4 presents the classification results. Finally, the last section evaluates the results and concludes the paper.

2. Estimation of the distribution parameters

2.1. The GGD based modeling

The continuous wavelet transform (CWT) of a signal $s(t)$ is defined by [25]

$$CWT_s(a,b;\psi) = \int_{-\infty}^{\infty} s(t)\frac{1}{\sqrt{a}}\psi^*\left(\frac{t-b}{a}\right)dt, \tag{1}$$

where a and b are respectively the scaling and translation coefficients, the $*$ denotes the complex conjugate, and ψ is the mother wavelet function. In practice, the CWT is regularly sampled at discrete time and scale positions

$$\psi_{m,n}(t) = a_0^{-m/2}\psi\left(a_0^{-m}t - b_0 n\right) \qquad m,n \in Z \tag{2}$$

where the signal can be decomposed into approximation and detail coefficients. This can be performed by interpreting CWT as a filter bank based on the multi-resolution concept where the signal can be decomposed into many detail and approximation sub-band coefficients [26].

The distribution of wavelet coefficients in each sub-band of the musical instrument sample can be modeled using GGD [15] as

$$p(x_{i,j};\alpha_G,\beta_G) = \frac{\beta_G}{2\alpha_G\Gamma(1/\beta_G)}e^{-(|x_{i,j}|/\alpha_G)^{\beta_G}} \tag{3}$$

where p is the probability density function (pdf) of the wavelet coefficients $x_{i,j}$ for the i^{th} sub-band, $\Gamma(\cdot)$ is the Gamma function given by

$$\Gamma(z) = \int_0^{\infty} e^{-t}t^{z-1}dt, \quad z > 0 \tag{4}$$

and α_G is referred as the scale parameter while β_G is known as the shape parameter, where the subscript G denotes the generalized Gaussian case. The Gaussian distribution is a special case of GGD with $\beta_G = 2$. The GGD parameters can be estimated using the maximum likelihood (ML) estimator [15] where for a fixed $\beta > 0$, the scale parameter for the i^{th} sub-band can be found as

$$\hat{\alpha}_G = \left(\frac{\beta_G}{L}\sum_{j=1}^{L}|x_{i,j}|^{\beta_G}\right)^{1/\beta_G}, \tag{5}$$

where $|x_{i,j}|$, $(j = 1,\ldots,L)$ are the absolute value of the wavelet coefficients in the i^{th} sub-band. The shape parameter can be found as the solution of the following equation

$$1+\frac{\Psi(1/\hat{\beta}_G)}{\hat{\beta}_G}-\frac{\sum_{j=1}^{L}|x_{i,j}|^{\hat{\beta}_G}\log|x_{i,j}|}{\sum_{j=1}^{L}|x_{i,j}|^{\hat{\beta}_G}}+\frac{\log\left(\frac{\hat{\beta}_G}{L}\sum_{j=1}^{L}|x_{i,j}|^{\hat{\beta}_G}\right)}{\hat{\beta}_G}=0\,, \tag{6}$$

which can be solved numerically where $\Psi(\cdot)$ is the digamma function given as $\Psi(z)=\Gamma'(z)/\Gamma(z)$ with the prime sign $(')$ denoting the derivative. Therefore, the distribution of the wavelet sub-band coefficients can be completely defined via the two parameters α_G and β_G. Figure 1 displays the wavelet sub-band coefficients of a Cello note sample with the number of coefficients $L=9000$, and the corresponding histogram where the estimated parameters are $\alpha_G=0.15$ and $\beta_G=1.32$.

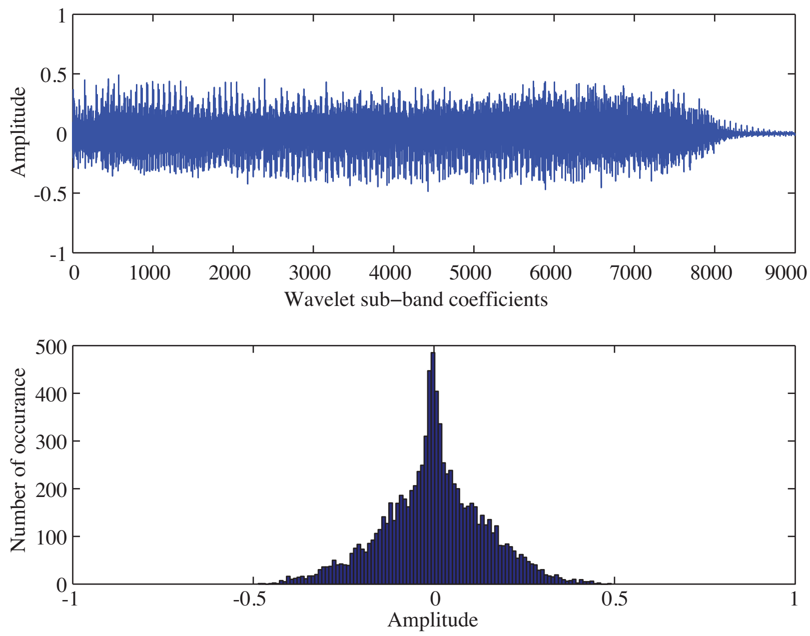

Figure 1. Wavelet sub-band coefficients of a Cello note sample and corresponding histogram.

2.2. Alpha-stable distribution based modeling

The alpha-stable distributions are generalizations of Gaussian distributions and share many properties. However an analytical expression for their pdf does not exist hence they are mostly described by their characteristic function [20]. The absence of the probability density function also explains the difficulty in building ML estimators for the alpha-stable distribution parameters [24].

One dimensional alpha-stable distribution is expressed by its characteristic function [20] as

$$\phi(t)=\begin{cases}\exp\left\{j\mu t-\gamma|t|^{\alpha}\left(1+j\beta sign(t)\tan(\frac{\alpha\pi}{2})\right)\right\}, & \text{if } \alpha\neq 1\\ \exp\left\{j\mu t-\gamma|t|^{\alpha}\left(1+j\beta sign(t)\frac{2}{\pi}\log|t|\right)\right\}, & \text{if } \alpha=1\end{cases} \tag{7}$$

where $\alpha \in (0,2]$, $\beta \in [-1,1]$, $\gamma > 0$, and $\mu \in (-\infty,\infty)$. The parameters given in Equation (7) characterize the pdf where the characteristic exponent α determines the impulsiveness, the skewness parameter β represents the symmetry, the scale parameter γ corresponds to the variance, and the mean of the density is represented by the parameter μ.

The parameters of the alpha-stable distribution for a data sequence $x_{i,j}$ is computed by first evaluating the p^{th} order fractional moments A_{p_i} and S_{p_i} as

$$A_{p_i} = \frac{1}{L}\sum_{j=1}^{L}|x_{i,j}|^p, \quad S_{p_i} = \frac{1}{L}\sum_{j=1}^{L} sign(x_{i,j})|x_{i,j}|^p. \tag{8}$$

The detailed selection criteria for an appropriate p value was given in [24] where our selection was explained in the following section. Then the estimated characteristic exponent parameter $\hat{\alpha}$ can be obtained from the measurements of sequence $x_{i,j}$ by solving

$$sinc(\frac{p\pi}{\hat{\alpha}}) = \left[q\left(\frac{A_{p_i}A_{-p_i}}{\tan(q)} + S_{p_i}S_{-p_i}\tan(q)\right)\right]^{-1}, \tag{9}$$

where $q = (p\pi)/2$. The skewness parameter can be estimated using

$$\hat{\beta} = \frac{\tan\left(\frac{\hat{\alpha}}{p}\arctan\left[\frac{S_{p_i}}{A_{p_i}}\tan\left(\frac{p\pi}{2}\right)\right]\right)}{\tan\left(\frac{\hat{\alpha}\pi}{2}\right)}, \tag{10}$$

and the scale parameter of the alpha-stable distribution can be estimated as

$$\hat{\gamma} = |\cos(\theta)|\left(\frac{\Gamma(1-p)}{\Gamma\left(1-\frac{p}{\hat{\alpha}}\right)}\frac{\cos\left(\frac{p\pi}{2}\right)}{\cos\left(\frac{p\theta}{\hat{\alpha}}\right)}A_{p_i}\right)^{\hat{\alpha}/p}, \tag{11}$$

where $\theta = \arctan\left(\hat{\beta}\tan\left(\frac{\hat{\alpha}\pi}{2}\right)\right)$. Similar to the GGD model, the distribution of the wavelet sub-band coefficients can be completely defined via the parameters α, β, and γ. For the same wavelet sub-band coefficients of the instrument note sample given in Figure 1, the estimated parameters of the alpha-stable model are $\alpha = 1.34$, $\beta = 0.01$, and $\gamma = 0.03$.

3. Experiments

We performed experiments using the instrument samples of University of Iowa Electronic Music Studios [27] which were recorded in an anechoic chamber, 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, and labeled accordingly [28]. We built the features from the GGD and alpha-stable distribution parameters and extracted the state-of-the-art features using the 20 instruments where the family and instrument names are presented in Table 1, based on the categories given in [29]. The wind instruments recordings include samples with and without vibrato whereas we selected the string instrument recordings played with bowing. Each of the instrument samples is in one of the three dynamic ranges: fortissimo (ff), mezzo forte (mf), and pianissimo (pp).

Table 1. Classification of the instruments.

Category	Sub-category	Instrument
String	Bow-string	Bass, Cello, Viola, Violin
	Keyboard	Piano
Wind	Brass	Alto Saxophone, Bass Trombone, Horn, Soprano Saxophone, Tenor Trombone, Trumpet, Tuba
	Woodwind	Alto Flute, Bass Clarinet, Bassoon, B♭ Clarinet, E♭ Clarinet, Flute, Oboe

We applied three-level one-dimensional wavelet decomposition to each of the instrument note samples as in [17, 18, 19] however with a Daubechies (db1) mother wavelet function. Consequently, we obtained the four sub-band coefficients D_1, D_2, D_3, A_3 where D_i, A_i denote the i^{th}-level detail and approximation sub-band coefficients, respectively. For each sub-band, we estimated the GGD model parameters α_G and β_G as well as the alpha-stable distribution parameters. In our study, the μ parameter was observed to have very small values close to zero and also similar for the same wavelet sub-band for different note samples. Therefore, we obtained the parameters of the alpha-stable distribution for a given musical instrument note sample wavelet sub-band sequence $x_{i,j}$, by computing the approximate values of α, β, and γ parameters by first evaluating the p^{th} order fractional moments A_{p_i} and S_{p_i} as explained in the previous section.

The selection of the fractional moment order (p) for the Equation (8) is crucial for an appropriate estimation of the distribution parameters. It has to be sufficiently smaller than the α parameter which is not known a priori. Inherently having multiple partials or harmonic frequencies, the wavelet sub-band coefficients of the musical instrument sounds can not accumulate to a constant value. Therefore, they may not have a very impulsive distribution leading to small α values (e.g., $\alpha < 0.5$). Besides, the parameter estimation performance has been verified on the generated pure skewed stable random series having at least 10000 samples with different density parameters and p values. Nevertheless, we have conducted many observations on the wavelet sub-band coefficient distributions of the musical instrument signals and selected the value of p as 0.1 which satisfies the recommended criterion.

Figure 2 presents the parameters of GGD model where each point in the figure corresponds to the parameters of GGD in a single sub-band for a specific note sample of a single instrument. The parameter values of the Violin note samples form a cluster around $\beta_G = 0.5$, $\alpha_G = 0.05$ whereas the values corresponding to the Oboe note samples are more spread. On the other hand, the parameter distributions of alpha-stable model in three dimensions, i.e., α, β, and γ, for the notes of the same instrument samples given in Figure 3 displays the clusters of both instruments which may be discriminated easier.

In this work, we built features from the parameters of GGD and alpha-stable distributions for each of the detail and approximation sub-bands. We concatenated the parameters according to the sub-bands as D_1, D_2, D_3, A_3, resulting with $\{(\alpha_{G,D_1}, \beta_{G,D_1}), (\alpha_{G,D_2}, \beta_{G,D_2}), (\alpha_{G,D_3}, \beta_{G,D_3}), (\alpha_{G,A_3}, \beta_{G,A_3})\}$, an 8-length feature vector for GGD model. Similarly, we formed a 12-length feature vector for alpha-stable (AS) model with the values of $\{(\alpha, \beta, \gamma)\}$ for the same four sub-bands. Furthermore, in order to evaluate the performance of skewed alpha-stable based features, we compared with not only the features built from GGD parameters but also with the state-of-the-art features used in musical signal processing and especially for musical instrument classification. These features with their descriptions and dimensions were presented in Table 2. The MFCCs represent the shape of the spectrum with few coefficients computed from the logarithm of the mel spectrum

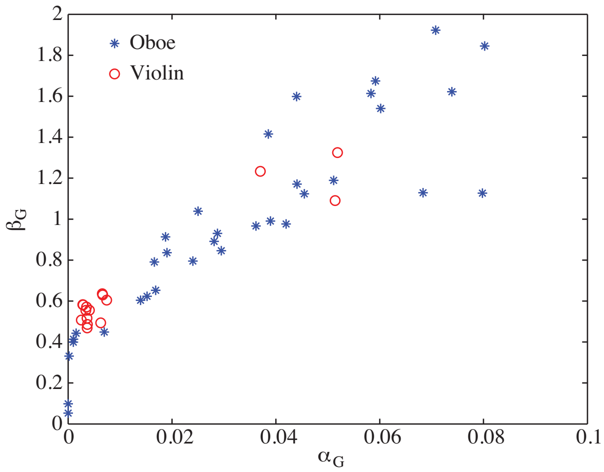

Figure 2. The representation of the generalized Gaussian density parameters.

which generally compacted to 13 coefficients using discrete cosine transform. The spectral centroid represents the mean value of the spectrum while the spectral spread is the variation of spectrum around this mean value. The spectral skewness gives a measure on the symmetry of the spectral distribution whereas the kurtosis measure the flatness. The change in the spectrum between each computed frame is represented with the mean and variance values of the spectral flux. The noisiness of the signal is measured with the zero crossing rate measure by counting the number of times that the signal passes the zero axis. On the other hand, harmonic features can be obtained similarly by computing the spectrum with harmonic frequency contents. Others, such as the temporal centroid gives the average over the energy envelope of the signal while the logarithm of the attack time is one of the important perceptually descriptors which computed as the logarithm of the time of the signal starts and reaches its stable part.

The MFCC and MIR features were extracted using the MIR Toolbox [30], while the MPEG-7 audio descriptors were obtained from an implementation [31] of MPEG-7 standard [32]. The algorithms were used with their default parameter settings. For the selection of the MIR features, we benefited from the feature ranking presented in [7] while for the MPEG-7 features we selected the descriptors of temporal and spectral timbre, although there is not yet consensus on the choice of parameters for musical instrument sound description [29].

We carried out experiments in five cases with the isolated note samples of bow-string, brass, woodwind, wind, and finally with all of the instruments. For each of the cases, we performed multi-class classifications using SVM classifiers. The SVMs have been developed based on statistical learning theory [33] and are widely used in classification problems because of their generalization ability. Although they were originally designed for solving two-class classification problems, two common methods: one-vs-all and one-vs-one, consider the multi-class classification case as a collection of two-class classifications. In one-vs-all method, k classifiers are constructed between one class and the rest $k-1$ number of classes for a k-class classification problem. For the one-vs-one method, $k(k-1)/2$ classifications are constructed between each possible class pairs. The decision is taken over all possible pairs using a majority vote. Although the choice of the method depends on the problem, the one-vs-all method often produces acceptable results [34]. Therefore, we chose to implement the SVM classifiers using one-vs-all method. Each classifier is built as a hard margin classifier and the simulations are

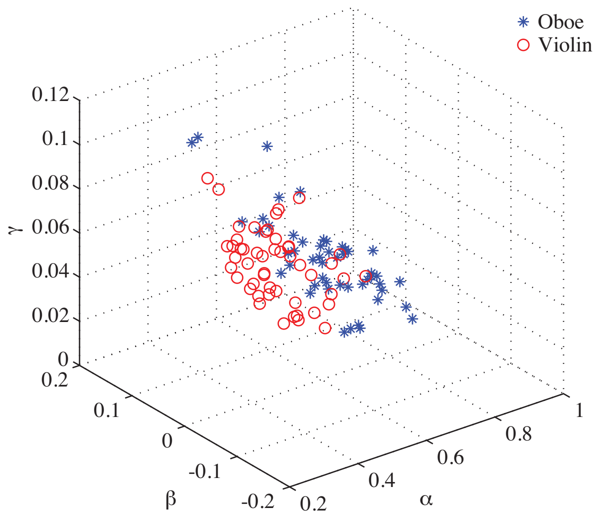

Figure 3. The representation of the skewed alpha-stable distribution parameters.

Table 2. The features, descriptions, and dimensions.

Scheme	Name/Description	Length
AS	Skewed alpha-stable distribution parameters for 3-level wavelet decomposition	12
GGD	Generalized Gaussian density parameters for 3-level wavelet decomposition	8
MFCC	Mel frequency cepstral coefficients	13
MIR	Spectral centroid, Mean and standard deviation of spectral flux, Spectral kurtosis, Spectral skewness, Spectral spread, Zero crossing rate	7
MPEG-7	Harmonic spectral centroid, Harmonic spectral deviation, Harmonic spectral spread, Harmonic spectral variation, Log attack time, Spectral centroid, Temporal centroid	7

performed using the radial basis function (RBF) kernel. The presented results are the average values obtained with the kernel parameter $\sigma = 1$, after a 10-fold stratified cross-validation scheme.

4. Results

We present the classification performance of our proposed skewed alpha-stable based feature for each of the five cases in confusion matrices where all the numbers are given in percentage. The average performance of the cases for each of the feature groups, i.e., AS, GGD, MFCC, MIR, and MPEG-7, are displayed in bar plots.

Table 3 presents the confusion matrix obtained for the classification of bow-string instruments. Cello was found to have the highest performance while the overall classification performance was found to be over 90% for all of the four instruments.

For the seven instruments of brass family, the confusion matrix is shown in Table 4. The highest performance with over 98% was found to be of Trumpet while Bass Trombone had the least accuracy among brass instruments. The closeness of the instruments, e.g., Alto Saxophone and Soprano Saxophone, can be seen from the misclassification ratios. Moreover, Bass Trombone was highly confused by Soprano Saxophone whereas

Table 3. Confusion matrix for the string instruments.

Instrument	Classified as			
	Bass	Cello	Viola	Violin
Bass	**91.8**	0.9	4.1	3.2
Cello	0.4	**98.2**	0.2	1.2
Viola	0.9	0.9	**95.9**	2.3
Violin	0.8	0.8	2.5	**95.9**

Tenor Trombone was misclassified as Trumpet.

Table 4. Confusion matrix for the brass instruments.

Instrument	Classified as						
	b	g	l	o	p	q	r
b = Alto Saxophone	**87.2**	1.1	0.1	9.6	0.7	0.9	0.4
g = Bass Trombone	2.6	**77.4**	7.7	9.1	1.1	2.1	0.0
l = Horn	0.2	6.2	**77.9**	8.8	4.6	0.4	1.9
o = Soprano Saxophone	5.3	1.5	2.9	**89.1**	0.6	0.6	0.0
p = Tenor Trombone	0.6	2.2	3.0	3.0	**85.8**	5.4	0.0
q = Trumpet	0.0	0.0	0.2	1.0	0.0	**98.8**	0.0
r = Tuba	4.4	0.4	0.0	0.7	3.2	0.0	**91.3**

When the classification of eight instruments of the woodwind family was considered, the corresponding confusion matrix was obtained as given in Table 5. Bassoon was found to have the best performance while Alto Flute had the worst ratio. Again, the similarity of the instruments can be followed through the table where every instrument in the family seems to be mainly misclassified as Flute.

Table 5. Confusion matrix for the woodwind instruments.

Instrument	Classified as							
	a	d	e	f	h	j	k	m
a = Alto Flute	**73.2**	1.2	0.6	1.6	1.8	6.8	9.2	5.6
d = Bass Clarinet	0.0	**81.9**	0.1	0.0	2.1	1.9	14.0	0.0
e = Bass Flute	5.9	1.7	**82.0**	0.2	6.9	0.4	2.9	0.0
f = Bassoon	0.8	0.0	0.8	**94.4**	1.5	1.2	1.3	0.0
h = B♭ Clarinet	4.1	1.9	1.3	3.7	**77.7**	5.6	4.1	1.6
j = E♭ Clarinet	2.7	2.1	0.8	0.7	2.7	**79.3**	8.0	3.7
k = Flute	1.8	3.8	0.1	1.5	2.2	5.5	**80.1**	5.0
m = Oboe	0.6	0.0	1.2	0.0	6.8	4.2	8.7	**78.7**

We combined all the brass and woodwind instruments and performed multi-class classification considering all of the wind instruments. The confusion matrix is shown in Table 6. Having the best performance of woodwind family, Bassoon was found to have the best ratio among the wind instrument family with over 97% correct classification ratio. However, the low performances were found to have E♭ Clarinet and Oboe with accuracy ratios lower than 70%.

We presented the confusion matrix for all of the 20 instruments in Table 7. Bassoon was again found to have the best performance among all instruments while Oboe had the worst ratio, similar to the results of wind

Table 6. Confusion matrix for the wind instruments.

Instrument	Classified as														
	a	b	d	e	f	g	h	j	k	l	m	o	p	q	r
a = Alto Flute	**71.6**	0.4	2.4	0.0	0.8	2.2	3.0	3.4	10.0	1.6	2.4	0.8	1.0	0.2	0.2
b = Alto Saxophone	1.5	**80.0**	2.3	2.0	4.3	1.8	0.1	0.1	4.7	0.0	0.0	3.2	0.0	0.0	0.0
d = Bass Clarinet	0.0	2.6	**81.9**	0.1	0.1	0.7	0.6	2.7	3.3	0.9	0.0	4.1	0.1	2.6	0.3
e = Bass Flute	4.9	1.7	1.2	**88.6**	0.0	0.2	0.8	0.0	0.2	0.0	0.2	1.0	0.0	0.6	0.6
f = Bassoon	0.2	0.0	0.0	0.0	**97.2**	0.5	0.0	0.3	0.2	0.0	1.3	0.3	0.0	0.0	0.0
g = Bass Trombone	2.0	2.9	2.9	0.0	0.0	**77.9**	0.0	0.0	2.3	3.6	0.0	6.3	0.3	0.0	1.8
h = B♭ Clarinet	0.6	3.8	0.0	1.0	0.9	0.0	**78.7**	9.9	2.4	0.1	0.3	1.9	0.3	0.1	0.0
j = E♭ Clarinet	0.5	1.3	1.3	0.5	1.8	0.0	5.2	**64.0**	7.7	0.0	1.8	8.9	5.0	1.8	0.2
k = Flute	2.3	1.1	0.6	0.4	0.8	0.9	3.0	3.3	**77.5**	2.1	1.3	1.6	0.0	4.7	0.4
l = Horn	1.0	0.4	0.0	0.0	0.0	11.6	0.0	1.7	6.3	**74.2**	0.2	3.8	0.6	0.2	0.0
m = Oboe	4.0	0.8	0.2	0.0	0.0	0.0	1.9	0.4	7.7	1.5	**67.9**	6.9	0.4	7.9	0.4
o = Soprano Saxophone	0.4	4.6	5.5	0.0	0.1	3.2	0.3	2.1	1.8	2.4	3.7	**71.6**	3.6	0.7	0.0
p = Tenor Trombone	0.4	0.0	0.4	0.0	0.0	2.6	0.8	7.2	9.0	0.0	0.0	1.8	**76.8**	0.8	0.2
q = Trumpet	0.2	0.1	0.0	0.4	0.0	0.1	3.0	0.0	1.3	0.9	3.1	1.4	0.5	**89.0**	0.0
r = Tuba	2.5	0.5	1.6	2.5	0.4	0.2	3.8	0.9	5.3	0.0	0.5	1.8	2.0	0.0	**78.0**

instrument classification. Again, E♭ Clarinet and Oboe had the lowest classification ratios which was also the case in [7] for 20 instruments classification where a subset of 21 features among 44 have been used. Piano was found to have a high ratio over 90%, in accordance with the previous studies [2, 3, 7].

Table 7. Confusion matrix for all of the instruments.

Instrument	Classified as																			
	a	b	c	d	e	f	g	h	i	j	k	l	m	n	o	p	q	r	s	t
a = Alto Flute	**65.6**	3.0	0.2	0.0	4.4	0.4	0.2	1.8	6.0	4.6	5.6	0.0	4.0	0.0	0.0	0.0	0.4	1.6	1.8	0.4
b = Alto Saxophone	2.2	**74.3**	3.5	1.8	0.2	1.1	1.9	0.2	1.0	0.9	4.9	1.9	0.0	0.0	4.0	0.6	0.0	0.0	1.5	0.0
c = Bass	0.1	1.3	**89.0**	0.2	0.1	0.1	0.0	0.0	0.3	0.1	1.3	0.0	0.7	0.8	3.0	0.0	0.0	0.1	1.7	1.2
d = Bass Clarinet	0.0	0.9	1.6	**70.0**	0.0	1.6	2.4	1.4	9.7	0.4	1.9	0.4	0.0	0.2	2.1	1.1	0.6	3.4	2.3	0.0
e = Bass Flute	0.0	9.0	2.0	0.2	**78.8**	0.0	0.2	0.0	3.1	0.2	0.4	0.0	0.0	0.0	1.6	0.0	0.2	3.5	0.6	0.2
f = Bassoon	1.5	2.1	0.3	0.2	1.3	**92.9**	0.2	0.0	0.2	0.0	1.3	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
g = Bass Trombone	0.0	5.9	1.1	4.4	0.4	0.0	**70.3**	0.0	3.0	0.0	0.8	5.5	1.0	0.0	3.2	1.7	0.0	1.2	0.0	1.5
h = B♭ Clarinet	0.7	1.7	0.0	1.3	0.1	0.9	0.0	**69.7**	3.0	10.7	2.1	0.0	1.3	0.0	2.7	0.9	1.3	2.7	0.9	0.0
i = Cello	1.1	1.5	0.1	0.0	0.3	0.6	0.0	2.3	**90.7**	0.0	0.7	0.3	0.1	0.0	0.8	0.1	0.1	0.1	0.0	1.2
j = E♭ Clarinet	0.0	0.3	0.0	2.1	0.0	0.2	0.0	3.8	1.7	**68.5**	6.3	0.3	2.5	1.7	6.7	3.5	1.7	0.0	0.0	0.7
k = Flute	2.7	2.1	0.0	0.6	0.0	1.0	1.0	2.2	2.7	3.0	**73.2**	1.2	3.1	0.5	1.6	0.1	1.6	1.0	0.0	2.4
l = Horn	0.0	2.5	0.2	0.2	0.0	0.0	4.2	0.0	0.2	2.7	5.8	**73.6**	2.9	0.2	4.6	0.4	0.0	0.0	0.0	2.5
m = Oboe	1.9	2.5	0.0	0.0	0.0	0.0	2.5	2.1	0.0	0.6	3.1	2.1	**60.4**	0.2	16.9	1.0	4.0	0.0	0.6	2.1
n = Piano	0.9	0.0	0.0	0.0	0.0	0.0	0.0	2.0	0.0	1.0	0.4	0.0	0.8	**91.6**	0.1	0.6	0.5	0.7	0.2	1.2
o = Soprano Saxophone	0.1	8.0	2.6	5.1	0.2	0.5	6.6	0.1	0.8	1.7	3.0	0.2	5.7	0.0	**61.4**	1.6	0.3	0.0	0.4	1.7
p = Tenor Trombone	0.0	0.2	0.0	0.0	0.0	0.2	1.4	0.0	0.0	0.8	1.8	3.2	4.6	0.2	2.8	**78.8**	6.0	0.0	0.0	0.0
q = Trumpet	2.0	0.0	0.1	0.0	0.0	0.0	0.9	1.1	0.0	0.3	1.8	1.3	3.3	0.0	2.2	0.3	**84.7**	0.7	0.4	0.9
r = Tuba	3.9	0.2	0.7	1.1	0.9	0.0	0.2	0.3	3.0	0.2	2.3	0.9	0.2	1.1	0.0	1.8	0.0	**80.9**	1.2	1.1
s = Viola	0.0	0.1	6.2	0.4	0.0	0.0	0.0	0.0	1.3	0.1	1.8	1.0	0.7	0.1	1.8	0.0	0.7	0.2	**81.6**	4.0
t = Violin	0.2	0.2	0.6	0.0	0.3	0.9	0.4	0.0	0.3	0.7	2.0	1.8	2.6	0.5	1.0	0.0	0.4	0.2	1.1	**86.8**

The main results of this work are summarized in the average classification performance ratios computed for each of the five feature sets, i.e, AS, GGD, MFCC, MIR, and MPEG-7. Figure 4 displays the classification performance results of different feature sets for the brass and woodwind instruments. The performance of our proposed skewed alpha-stable based feature set (AS) is almost equal to the MFCC based feature set for brass instruments while outperforms all of the four feature sets in the classification of the woodwind instruments. The decrease in the performances of all the feature sets is obvious for the woodwind instruments case whereas dramatically for the GGD based feature set. The MPEG-7 features were found not to be effective for musical instrument classification especially in identifying woodwind instruments, confirming the remarks given in the literature [3, 7, 29].

Figure 5 presents the average classification performance for string, wind, and all of the instruments for

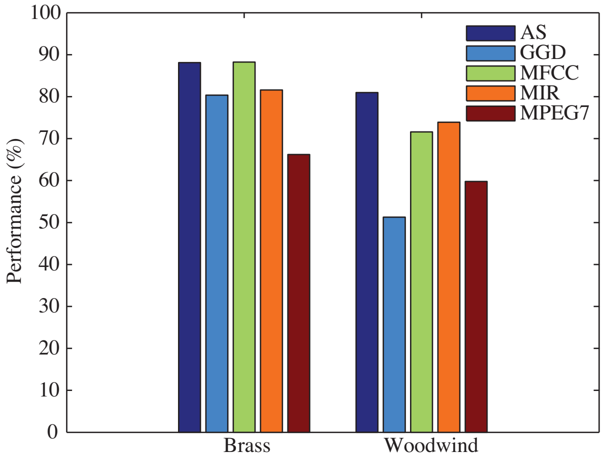

Figure 4. Classification performance for the brass and woodwind instruments.

the five feature sets. Our proposed skewed alpha-stable based feature set are found to have higher ratios than the other feature sets. The comparably higher ratios of string instruments seems to be based on the less number of classes to be discriminated whereas the highest performance for all of the instruments is lower than 80%. Note that the string instrument sounds contained only the bowed-string samples, yet they had higher number of samples than most of the other instruments. Obviously, the lower performance of the MFCC based feature set for the string instruments effected the classification performance in the case of all of the instruments, although it is higher than the other feature sets in the case of wind instruments. Nevertheless, the order of the performance ratios of the feature sets remained the same, changes occurred only for close values.

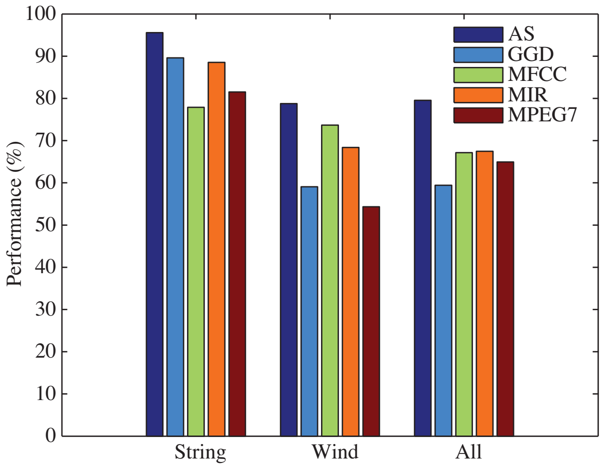

Figure 5. Classification performance for the string, wind, and all of the instruments.

The proposed method achieved better performance results than the GGD feature in all considered situations although the feature vector was composed similarly from the same wavelet sub-band coefficient distribution. The performance results for the woodwind instruments given in Figure 4 reflect that Gaussian representation may not be sufficient to discriminate between instruments based on the energy distributions. One of the reasons for these results may be the stable frequency components of the woodwind instruments making their wavelet sub-band distributions more peaky. This can also be seen from the lower individual performance of E♭ Clarinet and Oboe instruments given in Table 7. On the other hand, similar performance results for the remaining instruments denote that the proposed feature is capable of representing the discriminative property of the instruments where their sub-band energies are distributed close to a Gaussian distribution.

A statistical evaluation of the confusion matrices was performed with sensitivity and specificity scores. The sensitivity score is given by $TP/(TP + FN)$ and the specificity rate is calculated by $TN/(TN + FP)$, where TP is true positive, FN is false negative, TN is true negative, and FP is false positive, which of each denotes the number of corresponding identifications. Table 8 displays the sensitivity and specificity scores in percentages for all of the classification groups which statistically confirm the performance.

Table 8. Sensitivity and specificity scores for the classification performances.

Group	Sensitivity (%)	Specificity (%)
String	95.7	98.5
Brass	89.4	98.1
Woodwind	80.9	97.3
Wind	79.2	98.5
All	82.0	98.9

5. Conclusion

In this paper, we have presented a feature extraction method for musical instrument classification problem. The main contribution is to build features to represent the musical instrument samples by the parameters of skewed alpha-stable distribution model of their wavelet sub-band coefficient distributions. The multi-class classification performance of isolated note samples of musical instruments was evaluated using SVM classifiers with five different feature sets. The proposed skewed alpha-stable based feature outperformed the GGD based feature in all cases, and demonstrated the effectiveness of selecting the alpha-stable distribution for modeling the wavelet sub-band coefficients.

The performance was verified using the feature schemes extracted from the isolated note samples. The string instruments were found to have better average classification ratios while the performances were degraded with the increasing number of instruments. In our findings, some of the lower performance values for the state-of-the-art features than the ones in the literature were mainly due to the available number of musical instrument samples which makes fair comparison difficult. However, the classification performance of MFCC, MIR, and MPEG-7 feature sets were consistent with the works in the literature which were reflected in the order of performance ratios for feature sets, such as the lower performance of MPEG-7 features. For the individual instrument classification performance ratios, the lower confusion ratios of E♭ Clarinet and Oboe, and the higher ratio of Piano were obtained with the AS feature set similar to the performance of other feature sets in the literature. While our proposed AS feature scheme is as efficient as the other feature schemes only for this sample database case, there were no feature selecting and/or ranking schemes, no dimension reduction techniques, and no combination of the parameter sets as in some of the results given in literature. Therefore, it would be interesting to investigate the proposed skewed alpha-stable based feature for recognition of instruments from the CD excerpts which will be the subject of future research efforts.

References

[1] A. Klapuri, M. Davy (Eds.), *Signal Processing Methods for Music Transcription*, Springer, 2006.

[2] S. Essid, G. Richard, B. David, "Instrument recognition in polyphonic music based on automatic taxonomies", IEEE Trans. on Audio, Speech, and Language Processing, vol. 14, no. 1, pp. 68-80, 2006.

[3] B. Kostek, "Musical instrument classification and duet analysis employing music information retrieval techniques", Proceedings of IEEE, vol. 92, no. 4, pp. 712-729, 2004.

[4] P. Herrera-Boyer, G. Peeters, S. Dubnov, "Automatic classification of musical instrument sounds", Journal of New Music Research, vol. 32, no. 1, pp. 3-21, 2003.

[5] W. J. Pielemeier, G. H. Wakefield, M. H. Simoni, "Time-frequency analysis of musical signals", Proc. of IEEE, vol. 84, no. 9, pp. 1216-1230, 1996.

[6] I. Guyon, A. Elisseeff, "An introduction to variable and feature selection", Journal of Machine Learning Research, vol. 3, pp. 1157-1182, 2003.

[7] J. D. Deng, C. Simmermacher, S. Cranefield, "A study on feature analysis for musical instrument classification", IEEE Trans. on Systems, Man, and Cybernetics - Part B: Cybernetics, vol. 38, no. 2, pp. 429-438, 2008.

[8] S. Wegener, M. Haller, J. Burred, T. Sikora, S. Essid, G. Richard, "On the robustness of audio features for musical instrument classification", In: 16^{th} European Signal Processing Conference, Lausanne, Switzerland, 2008.

[9] Q. Ding, N. Zhang, "Classification of recorded musical instruments sounds based on neural networks", In: Proc. of the IEEE Symposium on Computational Intelligence in Image and Signal Processing, Honolulu, Hawaii, USA, pp. 157-162, 2007.

[10] J. Eggink, G. J. Brown, "Instrument recognition in accompanied sonatas and concertos", In: Proc. of IEEE International Conference on Acoustics, Speech and Signal Processing, Montreal, Canada, vol. 4, pp. 217-220, 2004.

[11] J. R. De Gruijl, M. A. Wiering, "Musical instrument classification using democratic liquid state machines", In: Y. Saeys, E. Tsiporkova, B. D. Baets, Y. V. de Peer (Eds.), 15^{th} Belgian-Dutch Conference on Machine Learning, pp. 33-40, 2006.

[12] J. Marques, P. J. Moreno, "A study of musical instrument classification using Gaussian mixture models and support vector machines". Technical Report Series CRL 99/4, Compaq Corporation, Cambridge Research Laboratory, 1999.

[13] A. Wieczorkowska, A. Kubik-Komar, "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, vol.37, no. 3 pp. 293-314. doi:10.1007/s10844-010-0140-5, 2011.

[14] A. Wieczorkowska, "Musical sound classification based on wavelet analysis", Fundamenta Informaticae, vol. 47, no. 1-2, pp. 175-188, 2001.

[15] M. N. Do, M. Vetterli, "Wavelet-based texture retrieval using generalized Gaussian density and Kullback-Leibler distance", IEEE Trans. on Image Processing, vol. 11, pp. 146-158, 2002.

[16] T. Li, M. Ogihara, "Toward intelligent music information retrieval", IEEE Trans. on Multimedia, vol. 8, no. 3, pp. 564-574, 2006.

[17] C. Tzagkarakis, A. Mouchtaris, P. Tsakalides, "Musical genre classification via generalized Gaussian and alpha-stable modeling", In: IEEE International Conference on Acoustics, Speech, and Signal Processing, Toulouse, France, vol. 5, pp. 217-220, 2006.

[18] M. I. Taroudakis, G. Tzagkarakis, P. Tsakalides, "Classification of shallow-water acoustic signals via alpha-stable modeling of the one-dimensional wavelet coefficients", Journal of Acoustical Society of America, vol. 119, no. 3, pp. 1396-1405, 2006.

[19] M. E. Özbek, F. A. Savacı, "Music instrument classification using Generalized Gaussian density modeling", In: IEEE 15^{th} Signal Processing and Communications Applications Conference, Eskişehir, Turkey, 2007.

[20] G. Samorodnitsky, M. S. Taqqu, *Stable Non-Gaussian Random Processes: Stochastic models with infinite variance*, Chapman and Hall/CRC, 2000.

[21] P. Kidmose, "Alpha-stable distributions in signal processing of audio signals", In: Proc. of the 41^{st} Conference on Simulation and Modelling, Scandinavian Simulation Society, pp. 87-94, 2000.

[22] E. E. Kuruoğlu, J. Zerubia, "Skewed α-stable distributions for modelling textures", Pattern Recognition Letters, vol. 24, pp. 339-348, 2003.

[23] D. Salas-Gonzalez, E. E. Kuruoğlu, D. P. Ruiz, "A heavy-tailed empirical Bayes method for replicated microarray data", Computational Statistics and Data Analysis, vol. 53, pp. 1535-1546, 2009.

[24] E. E. Kuruoğlu, "Density parameter estimation of skewed α-stable distributions", IEEE Trans. on Signal Processing, vol. 49, no. 10, pp. 2192-2201, 2001.

[25] S. G. Mallat, *A Wavelet Tour of Signal Processing: The Sparse Way* (3rd Edition), Academic Press, 2009.

[26] M. Vetterli, J. Kovačević, *Wavelets and Subband Coding*, Prentice Hall, 1995.

[27] L. Fritts, "The University of Iowa Electronic Music Studios Musical Instrument Samples", http://theremin.music.uiowa.edu, 1997.

[28] M. E. Özbek, C. Delpha, P. Duhamel, "Musical note and instrument classification with likelihood-frequency-time analysis and support vector machines", In: 15^{th} European Signal Processing Conference, Poznan, Poland, pp. 941-945, 2007.

[29] B. Kostek, *Perception-Based Data Processing in Acoustics*, Springer, 2005.

[30] O. Lartillot, P. Toiviainen, "A MATLAB toolbox for musical feature extraction from audio", In: Proc. of the 10^{th} International Conference on Digital Audio Effects, Bordeaux, France, pp. 237–244, 2007.

[31] M. Casey, "MPEG-7 sound-recognition tools", IEEE Trans. on Circuits and Systems for Video Technology, vol. 11, no. 6, pp. 737-747, 2001.

[32] ISO/IEC Working Group, "MPEG-7 overview", http://www.chiariglione.org/mpeg/standards/mpeg-7/mpeg-7.htm, 2004.

[33] V. N. Vapnik, *The Nature of Statistical Learning Theory*, Springer-Verlag, New York, 1995.

[34] B. Schölkopf, A. J. Smola, *Learning with Kernels*, MIT Press, 2002.