Music Recommendation System

Piotr Hoffmann[1], Andrzej Kaczmarek[2], Paweł Spaleniak[1], and Bożena Kostek[1]

[1] *Faculty of Electronics, Telecommunications and Informatics, Gdańsk University of Technology, Gdańsk, Poland*
[2] *Multimedia Systems Department, Gdańsk University of Technology, Gdańsk, Poland*

Abstract—The paper focuses on optimization vector content feature for the music recommendation system. For the purpose of experiments a database is created consisting of excerpts of music files. They are assigned to 22 classes corresponding to different music genres. Various feature vectors based on low-level signal descriptors are tested and then optimized using correlation analysis and Principal Component Analysis (PCA). Results of the experiments are shown for the variety of feature vectors. Also, a music recommendation system is presented along with its main user interfaces.

Keywords—feature vectors, music classification, music information retrieval, music parameterization, Principal Component Analysis.

1. Introduction

There are a few approaches to music classification within the Music Information Retrieval (MIR) area. The present prevailing MIR research study paradigm is based on low-level parametrization. For that purpose the MPEG 7 standard, Mel-frequency Cepstral Coefficients (MFCCs) or, finally, parameters suggested by researchers in [1]–[11] are commonly used. This concerns also music services which employ their own solutions for music parametrization [12]. The second approach, especially in many on-line services, i.e., social music networking systems in record labels as well as in record companies utilizes metadata such as for example: music genres, album name, date of the album release, names of artists, length of a particular song, lyrics, etc., to query by artists or music genres. The most important research in the MIR domain is related to the content-based retrieval. In particular, Query-by-Category utilizes musical style, genre, mood/emotion of a musical piece in music retrieval. Representation of such are social networking systems or music services as for example: iTunes [13], Amazon [14], Lastfm [15], Pandora [12].

However, even though music services deal effectively with music search, there is no common parametrization and classification system underlying these services. Moreover, in research, there exists a problem with comparing various algorithms and the low-level parameters effectiveness in this field since authors/services use different music databases, different taxonomy and music excerpts of different lengths, etc.

The paper first discusses the experiments and the effectiveness analysis of music genre classification. For the purpose of this research a music service was created with a database of more than 50,000 songs divided into 22 gen-

res. Music genre classification is based on low-level feature vectors, optimization using correlation analysis and Principal Component Analysis (PCA) and, on decision algorithms. The tests were performed with an external application that uses the algorithms implemented in the music recommendation service. Lastly, the music database along with its user interfaces which was made for the music service is described.

2. Experiments

2.1. Low-level Descriptors

The first step in experiments consisted in searching for low-level descriptors. Some of the previously by the authors used parameters [8], [15], [16] were reviewed, thus system original version includes the following parameters: 127 descriptors of the MPEG-7 standard: Audio Spectrum Centroid (ASC), Audio Spectrum Spread (ASS), Spectral Flatness Measure (SFM), Audio Spectrum Envelope (ASE), Mel-Frequency Cepstral Coefficients (MFCC) (40 descriptors: mean values and variances calculated from the 20 MFCC values) and 24 dedicated parameters: zero threshold-crossing rates, i.e., beat histogram. Parameter vector includes 191 descriptors. It should also be noted that all parameters are normalized for range (–1, +1) [8], [15], [16].

MFCC parameters utilize linear and logarithmic frequency scales that perceptually approximates frequency scales. It uses the range division from 0 to approx. 6700 Hz into 40 sub-bands with the first 13 being linear sub-bands with an equal bandwidth of 66.6667 Hz, and the remaining 27 being the logarithmic part of the scale with bandwidth ascending by 1.0711703 and increasing center frequencies (typically approx. linear $f < 1$ kHz; logarithmic above that). Therefore the number of the MFCC coefficients is 40: 20 PAR_MFCC parameters are calculated as arithmetic means, while 20 PAR_MFCCV parameters are calculated as variances of these coefficients calculated in all the segments.

2.2. Correlation Analysis

A large feature vector has some advantages, but also many disadvantages, such as for example the parameter separability problem. A variety of descriptors may allow for an easier differentiating between classifies genres. However an important aspect of the parametrization effectiveness analysis is reducing the feature vector redundancy. That's why

the analysis starting point was the examination of parameter vector separability. One of techniques used for this purpose is correlation analysis, which involves calculating the covariance matrix, and then the correlation matrix. The last step is interpreting the individual coefficients based on t-Student's statistics. With this one can determine which parameters can be considered redundant. By denoting the correlation coefficient calculated for $x_1, x_2, \ldots, x_n$ of parameter x and for $y_1, y_2, \ldots, y_n$ of parameter y with R_{xy}, it is possible calculate the statistics having the decomposition of t-Student with n-2 degree of freedom, using the equation:

$$t = \frac{R_{xy}}{\sqrt{1 - R_{xy}^2}} \cdot \sqrt{n - 2}, \qquad (1)$$

where: n – the number of parameter vectors.

A smaller vector, resulting from the correlation analysis was checked on the ISMIS database serving for the ISMIS 2010 conference contest. The contest results for classification of music genres were encouraging, with effectiveness approx. 80–90% [6]. The next step was to test the effectiveness of this vector, containing a number of features reduced to 173, because of the lower sampling frequency of 22050 Hz for the Synat database. In the first tests, the effectiveness was low. However, when the authors listened to random song excerpts in the Synat database, it turned out that the songs collected by the music robot is not uniform, which means that the songs are often not properly assigned to the appropriate genre. It was therefore decided to optimize database parametrization. To this end, a database of 11 of the most distinctive/characteristic music genres (Blues, Classical, Country, Dance&DJ, Jazz, Hard Rock & Metal, NewAge, Pop, Rap & Hip-Hop, R&B, and Rock) was prepared with three new parameter vectors proposed for them.
Parametrization included the following steps:

- downloading an excerpt of a recording with a duration of approximately 26 s,

- conversion to a 22050 Hz monophonic signal (right, left channel summing),

- segmentation into 8192 samples, i.e., 2 to the power of 13, due to the need for the FFT algorithm for spectral analysis,

- calculating Fourier spectra with a sample rate of 4410 samples (time equal to 0.2 s), and using a Blackman window, hence the overlap size equal to 3782 samples,

- calculating the power spectrum using a logarithmic scale for each segment.

  Due to the sampling frequency of 22050 Hz, the frequency resolution for the Fourier analysis is 2.692 Hz and the analysis band reaches 9986.53 Hz. The entire available frequency band is divided into sub-bands of increasing width directly proportional to

the centre frequencies. The first sub-band has a centre frequency of 100 Hz. The nonlinearity coefficient 1.194, borrowed from the König scale that defines the ratio of the width of the subsequent sub-bands in the entire analyzed frequency range was used.

- calculating the spectrograms in the above-described scale followed by cepstrograms by calculating cepstral coefficients Ci according to the discrete cosine transform [17]:

$$C_i = \sum_{j=1}^{N} \log E_j \cos\left(\frac{\pi i}{N}(j - 0.5)\right), \qquad (2)$$

where: i – number of a cepstral coefficient, E_j – energy of the j-th sub-band of the filter bank, N – number of (filter) channels in the filter bank.

- calculating statistical moments to 3, inclusive of individual cepstral coefficients representing sub-vectors being parts of the full 2048-long parameter vector.

Mean value m_i, i.e. the moment of the first order for the i-th cepstral coefficient is calculated with the Eq. (3) [18]:

$$m_i = \frac{1}{K} \sum_{k=0}^{K} C_i(k), \qquad (3)$$

where: K – number of segments, $C_i(k)$ – value of the i-th cepstral coefficient for segment k.

Variance and skewness, i.e. moments of the second and third order for the ith cepstral coefficient are calculated with the Eq. (4) [18]:

$$M_i(n) = \frac{1}{K} \sum_{k=0}^{K} \left[C_i(k) - m_i\right]^n, \qquad (4)$$

where: i – number of a cepstral coefficient, n – order of the moment (2nd or 3rd).

For cepstral parameters, each sub-vector has a length equal to the number of segments. However, their number is equal to the number of cepstral coefficients (cepstrum order) = 16. The vector length is 2048 which consists of 16 sub-vectors with 128 values each. The resulting cepstrogram can be converted into a shorter vector in several ways, such as by delta, trajectory or statistical transformation. The pilot studies showed that the transformation based on the statistical analysis proved to be the most effective. It should be noted that the full vector with a length of 2048 was also tested, but proved to be ineffective in the classification process. Also, based on the cepstrograms, the statistical parameters of individual cepstral coefficients was determined: statistical moments. The authors used three statistical parameters: average value (arithmetic mean), variance and skewness. This number is the optimization result. In this way, each sub-vector is described by three numbers, resulting in a major data reduction, that is, a shortening of

parameter vectors. Finally, parameter vectors have a length of $48 = 3 \times 16$.

Another tested feature vector was based on fuzzy logic, taking into account to what degree a particular song belongs to particular genres. In this case, each object in a database is described by x numbers, which are the distances from the centroids of all x classes. This is allows for further data reduction. Approximation of membership functions is based on the histograms of the objects distances from the intrinsic class centroid, according to the Eq. (5):

$$Fp(x) = 1 \quad \text{for} \quad x < g$$
$$Fp(x) = \exp\left[-d(x-g)\right] \quad \text{for} \quad x > g, \qquad (5)$$

where: g is limit value that can be interpreted as the distance value of overall (almost 100%) objects in a particular class from the centroid of this class, d – coefficient dependent (inversely proportional) on standard these distances deviation.

In this way, one can obtain the intrinsic membership function for each class. The membership function value can be interpreted as the quotient of the number of objects more distant from a given argument by the all objects of a given class number.

2.3. Testing Feature Vector Effectiveness

Experiments were performed with some available methods to improve the classification efficiency in the context of recognizing musical genres. That's why the part of a song influence to be analyzed was checked by employing various parameter k values in the kNN algorithm, on the other at the later stage hand the Principal Component Analysis was used to reduce the parameters redundancy.

The tests were performed with an external application that uses decision algorithms implemented in the Synat service:

– fuzzy logic,

– kNN, a minimum-distance classifier, where k value is the items number included in the decision-making process, thus using an 11-element vector parameter (kNN11) and a 48-element vector parameter (kNN48).

For the experiments, two database were used, namely: a 1100-song database (songs composed that are not present in the Synat database) containing 100 songs from each of the 11 selected most uniform genres, and the Synat database. Table 1 shows the songs number in the 1100 and the Synat databases for the analyzed genres.

Pilot tests were carried out using the "leave-one-out" procedure and NN classifier, following prior normalization of parameters $\langle -1, +1 \rangle$. The tests focused on the selected parameters from vector 173. The authors tested selected parameter groups (ASE, SFM, MFCC, dedicated parameters: beat histogram), as well as other individual parameters such as centroids and spectrum centroid variances.

Table 1
Number of songs in the databases: "1100" and Synat

Genre	Size of the song databases	
	1100	Synat
Pop (P)	100	5976
Rock (Ro)	100	4957
Country (Co)	100	3007
R&B (RB)	100	2907
Rap & Hip-Hop (Ra)	100	2810
Classical (Cl)	100	2638
Jazz (J)	100	2543
Dance & Dj (D)	100	2289
New Age (N)	100	2122
Blues (B)	100	1686
Hard Rock & Metal (H)	100	1175
Total	1100	32110

The tests results showed that the best classification performance is displayed by the MFCC parameters group [19], [20]. These were the mean values and variances of 20 Mel-Frequency Cepstral Coefficients calculated from 40 power spectrum sub-bands distributed over the frequency scale in a combined linear and logarithmic manner. The linear part related to the band lower part, while the logarithmic part concerned the frequency band upper part. This led us to checking how a division of the given band influences the classification mel-frequency cepstral parameters performance. Furthermore, it should be verified whether the higher statistical moments parameters such as skewness and kurtosis can be included in the resulting vector. Pilot tests confirmed that the best frequency band division is a fully logarithmic one, and the maximum statistical moment order should be 3 (skewness). Therefore, for further research the 48-element vector containing 16 frequency sub-bands is used.

Further experiments were designed to identify which of 30-second song fragments provides the most characteristic information for the entire song for effective genre recognition. Four song fragments was analyzed: intro, the first half middle (middle1), the second half middle (middle2) and the song middle (middle). Doing parametrization for the song final section was deemed irrelevant since songs are faded out coming to the end. It should be noted that this is a very important research part due to the fact that the music databases often store 30-second song fragments (the copyright aspect). For this experiments part, a database with 1100 songs is used.

Figure 1 shows the effectiveness of music genre recognition within the collection of 1100 songs for the analyzed song fragment. For comparison of results, the kNN algorithm was tested using two parameter vectors: 11 and 48. The best results were achieved for the middle of the songs first half. The maximum increase in the classification efficiency was 56% compared with the song intro. All classifiers achieve higher scores when testing further song fragments

in relation to the initial section. Using the shorter parameter vector for the kNN algorithm reduces the classification efficiency by 7 percentage points, on average. In the best case, the song classification efficiency was 45%, so in comparison with common test sets the score should be considered low [21]. For further experiments, a 30-second fragment of the first part song middle.
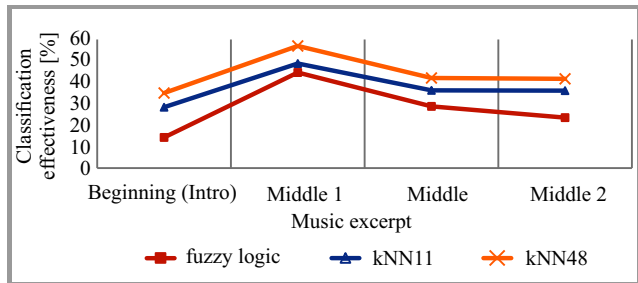

Fig. 1. Genre classification results for the song fragments.

Another aspect leading to improving the classification effectiveness is the resulting vector optimization due to the minimum-distance classifiers used. As mentioned earlier, in the first run, min-max parameters for range $\langle -1, +1 \rangle$ were normalized. This means that the parameters weights are aligned, but this assumption should be regarded only as a starting point for optimization activities.

Proper selection of parameter weights is key to maximizing the effectiveness of kNN classifier. Therefore, an optimization procedure for determining feature vector weights is developed. It involves kNN classifier multiple use on the parameters basis with simultaneous systematic changes of weight values. Weight vector at this procedure starting point is aligned, and all weights are equal to 1. The next optimization steps involve changing the individual weights values by in plus and in minus increments assumed apriori. Resulting for this, kNN classifier forms a confusion matrix, and in each optimization step it is necessary to assess whether there was an in plus or an in minus change. In the described optimization system, these criteria were applied to the value of the κ coefficient with the following equation [22]:

$$
\kappa = \frac{N \cdot \sum\limits_{i=1}^{r} x_{ii} - \sum\limits_{i=1}^{r} \left(x_{i+} \cdot x_{+i} \right)}{N^2 - \sum\limits_{i=1}^{r} \left(x_{i+} \cdot x_{+i} \right)}, \tag{6}
$$

where: N – the number of all the objects in the database, r – number of classes, x_{ii} – values on the main diagonal, x_{i+} – sum in rows (TP + FN), x_{+i} – sum in columns (TP + FP).

Acronims used: TP – correct indication of the distinguished class (TP – *true positive*), FP – incorrect indication of the distinguished class (FN – *false positive*).

The aim of the optimization is to maximize the κ value, which is a classifier quality measure. Weight values are changed up to the moment when subsequent small changes

of the weight do not cause improvement. Therefore, a series of optimization is used where weight increases and decreases become smaller in the consecutive steps. Optimization begins with changes +100% and –50%, and successively after reduction by a factor of $\frac{2}{3}$ in each step, stop at 3%. The algorithm is implemented in 10 steps.

Further experiments were divided into three parts:

- examining the classification effectiveness of selected decision algorithms using the Synat database,

- examining the effects of k parameter changes in the kNN algorithm for the 1100-song database,

- examining the classification effectiveness for 6 music genres for both Synat and 1100-song databases.

For the classifiers training phase, 70% of the collections in Table 1 was used, while for the testing phase 30%. Song fragments used in the study were taken from the first part of the song middle.

Below are the test results for the following classifiers: fuzzy logic, kNN11, kNN48. The classification results are provided in Figs. 2–9. The first one of the analyzed algorithms, fuzzy logic, draws on the membership function. During the tests, each 11 classes (music genres) received its membership function by which the elements of test sets are assigned. Classification effectiveness before and after optimization for the fuzzy logic classifier are presented in Fig. 2. Average gain resulting from the optimization is 2.5%. For most genres, the average gain ranges between 2% to 3%. The best recognition performance is obtained for Classical (64.44%), and the worst for Blues (16.61%). The biggest gain after optimization was recorded for Hard Rock & Metal, and is 5.5%.
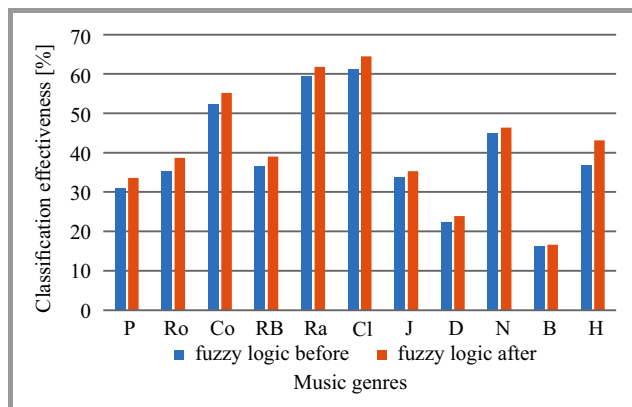

Fig. 2. Classification effectiveness for fuzzy logic based classifier before and after optimization for the Synat database, denotation as shown in Table 1.

For the tests with the kNN algorithm, $k = 15$ is used. The algorithm was tested in two parameter vector variants: 11 and 48. Figure 3 shows test results for vector 11, while Fig. 4 shows test results for vector 48. The total classification kNN11 algorithm efficiency is 38%, which makes it comparable with the fuzzy logic algorithm. Gain after

optimization is very low, at a negligible 0.5% level. For the short parameter vector, optimization gives less gain. The best recognition performance is obtained for Classical (63.68%), and the worst for Pop (14.89%) and Blues (15.03%), see Fig. 3.
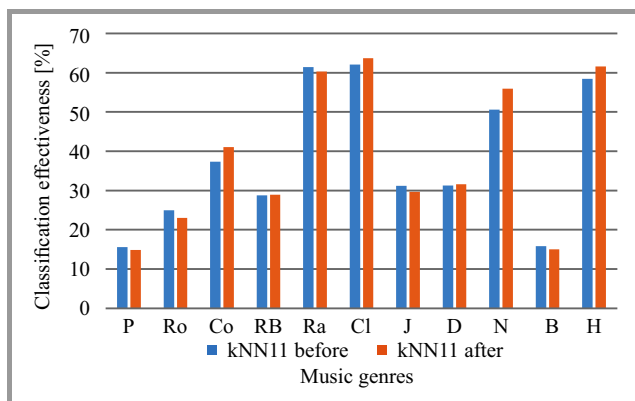

Fig. 3. Classification effectiveness for the kNN11 classifier $k=15$ before and after optimization for the Synat database, denotation as in Table 1.
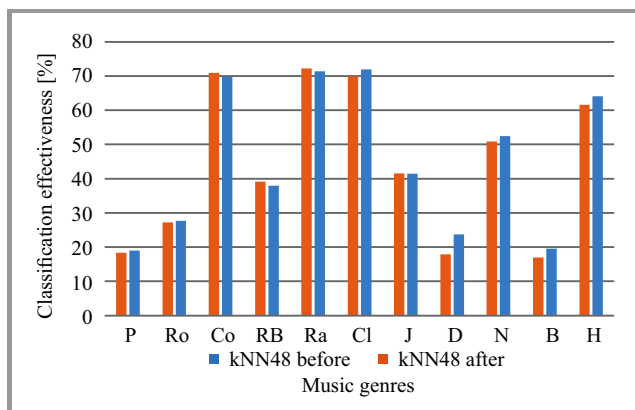

Fig. 4. Classification effectiveness for the kNN48 classifier $k=15$ before and after optimization for the Synat database, denotation as in Table 1.

The classifier variant with the longer parameter vector is the most effective decision algorithm among the ones tested. The total classification efficiency is 45%. The optimizations give percentage gain in genre recognition performance. As for kNN11 and fuzzy logic, the best results are obtained for Classical (71.9%), and the worst for Pop (19.02%) and Blues (19.57%).

The first stage of tests on the parameter vectors optimization confirmed that changing the parameter vectors weights was appropriate. The gain in relation to the main results is most visible for the best classification results.

Overall, the best of the tested algorithms is kNN48. However, its characteristic feature is a large values spread. There are genres that are recognized with the efficiency of 20% and 70% in a single test set. The genre that is recognized best of all in the Synat database is Classical. For further work on the optimization of parameter vectors, kNN48 algorithm is to be used.

Subsequent tests were designed to determine the optimal value for k parameter in kNN algorithm with the 48-element parameter vector. The tests were carried out for the 1100-song database for three k values: 4, 11, 15. Using the reduced song database allowed for comparing optimization performance for smaller collections of songs that authors listened to. Figure 5 shows classification results for this test. Using the reduced training database improved the optimization performance by 56.97%, a 7% improvement in relation to the Synat database. In relation to the Synat database, the best recognizable music genres changed. This may indicate improper assignment. During the experiments, the effects of the k parameter on the optimization performance result was examined. As the parameter value decreases, the music genre classification performance increases after parameter vector optimization. For $k=4$, the overall performance is 56.97%.
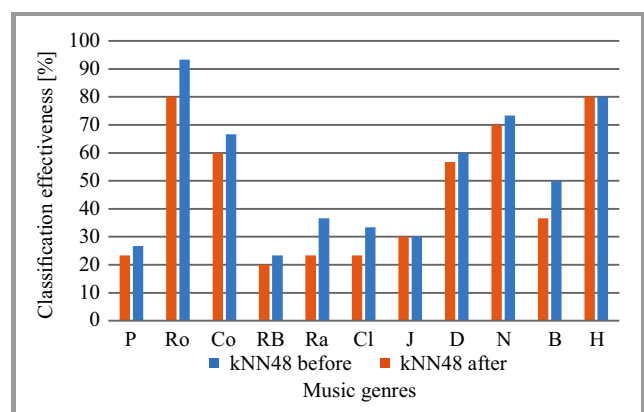

Fig. 5. Classification effectiveness for the kNN48 classifier $k=15$ before and after optimization for the 1100-song database, denotation as in Table 1.

The results show that parameter optimization weights is a way to improve the classification results. Genres recognized more effectively achieve better scores after result optimization. The Synat database contains over 32,000 songs, making it impossible to check (by listening to them) the assignment accuracy of all the songs to corresponding genres. This is reflected in the training process conducted incorrectly for classifiers in the decision algorithms. The large set size has a negative impact on the optimization of parameter vectors.

Table 2 shows a results summary for variable value of k parameter. The genre that is recognized best of all is Rock and Hard Rock & Metal. The performance over 90% is comparable to the results which were achieved for common test sets [23]. This shows that there is great potential in genre classification solutions that use minimum-distance classifiers coupled with parameter weights optimization. The optimal the k parameter value which yields the best weight optimization results is 4.

Next tests aimed at investigating the reducing effects the number of classes (music genres) on the classification effectiveness. The initial set of 22 genres had been reduced

Table 2

Percentage of recognition efficiency for music genres
in the 1100-song database using k parameter
of the kNN algorithm

Genre	$k = 15$		$k = 4$	
	Before	After	After	After
Pop	23.33	26.67	40.00	50.00
Rock	80.00	93.33	96.67	93.33
Country	60.00	66.67	70.00	66.67
R&B	20.00	23.33	36.67	50.00
Rap & Hip-Hop	23.33	36.67	40.00	53.33
Classical	23.33	33.33	40.00	40.00
Jazz	30.00	30.00	36.67	36.67
Dance & DJ	56.67	60.00	60.00	66.67
New Age	70.00	73.33	70.00	66.67
Blues	36.67	50.00	46.67	36.67
Hard Rock & Metal	80.00	80.00	73.33	66.67
Total	45.76	52.12	55.45	56.97

during the pilot tests to 11. Further tests were conducted using 6 genres. The sets size is presented in Table 1.
For further tests, the following genres was selected: Classical, Dance & DJ, Hard Rock & Metal, Jazz, Rap & Hip-Hop, and Rock. The authors compared 1100-song and Synat databases using the kNN48 classifier. The results are shown in Figs. 6 and 7.
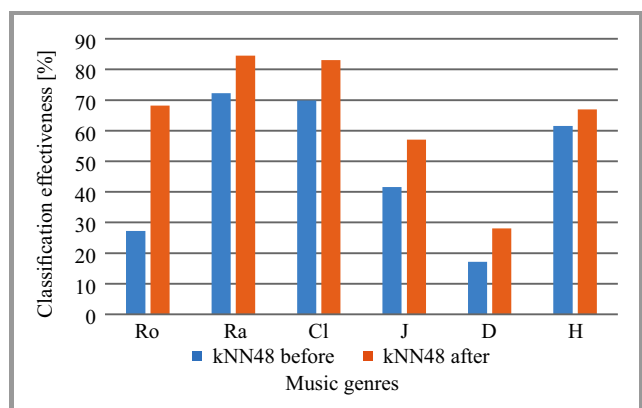

Fig. 6. Classification effectiveness for the kNN48 classifier k=15 before and after optimization for the Synat database for the reduced number of classes (genres), denotation as in Table 1.

Classification effectiveness for songs in the Synat database including 6 genres is 16% higher after optimization than for 11 genres with optimization (see Fig. 6). Rock – poor recognition performance in the Synat database (27.23%); after optimizing the classes number and parameter weights, its recognition performance amounts to 70%. The genres that were effectively recognized before changing the classes number continue to be properly classified after optimization. Similar results as for the Synat database can be observed in tests carried out for the 1100-song database. The average classification effectiveness improved in rela-
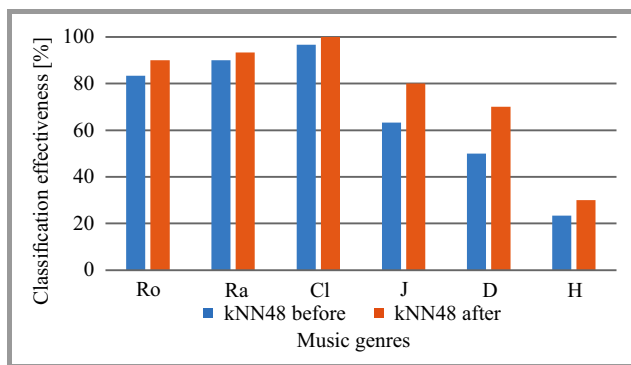

Fig. 7. Classification effectiveness for the kNN48 classifier k=15 before and after optimization for the 1100-song database with 6 genres, denotation as in Table 3.

tion to the tests without optimization by 32% and amounts to 77.22%. The overall result is similar to those obtained for common test databases [21]. After optimizations, Classical genre recognition performance is 100%. The summary comparing the resulting classification effectiveness for 6 genres is shown in Table 3.

Table 3

Percentage of recognition efficiency for music genres
in the 1100-song database using k parameter
of the kNN algorithm

Genre	Synat		1100	
	Before	After	After	After
Rock (Ro)	27.23	68.19	83.33	90.00
Rap & Hip-Hop (Ra)	72.24	84.46	90.00	93.33
Classical (Cl)	69.88	83.02	96.67	100.00
Jazz (J)	41.55	57.02	63.33	80.00
Dance & DJ (D)	17.19	28.11	50.00	70.00
Hard Rock & Metal (H)	61.56	66.95	23.33	30.00
Total	48.27	64.62	67.78	77.22

The experiments conducted confirm the need for optimization of data to be classified. In the tests, weights in parameter vectors and the number of music genres (classes) have been optimized. Another important aspect in developing automatic music genre recognition systems is to prepare a training set accurately.

2.4. PCA-based Feature Vector Optimization

The last method for increasing the efficiency of the music genre classification, which was used in the tests is the Principal Component Analysis (PCA) method, which operates on the data variances. This method reduces the data amount to be classified on the basis of their correlation. As a result, this method creates a smaller parameters number, which is a linear old data combination and the PCA values. Principal components can be determined as a linear observable variables combination. Successive components explain the decreasing amount of the total variables variance. Reduc-

ing the feature space dimension and organizing them into subsets is useful mainly because of the reducing possibility the variables number. Also this allows for a simplified interpretation of the relationship between the components and the features adopted ordering [25].
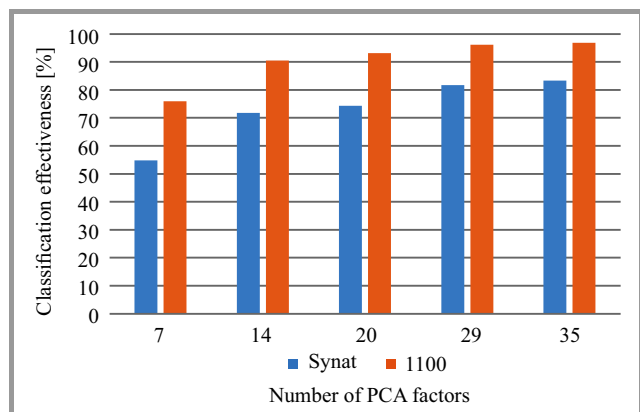

Fig. 8. Classification effectiveness for the *k*NN48 classifier *k*=15 after PCA method optimization for the Synat database for a reduced number of classes (genres).
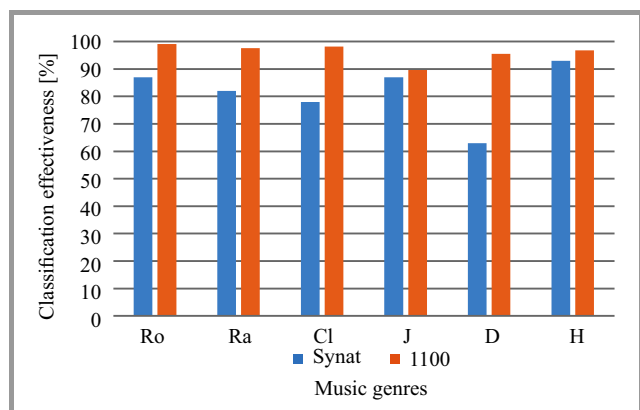

Fig. 9. Classification effectiveness for the *k*NN48 classifier *k*=15 after PCA method optimization for the 1100 database for a reduced number of classes (genres), denotation as in Table 3.

The tests were carried out as before with the most effective *k*NN algorithm use on both Synat and 1100 databases. Six music genres were used in experiments, as before. The five PCA vectors were of size: 7, 14, 20, 29 and 35. The application of the PCA significantly improved the efficiency of musical genres classification. The detailed results are presented in Figs. 8 and 9. The use of the redundancy reduction of the data makes the classification efficiency is increased by 20 pp. The effectiveness of recognition the Synat set is lower than the 1100 by 20 pp. The increase classification efficiency is consistent with the increasing number of parameters PCA (data). It should be noted that when employing 14 PCA parameters, this achieved 90% in classification for the 1100 data set.
In Figs. 10 and 11 the results divided by music genres are presented. Currently, all genres are recognized with at least 70% efficiency for the tested data set. The highest

increase in the classification effectiveness was achieved for the Dance&DJ and Hard Rock & Metal genres. This is reflected by better these classes separability after performing the linear combination reduction. In Fig. 11, the classification efficiency increase of the PCA method compared to weight parameter value regardless of the tested data set can be observed. For most genres this is clearly visible.
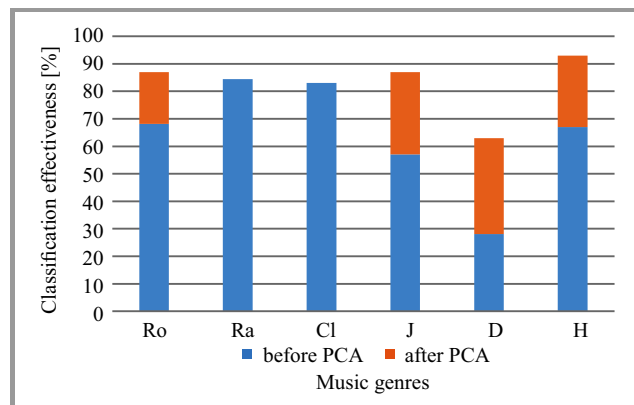

Fig. 10. Classification effectiveness rise for the *k*NN48 classifier *k* = 15 after PCA method optimization for the Synat database for the reduced number of classes (genres), denotation as in Table 3.
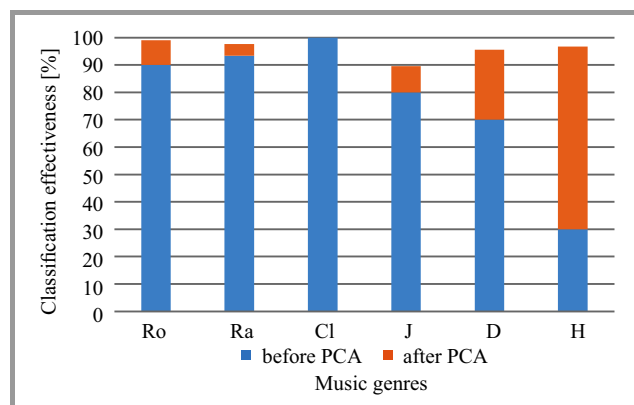

Fig. 11. Classification effectiveness rise for the *k*NN48 classifier *k* = 15 after PCA method optimization for the 1100 data set for the reduced number of classes (genres), denotation as in Table 3.

The tests performed using the PCA method show that one can effectively identify genres, on a large number database (Synat). After using the PCA method classification effectiveness for the Synat database is above 80% and for the 1100 data set above 95%. The result should be classified as very good. All of the tested music genres after the PCA performance were recognized with a very good effectiveness.

3. Music Service

3.1. Description of the System

As mentioned before, the Synat music service encompasses approximately 32,000 30-second song fragments allo-
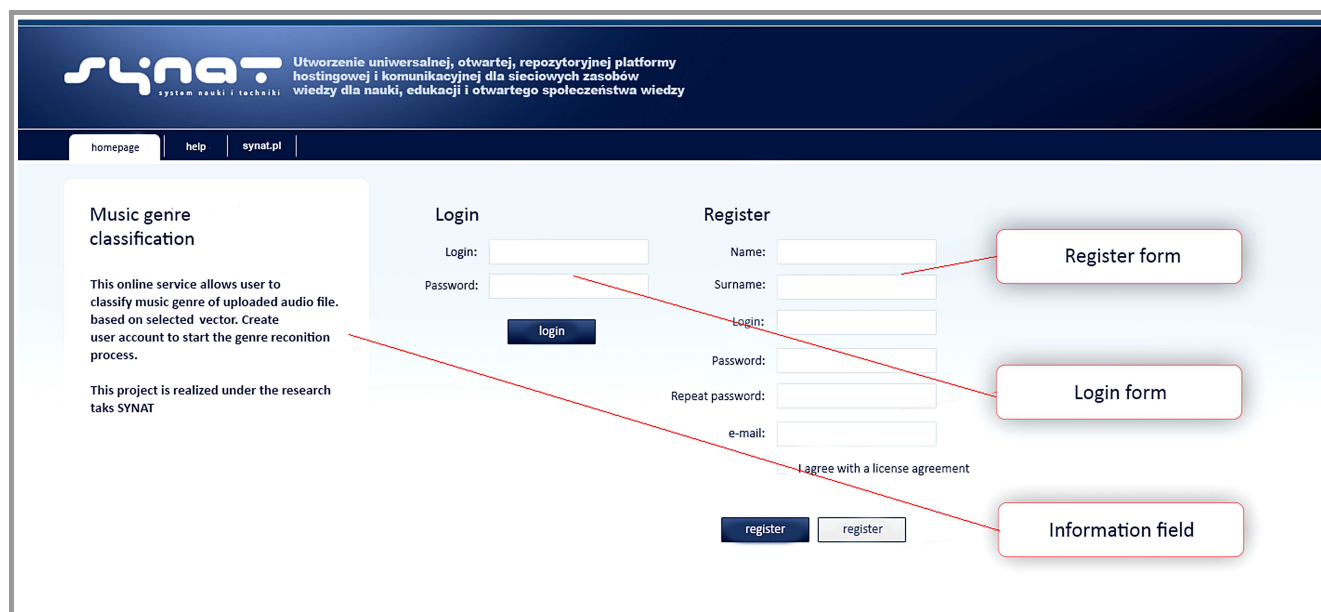
Fig. 12. Home page of the Synat database website.

cated to 22 music genres: Alternative Rock, Blues, Broadway&Vocalists, Children's Music, Christian&Gospel, Classic Rock, Classical, Country, Dance&DJ, Folk, Hard Rock&Metal, International, Jazz, Latin Music, Miscellaneous, New Age, Opera&Vocal, Pop, Rap&Hip-Hop, Rock, R&B, and Soundtracks. The database contains additional metadata, such as: artist name, album name, genre and song title. In addition to the items listed in the database, songs include also track number, recording year and other parameters typically used for annotation of recordings.

The music service utilizes a number of technologies, among which are:

– Debian 6.0 Squeeze AMD64 – for server software and basic programming libraries;

– IPtables Guarddog – protects the system against unauthorized access from the Internet;

– Apache httpd FastCGI mod_fcgid – the service is implemented as a fcgid application running in the popular Apache http server environment;

– Firebird 2.5/3.0 – database server stores information on song metadata, as well as their parameters and metrics;

– FFmpeg – converts audio files to wav format. This is required for the libsndfile library to read samples for subsequent song parametrization, etc.

3.2. User Interfaces

This section presents graphical user interfaces of the music recognition system with the associated music genres. Home page (Fig. 12), available to all users visiting the website, allows for:

– getting acquainted with the idea of the project/service,

– signing in for users with an active account,

– signing up for new users,

– going to the Synat project website.

The signed/logged-in user will be redirected to the subpage where they can select an audio file for music classification analysis (Fig. 13). One can choose a file either from their own computers or from the Synat music database. Synat service accepts all of audio files to be analyzed. After loading, the file can be played back with a simple player supporting playback. Another useful system feature is that the user can choose the feature vector that is to be used in further processing. There is information about the recommended audio file format to be uploaded. The page showing the results allows for playing the song and reading the pie chart showing the degree to which the song belongs to the given genre (Fig. 13). The page displays also a list with recommended songs. A recommended songs list contains files that are to some extent similar to the uploaded file. The similarity is determined and based on the parametric distance between files. This recommended songs list sorts in order from closest to farthest.

The list returns results according to the k value in the kNN algorithm. In addition to the track name meta data are also shown. The user can also explore and search songs within the Synat database only. The inquiry to the system can either be made by a favorite genre or a song. The user can listen to the chosen song or download the feature vector assigned to it. The Synat service includes also the extensive help site, which contains answers to frequently asked questions.
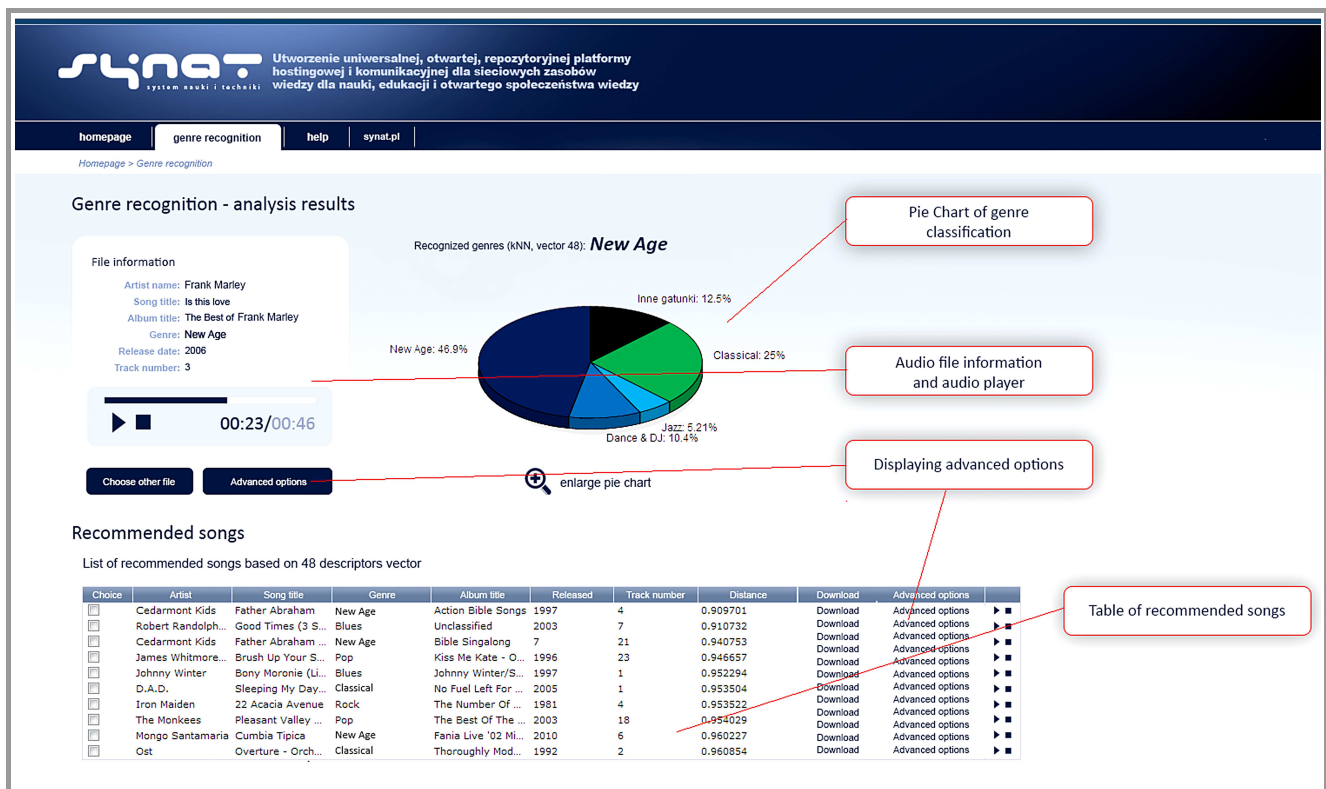
Fig. 13. Subpage with the music genre recognition and analysis result.

4. Conclusions

The main conclusion for the experiments performed is that optimizing both the feature vector and classification algorithm is essential, specifically for a larger music databases. Music genre classification solutions cover a number of systems that when used with common test sets (GZTAN [23], ISMIR [11], MIREX [10]) achieve efficiency above 80%. However, the size of these sets is approximately 1000 songs. Therefore, it is difficult to compare the proposed solutions effectiveness within the study performed by the authors since one needs to consider the analyzed music fragment length, a test set and the capacity of such a classifier to learn and improve. Moreover, in most cases, music databases contain 20- to 30-second recordings fragments. The studies have shown that the optimal fragment for testing the genre recognition effectiveness is a song first part middle.

At the current state of the research, the best results with the Synat database were achieved for the vector containing 48 parameters. The tests confirmed that optimizing input data is beneficial. Assigning weights to parameters, reducing the classes number and using the PCA method substantially increased the classification the kNN algorithm effectiveness. This resulted in approximately 90% genre classification effectiveness. PCA significantly reduced the data number and was easy to implement. In the presented solution PCA is incorporated as the music recognition system part. The current automatic music genre recognition effectiveness achieved in the experiments final phase is acceptable, and further work should aim to increase the classification effectiveness for much larger music databases.

Acknowledgements

References

[1] J.-J. Aucouturier and F. Pachet, "Representing musical genre: A state of art", *J. New Music Res.*, vol. 32, no. 1, pp. 83–93, 2003.

[2] J. Benesty, M. Mohan Sondhi, and H. Yiteng, *Springer Handbook of Speech Processing*. Berlin Heidelberg: Springer, 2008.

[3] S. Ewert, "Signal processing methods for music, synchronization, audio matching, and source separation", Ph.D. thesis, Friedrich-Wilhelms University, Bonn, Germany, 2012.

[4] P. Hoffmann, B. Kostek, A. Kaczmarek, and P. Spaleniak, "Creating a reliable music discovery and recommendation system", in *Synat Workshop SYNAT 2013*, Warsaw, Poland, 2013, Studies in Computational Intelligence, Springer (in print).

[5] P. Hoffmann, B. Kostek, A. Kaczmarek, and P. Spaleniak, "Wyszukiwarka Nagrań muzycznych – Serwis muzyczny Synat", *Telecom. Review*, no. 8–9, 2013 (in Polish).

[6] A. Holzapfel and Y. Stylianou, "Musical genre classification using nonnegative matrix factorization-based features", *IEEE Trans. Audio, Speech, Lang. Proces.*, vol. 16, no. 2, pp. 424–434, 2008.

[7] K. Hyoung-Gook, N. Moreau, and T. Sikora, *MPEG-7 Audio and Beyond: Audio Content Indexing and Retrieval*. Wiley, 2005.

[8] J. P. Bello, "Low-level features and timbre, MPATE-GE 2623 Music Information Retrieval", New York University [Online]. Available: http://www.nyu.edu/classes/bello/MIR_files/timbre.pdf

[9] D. Jang, M. Jin, and C. D. Yoo, "Music genre classification using novel features and a weighted voting method", in *Proc. IEEE Int. Conf. Multimed. Expo ICME 2008*, Hanover, Germany, 2008, pp. 1377–1380.

[10] Kappa Coefficient [Online]. Available: http://statsoft.pl/czytelnia/artykuly/Krzywe_ROC_czyli_ocena_jakosci.pdf

[11] B. Kostek, "Content-based approach to automatic recommendation of music", in *Proc. 131st Audio Engin. Soc. Convention*, New York, NY, USA, 2011.

[12] B. Kostek, "Music information retrieval in music repositories", in *Rough Sets and Intelligent Systems – Professor Zdzisław Pawlak in Memoriam*, Z. Suraj and A. Skowron, Eds. Intelligent Systems Reference Library, vol. 42, pp. 463–489. Berlin Heidelberg: Springer, 2013.

[13] B. Kostek, *Perception-Based Data Processing in Acoustics – Applications to Music Information Retrieval and Psychophysiology*. Studies in Computational Intelligence, vol. 3. Berlin Heidelberg: Springer, 2005.

[14] B. Kostek, *Soft Computing in Acoustics, Applications of Neural Networks, Fuzzy Logic and Rough Sets to Musical Acoustics. Studies in Fuzziness and Soft Computing*. Heildelberg New York: Physica Verlag, 1999.

[15] B. Kostek and A. Czyzewski, "Representing musical instrument sounds for their automatic classification", *J. Audio Engin. Soc.*, vol. 49, pp. 768–785, 2001.

[16] B. Kostek and Ł. Kania, "Music information analysis and retrieval techniques", *Archiv. of Acoust.*, vol. 33, no. 4, pp. 483–496, 2008.

[17] B. Kostek *et al.*, "Report of the ISMIS 2011 Contest: Music Information Retrieval, Foundations of Intelligent Systems", in *Proc. 19th Int. Symp. Methodol. Intell. Sys. ISMIS 2011*, Warsaw, Poland, 2011. Springer, 2011, pp. 715–724.

[18] Lastfm [Online]. Available: http://www.last.fm/

[19] T. Li, M. Ogihara, and Q. Li, "A comparative study on content-based music genre classification", in *Proc. 26th Ann. Int. ACM SIGIR Conf. Res. Develop. Inform. Retriev.*, Toronto, Canada, 2003, pp. 282–289.

[20] T. Lidy, A. Rauber, A. Pertusa, and J. Inesta, "Combining audio and symbolic descriptors for music classification from audio", Music Information Retrieval Information Exchange (MIREX), 2007.

[21] A. Lindsay and A. Herre, "MPEG-7 and MPEG-7 audio – an overview", *J. Audio Engin. Soc.*, vol. 49, no. 7–8, pp. 589–594, 2001.

[22] M. Mandel and D. Ellis, "LABROSA's audio music similarity and classification submissions", Music Information Retrieval Information Exchange (MIREX), 2007.

[23] Music store Amazon [Online]. Available: http://www.amazon.com/

[24] Music store Itunes [Online]. Available: https:// www.apple.com/pl/itunes/

[25] E. Panagakis, E. Benetos, and C. Kotropoulos, "Music genre classification: a multilinear approach", in *Proc. 9th Int Conf Music Inform Retriev ISMIR 2008*, Philadelphia, PA, USA, 2008, pp. 583–588.

[26] Pandora [Online]. Available: http://www.pandora.com

[27] P. Symeonidis, P. Ruxanda, A. Nanopoulos, and Y. Manolopoulos, "Ternary semantic analysis of social tags for personalized music recommendation", in *Proc. 9th Int Conf Music Inform Retriev ISMIR 2008*, Philadelphia, PA, USA, 2008, pp. 219–224.

[28] ISMIR – The International Society for Music Information Retrieval, International Conference on Music Information Retrieval [Online]. Available: http://www.ismir.net/

[29] G. Tzanetakis and P. Cook, "Musical genre classification of audio signal", *IEEE Trans Speech and Audio Proces.*, vol. 10, no. 5, pp. 293–302, 2002.

[30] G. Tzanetakis, G. Essl, and P. Cook, "Automatic musical genre classification of audio signals", in *Proc. Int. Symp. Music Inform. Retriev. ISMIR 2001*, Bloomington, USA, 2001.

[31] Varence website [Online]. Available: http://www.verance.com/

[32] L. J. Williams and H. Abdi, "Principal component analysis", *Wiley Interdiscip. Rev. Computat. Statis.*, vol. 2, no. 4, pp. 433–459, 2010.

[33] P. Żwan and B. Kostek, "System for automatic singing voice recognition", *J. Audio Engin. Soc.*, vol. 56, no. 9, pp. 710–723, 2008.

Bożena Kostek holds professorship at the Faculty of Electronics, Telecommunications and Informatics, Gdańsk University of Technology (GUT), Poland. She is Head of the Audio Acoustics Laboratory. She received her M.Sc. degrees in Sound Engineering (1983) and Organization and Management (1986) from GUT. She also received postgraduate DEA degree (1988) from Toulouse University, France. In 1992 she supported her Ph.D. thesis with honors at GUT, and in 2000 her D.Sc. degree at the Research Systems Institute, Polish Academy of Sciences. In 2005 President of Poland granted her the title of Professor. In 2013 Prof. Kostek has been elected as a member of the Polish Academy of Sciences. She published over 450 scientific papers in journals and at international conferences, she also published three books. She led a number of research projects supported by the Polish Ministry of Science, took part in many Polish and European research projects. She serves as the Editor-in-Chief of the Journal of the Audio Eng. Soc. since 2011. She was also the Editor-in-Chief of Archives of Acoustics (2007–2012). She was the recipient of many prestigious awards for research, including those of the Prime Minister of Poland for outstanding research achievements, prizes of the Polish Academy of Sciences and Ministry of Science and the Bachelor Cross of the Polonia Restituta Order (2011). She also received the Audio Eng. Soc. Fellowship Award in 2010 and Citation award in 2013. Her research activities are interdisciplinary, however the main research interests focus on audio signal processing, acoustical monitoring, music information retrieval, cognitive bases of sound and vision processing, human-computer interaction, as well as applications of soft computing methods to the mentioned domains.
E-mail: bokostek@audioakustyka.org
Laboratory of Audio Acoustics
Faculty of Electronics, Telecommunications
and Informatics
Gdańsk University of Technology
Narutowicza st 11/12
80-233 Gdańsk, Poland

Andrzej Kaczmarek obtained the M.Sc. degree in Computer Science in 1972. He works from 1975 in Gdańsk University of Technology, and from 1985 in Sound Engineering Department (today Multimedia Systems Department on Faculty of Electronics, Telecommunication, and Informatics). He received his Ph.D. degree in Telecommunications in 1995 from Gdańsk University of Technology. Thesis was based on research on classification of organ pipes sound. He has participated in several national and European projects, including grants KBN (last Security and Synat). His scientific interests concentrate on speech signal analysis (speaker and speech recognition, biometrics, medical applications), musical signal analysis, data mining and statistical methods. He is author or co-author of about 50 scientific papers.
E-mail: akacz@sound.eti.pg.gda.pl
Multimedia Systems Department
Gdańsk University of Technology
Narutowicza st 11/12
80-233 Gdańsk, Poland

Paweł Spaleniak graduated from Gdańsk University of Technology, Faculty of Electronics, Telecommunications and Informatics in the Multimedia Systems Department (2011), obtaining a master's degree in Engineering. The aim of his thesis was software development capable to automatically analyze the emitted commercials level. His interests include audio and image analysis and processing using C++ language. He is a graphic designer, mainly for Web applications. Moreover, he is interested in room acoustics, as well as audio recording, mixing and mastering. Paweł Spaleniak received a Best paper award in the contest for young scientists for the paper entitled "Automatic analysis system of TV commercial emission level" during the NTAV/SPA 2012 conference which took place in Łódź, 26–29 September, 2012. The contest was organized by the Foundation for Supporting the Development of Radio Communications and Multimedia Techniques.
E-mail: papol@sound.eti.pg.gda.pl
Laboratory of Audio Acoustics
Faculty of Electronics, Telecommunications
and Informatics
Gdańsk University of Technology
Narutowicza st 11/12
80-233 Gdańsk, Poland
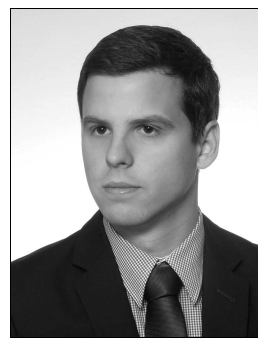
Piotr Hoffmann graduated from the Faculty of Electronics, Telecommunications and Informatics at Gdańsk University of Technology (2012). The subject of engineer's thesis was comparative analysis of decision algorithms in the context of effectiveness speakers recognition. The subject of his master's thesis was sound reinforcements system project of Gdańsk Shakespeare Theater. Piotr Hoffmann received a best master thesis award at IEEE Computer Science contest. He is interested in sound reinforcements systems, sound perception and intelligent processing signals.
E-mail: phoff@sound.eti.pg.gda.pl
Laboratory of Audio Acoustics
Faculty of Electronics, Telecommunications
and Informatics
Gdańsk University of Technology
Narutowicza st 11/12
80-233 Gdańsk, Poland