PAN ARCHIVES OF ACOUSTICS
Vol. **42**, No. 2, pp. 223–233 (2017)

DOI: 10.1515/aoa-2017-0025

Music Performers Classification by Using Multifractal Features: A Case Study

Natasa RELJIN[(1)], David POKRAJAC[(2)]

[(1)] *University of Connecticut*
260 Glenbrook Road, Unit 3247, Storrs, CT 06269, USA; e-mail: natasa.reljin@gmail.com

[(2)] *Delaware State University*
1200 North DuPont Hwy, Dover, DE 19901, USA

(*received September 27, 2015; accepted February 28, 2017*)

In this paper, we investigated the possibility to classify different performers playing the same melodies at the same manner being subjectively quite similar and very difficult to distinguish even for musically skilled persons. For resolving this problem we propose the use of multifractal (MF) analysis, which is proven as an efficient method for describing and quantifying complex natural structures, phenomena or signals. We found experimentally that parameters associated to some characteristic points within the MF spectrum can be used as music descriptors, thus permitting accurate discrimination of music performers. Our approach is tested on the dataset containing the same songs performed by music group ABBA and by actors in the movie *Mamma Mia*. As a classifier we used the support vector machines and the classification performance was evaluated by using the four-fold cross-validation. The results of proposed method were compared with those obtained using mel-frequency cepstral coefficients (MFCCs) as descriptors. For the considered two-class problem, the overall accuracy and *F-measur*e higher than 98% are obtained with the MF descriptors, which was considerably better than by using the MFCC descriptors when the best results were less than 77%.

Keywords: music classification; multifractal analysis; support vector machines; cross-validation; mel-frequency cepstral coefficients.

Notations

F – first point in multifractal spectrum,
FM – first point and point of maximum in multifractal spectrum,
FML – first point, point of maximum and last point in multifractal spectrum,
FN – false negative,
FP – false positive,
FV – feature vector,
L – last point in multifractal spectrum,
M – point of maximum in multifractal spectrum,
ML – point of maximum and last point in multifractal spectrum,
MF – multifractal,
MFCC – mel-frequency cepstral coefficients,
OA – overall accuracy,
RBF – radial basis function,
SVM – support vector machines,
TN – true negative,
TP – true positive.

1. Introduction

Explosive growth of inexpensive but highly powerful multimedia devices enables the production of extremely huge collection of various audio-visual data. Indexing and browsing such data, searching for desired files, and classifying existing material, have become very difficult tasks. Many efforts have been made to resolve these problems. Except for images and video, significant attention was devoted to the automatic analysis and classification of music content and audio information, having numerous potential applications, such as genre classification, musical instrument classification, indexing of audio databases, etc., which are usually referred as music information retrieval (MIR), as reported in (Wold *et al.*, 1996; Li *et al.*, 2001; Tzanetakis, Cook, 2002; Guo, Li, 2003; Kostek, 2004; Barbedo, Lopes, 2007; Feng *et al.*, 2008; Lee *et al.*, 2009; Jensen *et al.*, 2009).

Humans, especially those who are musically educated and/or gifted, are capable of making accurate distinctions between different music pieces and separating sounds originated from different sources. Recognition and classification of sounds is performed spontaneously by means of subjective auditory sensation generated in human brain. Regarding machine process-

ing, description and classification of sounds are very hard and challenging tasks, because a device (or an algorithm) needs appropriate descriptions of perceptual features. The main problem in automatic classification of sounds is to find suitable objective descriptors in good correspondence to subjective sensation of sounds. Usually, descriptors are expressed as numerals and are arranged in the form of an appropriate feature vector (FV). By comparing FVs of different sounds their similarity/dissimilarity can be evaluated.

Standard approach for describing audio content uses temporal and/or spectral features (Tzanetakis, Cook, 2002; Guo, Li, 2003; McKinney, Breebaart, 2003; Rein, Reisslein, 2006; Chudy, 2008). Among various spectral features, the most frequently used are *mel-frequency cepstral coefficients* (MFCCs), which are perceptually motivated and well suited to the human auditory system. The Mel scale[1] was introduced by (Stevens *et al.*, 1937) as a scale of pitches which are subjectively equal in distance from one another. The MFCCs were proven as an efficient tool for speaker identification (Davis, Mermelstein, 1980) and have widely been used in different systems for automatic speech recognition and speaker classification, for instance (Huang *et al.*, 2001; Muda *et al.*, 2010). Later, the use of MFCC was extended to music analysis and classification (Logan, 2000; Berenzweig *et al.*, 2002; Tsai, Wang, 2006; Feng *et al.*, 2008). Basic audio descriptors are even standardized and embedded into the MPEG-7 standard (Kostek, 2004; Lyndsay, 2011; Gomez, 2013;). Note that for the two challenging problems in MIR: recognition of music genres and recognition of instruments playing together in a given music sample, the data mining contest was organized in 2011, in conjunction with the 19th Int. Symposium on Methodologies for Intelligent Systems, ISMIS 2011 (Kryszkiewicz *et al.*, 2011). In this contest, competitors were requested to use feature vectors with a defined number of 171 descriptors from which the first 147 were 'standard' descriptors: MPEG-7 (127 descriptors), and MFCC (20), while additional 24 were related to time domain and have been a free choice of competitors being their original contribution (Kostek, 2011). In the paper by (Schedl *et al.*, 2013) very interesting study regarding system-based and user-centric MIR was derived. The authors pointed out the problems with subjective judgment of similarity of songs/music and hence difficulties at user-centric evaluation in fields related to MIR.

In (Mandelbrot, 1967) Mandelbrot introduced a new kind of geometry, called the *fractal geometry*, which has been proven as an efficient way for describing the complexity of structures, objects, systems, or phenomena. The fractal concept is one of the most important developments in mathematics in the second half of the 20th century. Fractals are central to understanding and quantitatively evaluating a wide variety of complex structures (for instance, the shape of clouds, structure of a tree or a snowflake, etc.) as well as chaotic, non-stationary and nonlinear systems, in cases when Euclidean geometry falls down. Such structures and systems can be described quantitatively by a *fractal dimension* (FD) which is usually a non-integer number (thus Mandelbrot coined the term *fractal*, meaning fractional or broken). By using the FD, objective description, characterization, comparison and classification of complex and irregular structures is enabled. This concept was very successful in describing events, signals, structures or phenomena, characterized by a fundamental feature known and referred to as *self-similarity*. This property means that by observing the structure of the object in different scales, for instance by zooming part of the structure, (almost) the same shape arises, i.e., it seems that the structure is composed of smaller versions of itself. Artificially generated self-similar objects by applying some predefined rules, for instance, the Cantor set, the von Koch's curve and snowflake, the Sierpinski carpet, etc. (Peitgen *et al.*, 2004), have *exactly the same* FD in all scales. Such objects are characterized by an unique FD and are known as *monofractals*. Conversely, a large scale of (mainly natural) objects, for instance, a coastline, structure of a tree, venous, arterial or nervous system, some vegetables (cauliflower, broccoli), even trends in economy, structure of vocal sounds, music, etc., exhibit some kind of self-similarity, but not in strict sense, meaning these objects have different FD at different scales. Such objects cannot be described by an unique FD. Instead, the distribution of FDs over different scales is used to provide even deeper insight into the structure. This is a simple explanation of *multifractal concept*, as an extension of fractal geometry (Mandelbrot, 1982). Since FDs of such objects differ at various scales, these objects are known as *multifractals*. The distribution of FDs can be expressed in the form of so-called *multifractal spectrum*, which will be described in Sec. 2.

The interdisciplinary nature of fractal geometry and multifractals has found a broad spectrum of applications, for instance, in the classification of natural objects (Mandelbrot, 1982; Stanley, Meakin, 1988), in the analysis of nonlinear and chaotic physical phenomena (Grassberger, 1983; Hentschel, Procaccia, 1983), in the description of biology structures (Buldyrey *et al.*, 1994; Iannaccone, Khokha, 1996), in medicine (Sedivy, Mader, 1997; Vasiljevic *et al.*, 2012; Reljin *et al.*, 2015), in economics (Falconer, 2003), even in arts (Bovill, 1996; Peitgen *et al.*, 2004). Moreover, these techniques found significant applications in signal and image analyses

[1]The word **mel** comes from the word **melody**, indicating to the pitch comparison.

and processing: the reader can find many examples, for instance, in (Vehel, Mignot, 1994; Vehel, 1996; 1998; Reljin *et al.*, 2000).

Regarding the audio, fractal and multifractal analyses were also used, mainly for speech analysis and recognition, but also for music analysis and classification. For instance, the paper (Sabanal, Nakagawa, 1996) considered fractal properties of vocal sounds and results were applied to the speech recognition model. Further, the authors of (Maragos, Potamianos, 1999; Pitsikalis, Maragos, 2009) applied the fractal dimension of speech signals to their automatic recognition and classification. The Higuchi fractal dimension (Higuchi, 1988), in combination with MFCC, was used in (Ezeiza *et al.*, 2011) in order to improve the correct word rate for automatic speech recognition. The authors of (Krajewski *et al.*, 2012) evaluated fractal features, among other nonlinear dynamics features, for speech based sleepiness detection, while the authors of (Gonzales *et al.*, 2012) explored the fractal and multrifractal nature of speech signals from two different Portuguese speech databases and found that, in general, all analyzed signals revealed multifractal behavior under a time frame analysis ranging from 50 ms to 100 ms.

In (Hsu, Hsu, 1990), the authors suggested the methodology for describing and characterizing music pieces using fractal dimension. Their initial results were derived from some Bach's and Mozart's music pieces and Swiss children's songs. In the study (Bigerelle, Iost, 2000) was shown that fractal dimension can be used to discriminate different music genres. In the papers (Su, Wu, 2006; 2007) authors found that sequences of musical notes exhibit fractal nature, and demonstrated the applicability of MF analysis to distinguish between different styles of music. Music search and playlist generation based on fractal dimensions of music were presented in the paper (Hughes, Manaris, 2012). Also, the authors of (Zlatintsi, Maragos 2013) proposed the use of multiscale fractal dimension for recognizing musical instruments.

The main benefit of using MF in signal processing is that this concept enables both local and global analyses of an observed signal. Hence, using MF analysis it is possible to find and extract details from the signal under consideration, which carries some hidden and subtle information thus enabling the recognition, selection and classification of complex signals. For instance, MF was applied to find clicks in heart sounds characterizing pathological syndrome so-called the MVP (*mitral valve prolapse*, or *click-murmure syndrome*), as reported in (Gavrovska *et al.*, 2013). Moreover, it was shown that some MF parameters can be used as characteristic features for discriminating malignant from benign cases from appropriate medical signals, as shown in (Reljin *et al.*, 2008), or for identifying the primary cancer from biopsy images of bone metastases (Vasiljevic *et al.*, 2012).

In music/sound analysis and classification particular attention should be addressed to the problem of distinguishing and classifying performers (or music groups) playing the same music piece(s) in the same manner, using the same types of instruments, similar vocals, and under the same arrangements. In this case, performed music pieces are subjectively similar thus making their recognition and classification very difficult even for musically skilled persons. This problem, which may be of interest in forensic and/or copyright issues, is the goal of our research. Since the MF analysis has been proven as an efficient tool for describing signals and finding fine details and characteristic parts within signals, we investigated the use of MF for resolving this problem. Although the MF analysis was applied to some of audio related problems, to the best of our knowledge this approach was not used for given task of music performers classification. We found that only a few MF parameters, which correspond to the characteristic points within the MF spectrum, could be used as features, thus enabling successful classification of music performers. As a classification tool we used support vector machines (SVM), while four-fold cross-validation was used to determine optimal values of the SVM hyperparameters and to evaluate the accuracy of the proposed classifier.

The paper is organized as follows. Section 2 considers the basics of multifractal analysis. In Sec. 3 a concept of proposed classifier based on MF parameters as relevant features is presented. Section 4 describes experimental setup for music performer classification based on MF features and presents classification results derived on the dataset containing the same songs performed by music group ABBA and by actors in the movie *Mamma Mia*. Results are compared with those derived on the same dataset by applying MFCC as characteristic features and using the same SVM classifiers with linear, polynomial and RBF kernel functions. Concluding remarks are given in Sec. 5.

2. Basics of the multifractal analysis

Multifractal spectrum can be derived in several ways, as reported in literature (Harte, 2001; Peitgen *et al.*, 2004). The *histogram method* is very popular due to its simplicity and possibility to determine MF spectrum directly from measured data (Chhabra, Jensen, 1989; Vehel, 1998). Main steps for deriving the MF spectrum using this method will be briefly explained as follows.

Signal is covered by non-overlapping boxes B_i of side width ε. Within the box, a signal is characterized by some attribute, called a *measure*, $\mu(B_i)$ (Vehel, Mignot, 1994). Different measures may be used, for instance, the *maximum* of the signal's intensity, the

minimum value, the *sum* of intensities, etc. (Vehel, 1998). It is common to assume normalized space, i.e., ε, $\mu \in [0, 1]$.

For audio signals, which are considered in this paper, signal is time-dependent and boxes are represented by time intervals (windows) of width ε. Among different measures that can be used, we found that measure maximum provided the best results for considered classification problem. For a particular time interval B_i the *coarse Hölder exponent* α_i is calculated as (Vehel, Mignot, 1994):

$$\alpha_i = \frac{\log\left(\mu(B_i)\right)}{\log \varepsilon}. \tag{1}$$

By using a sequence of time intervals $B_i^{(k)}$, with descending widths $\varepsilon^{(k)}$, $k = 1, 2, \ldots$, i.e., $\varepsilon^{(k)} > \varepsilon^{(k+1)}$, and $B_i^{(k)} \supset B_i^{(k+1)}$, the corresponding values of parameter α_i will differ, but will approach to limit value α, known as the Hölder exponent (Vehel, 1998). Due to practical reasons, the value of α is usually estimated as a slope of linear regression line in the log-log diagram: $\log(\mu(B_i))$ vs. $\log(\varepsilon)$, for several values of ε (Su, Wu, 2006).

After estimating Hölder exponents for all signal samples the α-representation of the signal is obtained – each signal sample is characterized by its α value. Since α is calculated from the measure μ around a particular sample of the signal, this parameter describes *local singularity* (*regularity*) of the signal. For samples where the signal is smooth (slow varying with respect to neighbor samples) the value of α is small, while samples within regions with sudden changes are characterized by high values of α (Vehel, 1998). The Hölder exponent has finite limits $\alpha_{\min}$ and $\alpha_{\max}$. In the whole signal, many samples can have the same value of α, i.e., they can have the same local regularity. After estimating the Hölder exponents of all samples within the observed signal, we find their distribution. All obtained values of Hölder exponents can be considered as the α-space. The continuous α-space is divided into R values as follows:

$$\alpha_r = \alpha_{\min} + (r-1)\Delta\alpha, \qquad r = 1, 2, \ldots, R, \tag{2}$$

$$\Delta\alpha = (\alpha_{\max} - \alpha_{\min})/R, \tag{3}$$

and the histogram of α_r values is calculated: if the actual value of α falls within the subrange $[\alpha_r,\ \alpha_{r+1})$, this value is replaced by α_r. The number of subranges R has to be determined empirically. Small value of R behaves as low-frequency filtering: MF spectrum will be smooth but with small resolution, which reduces discriminative capabilities. As opposed to when using high R value, more details can be extracted but spectrum becomes irregular (saw-toothed). A compromise solution could be to choose R between 50 and 100.

In the next step, the α-space is covered by boxes of width δ ($\delta < 1$) and the number of boxes, $N_\delta(\alpha)$, containing given value of Hölder exponent, $\alpha = \alpha_r$, is counted. Furthermore, the *Hausdorff dimension* of the distribution of α, also known as the *multifractal singularity spectrum* (or simply, the *MF spectrum*), is determined as (Vehel, 1998):

$$f(\alpha) = -\lim_{\delta \to 0} \frac{\log(N_\delta(\alpha))}{\log \delta}. \tag{4}$$

In practice, similar to determining Hölder exponents, the values of $f(\alpha)$ are estimated by a linear regression in $(\log(\delta), \log(N_\delta))$ for several box sizes δ. The plot of MF spectrum is usually parabola shaped, as shown in Fig. 1, with finite values, $\alpha_{\min}$; $\alpha_{\max}$; $f_{\min}(\alpha)$; $f_{\max}(\alpha)$.

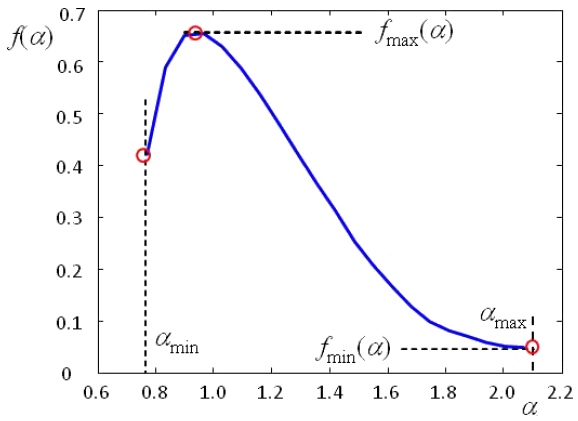

Fig. 1. Typical shape of MF spectrum.

Values of $f(\alpha)$ describe *global regularity* of the signal. Small values of $f(\alpha)$ correspond to rare events, meaning that a small number of points in the original space (amplitude-time space) is characterized by this particular value of α. The opposite is true for high values of $f(\alpha)$. By combining the pair $(\alpha,\ f(\alpha))$, both local regularity (via α) and global behavior (via $f(\alpha)$) can be described, thus permitting fine analysis and classification of signals, both images (Vehel, 1998; Reljin *et al.*, 2000) and music (Su, Wu, 2006; 2007).

3. Music performer classifier based on MF features

3.1. Creation of audio database

For given problem, distinguishing and classifying music performers playing the same melodies in quite similar manner, our initial assumption was that irrespective of subjective similarity, each performer or music group has some characteristic (individual) features. Since the MF analysis has been proven as a powerful method for finding subtle details within signals (Gavrovska *et al.*, 2013) and for characterizing and distinguishing different signals (Reljin *et al.*, 2008; Vasiljevic *et al.*, 2012) we investigated the MF approach for resolving given problem. As examples we

will observe the same 14 songs, listed in Table 1, performed by famous Swedish group ABBA, producing mega hits from 1974 to 1982 (Sheridan, 2012), and by actors in the movie *Mamma Mia* (released in 2008). These songs are denoted here as ABBA and MOVIE. These melodies were composed by the same composers and arrangers (Benny Anderson and Björn Ulvaeus) and were performed in a very similar way, being difficult for distinguishing even for musically educated persons. We have created a database as follows.

Table 1. Songs considered for classification.

Pieces ABBA	Title of the song	Pieces MOVIE
1, 2	Dancing Queen	29, 30
3, 4	Does Your Mother Know	31, 32
5, 6	Gimme Gimme Gimme	33, 34
7, 8	I Have a Dream	35, 36
9, 10	Lay All Your Love on Me	37, 38
11, 12	Mamma Mia	39, 40
13, 14	Money Money Money	41, 42
15, 16	SOS	43, 44
17, 18	Super Trouper	45, 46
19, 20	Take a Chance on Me	47, 48
21, 22	Thank You for the Music	49, 50
23, 24	The Name of the Game	51, 52
25, 26	The Winner Takes It All	53, 54
27, 28	Voulez Vous	55, 56

Prior to further processing and classification, all the songs are preprocessed. First, recordings are converted from stereo to mono, and downsampled to 8 kHz, with the *Audacity software* (Audacity, 2015). Although music is characterized by wide bandwidth, and nowadays the sampling frequency is usually 44.1 kHz, the reason for using 8 kHz is based on several facts. In the paper (Rein, Reisslein, 2006) was shown that sampling frequency of 8 kHz is sufficient to identify classical music compositions, although such music is characterized by spectra rich in harmonics. Further, the authors of (Jensen *et al.*, 2009) derived deep quantitative analysis of a MFCC, as a common audio similarity measure. The authors have shown that if all songs have the same sampling frequency of 8 kHz, the classification accuracy decreased only by few percents compared to when higher sampling frequency is used. Certainly, extracting MFCCs from downsampled songs is computationally easier and cheaper, and since classification accuracy is not noticeably degraded, authors suggested the use of homogeneous music collection downsampled to 8 kHz. This is of particular interest in cases when songs in actual database do not have the same sampling rate.

After downsampling, songs were normalized with respect to their amplitudes. Each song from the considered audio collection has parts that repeat; parts with the same melody but different lyrics – verses, and parts with the same melody and the same lyrics – choruses. Hence, by selecting just these parts of the songs, we obtain parts that are subjectively similar and have all the necessary information for representing the basic music characteristics of the song (melodic line and harmony), as well as characteristics of performers. By following this assumption, we constructed two music sequences per song (in the text we will call these sequences *pieces*): the first verse and chorus (denoted by odd numbers 1–27 for ABBA and 29–55 for MOVIE), and the second verse and chorus (denoted by even numbers 2–28 and 30–56 for ABBA and MOVIE, respectively). Each piece lasts about 40 seconds. This way we constructed the music dataset with 28 pieces per performer (music group ABBA and MOVIE), i.e., 56 pieces in the whole dataset, as indicated by ordinal numbers 1 to 56 in Table 1.

Next step in signal preprocessing is to divide each music piece into short overlapping blocks called *frames* as is common in audio analysis (Logan, 2000; Berenzweig *et al.*, 2002; Barbedo, Lopez, 2007; Zlatinsi, Maragos, 2013). The reason for using short sequences, of length 20–50 ms, is to assure the stationarity of the signal. Namely, as shown in (Rabiner, Juang, 1993, p. 17), speech signal is almost stationary over a sufficiently short period of time (between 5 and 100 ms), and similar conclusion is derived for music instruments (Zlatintsi, Maragos, 2013). In our study, we used frames of 32 ms of length overlapped by 50%. In order to reduce the discontinuities at the edges of the frames, and thus to reduce the spectral leakage, the Hamming tapered window is applied.

3.2. MF feature extraction

For each particular frame within our dataset, we estimated its MF spectrum using custom developed software based on histogram method (Reljin *et al.*, 2000). Then we determined the MF spectra for every music piece (28 of ABBA and 28 of MOVIE) as a mean of MF spectra of their frames. The two illustrative examples are depicted in Figs. 2–5. Figures 2 and 3 respectively relate to first and second pieces of song *Dancing Queen* (pieces denoted by numerals 1 and 2 for ABBA, and by 29 and 30 for MOVIE, according to Table 1), while Figs. 4 and 5 relate to song *Does Your Mother Know* (pieces 3 and 4 for ABBA, and 31 and 32 for MOVIE). Characteristic points within MF spectra plots: the first point (F), the last point (L), and the point of maximum (M), are denoted, and corresponding values of α and $f(\alpha)$ are inserted into Figs. 2–5. Although MF spectra are visually similar, the three significant conclusions may be derived:

1. For the same music group playing the same song, for instance *Dancing Queen*, Figs. 2 and 3, the first

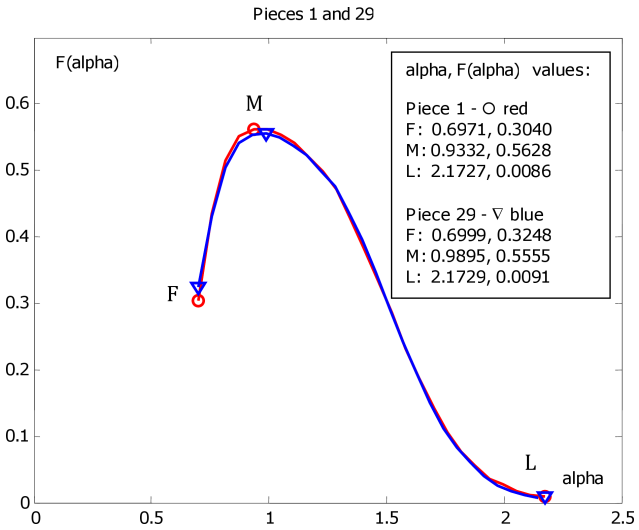

Fig. 2. MF spectra of the first pieces of the song *Dancing Queen*: 1 (ABBA) and 29 (MOVIE).

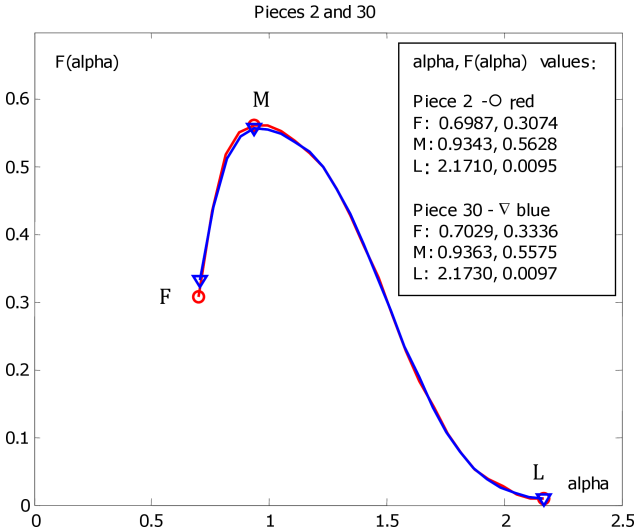

Fig. 3. MF spectra of the second pieces of the song *Dancing Queen*: 2 (ABBA) and 30 (MOVIE).

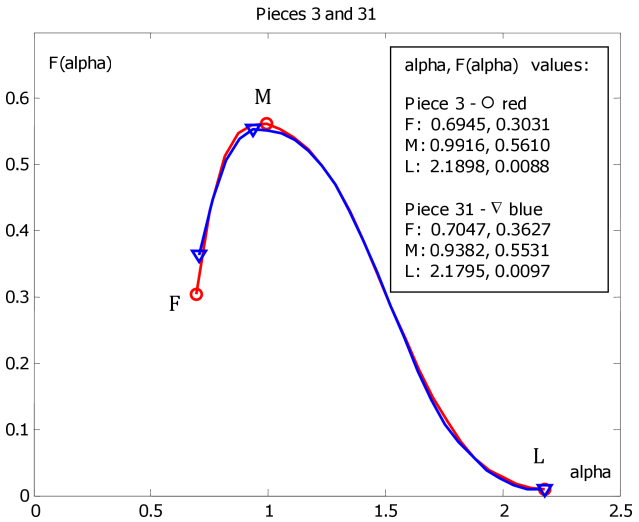

Fig. 4. MF spectra of the first pieces of song *Does Your Mother Know*: 3 (ABBA) and 31 (MOVIE).

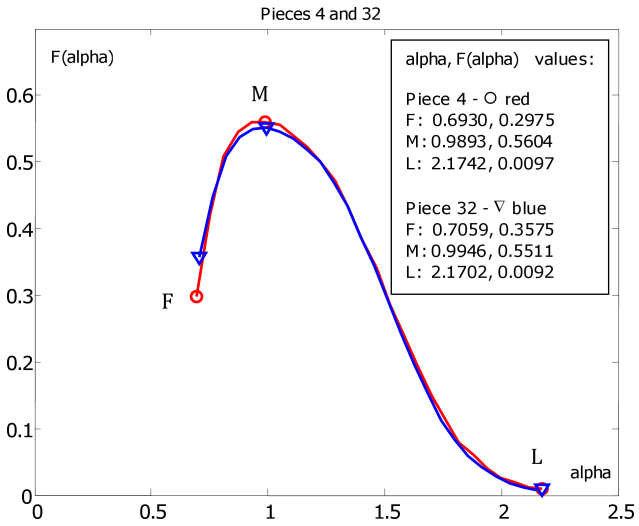

Fig. 5. MF spectra of second pieces of song *Does Your Mother Know*: 4 (ABBA) and 32 (MOVIE).

and second pieces (denoted as 1–2 for ABBA; and 29–30 for MOVIE) have very similar values (α, $f(\alpha)$) at characteristic points F, M and L.

2. For the same song performed by different groups, characteristic values (α, $f(\alpha)$) differ, as noted in Figs. 2 and 3. By comparing characteristic values for pieces 1 and 29; and 2 and 30, we can note that although differences are not so intensive, they can be used for classification of music performers.

3. For different songs performed by the same group, characteristic values (α, $f(\alpha)$) differ, as can be noted when comparing corresponding values for pieces 1 and 3, and 2 and 4 (ABBA); and 29 and 31, 30 and 32 (MOVIE).

Similar conclusions can be derived for all considered songs in our dataset.

3.3. Feature vectors

Based on conclusions 1 to 3 from previous section, we proposed the use of MF parameters: values of α and $f(\alpha)$ describing characteristic points as components of feature vectors. We explored different combinations of characteristic points from MF spectra: (i) only single points (F, M or L), (ii) combinations of two points (F+M, F+L, M+L), and (iii) all three points (F+M+L). Hence, obtained feature vectors associated to each of music pieces contain 2, 4 and 6 components for cases (i), (ii) and (iii), respectively.

The proposed method for music classification using MF features is compared with method that uses MFCCs as characteristic features. To this end we calculated MFCCs for the same frames of all 56 music pieces, using publicly available software the *Auditory toolbox* (Slaney, 1998). By finding MFCCs for the whole dataset, we selected their dominant components

supporting more than 98% of signal energy (in our case we selected first 13 components).

3.4. Classification and performance evaluation

In this study we used support vector machines as a classifier. This method was originally developed for binary (two-class) problem (Vapnik, 1998) and later was extended to multiclass problems as well (Hsu, Lin, 2002). The SVMs are widely and successfully used in many classification problems, including the music information retrieval. Moreover, as noted in (Rosner *et al.*, 2014), the SVM algorithm is even better choice for music genre classification than, for instance, a very popular k-nearest neighbor (k-NN) method. The goal of SVMs is to construct a hyperplane in the space of transformed input vectors, which will separate observations from different classes such that the minimal distance between observations and the separation hyperplane is maximized (Kecman, 2001; Bishop, 2006). We used SVMs with different kernels: linear, polynomial and radial basis function (RBF) (Kecman, 2001; Chang, Lin, 2011).

The performance of the classification model can be measured in several ways. The confusion matrix is frequently used, and for the two-class problem has the form as given in Table 2 (Tan *et al.*, 2005). Classes are denoted as $+1$ and -1: in our case classes correspond to music pieces ABBA and MOVIE. The entries of the confusion matrix have the following meaning: the true positive (TP) value denotes the number of music pieces belonging to the class $+1$ which are correctly classified as $+1$, while false negative (FN) is the number of pieces from class $+1$ which are incorrectly predicted as class -1. Similarly, false positive (FP) represents the number of pieces from the class -1 which are incorrectly classified, and true negative (TN) is the number of correctly classified pieces from the class -1.

Table 2. Confusion matrix for our two-class problem.

	Predicted class $+1$	Predicted class -1
True class $+1$ (ABBA)	True Positive (TP)	False Negative (FN)
True class -1 (MOVIE)	False Positive (FP)	True Negative (TN)

By combining entries from the confusion matrix, several performance measures can be derived (Tan *et al.*, 2005). These measures are *Overall Accuracy* (OA) and *F-measure* (Tan *et al.*, 2005), which are defined as:

$$OA = \frac{TP + TN}{TP + TN + FP + FN}, \tag{5}$$

$$F\text{-}measure = \frac{2 \cdot TP}{2 \cdot TP + FP + FN}. \tag{6}$$

These two measures are compact, describing classifier's performance with only one value, thus being of high practical use.

To determine the optimal values of the SVM hyperparameters as well as to evaluate the classifier's performance, we utilized a four-fold cross validation method (Bishop, 2006), recommended for small data sets.

4. Experimental results and discussion

Based on previous considerations we developed experimental setup for music performer classification, as depicted in Fig. 6. For every music piece from our dataset we created two groups of feature vectors: based on MF parameters and on MFCC. Feature vectors are determined from MF spectra using custom developed software based on histogram method (Reljin *et al.*, 2000), while MFCCs are calculated following the procedure in (Slaney, 1998).

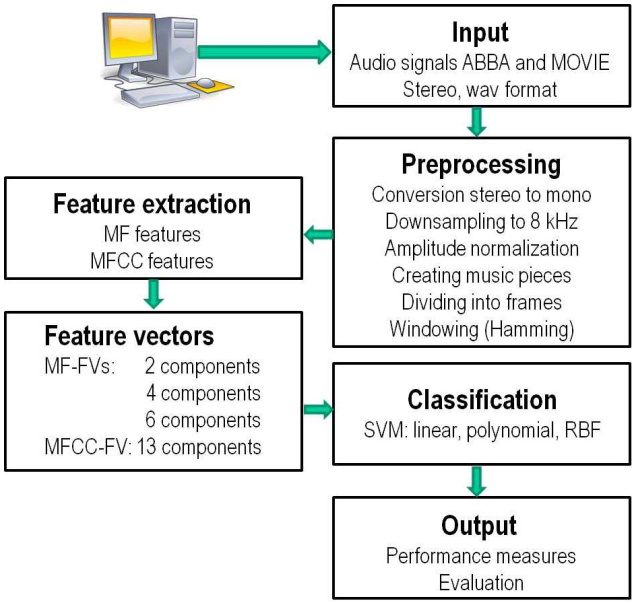

Fig. 6. Block scheme of experimental setup used for music performer classification.

As already noted, feature vectors created from MF analysis consist of pairs $(\alpha,\ f(\alpha))$ from characteristic points in MF spectra, F, M, L and their combinations with two points (F+M, M+L, or F+L), and all three points (F+M+L). This way MF feature vectors are very low-dimensional, containing $n = 2, 4$ and 6 components, for single points, combination of two points, and all three points, respectively. In Fig. 7, plots of pairs $(\alpha,\ f(\alpha))$ associated to characteristic points: F, M and L of MF spectra, for the whole dataset of 56 music pieces (28 ABBA and 28 MOVIE), are depicted. Points related to ABBA are presented by circles, while triangles are related to MOVIE pieces.

Using feature vectors based on the described MF and MFCC features, the classification of music groups

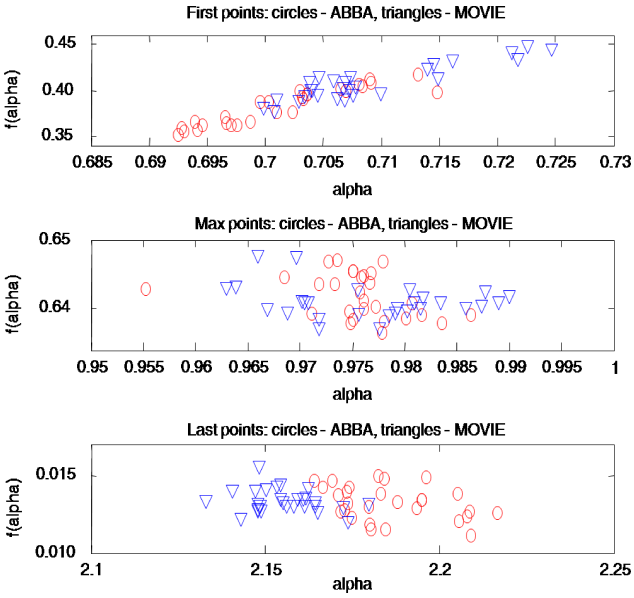

Fig. 7. Plots of characteristic points $(\alpha, f(\alpha))$ of MF spectra for the music dataset with 56 pieces: 28 pieces denoted as ABBA and 28 denoted as MOVIE in Table 1. Characteristic points: the first point (F), points of maxima (M), and the last point (L), are depicted in upper, middle and lower plots, respectively.

from our dataset was performed by applying the support vector machines algorithms embedded in an integrated software LIBSVM (Chang, Lin, 2011). We used linear SVMs and SVMs with polynomial and RBF kernel functions (Kecman, 2001; Bishop, 2006). To assure fair classification and comparison, the four-fold cross-validation was applied, enabling also the determination of optimal values of relevant SVM parameters for the best possible classification.

Classification results, expressed by *Overall Accuracy* and *F-measure* (given by Eqs. (5) and (6)), are presented in Table 3. Note that the MF feature vectors, labelled in Table 3 as **MF_type-n**, are low-dimensional, containing only $n = 2, 4$ or 6 components, respectively, where **type** relates to characteristic point(s): single point (F, M, or L), combination of two points (FM, ML, or FL), and all three points (FML). Since MFCC method is widely used and well described in literature we only note here that first 13 dominant components, supporting more than 98% of signal energy, are used for feature vectors, denoted in Table 3 as **MFCC-13**.

As can be noted from results presented in Table 3, our general observation is that MF features provide better classification than MFCC. The best classification result is obtained with MF-FLM feature vector with polynomial SVM (row 7 in Table 3): obtained results for *OA* and *F-measure* were 98.21% and 98.46%, respectively, and very close were results for the MF-FL (row 6). On the other hand, the best results with MFCCs were 71.43% and 76.50%, respectively for the *OA* and *F-measure* (last row, for RBF SVM).

Note that the single point L (corresponding to high local changes), having only 2 components in FVs, exhibits (in general) the most relevant discriminative capabilities with *F-measure* and Overall Accuracy in the range between 91.07% and 97.13%, depending on the SVM used, as shown in third row in Table 3. The F point (first row) achieves the second best result, while the M point (second row) is the worst case. By combining the two characteristic MF points, classification results are improved. For instance, even M with F produces slightly better results than those of L point alone (row 4, case with linear SVM), while combining M with L (row 5, for all three SVMs) classification capabilities are improved significantly, being even better in comparison to only L point.

These results can be explained by observing plots of $(\alpha, f(\alpha))$ pairs of characteristic points, as depicted in Fig. 7. As evident from plots in Fig. 7, points characterizing music pieces ABBA and MOVIE are clustered in some way, thus permitting their distinction and classification. The best discriminative behavior is that of last points (lowest plot in Fig. 7) – the overlapping of classes is minimal for the case of L points for ABBA and MOVIE, while for the M points (middle plot) classes ABBA and MOVIE are very interweaved thus being difficult to separate.

Table 3. Classification results for different feature vectors and different SVMs.

Feature vector	Linear SVMs		Polynomial SVMs		RBF SVMs	
	OA [%]	*F-meas* [%]	*OA* [%]	*F-meas* [%]	*OA* [%]	*F-meas* [%]
MF_F-2	73.21	71.96	76.79	79.57	69.64	74.32
MF_M-2	55.36	59.36	53.57	65.64	55.36	61.19
MF_L-2	91.07	92.07	96.43	97.13	94.64	94.92
MF_FM-4	91.07	92.59	92.86	93.10	92.86	91.59
MF_ML-4	92.86	93.90	96.43	97.13	96.43	96.48
MF_FL-4	**98.21**	97.87	**98.21**	97.87	**98.21**	98.41
MF_FLM-6	**98.21**	97.87	**98.21**	**98.46**	**98.21**	98.18
MFCC-13	71.43	73.56	71.43	76.28	**71.43**	**76.50**

5. Conclusions

In this paper the use of MF features for music performers classification is proposed. Our study indicates that features obtained from characteristic points of multifractal spectra may be promising for audio classification tasks, and are more suitable for the classification of different music performers playing the same songs than well-adopted mel-frequency cepstral coeffcients. By considering the same songs performed by different music groups (14 songs from the music group ABBA and the same songs performed by the cast of the movie *Mamma Mia*) in quite similar way, the *F-measure* and the Overal Accuracy were about 98% (or slightly better, depending on the SVM used) with MF-based features, which were notably better than the best result with MFCC features (less than 77%). Moreover, by using low-dimensional feature vectors with only two components (containing values of α and $f(\alpha)$ from last points in MF spectra), very good classification was obtained: *Fmeasure* and Overall Accuracy were between 91% and 97% (depending on the SVM used). These results could be explained by the fact that multifractal analysis captures both local regularity and global behavior of the observed signal, permitting better distinguishing of subtle details within signals.

The proposed methodology was evaluated on a limited dataset with 28 songs. A part of our work in progress is the validation of the proposed method using MF features on larger audio datasets. In addition, we plan to compute the MF spectra over shorter music pieces (for instance, 5 to 10 seconds, instead of the currently used 40 second pieces), and explore if performance could be further improved. While the principal aim of the study is music performer classification, it can be used as a case study for a much broader problem of utilizing features based on multifractal analysis for various kinds of music (or audio in general) classification tasks.

Competing interests

The authors do not have any competing interests.

Acknowledgment

The authors would like to thank Dr. Tia L. Vance, University of Maryland Eastern Shore, for her valuable comments.

This work was supported in part by the Delaware IDeA Network of Biomedical Research Excellence grant, US Department of Defense funded "Center for Advanced Algorithms" grant (W911NF-11-2-0046), the US Department of Defense Breast Cancer Research Program (HBCU Partnership Training Award #BC083639), the US National Science Foundation (CREST grant #HRD-1242067), and the US Department of Defense/Department of Army (#54412-CI-ISP).

References

1. *Audacity software*, http://audacityteam.org/, Retrieved Feb. 09, 2017.
2. Barbedo J.G.A., Lopes A. (2007), *Automatic genre classification of musical signals*, EURASIP Journal of Advances in Signal Processing, Article ID 64960.
3. Berenzweig A.L., Ellis D.P.W., Lawrence S. (2002), *Using voice segments to improve artist classification of music*, Proceedings of the Audio Engineering Society (AES) 22nd International Conference on Virtual, Synthetic and Entertainment Audio, pp. 119–122, Espoo, Finland.
4. Bigerelle M., Iost A. (2000), *Fractal dimension and classification of music*, Chaos, Solitons and Fractals, **11**, 14, 2179–2192.
5. Bishop C.M. (2006), *Pattern recognition and machine learning*, Springer, New York.
6. Bovill C. (1996), *Fractal geometry in architecture and design*, Springer Science & Business Media, Boston: Birkhauser.
7. Buldyrev S., Goldberger A., Havlin S., Peng C., Stanley H. (1994), *Fractals in biology and medicine: From DNA to heartbeat*, [in:] *Fractals in science*, Bunde A., Havlin S. [Eds.], pp. 49–87, Berlin: Springer-Verlag.
8. Chang C.-C., Lin C.-J. (2011), *LIBSVM: A library for support vector machines*, ACM Transactions on Intelligent Systems and Technology, **2**, 3, 27:1–27:27, http://www.csie.ntu.edu.tw/∼cjlin/libsvm, Retrieved Feb. 09, 2017.
9. Chhabra A., Jensen R.V. (1989), *Direct determination of the $f(\alpha)$ singularity spectrum*, Physical Review Letters, **62**, 12, 1327–1330.
10. Chudy M. (2008), *Automatic identification of music performer using the linear prediction cepstral coefficients method*, Archives of Acoustics, **33**, 1, 27–33.
11. Davis S., Mermelstein P. (1980), *Comparison of parametric representations for monosyllabic word recognition in continuously spoken sentences*, IEEE Transactions on Acoustics, Speech, and Signal Processing, **28**, 4, 357–366.
12. Ezeiza A., de Ipina K.L., Hernandez C., Barroso N. (2011), *Combining mel frequency cepstral coefficients and fractal dimensions for automatic speech recognition*, Advances in Nonlinear Speech Processing (NOLISP 2011), Lecture Notes in Computer Science (LNAI), **7015**, pp. 183-189, Las Palmas de Gran Canaria, Spain.
13. Falconer K. (2003), *Fractal geometry: Mathematical foundations and application.* 2nd ed., John Wiley & Sons, Ltd.
14. Feng L., Nielsen A.B., Hansen L.K. (2008), *Vocal segment classification in popular music*, Proceedings

of 9th International Symposium on Music Information Retrieval (ISMIR08), pp. 121–126, Philadelphia, PA, USA, 2008.

15. Gavrovska A., Zajic G., Reljin I., Reljin B. (2013), *Classification of prolapsed mitral valve versus healthy heart from phonocardiograms by multifractal analysis*, Computational and Mathematical Methods in Medicine, Article ID 376152.
16. Gomez E., Gouyon F., Herrera P., Amatrian X. (2013), *MPEG-7 for content-based music processing*, Digital Media Processing for Multimedia Interactive Services: Proceedings of the 4th European Workshop on Image Analysis for Multimedia Interactive Services: Queen Mary, University of London.
17. Gonzalez D.C., Ling L.L., Violaro F. (2012), *Analysis of the multifractal nature of speech signals*, Progress in Pattern Recognition, Image Analysis, Computer Vision, and Applications, Lecture Notes in Computer Science, **7441**, 740–748.
18. Grassberger P. (1983), *Generalized dimensions of strange attractors*, Physics Letters A, **97**, 6, 227–230.
19. Guo G., Li S.Z. (2003), *Content-based audio classification and retrieval by support vector machines*, IEEE Transactions on Neural Networks, **14**, 1, 209–215.
20. Harte D. (2001), *Multifractals: Theory and applications*, Chapman and Hall.
21. Hentschel H.G.E, Procaccia I. (1983), *The infinite number of generalized dimensions of fractals and strange attractors*, Physica D: Nonlinear Phenomena, **8**, 3, 435-444.
22. Higuchi T. (1988), *Approach to an irregular time series on the basis of the fractal theory*, Physica D, **31**, 277–283.
23. Huang X., Acero A., Hon H. (2001), *Spoken language processing – A guide to theory, algorithm, and system development*, Prentice Hall PTR, New Jersey.
24. Hsu K.J., Hsu A.J. (1990), *Fractal geometry of music*, Proceedings of the National Academy of Sciences of the USA, **87**, 3, 938–341.
25. Hsu C.-W., Lin C.-J. (2002), *A comparison of methods for multiclass support vector machines*, IEEE Transactions on Neural Networks, **13**, 2, 415–425.
26. Hughes D., Manaris B. (2012), *Fractal dimensions of music and automatic playlist generation*, Proceedings of the Eighth International Conference on Intelligent Information Hiding and Multimedia Signal Processing, pp. 436–440, Piraues, Greece.
27. Iannaccone P.M., Khokha M.K. (1996), *Fractal geometry in biological systems*, CRC Press, Boca Raton, FL.
28. Jensen J.H., Christensen M.G., Ellis D.P.W., Jensen S.H. (2009), *Quantitative analysis of a common audio similarity measure*, IEEE Transactions on Audio, Speech, and Language Processing, **17**, 4, 693–703.
29. Kecman V. (2001), *Learning and soft computing: Support vector machines, neural networks, and fuzzy logic models*, The MIT Press, Cambridge, MA, USA.
30. Kostek B. (2004), *Musical instrument classification and duet analysis employing music information retrieval techniques* (Invited Paper), Proceedings of IEEE, **92**, 4, 712–729.
31. Kostek B. (2011), *Report of the ISMIS 2011 contest: Music information retrieva*, Proceedings of 19th International Symposium ISMIS, pp. 715–724, Warsaw, Poland.
32. Krajewski J., Schnieder S., Sommer D., Batliner A., Schuller B. (2012), *Applying multiple classifiers and non-linear dynamics features for detecting sleepiness from speech*, Neurocomputing, **84**, 65–75.
33. Kryszkiewicz M., Rybinski H., Skowron A., Raś Z.W. [Eds.] (2011), *Foundations of Intelligent Systems – 19th International Symposium*, ISMIS 2011, Warsaw, Poland, June 28–30, Proceedings, Springer series: Lectures Notes in Artificial Intelligence, Vol. 6804, ISBN 978-3-642-21915-3.
34. Lee C.-H., Shih J.-L., Yu K.-M., Lin H.-S. (2009), *Automatic music genre classification based on modulation spectral analysis of spectral analysis of spectral and cepstral features*, IEEE Transactions on Multimedia, **11**, 4, 670–682.
35. Li D., Sethi I.K., Dimitrova N., McGee T. (2001), *Classification of general audio data for content-based retrieval*, Pattern Recognition Letters, **22**, 5, 533–544.
36. Lindsay A., Herre J. (2011), *MPEG-7 and MPEG-7 audio – An overview*, AES Journal, **49**, 7–8, 589–594.
37. Logan B. (2000), *Mel frequency cepstral coefficients for music modeling*, Proceedings of 1st International Symposium on Music Information Retrieval (ISMIR00), pp. 5–11, Plymouth, MA, USA.
38. Mandelbrot B.B. (1967), *How long is the coast of Britain? Statistical self-similarity and fractional dimension*, Science, **156**, 636–638.
39. Mandelbrot B.B. (1982), *The fractal geometry of nature*, W.H. Freeman, Oxford.
40. Maragos P., Potamianos A. (1999), *Fractal dimensions of speech sounds: Computation and application to automatic speech recognition*, Journal of Acoustical Society of America, **105**, 3, 1925–1932.
41. McKinney M.F., Breebaart J. (2003), *Features for audio and music classification*, Proceedings of 4th International Symposium on Music Information Retrieval (ISMIR03), pp. 151–158, Baltimore, MD, USA.
42. Muda L., Begam M., Elamvazuthi I. (2010), *Voice recognition algorithms using mel frequency cepstral coefficient (MFCC) and dynamic time warping (DTW) techniques*, Journal of Computing, **2**, 3, 138–143.
43. Peitgen H.-O., Juergens H., Saupe D. (2004), *Chaos and fractals*, 2nd Ed, Springer.
44. Pitsikalis V., Maragos P. (2009), *Analysis and classification of speech signals by generalized fractal dimension features*, Speech Communication, **51**, 12, 1206–1223.
45. Rabiner L., Juang B-H. (1993), *Fundamentals of Speech Recognition*, Prentice-Hall.

46. Rein S., Reisslein M. (2006), *Identifying the classical music composition of an unknown performance with wavelet dispersion vector and neural nets*, Elsevier-Information Sciences, **176**, 12, 1629–1655.

47. Reljin I., Reljin B., Pavlović I., Rakočević I. (2000), *Multifractal analysis of gray-scale images*, Proceedings of 10th IEEE Mediterranean Electrotechnical Conference (MELECON-2000), pp. 490–493, Lemesos, Cyprus.

48. Reljin I., Reljin B., Avramov-Ivic M., Jovanovic D., Plavec G., Petrovic S., Bogdanovic G. (2008), *Multifractal analysis of the UV/VIS spectra of malignant ascites: Confirmation of the diagnostic validity of a clinically evaluated spectral analysis*, Physica A: Statistical Mechanics and its Applications, **387**, 14, 3563–3573.

49. Reljin N., Reyes B.A., Chon K.H. (2015), *Tidal volume estimation using the blanket fractal dimension of the tracheal sounds acquired by smartphone*, Sensors, **15**, 5, 9773–9790.

50. Rosner A., Schuller B., Kostek B. (2014), *Classification of music genres based on music separation into harmonic and drum components*, Archives of Acoustics, **39**, 4, 629–638.

51. Sabanal S., Nakagawa M. (1996), *The fractal properties of vocal sounds and their application in the speech recognition model*, Chaos, Solitons and Fractals, **7**, 11, 1825–1843.

52. Sedivy R., Mader R. (1997), *Fractals, chaos and cancer: Do they coincide?*, Cancer Investigation, **15**, 6, 601–607.

53. Schedl M., Flexer A., Urbano J. (2013), *The neglected user in music information retrieval research*, Journal of Intelligent Information Systems, **41**, 523–539.

54. Sheridan S. (2012), *The complete ABBA*, 2nd Ed, Titan Books, London, UK.

55. Slaney M. (1998), *Auditory toolbox, version 2, Technical report 1998-010*, Interval Research Corporation.

56. Stanley H.E., Meakin P. (1988), *Multifractal phenomena in physics and chemistry*, Nature, **335**, 405–409.

57. Stevens S.S., Volkmann J., Newman E.B. (1937), *A scale for the measurement of the psychological magnitude pitch*, Journal of the Acoustical Society of America, **8**, 3, 185–190.

58. Su Z.-Y., Wu T. (2006), *Multifractal analyses of music sequences*, Physica D: Nonlinear Phenomena, **221**, 2, 188–194.

59. Su Z.-Y., Wu T. (2007), *Music walk, fractal geometry in music*, Physica A: Statistical Mechanics and its Applications, **380**, 418–428.

60. Tan P.-N., Steinbach M., Kumar V. (2005), *Introduction to data mining*, Addison-Wesley, Upper Saddle River, NJ, USA.

61. Tsai W.-H., Wang H.-M. (2006), *Automatic singer recognition of popular music recordings via estimation and modeling of solo vocal signals*, IEEE Transactions on Audio, Speech, and Language Processing, **14**, 1, 330–341.

62. Tzanetakis G., Cook P. (2002), *Musical genre classification of audio signals*, IEEE Transactions on Speech and Audio Processing, **10**, 5, 293–302.

63. Vapnik V.N. (1998), *Statistical learning theory*, John Wiley & Sons, New York.

64. Vasiljevic J., Reljin B., Sopta J., Mijucic V., Tulic G., Reljin I. (2012), *Application of multifractal analysis on microscopic images in the classification of metastatic bone disease*, Biomedical Microdevices, **14**, 3, 541–548.

65. Véhel J.L., Mignot P. (1994), *Multifractal segmentation of images*, Fractals, **2**, 3, 379-382.

66. Véhel J.L. (1996), *Fractal approaches in signal processing*, [in:] *Fractal geometry and analysis: The Mandelbrot festschrift*, Evertsz C.J.G., Peitgen H.-O., Voss R.F. [Eds.], World Scientific.

67. Véhel J.L. (1998), *Introduction to the multifractal analysis of images*, Fractal Image Encoding and Analysis, **159**, 299–341.

68. Wold E., Blum T., Keislar D., Wheaton J. (1996), *Content-based classification, search, and retrieval of audio*, IEEE Multimedia, **3**, 3, 27–36.

69. Zlatintsi A., Maragos P. (2013), *Multiscale fractal analysis of musical instrument signals with application to recognition*, IEEE Transactions on Audio, Speech, and Language Processing, **21**, 4, 737–748.