Application of analysis of variance and post hoc comparisons to studying the discriminative power of sound parameters in distinguishing between musical instruments

Alicja Wieczorkowska · Agnieszka Kubik-Komar

Received: 18 January 2010 / Revised: 26 July 2010 / Accepted: 19 October 2010 /
Published online: 10 November 2010
© The Author(s) 2010. This article is published with open access at Springerlink.com

Abstract In this paper, the influence of the selected sound features on distinguishing between musical instruments is presented. The features were chosen basing on our previous research. Coherent groups of features were created on the basis of significant features, according to the parameterization method applied, in order to constitute small, homogenous groups. In this research, we investigate (for each feature group separately) if there exist significant differences between means of these features for the studied instruments. We apply analysis of variance along with post hoc comparisons in the form of homogeneous groups, defined by mean values of the investigated features for our instruments. If a statistically significant difference is found, then the homogenous group is established. Such a group may consist of only one instrument (distinguished by this feature), or more (instruments similar with respect to this feature). The results show which instruments can be best discerned by which features.

Keywords Music information retrieval · Analysis of variance · Homogeneous groups

1 Introduction

Automatic classification of musical instruments in audio recordings is an example of research on music information retrieval. Huge repositories of audio data available

A. Wieczorkowska (✉)
Polish-Japanese Institute of Information Technology,
Koszykowa 86, 02-008 Warsaw, Poland
e-mail: alicja@poljap.edu.pl

A. Kubik-Komar
University of Life Sciences in Lublin,
Akademicka 13, 20-950 Lublin, Poland
e-mail: agnieszka.kubik@up.lublin.pl

for users are challenging from the point of view of content-based retrieval. The users can be interested in finding melodies sung to the microphone (queries by humming), identify the title and the performer of a piece of music submitted as an audio input containing a short excerpt from the piece (query by example), or finding pieces played by their favorite instruments. Browsing audio files by the users is a tedious task. Therefore, any automation comes very handy. If the audio data are labeled, searching is easy, but usually the text information added to the audio file is limited to the title, performer etc. In order to perform automatic content annotation, sound analysis is usually performed, sound features extracted, and then the contents can be classified into various categories, in order to fulfill the user's query and find the contents specified.

The research presented in this paper is an extended version of an article presented at the ISMIS'09 conference (Wieczorkowska and Kubik-Komar 2009a), addressing the problem of instrument identification in sound mixes. The identification of instruments playing together can aid automatic music transcription with assigning recognized pitches to instrument voices (Klapuri 2004). Also, finding pieces of music with excerpts played by a specified instrument can be desirable for many users of audio repositories. Therefore, investigating the problem of automated identification of instruments in audio recordings is vital for music information retrieval tasks.

In our earlier research (Wieczorkowska et al. 2008), we performed automatic recognition of predominant instrument in sound mixes using SVM (Support Vector Machines). The feature vector applied was used before in research on automatic classification of instruments (NSF 2010; Zhang 2007), and it contains sound attributes commonly used for timbre identification purposes. Most of the attributes describe low-level sound properties, based on MPEG-7 audio descriptors (ISO/IEC JTC1/SC29/WG11 2004), and since many of them are multi-dimensional, derivative features were used instead (minimal/maximal value etc.). Still, the feature vector is quite long, and it contains groups of attributes that can constitute a descriptive feature set themselves. In this research, we decided to compare descriptive power of these groups. In-depth statistical analysis of the investigated sets of features for the selected instruments is presented in this paper.

Our paper is organized as follows. In Section 2 we briefly familiarize the reader with task and problems related to the automatic identification of musical instrument sounds in audio recordings. The feature groups used in our research are also presented here. In the next section, we describe settings and methodology of our research, as well as the audio data used to produce the feature vectors. In Section 4 we describe in depth the results of the performed analyzes. The last section concludes our paper.

2 Automatic identification of musical instruments based on audio descriptors

Since audio data basically represent sequences of samples encoding the shape of the sound wave, these data usually are processed in order to extract feature vectors, and then automatic classification of audio data is performed. Sound features used for musical instrument identification purposes include time domain descriptors of sound, spectral descriptors, time-frequency descriptors and can be based on Fourier or wavelet analysis etc. Feature sets applied in research on instrument recognition

include MFCC (Mel-Frequency Cepstral Coefficients), Multidimensional Analysis Scaling trajectories of various sound features, statistical properties of spectrum and so on; more details can be found in Herrera et al. (2000). Many sound descriptors were incorporated into MPEG-7 standard for multimedia (including audio) content description (ISO/IEC JTC1/SC29/WG11 2004), as they are commonly used in audio research.

Various classifiers can be applied to the recognition of musical instruments. Research performed on isolated monophonic (monotimbral) sounds so far showed successful application of k-nearest neighbors, artificial neural networks, rough-set based classifiers (Wieczorkowska and Czyzewski 2003), SVM, and so on (Herrera et al. 2000). Research was also performed on polyphonic (polytimbral) data, when more than one instrument sound is present in the same time. In this case, researchers may also try to separate these sounds from the audio source. Outcome of research on polytimbral instrumental data can be found in Dziubinski et al. (2005), Itoyama et al. (2008), Little and Pardo (2008), Viste and Evangelista (2003), Wieczorkowska et al. (2008), Zhang (2007). The results of research in this area are rather difficult for comparison, since various scientists utilize different data sets: of different number of classes (instruments and/or articulation), different number of objects/sounds in each class, and basically different feature sets. The recognition of instruments for isolated sounds can reach 100% for a small number of classes, more than 90% if the instrument or articulation family is identified, or about 70% or less for recognition of an instrument when there are more classes to recognize. The identification of instruments in polytimbral mixes is lower than that, even below 50% for same-pitch sounds. More details can be found in the previous paper, focusing on this research (Wieczorkowska and Kubera 2009).

Recognition for monotimbral data is relatively easy, in particular for isolated sounds, and more challenging for polytimbral data. The research discussed in this paper aims at the identification of predominant instrument in mixes of sounds of the

Fig. 1 Sounds of the same pitch and their mixes. On the left hand side, time domain representation of sound waves is shown; on the right hand side, spectrum of these sounds is plotted. Triangular wave and flute sound are shown, both of frequency 440 Hz (i.e., A4 in MIDI notation). After mixing, spectral components (harmonic partials) overlap. The diagrams were prepared using Adobe Audition (Adobe 2003)

same pitch, as this is the most difficult case (harmonic partials in spectra overlap). Exemplary mix of two sound waves of the same pitch is shown in Fig. 1. As we can see, the flute sound is much more difficult to recognize after adding another sound with overlapping spectrum.

2.1 Feature groups

In the previous research, we have been investigating automatic identification of predominant instrument in same-pitch mixes (Wieczorkowska et al. 2008; Wieczorkowska and Kubik-Komar 2009b). The feature vector used in this research consisted of 219 features, based on MPEG-7 audio descriptors and other parameters used in automatic sound classification (ISO/IEC JTC1/SC29/WG11 2004; NSF 2010; Zhang 2007). Although these features were used before in various configurations in similar research, this feature vector was arbitrary chosen. Therefore, we decided to check if it could be limited. Actually, the feature vector contains (among others) a few groups of descriptors that alone can be applied to sound recognition (Wieczorkowska and Kubera 2009):

- *AudioSpectrumBasis*: $basis_1, \dots, basis_{165}$ —parameters of the spectrum basis functions, used to reduce the dimensionality by projecting the spectrum (for each frame) from high dimensional space to low dimensional space with compact salient statistical information. The spectral basis descriptor is a series of basis functions derived from the Singular Value Decomposition (SVD) of a normalized power spectrum. The total number of sub-spaces in basis functions in our case was 33, and for each sub-space, minimum/maximum/mean/distance/standard deviation were extracted, yielding 33 subgroups, five elements in each group. The obtained values were averaged over all analyzed frames of the sound;
- *MFCC*—minimum, maximum, mean, distance, and standard deviation of the MFCC vector, averaged through the entire sound. In order to extract MFCC, Fourier transform is calculated for the analyzed sound frames, then logarithm of the amplitude spectrum is taken. Next, spectral coefficients are grouped into 40 groups according to mel scale (perceptually uniform frequency scale). For the obtained 40 coefficients, Discrete Cosine Transform is applied, yielding 13 cepstral features per frame (Logan 2000). Distance of the vector is calculated as the sum of dissimilarity (absolute difference of values) of every pair of coordinates in the vector;
- *energy*—average energy of spectrum in the parameterized sound;
- *Tris*: $tris_1, \dots, tris_9$ —various ratios of harmonics (i.e. harmonic partials) in the spectrum; $tris_1$: ratio of the energy of the fundamental to the total energy of all harmonics, $tris_2$: amplitude difference [dB] between 1st and 2nd partial, $tris_3$: ratio of the sum of 3rd and 4th partial to the total energy of harmonics, $tris_4$: ratio of partials no. 5–7 to all harmonics, $tris_5$: ratio of partials no. 8–10 to all harmonics, $tris_6$: ratio of the remaining harmonic partials to all harmonics, $tris_7$: brightness - gravity center of spectrum, $tris_8, tris_9$: contents of even and odd (without fundamental) harmonics in spectrum, respectively;
- *AudioSpectrumFlatness*, $flat_1, \dots, flat_{25}$ —vector describing the flatness property of the power spectrum within a frequency bin for selected bins, averaged for the entire sound; flatness values describe 25 frequency bands. According to

MPEG-7 recommendations, audible range is divided into eight octaves covering 62.5 Hz–16 kHz with 1/4-octave resolution, and additional two bands covering the lowest (below 62.5 Hz) and the highest (above 16 kHz) frequencies; our 25 bands cover 24 bands with 1/4-octave resolution, starting from approximately octave no. 4 (in MIDI notation), as usually implemented in MPEG-7 feature vectors, plus additional band covering the highest frequency range.

When investigating the significance for the set of 219 sound parameters (including the ones mentioned above), used in our previous research, the attributes representing the above groups were often pointed out as significant, i.e. of high discriminant power (Wieczorkowska and Kubik-Komar 2009b). Therefore, it seems promising to perform investigations for the groups mentioned above.

Since *AudioSpectrumBasis* group presents a high-dimensional vector itself, and the first subgroup $basis_1, \dots, basis_5$ turned out to have high discriminant power, we decided to limit the *AudioSpectrumBasis* group to $basis_1, \dots, basis_5$. In *AudioSpectrumFlatness* group, $flat_{10}, \dots, flat_{25}$ had high discriminant power, whereas $flat_1, \dots, flat_9$ had not, as we have observed in our previous research (Wieczorkowska and Kubik-Komar 2009b). Also, we decided to investigate *energy* as a single conditional attribute, and the *Tris* group, as well as *MFCC* group. One could discuss whether such parameters as minimum or maximum of MFCC ($MFCC_{min}$, $MFCC_{max}$) are meaningful, but since these parameters yielded high discriminant power, we decided to investigate them.

Altogether, the following groups (feature sets) were investigated in this paper:

- *AudioSpectrumBasis* group: $basis_1, \dots, basis_5$;
- *MFCC* group: $MFCC_{min}$, $MFCC_{max}$, $MFCC_{mean}$, $MFCC_{dist}$, and $MFCC_{sd}$ parameters;
- *Energy*: single parameter, *energy*;
- *Tris* group: $tris_1, \dots, tris_9$ parameters,
- *AudioSpectrumFlatness* group: $flat_{10}, \dots, flat_{25}$.

3 Research settings

In order to check if the particular groups of sound features can discriminate instruments, we performed multivariate analysis of variance (MANOVA). Next, we analyzed the parameters from this group using univariate method (ANOVA). In case of rejecting the null hypothesis about the equality of means between instruments for a given feature group, we used post hoc comparisons to find out how we can discriminate particular instruments on the basis of parameters included in this feature group, i.e. which sound attributes from this group are best suited to recognize a given instrument (discriminate it from other instruments).

3.1 Audio data

Our data represented sounds of 14 instruments from MUMS CDs (Opolko and Wapnick 1987): B-flat clarinet, flute, oboe, English horn, trumpet, French horn, tenor trombone, violin (bowed vibrato), viola (bowed vibrato), cello (bowed vibrato), piano, marimba, vibraphone, and tubular bells. Twelve sounds, representing octave

no. 4 (in MIDI notation) were used for each instrument, as a target sound to be identified in classification. Additional sounds were mixed with the main sounds, both for training and testing of the classifiers in further experiments with automatic classification of musical instruments (Kursa et al. 2009). The level of added sounds was adjusted to 6.25%, $12.5/\sqrt{2}\%$, 12.5%, $25/\sqrt{2}\%$, 25%, $50/\sqrt{2}\%$, and 50% of the level of the main sound, since our goal was to identify the predominant instrument. For each main instrumental sound, additional four mixes with artificial sounds were prepared for each level: with white noise, pink noise, with triangular wave and with saw-tooth wave (i.e., both of harmonic spectrum) of the same pitch as the main sound. This set was prepared to be used as a training set for classifiers, subsequently tested on musical instrument sound mixes, using same pitch sounds. Each sound to be identified was mixed with 13 sounds of the same pitch representing the remaining 13 instruments from this data set. Again, the sounds added in mixes were adjusted in level, at the same levels as in training. Results of these experiments can be found in Kursa et al. (2009).

In this research, we investigated data representing musical instrument sounds, as well as mixes with artificial sounds. All these data were parameterized using feature sets as described in Section 3.

3.2 Materials and methods

In the described experiments, our target was to recognize instrument as a class. For each instruments, sound samples represented various pitch (12 notes from one octave) and various levels of added sounds. Altogether, each instrument was represented by 420 samples. Since our goal was to identify instrument, we omitted distinguishing for particular levels or pitch values.

MANOVA was used in our research to verify the hypothesis about the lack of differences (between the instruments) for vectors of mean values of the selected features. The test statistic based on Wilks Λ (Morrison 1990) was applied, which can be transformed to a statistical test having approximately an F distribution. A transformation for Wilks' lambda was given by Rao (Finn 1974; Rao 1951):

$$F = \frac{1 - \Lambda^{1/s}}{\Lambda^{1/s}} \cdot \frac{ms + 1 - d_h p / 2}{d_h p}$$

where

$$s = \sqrt{\frac{p^2 d_h^2 - 4}{p^2 + d_h^2 - 5}}$$

$$m = d_e - (p + 1 - d_h) / 2$$

- p number of variables,
- d_h number of degrees of freedom for the hypothesis,
- d_e number of degrees of freedom for the error.

Using Rao's transformation, the hypothesis H_0 is rejected with confidence $1 - \alpha$ if F exceeds the 100α upper percentage point of the F distribution with $d_h p$

and $ms + 1 - d_h p/2$ degrees of freedom (Finn 1974). In our case, $p = 5 + 5 + 1 + 9 + 16 = 36$ parameters, $d_h = k - 1$, where k —number of groups (instruments), therefore $k = 14$, and $d_e = N - k$, N —total sample size, so $N = 14 \cdot 420 = 5,880$.

This form of MANOVA results make it easier to obtain p -value and is definitely preferred (Bartlett et al. 2000). In case of rejecting this hypothesis, i.e. in case of finding out that there existed significant differences of means between instruments, we applied the post hoc comparisons between average values of the studied features, based on HSD Tukey test (HSD—Honestly Significant Difference) (Tukey 1993; Winer et al. 1991), preceded by univariate analysis of variance (ANOVA).

As assumptions of the analysis of variance, normality of distribution of the studied features is required, as well as homogeneity of variance between groups (instruments in our case). However, if the number of observations per group is fairly large, then deviations from normality do not really matter (StatSoft, Inc. 2001). This is because of the central limit theorem, according to which the sampling distribution of the mean approximates the normal distribution, irrespective of the distribution of the variable in the population. More detailed discussion of the robustness of the F statistic can be found in Box and Andersen (1955), or Lindman (1974). In our research, all features we used represented means, calculated for a sequence of frames for a given sound, thus improving normality of the distribution. As far as homogeneity of variance between groups is concerned, Lindman (1974, p. 33) shows that the F statistic is quite robust against violations of the homogeneity assumption, i.e. in case of heterogeneity of variances (Box 1954a, b; StatSoft, Inc. 2001). We might use two powerful and commonly applied tests of homogeneity of variance, namely the Levene test, and the Brown–Forsythe modification of this test. However, as mentioned above, we realize that the assumption of the homogeneity of variances is not crucial for the analysis of variance, in particular in the case of balanced (equal numbers of observations) designs. Moreover, these tests are not necessarily very robust themselves. For instance, in Glass and Hopkins (1995) the authors state that these tests are fatally flawed (1995, p. 436). Taking into consideration the above explanations, and the large number of observations in our research, each one representing a mean of frame-based features (as a single observation), as well as equality of groups, we did not pay too much attention regarding the assumptions of the analysis of variance.

The results of ANOVA give us the information if there are significant differences of instruments' means separately for each of studied features in a given group; these calculations are performed before the post-hoc comparisons are made. Whilst changing the multivariate into univariate analysis, we sacrifice the information about the relationships within the studied features, but, on the other hand, we obtain very useful information about discriminative power of each parameter.

The post hoc comparisons are presented in the form of homogeneous groups defined by mean values of a given feature, and consisting of the instruments which are not significantly different with respect to this feature.

Therefore, mean values of each feature defined homogenous groups of instruments. If differences between means for some instruments were not statistically significant, they constituted a group. The less instruments (sometimes even only one) in such a homogenous group, the higher the discerning power of a given feature.

All statistical calculations presented in this paper were obtained using STATISTICA software (StatSoft, Inc. 2001).

4 Results of analyzes of feature groups

In this section, we present results of analyzes of feature groups, as defined in Section 3. Each group represents an approximately uniform set, meaningful from the point of view of identification of particular sound timbers, thus aiding recognition of particular musical instruments.

4.1 Analysis of the *AudioSpectrumBasis* group

The results of MANOVA show that the vector of mean values of the analyzed *AudioSpectrumBasis* features significantly differed between the instruments ($F = 188.0$, $p < 0.01$). On the basis of the univariate results (ANOVA) we conclude that the means for instruments differed significantly for each parameter separately. The statistics F having a Fisher distribution with parameters 13 and 5,866 (i.e.

Instrument	mean basis1	1	2	3	4	5	6	7	8	9
marimba	0.078112	x								
flute	0.084683		x							
English horn	0.086258		x	x						
oboe	0.088256		x	x	x					
violin	0.090201			x	x	x				
trombone	0.092303			x	x	x	x			
viola	0.093202				x	x	x			
vibraphone	0.093699				x	x	x			
cello	0.094367				x	x	x			
clarinet	0.096370					x	x			
French horn	0.097365						x			
ctrumpet	0.103000								x	
tubular bells	0.104220								x	
piano	0.120136									x

Instrument	mean basis2	1	2	3	4	5	6
marimba	0.192175	x					
piano	0.195695	x	x				
vibraphone	0.198370		x				
tubular bells	0.204552			x			
French horn	0.205662			x			
trombone	0.211415				x		
clarinet	0.212789				x		
English horn	0.213999				x	x	
trumpet	0.214331				x	x	
flute	0.214437				x	x	
cello	0.215237				x	x	
oboe	0.217359					x	x
violin	0.219894						x
viola	0.220354						x

Instrument	mean basis3	1	2
marimba	0.160810	x	
violin	0.169641		x
oboe	0.169678		x
viola	0.170077		x
trumpet	0.170234		x
English horn	0.170362		x
clarinet	0.170439		x
flute	0.170605		x
trombone	0.170889		x
cello	0.171417		x
French horn	0.171706		x
tubular bells	0.171743		x
vibraphone	0.172575		x
piano	0.173102		x

Instrument	mean basis4	1	2	3	4	5	6	7	8	9
piano	10.33462	x								
vibraphone	11.79222		x							
marimba	13.07193			x						
French horn	15.82109				x					
tubular bells	16.05066				x					
cello	17.58667					x				
trombone	18.72267						x			
flute	19.64378						x			
English horn	20.70552							x		
clarinet	20.81804							x		
trumpet	21.46201								x	
viola	22.19190								x	x
oboe	22.73954									x
violin	23.05738									x

Instrument	mean basis5	1	2	3	4	5	6	7	8	9	10
piano	0.017810	x									
vibraphone	0.022358		x								
marimba	0.024512			x							
tubular bells	0.027763				x						
French horn	0.027912					x					
cello	0.029870						x				
trombone	0.032427							x			
flute	0.033932								x		
clarinet	0.034693								x	x	
English horn	0.035160								x	x	x
trumpet	0.035591									x	x
viola	0.036379										x
oboe	0.038118										x
violin	0.038379										x

Fig. 2 Homogenous groups of instruments for *AudioSpectrumBasis* parameters. The columns labeled as 1, ..., 9 represent homogenous groups with respect to a given parameter (feature)

$F(13, 5,866)$), was equal to 113.8, 115.0, 12.9, 371.78, and 405.7 for $basis_1, \dots, basis_5$, respectively, and each of these values produced p-value less than 0.01. These results allowed us to apply post-hoc comparisons for each of the *AudioSpectrumBasis* parameters, presented in form of tables consisting of homogenous groups of instruments (Fig. 2).

The results of post hoc analysis revealed that $basis_4$, $basis_5$ and $basis_1$ distinguish instruments to a large extent. The influence of $basis_2$ and $basis_3$ on differentiation between instruments is rather small. Marimba, piano, as well as vibraphone, and the pair of tubular bells and French horn, often determine separate groups. Piano,

Fig. 3 Means for the *AudioSpectrumBasis* group

vibraphone, marimba, cello and trombone are very well separated by *basis*₅, since each of these instruments constitute a 1-element group. Piano, vibraphone, marimba, and cello are separated by *basis*₄, too. Also, *basis*₁ separates marimba and piano. The *basis*₃ parameter only discerns marimba from other instruments (only two groups are produced); *basis*₂ does not separate any single instrument.

Whilst looking at the means of *basis* parameters (Fig. 3) we can indicate the parameters producing similar plots of means—these are *basis*₅, *basis*₄, and, to a lesser degree, *basis*₂. Similar values in plots of means indicate that the parameters

Fig. 4 Means for the MFCC group

producing these plots represent similar discriminative properties for these data. When mean values of some parameters are similar for several classes, i.e. instruments (which means that they are almost aligned in the plot, and distances between these values are short), then these instruments may be collocated in the same group, i.e. homogenous group with respect to these parameters. When a mean value for a particular instrument is distant from the other mean values, then this instrument is very easy to distinguish from the others with respect to this feature. For example, marimba is well distinguished from the other instruments with respect to $basis_3$ (the mean value for marimba is distant from the others), whereas mean values of $basis_4$ and $basis_5$ for English horn and trumpet are similar, and these instrument are situated in the same group, so these instruments are difficult to be differentiated basing on these features.

We can also see that despite the lowest number of groups produced by $basis_3$ the difference of means between marimba and other instruments is very high, so the influence of this parameter for such distinction (between marimba and others) is quite important.

The *AudioSpectrumBasis* group represents features extracted through SVD, so good differentiation of instruments was expected because SVD should yield the most salient features of the spectrum. On the other hand, these features might not be sufficient for discrimination of particular instruments. It is a satisfactory result that we can differentiate several instruments on the basis of three features ($basis_1$, $basis_4$, and $basis_5$). Especially, distinguishing between marimba, vibraphone and piano is a very good result, since these instruments pose difficulties in automatic instrument classification task. Their sounds (particularly marimba and vibraphone) are similar and have no sustained part, thus have no steady state, so calculation of spectrum is more difficult—but, as we can see, useful for instrument discrimination purposes.

4.2 Analysis of the MFCC group

The results of MANOVA indicate that the mean values of MFCC features differ significantly between the studied instruments ($F = 262.84$, $p < 0.01$).

The univariate results show that means of each parameter (Fig. 4) from this group were significantly different at the significance level p equal to 0.01 and $F(13, 5.866)$ values equal to 218.67, 329.92, 479.27, 698.8, and 550.9 for $MFCC_{min}$, $MFCC_{max}$, $MFCC_{mean}$, $MFCC_{dis}$ and $MFCC_{std}$ respectively.

The analysis of homogeneous groups see (Fig. 5) shows that $MFCC_{sd}$ and $MFCC_{max}$ yielded the highest difference of means, while $MFCC_{min}$ —the lowest one. Each feature defined six to nine groups, homogenous with respect to the mean value of a given feature.

The piano determined the separate group for every parameter from our MFCC feature set. The conclusion is that this instrument is very well distinguished by MFCC. However, there were no parameters here that would be capable to distinguish between marimba and flute. These two instruments were always situated in the same group, since the average values of studied parameters for these instruments do not differ too much. Vibraphone and bells were in different groups only for $MFCC_{mean}$.

Piano, cello, viola, violin, bells, English horn, oboe, French horn, and trombone constitute separate groups, so these instruments can be easily recognized on the basis

instrument	mfccmax	1	2	3	4	5	6	7	8	9
piano	3.5525	x								
vibraphone	7.8370	x								
tubularbells	8.3961	x	x							
ctrumpet	9.0100	x	x							
frenchhorn	9.4707		x							
bflatclarinet	10.6153			x						
tenorTrombone	10.7073			x						
marimba	11.9284				x					
flute	12.2110			x						
englishhorn	12.8648				x					
oboe	14.1156					x				
cello	14.7482						x			
viola	15.7892							x		
violin	18.6833								x	

instrument	mfccmin	1	2	3	4	5	6
viola	-15.1502	x					
oboe	-14.1205		x				
violin	-14.1015			x			
marimba	-13.9336		x				
englishhorn	-12.6130				x		
bflatclarinet	-12.5465				x		
cello	-12.5415					x	
ctrumpet	-12.5147					x	
tenorTrombone	-12.4508						x
flute	-12.4024					x	
vibraphone	-11.9982			x	x		
tubularbells	-11.2825					x	x
frenchhorn	-10.7254						x
piano	-4.9307						x

Instrument	mfccmean	1	2	3	4	5	6	7	8
violin	-2.4815	x							
flute	-2.2779		x						
viola	-2.2270		x						
marimba	-2.1736		x						
cello	-2.1386		x						
frenchhorn	-1.7498			x					
oboe	-1.6739			x					
tenorTrombone	-1.6233			x	x				
englishhorn	-1.5129				x	x			
vibraphone	-1.3626					x			
bflatclarinet	-0.7908						x		
ctrumpet	-0.7012							x	
tubularbells	-0.4874								x
piano	-0.2441								x

Instrument	mfccstd	1	2	3	4	5	6	7	8	9
piano	2.0407	x								
frenchhorn	4.9778		x							
tubularbells	5.0932		x							
vibraphone	5.2465		x	x						
ctrumpet	5.3980			x						
tenorTrombone	5.7191				x					
bflatclarinet	6.0528					x				
flute	6.1242					x	x			
marimba	6.4129						x			
englishhorn	6.8077							x		
cello	6.9213							x		
oboe	7.2157								x	
viola	7.5094									x
violin	7.747165									x

Instrument	mfccdis	1	2	3	4	5	6	7	8
piano	173.8251	x							
frenchhorn	433.2952		x						
tubularbells	461.3751			x					
vibraphone	471.0583			x					
ctrumpet	479.7590				x	x			
tenorTrombone	497.5668					x			
flute	532.8474						x		
bflatclarinet	553.1953							x	
marimba	555.1694							x	
cello	600.2499								x
englishhorn	623.1473								x
violin	647.3444								x
oboe	648.5515								x
viola	659.6647								x

Fig. 5 Homogenous groups of instruments for MFCC-based parameters

of MFCC. On the other hand, some groups overlap, i.e. the same instrument may belong to two groups.

The shape of plots of mean values of $MFCC_{std}$, $MFCC_{dis}$ and, to a lower extend, $MFCC_{max}$ (Fig. 4) is very similar, however the homogeneous groups, apart from piano, are different. As we mentioned before, piano is very well distinguished on the basis of all parameters from the $MFCC$ group, since it always constitutes a separate, 1-element group. This is because the means for piano and other instruments in most cases are extremely distant.

MFCC parameters described here represent general properties of the MFCC vector. We consider it a satisfactory result that such a compact representation turned out to be sufficient to discern between many instruments, mainly stringed instruments of sustained sounds, i.e. cello, viola, and violin, and wind instruments,

Fig. 6 Means for *Energy*

both woodwinds (of very similar timbre, i.e. oboe and English horn, which can be considered as a type of oboe) and brass (French horn and trombone). Even non-sustained sounds can be distinguished, i.e. tubular bells and piano, separated by every feature from our MFCC group.

4.3 Analysis of *energy*

In our previous paper (Wieczorkowska and Kubik-Komar 2009a) we concluded that *energy* yielded different results than *Tris* parameters. So in this paper we decided to analyze this parameter in a separate group.

The results of analysis of variance indicated the significant differences between mean values of studied instruments ($F = 1036.74$, $p < 0.01$). In (Fig. 6) we can notice the extremely low value for piano and the highest one for violin, so we can expect that these two instruments are well distinguished by this parameter.

Fig. 7 Homogenous groups of instruments for *energy*

Instrument	energy	1	2	3	4	5	6	7	8
piano	-8,7645	x							
vibraphone	-5,6405		x						
tubularbells	-5,1777			x					
frenchhorn	-5,0127			x					
ctrumpet	-4,3684				x				
tenorTrombone	-3,9975					x			
marimba	-3,8601					x			
bflatclarinet	-3,7595					x			
cello	-3,2123						x		
flute	-3,2120						x		
englishhorn	-3,1295						x		
viola	-2,5177							x	
oboe	-2,4680							x	
violin	-1,6318								x

Our presumption is confirmed by post-hoc results (Fig. 7). This parameter formed 8 homogeneous groups and 4 of them were determined by separate instruments such as piano, violin, vibraphone and trumpet.

Energy turned out to be quite discriminative as a single parameter. We are aware that if more input data are added (more recordings), our outcomes may need re-adjustment; still, discriminating four instruments here (piano, violin, vibraphone and trumpet) on the basis of one feature confirms high discriminative power of this attribute.

4.4 Analysis of the *tris* group

The results of MANOVA show that mean values of *tris* parameters were significantly different for the studied set of instruments ($F = 210.9$, $p < 0.01$)

The univariate results of $F(13, 5,866)$ test are as follows: $tris_1 - 352.08$ $tris_2 - 40.35$, $tris_3 - 402.114$, $tris_4 - 280.86$, $tris_5 - 84.14$, $tris_6 - 19.431$, $tris_7 - 12.645$, $tris_8 - 436.89$, $tris_9 - 543.39$ and all these values indicated significant differences between means of studied instruments at the significance level equal to 0.01. For the *tris* feature set consisting of $tris_1, \dots, tris_9$ parameters, each parameter defined from three to nine groups, as presented in Fig. 8.

As we can see, $tris_3$, $tris_8$, and $tris_9$ produced the highest number of homogeneous groups. Some wind instruments (trumpet, trombone, English horn, oboe), or their pairs, were distinguished most easily—determined separate groups for the features forming eight to nine homogeneous groups.

Taking into considerations the plots of mean values (Fig. 9) we can add some more information. Namely, the $tris_3$ parameter, in spite of constituting the lowest number of homogeneous groups, distinguishes piano very well. In most cases, the means of vibraphone and marimba, sometimes also piano, are similar, and when they are high, then at the same time the means for oboe and trumpet are low, and vice versa.

In case of the *Tris* group, we were expecting good results, since these features were especially designed for the purpose of musical instrument identification. For instance, clarinet shows low contents of even harmonic partials in its spectrum for lower sounds ($tris_8$). However, as we can see, other instruments—piano, vibraphone—also show low value of contents of even partials, and marimba even lower than these instruments. However, clarinet shows very high $tris_9$, i.e. amount of odd harmonic partials (excluding the fundamental, marked as no. 1) in the spectrum, which corresponds somehow to small amount of even partials—and this feature discriminates clarinet very well. The results for the *Tris* parameters, presented in Fig. 8, show that this set of features is quite well designed and can be applied as a helpful tool for musical instrument identification purposes.

4.5 Analysis of the *AudioSpectrumFlatness* group

AudioSpectrumFlatness feature set consisted of $flat_{10}, \dots, flat_{25}$ parameters. The vector of means for these parameters, similarly to other ones, significantly differed between the studied instruments ($F = 94.00$, $p < 0.01$).

All p -values for univariate variance test results were less than 0.01 with the following values of $F(13, 5,866)$: $flat_{10} - 437.101$, $flat_{11} - 386.281$, $flat_{12} - 351.645$, $flat_{13} - 273.743$, $flat_{14} - 255.16$, $flat_{15} - 223.367$, $flat_{16} - 194.828$, $flat_{17} - 138.04$,

Instrument	tris1	1 2 3 4 5 6 7 8
ctrumpet	0.166245	x
oboe	0.188429	x
englishhorn	0.259002	x x
tenorTrombone	0.307373	x x
bflatclarinet	0.327424	x x
piano	0.404295	x x
tubularbells	0.448198	x x
violin	0.468752	x
frenchhorn	0.477348	x
flute	0.537660	x
viola	0.575315	x
cello	0.616396	x
marimba	0.688202	x
vibraphone	0.727886	x

Instrument	tris2	1 2 3
oboe	40142	x
englishhorn	68672	x
flute	74392	x
violin	226792	x
ctrumpet	298678	x
viola	400077	x
cello	579224	x
bflatclarinet	663177	x
tenorTrombone	1414887	x
frenchhorn	1841127	x
marimba	4544664	x
tubularbells	11665139	x x
vibraphone	19891828	x
piano	76332434	x

Instrument	tris3	1 2 3 4 5 6 7 8 9
marimba	0.032259	x
piano	0.045117	x
flute	0.048801	x x
viola	0.076112	x x
vibraphone	0.080386	x
frenchhorn	0.084755	x
cello	0.096423	x x
tubularbells	0.116387	x x
violin	0.137848	x
tenorTrombone	0.171827	x
englishhorn	0.325869	x
oboe	0.343625	x x
bflatclarinet	0.368974	x x
ctrumpet	0.382305	x

Instrument	tris4	1 2 3 4 5 6 7 8
frenchhorn	0.011492	x
piano	0.011960	x
vibraphone	0.013794	x x
marimba	0.016000	x x
flute	0.019788	x x
tenorTrombone	0.025139	x x
englishhorn	0.029442	x x
tubularbells	0.043603	x
cello	0.046289	x
oboe	0.081428	x
viola	0.104379	x
violin	0.125048	x
ctrumpet	0.150252	x
bflatclarinet	0.196621	x

Instrument	tris5	1 2 3 4
frenchhorn	0.002759	x
piano	0.005540	x
flute	0.005567	x
englishhorn	0.006739	x
tenorTrombone	0.006820	x
marimba	0.007010	x
vibraphone	0.012816	x
oboe	0.014900	x
violin	0.023790	x
cello	0.023990	x
bflatclarinet	0.024949	x
viola	0.029246	x x
tubularbells	0.031247	x
ctrumpet	0.033919	x

Instrument	tris6	1 2 3 4 5
frenchhorn	0.021557	x
englishhorn	0.022592	x
tenorTrombone	0.037626	x x
flute	0.041990	x x
piano	0.042921	x x
oboe	0.046079	x x
vibraphone	0.054340	x x
bflatclarinet	0.055350	x x
ctrumpet	0.059698	x x x
viola	0.076991	x x x
marimba	0.081160	x x
violin	0.084221	x x
cello	0.086773	x
tubularbells	0.095496	x

Instrument	tris7	1 2 3 4 5 6
frenchhorn	1.730042	x
englishhorn	2.493493	x x
piano	2.602050	x x x
flute	3.025873	x x x
tenorTrombone	3.099660	x x x x
vibraphone	3.522109	x x x
viola	3.613917	x x x
cello	3.957027	x x x x
oboe	3.967683	x x x x
violin	4.039843	x x x
bflatclarinet	4.198013	x x
ctrumpet	4.462286	x x x
marimba	5.339934	x x
tubularbells	5.941845	x

Instrument	tris8	1 2 3 4 5 6 7 8 9
marimba	0.094389	x
piano	0.121932	x x
vibraphone	0.129555	x
bflatclarinet	0.144759	x
cello	0.240345	x
viola	0.258904	x x
frenchhorn	0.273323	x x
violin	0.281908	x
tubularbells	0.370951	x
flute	0.392088	x x
ctrumpet	0.408881	x
tenorTrombone	0.449997	x
englishhorn	0.510827	x
oboe	0.549077	x

Instrument	tris9	1 2 3 4 5 6 7 8 9
vibraphone	0.055816	x
piano	0.064567	x x
flute	0.069978	x x
marimba	0.077650	x x
frenchhorn	0.087050	x x
tubularbells	0.111881	x
cello	0.142859	x
viola	0.165404	x x
tenorTrombone	0.174318	x
englishhorn	0.226685	x
violin	0.242791	x x
oboe	0.262232	x
ctrumpet	0.386531	x
bflatclarinet	0.524384	x

Fig. 8 Homogenous groups of instruments for the *tris* parameters

$flat_{18} - 106.26$, $flat_{19} - 90.84$, $flat_{20} - 62.56$, $flat_{21} - 54.47$, $flat_{22} - 62.54$, $flat_{23} - 60.28$, $flat_{24} - 70.61$, $flat_{25} - 58.4$. The plots of means for these parameters are presented in Fig. 10. As we can see, the higher number of the element of this feature vector (the higher the frequency), the higher mean values of flatness parameters are obtained.

For the first four plots we can notice that most of values are at the similar level except marimba and vibraphone, which means are high compare to the others instruments. Than values for other instruments are getting higher and higher except clarinet, viola, oboe, English horn, cello or trumpet, which means are changing to a lesser degree. We can observe these changes also as the results of post hoc comparisons. They show the high discriminating power of $flat_{10}, \dots, flat_{14}$, distinguishing marimba, vibraphone, French horn (these instruments constitute separate,

Fig. 9 Means for the *tris* group

Fig. 10 Means for the *AudioSpectrumFlatness* group

Fig. 10 (continued)

Instrument	flat10	12345	Instrument	flat11	123456789	Instrument	flat12	1234567
oboe	0.014943	x	bflatclarinet	0.020055	x	bflatclarinet	0.024695	x
bflatclarinet	0.015010	x	ctrumpet	0.020987	x	ctrumpet	0.031724	x
ctrumpet	0.018232	x	oboe	0.024539	x	violin	0.040698	x
violin	0.038493	x x	violin	0.035421	x x	cello	0.040773	x
englishhorn	0.044830	x x	cello	0.046077	x x x	viola	0.041368	x
viola	0.050283	x	viola	0.048797	x x x	oboe	0.056885	x x
tubularbells	0.052140	x	flute	0.064166	x x x	flute	0.090423	x
cello	0.055557	x	englishhorn	0.072102	x x x	englishhorn	0.090802	x
flute	0.063592	x	tenorTrombone	0.094130	x x	tenorTrombone	0.137455	x
tenorTrombone	0.065250	x	tubularbells	0.099075	x	tubularbells	0.141556	x x
frenchhorn	0.163410	x	piano	0.174595	x	piano	0.176451	x
piano	0.174206	x	frenchhorn	0.207680	x	frenchhorn	0.272002	x
vibraphone	0.335266	x	vibraphone	0.348523	x	vibraphone	0.386760	x
marimba	0.467506	x	marimba	0.475329	x	marimba	0.483492	x

Instrument	flat13	12345678	Instrument	flat14	123456	Instrument	flat15	123456
violin	0.046914	x	viola	0.062364	x	bflatclarinet	0.066725	x
viola	0.049184	x	ctrumpet	0.063099	x	viola	0.080146	x
ctrumpet	0.052374	x	bflatclarinet	0.065469	x	ctrumpet	0.085848	x
bflatclarinet	0.058889	x	violin	0.077652	x x	oboe	0.105440	x
cello	0.067166	x x	oboe	0.086109	x x	violin	0.112405	x
oboe	0.112424	x x	cello	0.120477	x	englishhorn	0.170888	x
englishhorn	0.141665	x x	englishhorn	0.122819	x	cello	0.191760	x
flute	0.186744	x x	tubularbells	0.250101	x	piano	0.279251	x
tubularbells	0.188017	x	flute	0.260961	x	tubularbells	0.347861	x
piano	0.219529	x	piano	0.278654	x	tenorTrombone	0.359064	x
tenorTrombone	0.222159	x	tenorTrombone	0.285143	x	flute	0.362818	x
frenchhorn	0.387380	x	frenchhorn	0.422654	x	frenchhorn	0.436890	x
vibraphone	0.447296	x	vibraphone	0.475885	x	vibraphone	0.514401	x
marimba	0.549557	x	marimba	0.545634	x	marimba	0.517081	x

Instrument	flat16	12345678	Instrument	flat17	1234567	Instrument	flat18	123456
bflatclarinet	0.091968	x	bflatclarinet	0.181683	x	bflatclarinet	0.237189	x
viola	0.105800	x x	viola	0.212196	x x	viola	0.262198	x x
violin	0.143801	x x x	violin	0.239518	x x x	ctrumpet	0.315650	x x
ctrumpet	0.154708	x x	englishhorn	0.266345	x x x	cello	0.330315	x
englishhorn	0.170394	x x	ctrumpet	0.271978	x x x	englishhorn	0.346623	x
oboe	0.173960	x x	cello	0.295108	x x	oboe	0.353775	x
cello	0.212787	x	oboe	0.326106	x	violin	0.374986	x
piano	0.332213	x	piano	0.480787	x	piano	0.522027	x
tubularbells	0.409607	x	tenorTrombone	0.513049	x x	tenorTrombone	0.540645	x x
flute	0.412874	x	tubularbells	0.520715	x x	tubularbells	0.548416	x x
tenorTrombone	0.420878	x x	flute	0.558908	x x	flute	0.595037	x x
frenchhorn	0.473771	x x	frenchhorn	0.568925	x x	frenchhorn	0.604290	x x
marimba	0.519029	x	vibraphone	0.589326	x	vibraphone	0.614970	x
vibraphone	0.525915	x	marimba	0.599523	x	marimba	0.630014	x

Instrument	flat19	123456	Instrument	flat20	1234567	Instrument	flat21	123456
bflatclarinet	0.291931	x	viola	0.401988	x	cello	0.538826	x
viola	0.307426	x x	cello	0.404817	x x	viola	0.538837	x
cello	0.363246	x x	ctrumpet	0.438715	x x x	englishhorn	0.561118	x x
ctrumpet	0.372206	x x	englishhorn	0.461322	x x x x	ctrumpet	0.582784	x x x
englishhorn	0.407474	x x	violin	0.467159	x x x	violin	0.606290	x x
violin	0.409841	x x	bflatclarinet	0.482396	x x	oboe	0.625843	x x
oboe	0.445371	x	oboe	0.522661	x x	bflatclarinet	0.677625	x
tenorTrombone	0.570126	x	tenorTrombone	0.582274	x x	tenorTrombone	0.681152	x x
piano	0.586062	x x	frenchhorn	0.643193	x x	flute	0.736068	x x
frenchhorn	0.617847	x x	piano	0.644055	x x	frenchhorn	0.737435	x x
flute	0.621545	x x	marimba	0.662722	x	marimba	0.745671	x
tubularbells	0.622988	x x	tubularbells	0.664061	x	vibraphone	0.757439	x
vibraphone	0.633287	x x	flute	0.665833	x	tubularbells	0.763404	x
marimba	0.638555	x	vibraphone	0.672216	x	piano	0.775181	x

Fig. 11 Homogenous groups of instruments for *AudioSpectrumFlatness* parameters

Instrument	flat22	1 2 3 4 5 6	Instrument	flat23	1 2 3 4 5 6 7 8	Instrument	flat24	1 2 3 4 5
cello	0.546670	x	englishhorn	0.582572	x	englishhorn	0.616817	x
englishhorn	0.568780	x	cello	0.610367	x x	oboe	0.630461	x
viola	0.584822	x	oboe	0.635661	x x	cello	0.643463	x
oboe	0.665451	x	viola	0.650052	x x	ctrumpet	0.693591	x
ctrumpet	0.666899	x	flute	0.683919	x x	viola	0.721930	x
violin	0.702307	x x	ctrumpet	0.703451	x x	tenorTrombone	0.775755	x
tenorTrombone	0.711981	x x x	tenorTrombone	0.749812	x x	bflatclarinet	0.791207	x x
bflatclarinet	0.749391	x x x	violin	0.752233	x x	violin	0.791554	x x
frenchhorn	0.760477	x x x	bflatclarinet	0.775778	x x	frenchhorn	0.810691	x x x
marimba	0.767954	x x	frenchhorn	0.793408	x x x	tubularbells	0.811975	x x x
flute	0.779830	x x	marimba	0.796483	x x x	marimba	0.819187	x x x
vibraphone	0.780151	x x	tubularbells	0.799299	x x x	flute	0.828643	x x
tubularbells	0.785106	x x	vibraphone	0.809932	x x	vibraphone	0.835247	x x
piano	0.802569	x	piano	0.826572	x	piano	0.847027	x

Instrument	flat25	1 2 3 4 5 6 7
ctrumpet	0.768871	x
oboe	0.825799	x
violin	0.847598	x x
viola	0.852023	x x
cello	0.854737	x
englishhorn	0.861977	x x
bflatclarinet	0.887746	x x
tubularbells	0.895093	x x
tenorTrombone	0.898937	x x
frenchhorn	0.903252	x x x
marimba	0.906131	x x x
flute	0.915134	x x
vibraphone	0.918556	x x
piano	0.926591	x

Fig. 11 (continued)

1-element groups), and, to a lesser degree, piano (Fig. 11). For these features ($flat_{10}, \dots, flat_{14}$), some 1-element groups are produced, and for the next features from the *AudioSpectrumFlatness* set, the size of the homogenous groups is growing.

To be more precise, for increasing i in $flat_i$, the group consisting of marimba, vibraphone, and French horn was growing—other instruments were added. At the same time, homogeneous group determined by oboe, clarinet, trumpet, violin, and English horn was differentiating into separate groups.

AudioSpectrumFlatness is the biggest feature set analyzed here. High discriminative power of spectral flatness is confirmed by the results shown in Fig. 11, since in many cases 1-element groups are created wrt. particular elements of the flatness feature vector. This illustrates high descriptive power of the shape of the spectrum, represented here by the spectrum flatness.

5 Summary and conclusions

In this paper, we compared feature sets used for musical instrument sound classification. Mean values for data representing given instruments and statistical tests for these data were presented and discussed. Also, for each feature, homogeneous groups were found, representing instruments which are similar with respect to this

feature. Instruments, for which the mean values of a given feature were significantly different, were assigned to different groups, and instruments, for which the mean values were not statistically different, were assigned to the same group.

Sound features were grouped according to the parameterization method, including MFCC, proportions of harmonics in the sound spectrum, and MPEG-7 based parameters (*AudioSpectrumFlatness*, *AudioSpectrumBasis*). These groups were chosen as a conclusion of our previous research, indicating high discriminant power of particular features for instrument discrimination purposes.

Piano, vibraphone, marimba, cello, English horn, French horn, and trombone turned out to be the most discernible instruments. It is very encouraging, because marimba and vibraphone represent idiophones (a part of percussion group), so sound is played by striking and is not sustained (similarly for piano), so there is no steady state, thus making parameterization more challenging. Also, since the investigations were performed for small groups of features (up to 16), we conclude that these groups constitute a good basis for instrument discernment.

The results enabled us to indicate, for each instrument, which parameters within a given group represent the highest distinguishing power, and indicate which features are most suitable to distinguish this instrument. Following the earlier research based on sound features described here and SVM classifiers (Wieczorkowska et al. 2008), experiments on automatic musical instrument identification were also performed using random forests as classifiers (Kursa et al. 2009). The obtained results confirmed significance of particular features, and yielded very good accuracy.

Acknowledgements The presented work was partially supported by the Research Center of PJIIT, supported by the Polish National Committee for Scientific Research (KBN).

The authors would like to thank Elżbieta Kubera from the University of Life Sciences in Lublin for help with preparing the initial data for experiments and improving the description of features.

Open Access This article is distributed under the terms of the Creative Commons Attribution Noncommercial License which permits any noncommercial use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

References

- Adobe Systems Incorporated: Adobe Audition 1.0 (2003).
- Bartlett, H., Simonite, V., Westcott, E., & Taylor, H. R. (2000). A comparison of the nursing competence of graduates and diplomates from UK nursing programmes. *Journal of Clinical Nursing*, 9, 369–381.
- Box, G. E. P. (1954a). Some theorems on quadratic forms applied in the study of analysis of variance problems, I. Effect of inequality of variance in the one-way classification. *The Annals of Mathematical Statistics*, 25(2), 290–302.
- Box, G. E. P. (1954b). Some theorems on quadratic forms applied in the study of analysis of variance problems, II. Effects of inequality of variance and of correlation between errors in the two-way classification. *The Annals of Mathematical Statistics*, 25(3), 484–498.
- Box, G. E. P., & Andersen, S. L. (1955). Permutation theory in the derivation of robust criteria and the study of departures from assumption. *Journal of the Royal Statistical Society, Series B (Methodological)*, 17(1), 1–34.
- Dziubinski, M., Dalka, P., & Kostek, B. (2005). Estimation of musical sound separation algorithm effectiveness employing neural networks. *Journal of Intelligent Information Systems*, 24(2–3), 133–157.
- Finn, J. D. (1974). *A general model for multivariate analysis*. New York: Holt, Rinehart and Winston.

- Glass, G. V., & Hopkins, K. D. (1995). *Statistical methods in education and psychology*. Allyn & Bacon.
- Herrera, P., Amatriain, X., Batlle, E., & Serra, X. (2000). Towards instrument segmentation for music content description: A critical review of instrument classification techniques. In *International symposium on music information retrieval ISMIR*.
- ISO/IEC JTC1/SC29/WG11 (2004). *MPEG-7 overview*. Available at <http://www.chiariglione.org/mpeg/standards/mpeg-7/mpeg-7.htm>.
- Itoyama, K., Goto, M., Komatani, K., Ogata, T., & Okuno, H. G. (2008). Instrument equalizer for query-by-example retrieval: Improving sound source separation based on integrated harmonic and inharmonic models. In *9th international conference on music information retrieval ISMIR*.
- Klapuri, A. (2004). *Signal processing methods for the automatic transcription of music*. Ph.D. thesis, Tampere University of Technology, Finland.
- Kursa, M., Rudnicki, W., Wieczorkowska, A., Kubera, E., & Kubik-Komar, A. (2009). Musical instruments in random forest. In J. Rauch, Z. W. Ras, P. Berka, & T. Elomaa (Eds.), *Foundations of intelligent systems, 18th international symposium, ISMIS 2009, Prague, Czech Republic, 14–17 September 2009, Proceedings. LNAI 5722* (pp. 281–290). Berlin Heidelberg: Springer-Verlag.
- Lindman, H. R. (1974). *Analysis of variance in complex experimental designs*. San Francisco: W. H. Freeman & Co.
- Little, D., & Pardo, B. (2008). Learning musical instruments from mixtures of audio with weak labels. In *9th international conference on music information retrieval ISMIR*.
- Logan, B. (2000). Mel frequency cepstral coefficients for music modeling. In *International symposium on music information retrieval MUSIC IR*.
- Morrison, D. F. (1990). *Multivariate statistical methods* (3rd ed.). New York: McGraw Hill.
- NSF (2010). *Automatic indexing of audio with timbre information for musical instruments of definite pitch*. <http://www.mir.uncc.edu/>.
- Opolko, F., & Wapnick, J. (1987). *MUMS—McGill University master samples*. CD's.
- Rao, C. R. (1951). An asymptotic expansion of the distribution of Wilks' criterion. *Bulletin of the International Statistical Institute*, 33, 177–181.
- StatSoft, Inc. (2001). *STATISTICA, version 6*. <http://www.statsoft.com/>.
- Tukey, J. W. (1993). The problem of multiple comparisons. Multiple comparisons: 1948–1983. In H. I. Braun (Ed.), *The collected works of John W. Tukey* (vol. VIII, pp. 1–300). Chapman Hall.
- Viste, H., & Evangelista, G. (2003). Separation of harmonic instruments with overlapping partials in multi-channel mixtures. In *IEEE workshop on applications of signal processing to audio and acoustics WASPAA-03, New Paltz, NY*.
- Wieczorkowska, A., & Czyzewski, A. (2003). Rough set based automatic classification of musical instrument sounds. In *International workshop on rough sets in knowledge discovery and soft computing RSKD*. Warsaw, Poland: Elsevier.
- Wieczorkowska, A. A., & Kubera, E. (2009). Identification of a dominating instrument in polytimbral same-pitch mixes using SVM classifiers with non-linear kernel. *Journal of Intelligent Information Systems*. doi:10.1007/s10844-009-0098-3.
- Wieczorkowska, A., Kubera, E., & Kubik-Komar, A. (2008). Analysis of recognition of a musical instrument in sound mixes using support vector machines. In H. S. Nguyen & V.-N. Huynh (Eds.), *SCKT-08: Soft computing for knowledge technology workshop, Hanoi, Vietnam, December 2008, proceedings. Tenth pacific rim international conference on artificial intelligence PRICAI 2008* (pp. 110–121).
- Wieczorkowska, A., & Kubik-Komar, A. (2009a). Application of analysis of variance to assessment of influence of sound feature groups on discrimination between musical instruments. In: J. Rauch, Z. W. Ras, P. Berka, & T. Elomaa (Eds.), *Foundations of intelligent systems, 18th international symposium, ISMIS 2009, Prague, Czech Republic, proceedings. LNAI 5722* (pp. 291–300). Berlin Heidelberg: Springer-Verlag.
- Wieczorkowska, A., & Kubik-Komar, A. (2009b). Application of discriminant analysis to distinction of musical instruments on the basis of selected sound parameters. In: K. A. Cyran, S. Kozielski, J. F. Peters, U. Stanczyk, & A. Wakulicz-Deja (Eds.), *Man-machine interactions. Advances in intelligent and soft computing* (Vol. 59, pp. 407–416). Berlin Heidelberg: Springer-Verlag.
- Winer, B. J., Brown, D. R., & Michels, K. M. (1991). *Statistical principals in experimental design* (3rd ed.). New York: McGraw-Hill.
- Zhang, X. (2007). *Cooperative music retrieval based on automatic indexing of music by instruments and their types*. Ph.D thesis, Univ. North Carolina, Charlotte.