

ORIGINAL RESEARCH

Unsupervised analysis of similarities between musicians and musical genres using spectrograms

Joe George, Lior Shamir*

Lawrence Technological University, Southfield, United States

Received: March 29, 2015
DOI: 10.5430/air.v4n2p61

Accepted: May 26, 2015
URL: <http://dx.doi.org/10.5430/air.v4n2p61>

Online Published: June 1, 2015

Abstract

Since the early days of the information era, digital music has been becoming one of the most consumed types of media, introducing the need for content-based tools that can search, browse, and retrieve music. Here we describe a method that can quantify similarities between musical genres in an unsupervised fashion, and computes networks of similarities between different musicians or musical styles. The method works by converting each song to its 2D spectrogram, and then extracting a large set of 2883 2D numerical content descriptors. The descriptors are weighted by their informativeness, and then the similarities between the musical styles are measured using the weighted distances between the musical pieces of each pair of musicians or genres. The similarities between all pairs provide a similarity matrix, which is visualized by a phylogeny. Experiments using 23 well known musicians representing seven musical genres show that the algorithm was able to separate the artists into groups that are in agreement with their respective musical genres. The analysis was done in an unsupervised fashion, and without any human definition or annotation of the musical styles.

Key Words: Music, Machine perception

1 Introduction

The increasing availability of music databases powered by the rapid advancement in information systems, computer networks, and the World Wide Web has reinforced the need for effective algorithms that can be used for content-based search, retrieval, and discovery of music. The demand for music retrieval software is driven by the needs of music consumers, as music is one of the most popular and influential forms of art, and the vast majority of modern music data is stored in digital format. The increased accessibility of digital media has led to dramatic changes in the music consumption culture compared to the pre-information era, where radio playlists and record stores had a dominant role in the music culture, and has been replaced by on-line shopping malls (e.g. Amazon, iTunes), search and indexing engines,

on-line repositories (e.g., YouTube), and social media.^[1] Other uses of music information retrieval systems include identification, plagiarism detection, unauthorized use of intellectual property, Motion Picture Association of America (MPAA) rating, or basic quantitative music research in the sense of the emerging field of the digital humanities.

On-line music stores and multimedia streaming services often organize their data in a hierarchical fashion by genres, but this type of categorization is based on the subjective perception of those who organize the data, and when a more detailed and accurate system of music information retrieval is desired it also introduces difficulties in maintaining hundreds of categories and sub-categories.^[2] Another approach to organizing music data is by user recommendation, analysis of purchasing patterns, or user like/dislike social me-

*Correspondence: Lior Shamir; Email: lshamir@mtu.edu; Address: Lawrence Technological University, Southfield, United States

dia input to predict the preferences of music consumers and make recommendations based on these patterns. Recently, substantial research work has been invested in contextual approaches to music, based on information related to music that is not contained in the audio such as musical scores,^[3] keywords, metadata, playlists, user ranking, and others,^[4] as opposed to methods based on the analysis of the audio content.^[5,6] Hybrid methods for music retrieval combine both audio and contextual information.^[7]

In their most basic form, sounds can be described by their pitch, loudness, and timbre, which includes the amplitude envelope, harmonicity and spectral envelope, and can be used for music classification and retrieval.^[8] Some studies suggested that the musical piece should be segmented based on its content to allow more efficient analysis, and experimental results using 10 melodies showed a good correlation between 40 human participants and music chunking algorithms.^[9]

The popularity of music has motivated the development of computer music analysis methods in the past several decades.^[10] In the sense of automatic music analysis using pattern recognition, significant efforts have been invested in content-based music retrieval and query by example,^[1,11] as well as related tasks such as cover song detection,^[12] version identification,^[13] automatic music recommendation,^[14,15] and sound quality prediction.^[16] Methods for content-based music retrieval include editing distances,^[17] alignment,^[18] minimum area between polynomial chains,^[19,20] n-grams,^[21,22] or hybrid modeling of music descriptors.^[23] Semi-automatic methods for music similarity using textual descriptors labeled manually for each musical piece such as “male voice” or “romantic piano” were able to generate playlists automatically after the labels for each musical piece were defined.^[24]

Numerous studies on automatic music analysis focused on music classification,^[25-30] often performed by automatic estimation of the similarity between two musical pieces.^[31] Classification of music can be accomplished by genre,^[32-36] emotion,^[37,38] musical instruments,^[39] composer,^[40] scene,^[41] melodic motifs,^[42] and more.^[43] Multi-purpose systems such as jMIR provide a framework for basic music classification that can be applied to a broad range of musical styles.^[44] Other directions of research include automatic annotation of music,^[45] and detection of difficult chords.^[46]

While in the machine learning domain substantial efforts were focused on classification, in music cognition studies methods such as multidimensional scaling (MDS) have been used to analyze similarities between musical excerpts and genres.^[47-53] Here we propose an unsupervised machine learning approach to music analysis based on complex morphological analysis of the audio content reflected by its 2D spectrogram, and use the scheme to measure similarities be-

tween musicians and musical genres. That is, the goal of the study is not to automatically classify musical pieces into a pre-defined set of classes representing musical genres, but to automatically deduce links between these classes and provide a network of similarities between these classes based on the content of their audio. This approach can be used for automatic music discoverability and content-based searching and browsing of music databases.

2 Music data

Numerous music datasets have been compiled and used, also for the purpose of automatic music classification competitions.^[54-58] However, the primary purpose of most of these datasets was music classification such as automatic classification of songs to musical genres, rather than automatic analysis of similarities between different musicians and different musical styles. For analyzing similarities between musicians we designed a music dataset that includes 23 well-known musicians representing seven basic popular music genres, which are heavy metal, reggae, electronic music, rap, jazz, country, and 60's Britpop. The musicians are Deep Purple, Iron Maiden, and Black Sabbath representing heavy metal, Bob Marley, Toots and the Maytals, and Jimmy Cliff representing the reggae genre, Johnny Cash and Merle Haggard representing country, Eminem, Snoop Dog, and 2Pac representing the rap musical genre, and The Beatles, Herman's Hermits, and The Kinks representing 60's britpop. Jazz was represented by Miles Davis, Louis Armstrong, and Duke Ellington, and the electronic music was represented in the dataset by Depeche Mode, Pet Shop Boys, DeVision, New Order, Faithless, and Aphex Twin. The criteria for selecting these musicians was the clear and strong association between the musicians and the musical genre (e.g., Iron Maiden and heavy metal).

A few exceptions included The Beatles, in which only pieces released before 1965 were used due to the significant change in their musical style during the second half of the 60's, in which The Beatles musical style was shifted to different musical genres such as rock and psychedelic rock, and included the use of musical instruments such as classical orchestras and string ensembles that were not typical for the 60's britpop sound. Another exception is The Kinks, for which only songs released during the 60's (the band's more significant era) were used, as the band changed its musical style significantly and attempted different genres in the following years, where the band received limited public attention before splitting in 1996. The music that the band produced during and after the 70's cannot be considered representative of 60's Britpop, and therefore was not included in the music dataset.

Each musician in the dataset was represented by 95 songs. The audio samples were extracted from MP3 files, which were downmixed from stereo to mono with SoX (Sound Ex-

change)^[59] “remix” command, and converted to WAV format. The length of each track was normalized to, a 60-second segment starting 30 seconds from the beginning of the track to skip a possible intro, a few seconds of silence, or music played by a single instrument. That is, each segment used in the experiment is a 60-second long segment from 00:30 to 01:30 of the original track. The normalization for length is required for the analysis of the spectrogram that will be described in Section 3.

Unlike datasets used for assessing the performance of genre classification algorithms, the dataset used in this study requires classes with higher similarity between them compared to the other classes, so that the ability of the algorithm to identify similarities between musical styles in an unsupervised fashion can be tested. For that reason each genre is represented by several different musicians who share similar musical styles. The ability of the algorithm to group these musicians automatically can be tested by its ability to identify that the musicians that share the same musical genres are indeed closer to each other compared to the other musicians in the dataset.

In addition to the datasets of the musicians, another dataset included studio tracks separated into 10 different basic musical genres: Classical music, country, swing, funk, blues, jazz, rock, alternative rock, heavy metal, and rap. The tracks used in that dataset were recorded by different musicians, in different studios, and at different times, so that the data cannot be biased by recording hardware or certain vocals or musical instruments.

Each of the 60-second music samples was transformed using SoX into a 800×512 2D monochrome spectrogram, which is a visual representation of the audio and provides precise information of the recording.^[60] The vertical dimension of the spectrogram corresponds to frequency, and the horizontal axis represents the time. The amplitude of a certain frequency at a certain time is visualized with shades of gray. Figure 1 shows a 60-second example spectrogram created from the song “Ain’t no love in the heart of the city” original version by Bobby Bland (1978).

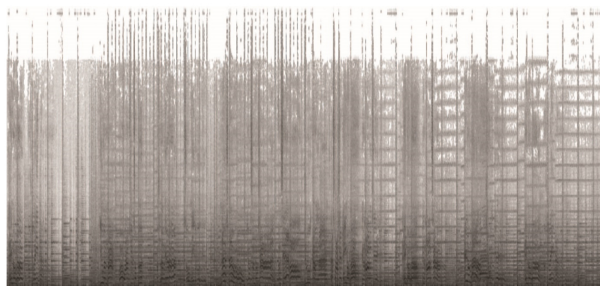


Figure 1: A spectrogram created from Bobby Bland’s “Ain’t no love in the heart of the city”

Analyzing audio using the spectrogram features has shown

to be effective in several different tasks related to music analysis.^[61–64] The two-dimensional representation of the data allows the application of the numerous mature multi-purpose image analysis algorithms proposed in the literature, providing more numerical content descriptors than descriptors extracted directly from the audio as will be described in this paper. In this study it allows the processing of the audio with the compound hierarchical algorithms that reflect complex morphology as will be described in Section 3.

3 Music analysis algorithm

The spectrograms were analyzed by the Wndchrm scheme,^[65–68] which quantifies numerous aspects of the two dimensional visual content. This type of analysis uses quantitative measurements of the visual content of the spectrograms such as textures,^[61] and it has been shown that numerical image content descriptors can be used to classify musical pieces into genres.^[63]

Wndchrm has been demonstrated to be effective in processing and identifying complex image morphology such as a broad range of microscopy and radiology images,^[69,70] as well as differences between artistic styles of paintings.^[71–73]

In audio analysis, it was used in combination with human “citizen scientists” in the WhaleFM project (part of Zooniverse) to analyze similarities between thousands of different whale calls, showing that whales have different dialects in different geographic locations.^[68] The observation was consistent for killer and pilot whales living in different geographical locations such as Norway, Iceland, and the Bahamas. In music, the method was applied to analyze the sound of popular music albums, and showed that in many cases the method could automatically sort albums in a fashion that is in high agreement with their chronological order.^[74]

Wndchrm is based on a large set of 2885 2D numerical content descriptors^[66] as described below:

- (1) Haralick texture features^[75] measuring the energy and entropy of the spectrogram’s co-occurrence matrix, and contribute 28 image descriptor values as described in.^[65]
- (2) Tamura texture features^[76] of contrast, directionality and coarseness, such that the coarseness 2D numerical content descriptors are the coarseness sum and its three-bin histogram, providing six 2D numerical content descriptors of the spectrogram.
- (3) Gabor Filters,^[78] with a convolution kernel of Gaussian harmonic function.^[79] Seven frequencies (1 through 7) are used, and each frequency adds a 2D numerical content descriptor.
- (4) Radon transform features,^[77] computed for angles 0, 45, 90, 135 degrees, and each of the resulting series

is then sampled into a 3-bin histogram, providing 12 image features.

- (5) Fractal features, based on box counting fractal analysis as thoroughly described in Ref.^[80]
- (6) Chebyshev statistics,^[81] computed using a 32-bin histogram of the 400-bin vector produced by the 2D Chebyshev transform of the spectrogram.
- (7) Multi-scale histograms computed on the pixel intensities of the spectrograms using different numbers of bins (3, 5, 7, and 9), as described in Ref.,^[82] providing $3+5+7+9=24$ 2D numerical content descriptors of the spectrogram.
- (8) First four moments (mean, standard deviation, skewness, kurtosis) computed on stripes of the spectrogram oriented in four different directions (0, 45, 90, 135 degrees). Each set of stripes is convolved into a three-bin histogram, providing $4 \times 4 \times 3=48$ 2D numerical audio content descriptors.^[66]
- (9) Edge features computed on the Prewitt gradient^[83] of the spectrogram, including the mean, median, variance, and eight-bin histogram of both the edge magnitude and the edge direction. Other edge features are the total number of edge pixels (normalized to the size of the spectrogram), the direction homogeneity, and the differences between direction histogram bins at a certain angle α and $\alpha + \phi$, convolved into a four-bin histogram.
- (10) High-contrast object statistics computed on all 8-connected 2D segments found in the Otsu binary mask of the spectrogram.^[84] Computed statistics include the Euler number,^[85] and the minimum, maximum, mean, median, variance, and a 10-bin histogram of both the objects areas and distances from the mass center of the spectrogram.
- (11) Zernike features are the absolute values of the coefficients of the Zernike polynomial approximation of the spectrogram,^[86] and providing 72 2D numerical audio content descriptors.
- (12) Chebyshev-Fourier features,^[87] implemented as a histogram with 32 bins sampling the polynomial coefficients of a Chebyshev-Fourier transform with highest polynomial order of $N=23$.

The features described above are computed from the original spectrogram, as well as several different 2D transforms of the music spectrograms and combinations of compound two-dimensional spectrogram transforms. The 2D transforms are FFT, Wavelet, Chebyshev transform, and a transform of the magnitude component of the edges. Further details about the 2D audio descriptors computed from 2D transforms of the spectrogram and compound spectrogram transforms is available in Ref.^[65-68, 70, 71]

The numerical content descriptors are computed such that the spectrogram is separated into 16 tiles, all have the same

size of 200×128 pixels. The set of numerical content descriptors described above is extracted from each tile independently as was done in Ref.^[66, 70] When a musical piece is assigned to the training or test set, all tiles of that song become part of that set, ensuring that tiles created from the same spectrogram will not be separated into both the test and training set, and therefore tiles that are part of the same musical piece cannot be matched with each other. The separation of each spectrogram to tiles allows analyzing different parts of the spectrogram separately rather than the entire spectrogram as a whole, providing more informative numerical content descriptors of each song.

The large feature set allows the algorithm to handle the complex nature of music data and reflect various aspects of the sound. However, not all of these 2D numerical content descriptors can be assumed to provide equally strong signal, and it is also expected that some of these descriptors make no or little contribution to the performance of the algorithm. To filter the least informative numerical audio descriptors, the values of each of the audio descriptors are normalized to the interval $[0, 1]$, and then a standard filtering approach is applied such that each 2D content descriptor is weighed by its power to distinguish between musical styles measured using its Fisher discriminant score,^[89] described by Equation 1,

$$W_f = \frac{\sum_{c=1}^N (\overline{T_f} - \overline{T_{f,c}})^2}{\sum_{c=1}^N \sigma_{f,c}^2} \quad (1)$$

where W_f is the computed Fisher discriminant score of feature f , $\overline{T_f}$ is the mean of the values of audio descriptor f in the entire training set, N is the number of musicians in the dataset, and $\overline{T_{f,c}}$ and $\sigma_{f,c}^2$ are the average and the variance of audio feature f among the training set samples of musician c . After the Fisher discriminants are computed, the 65% of the audio descriptors that have the lowest weights are removed from the analysis, providing a smaller set of 1009 spectrogram audio features. The optimal number of audio features used was determined experimentally as shown in Section 4.

After the weights of the numerical audio content descriptors described above are determined, the distance $d_{x,c}$ between a vector of numerical content descriptors x and a certain musician c is measured by Equation 2.

$$d_{x,c} = \frac{\sum_{t \in T_c} [\sum_{f=1}^{|x|} W_f (x_f - t_f)^2]^p}{|T_c|} \quad (2)$$

where T_c is the training set of musician c , t is a feature vector from T_c , $|x|$ is the size of the feature vector x , x_f is the value of audio descriptor f in the vector x , W_f is the weight of descriptor f computed by Equation 1, $|T_c|$ is the number of training samples of albums c , and p is the exponent, set to -5. The -5 exponent was determined experimentally.^[87]

The exponent -5 was deduced experimentally based on image data^[87] rather than spectrograms as was done in this study. However, that exponent has been shown to be optimal in a very broad range of image analysis problems ranging from face recognition^[87] to the analysis of microscopy images.^[69] The microscopy images contain mostly textures and low-level information rather than high-contrast features, and in that sense are more similar to the spectrograms. It was also tested with images of just blank textures, that do not contain any clear visible objects.^[65,90] It was also used for analyzing spectrograms of whale sounds.^[68]

Since the spectrogram of each song is divided into 16 tiles, the distance $D_{s,c}$ between song s to musician c is determined by the mean of the distances of the feature vectors $x_{1..16}$ to any of the other feature vectors of musician c in the training set as described by Equation 3.

$$D_{s,c} = \frac{\sum_{x \in S} d_{x,c}}{16} \quad (3)$$

The distance $D_{s,c}$ from the given track s (the mean distance of the 16 tiles of the track's spectrogram) can be used to determine the musician to which the track is predicted to belong. However, as mentioned in Section 1, the purpose of this study is not to classify songs to musicians or genres, but to deduce a network of similarities between the musical styles of different artists and identify musical genres in an unsupervised fashion. The similarity between two musicians c and q , each represented by a set of tracks, is determined by averaging the distance $D_{s,c}$ between all songs of a certain musician q in the test set and all songs of musician c in the training set as described by Equation 4.

$$M_{c,q} = \frac{\sum_{s \in q} D_{s,c}}{|q|} \quad (4)$$

The same task can be repeated for each pair of musicians in the dataset, providing a matrix that its cell values are the dissimilarities between the pairs of artists. That is, the value of the cell c, q in the dissimilarity matrix M is the computed distance between musician c and musician q . The dissimilarity matrix is then inverted into a matrix of similarities and normalized such that the cell $M_{c,r}$ in any column r is divided by the value of the cell $M_{r,r}$. Conceptually, that means that the similarity of each musician to all other musicians is divided by the self-similarity of that musician, so that the computed similarity of a certain musician to itself is always 1, and the computed similarity of that musician to all other values are greater than 0. Then, the similarity matrix is visualized using phylogenies by the Phylip package,^[91,92] used with randomize input order of sequences where 97 is the seed, 10 jumbles, and Equal-Daylight arc optimization.^[92]

The high-level algorithms of the computing of the feature weights through training and the generation of the phyloge-

gies are summarized by Algorithms 1 and 2, respectively.

Algorithm 1. Computing feature weights

```

For each musician c ∈ C
  For each audio sample a ∈ c
    Convert a to spectrogram s
    For i=1:4
      For j=1:4
        tile t ← s(i,j)
        Compute feature vector v from t
        Add vector v to matrix V
      End for
    End for
  End for
End for
For each column F of matrix V
  Wf ← Fisher (F)

```

Algorithm 2. Generating the phylogeny

```

For each musician c ∈ C
  For each audio sample s in class c
    For each tile t ∈ s
      For each musician r ∈ C
        Mc,r ← Mc,r+weighted_distance(t,r,W)/16
      End for
    End for
  End for
  Mc,r ← Mc,r/|c|
  Convert Mc,r to a phylogeny using phylip

```

4 Experimental results

The method described in Section 3 was applied to the music datasets described in Section 2. The first experiment was based on the dataset of 10 genres such that 40 songs from each of the seven genres were used for training and seven songs from each genre for testing. The accuracy by which the algorithm was able to associate a song with the genres was ~ 71%, which is significantly higher than 10% of mere chance accuracy, and therefore shows that the algorithm is informative in analyzing music data. The classification accuracy is not a goal in itself, but is used as a probe and indication that the algorithm is informative. As discussed in Section 1, the purpose of this study is not to classify song to genres or musicians, but to automatically identify and analyze the similarities between musical styles in an unsupervised fashion. The phylogeny that visualizes the similarities between the genres is displayed by Figure 2. A longer line between two genres on the graph reflects higher dissimilarity between these genres, so that genres placed on the graph close to each other are genres that are more similar to each

other.

As the figure shows, the classical music genre was positioned in the upper part of the phylogeny, with significant distance from the 20th century genres. After the classical music the algorithm positioned the jazz, swing, funk, blues, and country musical genres that developed during the first half of the 20th century. Further down in the phylogeny, the algorithm positioned genres developed primarily in the second half of the 20th century such as heavy metal, rap, rock and alternative, musical genres that were influenced from the musical styles of the early 20th century. The algorithm also placed the rock and alternative genres very close to each other, indicating that the algorithm could identify that these two genres are closely related, as widely agreed by music experts.^[93] The rap genre is also placed at the bottom of the phylogeny with the genres of the second half of the 20th century, but it is separated from the heavy metal, rock and alternative styles, an indication of the noticeable difference between these styles.

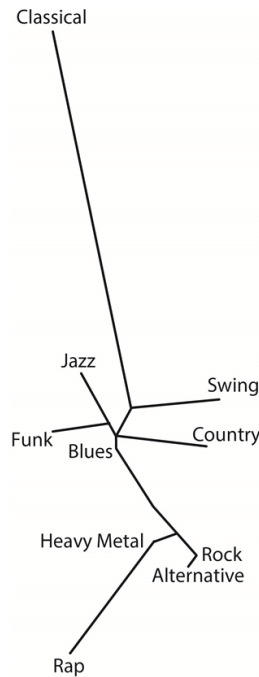


Figure 2: Automatically generated phylogeny of the 10 genres of music

The graph generated automatically by the method is in agreement with the human cognition and perception of music, as was determined experimentally by humans using multi-dimensional scaling.^[47] Like the music similarity analysis done by the algorithm, the classical music was separated by the human listeners from the other genres, indicating that like the proposed computer method, human listeners believed that classical music is substantially different from all other tested musical genres. The rap, rock, and heavy metal genres were placed close to each other, and the

funk, blues, jazz and country were also clustered.^[47] Overall, the phylogeny generated by the algorithm was similar to the similarity between musical genre annotated by humans, as described in Ref.^[47] A noticeable difference between the human and the computational analysis is the rap genre, which was placed close to the heavy metal, rock and alternative musical genres, but was not as close as these other three musical genres are close to each other. The human cognition experiment found that rap was perceived by humans to be similar to the heavy metal and rock genres, and clustered these three genres together.^[47]

In the second and more comprehensive experiment, the dataset of 23 musicians representing different musical genres was used. Ten songs from each musician were used for testing, and the remaining songs were used for training such that each musician had 85 songs in its training set. The experiment was repeated 40 times such that in each run different songs were randomly allocated to training and test sets, and the similarity matrices of the 40 runs were averaged into one similarity matrix that was used to create the phylogeny.

The classification accuracy of songs to the correct musician was ~81%, which is clearly far higher than ~4.3% of random guessing. However, the purpose of the proposed music analysis method is not to classify songs automatically by the musician, but to measure similarities between the musical styles of different artists. It should also be noted that the music dataset included more than one musician from each genre, so that songs from the same genre can add confusion to the automatic classification (e.g., Eminem songs can be associated with Snoop Dog, who shares the same genre and similar musical styles).

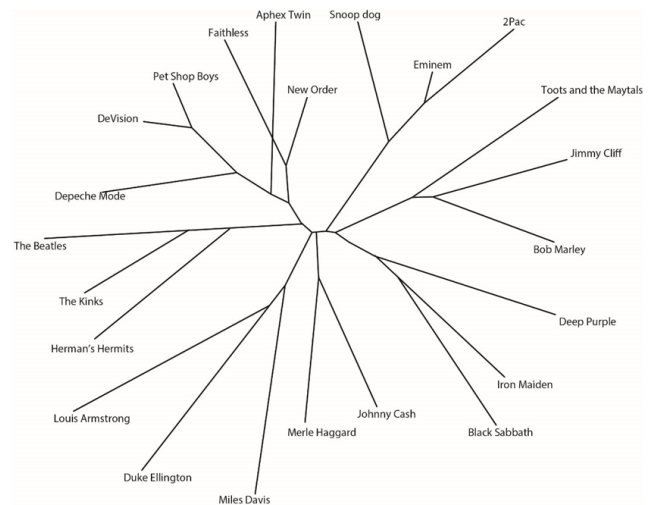


Figure 3: Automatically generated phylogeny of the similarities between the musical styles of the musicians based on analysis of the spectrograms of their songs. Artists of the same musical genres are clustered close to each other, on the same branch of the phylogeny.

Figure 3 displays the phylogeny of the musicians in the dataset described in Section 2. Longer paths between two musicians indicate that the musical styles of these two musicians are more similar compared to pairs of musicians with longer paths between them. As the figure shows, the heavy metal bands Deep Purple, Iron Maiden, and Black Sabbath are grouped on a single branch, while The Beatles (songs released before or during 1965), Herman’s Hermits, and The Kinks are clustered in another branch in the phylogeny. The same is the case with country music, reggae, rap, electronic music, and jazz. The musicians that represent each genre are located on a single branch in the phylogeny, which differentiates them from the other musical genres. That is, the method was able to separate the musicians into musical genres in an unsupervised fashion by analyzing the audio, and without using any human annotation of the musicians or other pre-defined knowledge that is not the content of the audio.

these numerical content descriptors are equally informative, and not all of them affect the analysis equally since the descriptors are weighted automatically by their informativeness, as evaluated by their Fisher discriminant scores^[89] described in Section 3. Figure 4 shows the sum of the Fisher discriminant scores of each group of features. The effect of each group of features is measured by the sum of the Fisher discriminant scores of all bins of that group, which directly reflects the impact of the group of features on the computed similarities between the different musicians, as explained in Section 3. For instance, the impact of the multi-scale histograms computed from the raw values of the spectrogram is the sum of the Fisher discriminant scores of the 24 bins of that feature.

As the graph shows, numerous 2D numerical content descriptors contribute to the analysis. These features are computed not merely from the raw pixel values of the 2D spectrograms, but also from transforms and multi-order transforms of the spectrograms. In particular, the Zernike polynomial features extracted from the Fourier transform of the spectrogram, the Fourier transform of the Wavelet (Symlet 5) transform, and the Fourier transform of the Edge magnitude transform were assigned with high Fisher discriminant scores and therefore had a high impact on the resulting phylogeny. Since these features are extracted from the transforms of the spectrograms, their ability to numerically reflect audio content is determined empirically by the Fisher discriminant scores. Other features that impacted the analysis are the Haralick texture features computed from the raw pixel values of the spectrograms, as well as several transforms and multi-order transforms.

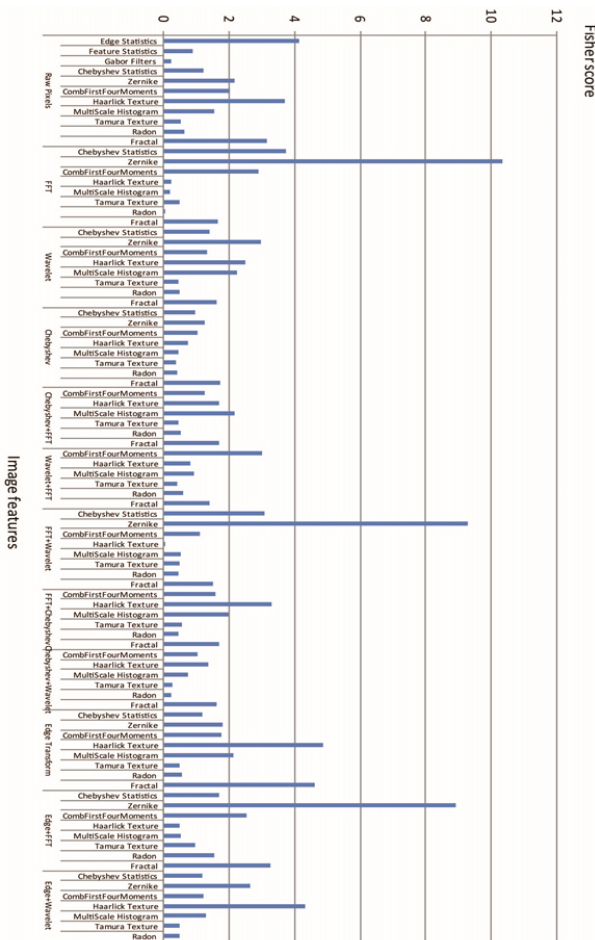


Figure 4: The fisher discriminant scores of the group of features used for analyzing the spectrograms

Each spectrogram is analyzed by extracting a very large and comprehensive set of 2D numerical content descriptors described in Section 3. As explained in Section 3, not all of

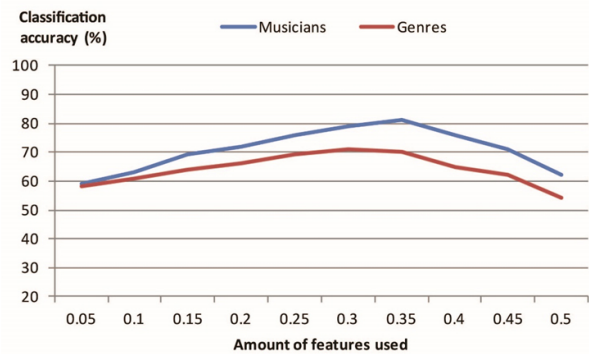


Figure 5: The classification accuracy of songs to musicians and songs to musical genres as a function of the amount of features used. Although classification is not the goal of this study, the classification accuracy can be used as a probe for adjusting the amount of features such that the system is informative.

The comprehensive analysis of the spectrograms is done using a very large set of numerical 2D content descriptors, weighted by their informativeness as described in Section 3. Since the feature set is large, many of the content de-

scriptors are not informative, and determining the optimal number of features was done empirically by using the classification accuracy as a probe for the informativeness of the feature set. Figure 5 shows how the classification accuracy of songs to musicians and songs to genres changes as the amount of features used increases.

As the figure shows, the classification accuracy increases as more features are used, but then starts to decrease when using more than 35% of the numerical spectrogram descriptors. Achieving high classification accuracy is not the goal of this study, but since the ability to correctly classify musical pieces indicates that the system can effectively analyze music, that simple measurement can be used as a probe for the informativeness of the system, and consequently its ability to cluster music classes by their similarities.

Comparison of the performance to classical general-purpose audio features

One of the key aspects of the proposed method is the use of numerical content descriptors computed from the spectrograms that were computed from the audio files. jAudio is a signal processing library that extracts a comprehensive set of audio features measuring different characteristics of the analyzed music sample. These include the area moments, beat histograms, zero crossing, spectral properties, harmonic properties, 1D and 2D moments, Mel-Frequency Cepstral Coefficients, and more,^[44] as described in the jAudio on-line documentation at <http://jaudio.sourceforge.net/jaudio10/features/feature.html>. jAudio computes a total number of 78 features.

The experiment was done in the same fashion described in Section 3, but the features were the 78 jAudio features instead of the spectrogram content descriptors. Repeating the experiment described above using the numerical audio content descriptors of jAudio provided the phylogeny displayed by Figure 6.

As the graph shows, some of the artists were grouped by the algorithm in a fashion that is in agreement with their musical genres. For instance, the rappers were all clustered on a single branch, as well as the three reggae musicians, three jazz musicians, and two country singers that were also grouped together in a separate branch in the phylogeny. However, the method failed to separate between the heavy metal bands, the 60's britpop bands, and the electronic music artists who are all positioned together in the upper part of the phylogeny and are not grouped in separate branches as the jazz, reggae, rap, and country musicians. These results show that the analysis of the spectrograms provides a more accurate analysis of the musical styles compared to the comprehensive set of audio descriptors of the jAudio package. The automatic classification accuracy of songs to musicians using the jAudio features was $\sim 47\%$, which is higher than $\sim 4.3\%$ of random guessing, but also substantially lower than the

classification accuracy when the spectrograms are used.

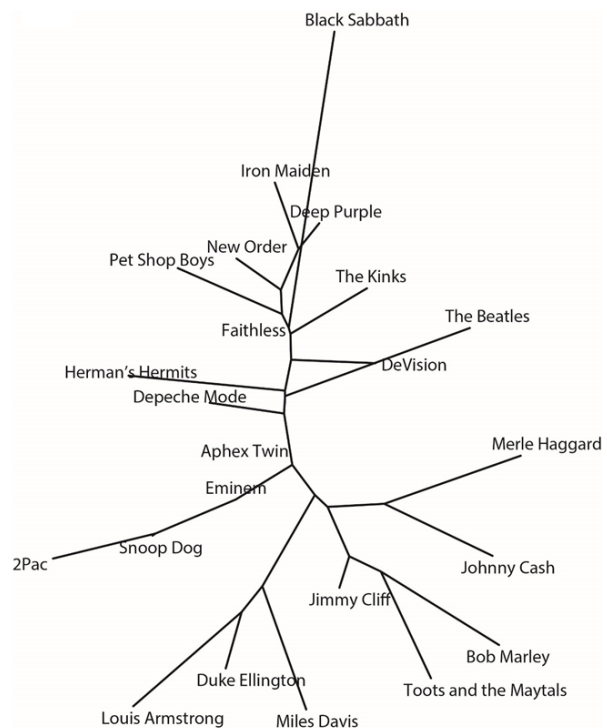


Figure 6: Automatically generated phylogeny of the similarities between the musical styles of the musicians based on analysis of the spectrograms of their songs

A downside of using 2D spectrograms compared to tools such as jAudio that extract numerical descriptors directly from the music files is computational complexity. While the generation of spectrograms from the audio files is performed very quickly and adds merely a negligible delay to the entire process, computing the features from the spectrograms for all 16 tiles requires about nine minutes using one core of an Intel Core-i7 processor. However, the algorithm can be parallelized with negligible overhead,^[65] what effectively accelerates the process by using multiple cores. The analysis of the music dataset described in Section 2 was performed with a computing cluster of 32 Intel Core-i7 cores, and a cluster of 320 AMD Opteron cores. That cluster is able to compute 1000 songs in less than 30 minutes.

Another computational bottleneck is the creation of the phylogenies, which can also become a computationally intensive task due to its complex grouping of sequences into branches (namely, the *fitch* program, with computational complexity of $O(n^5)$), as well as the intensive creation of trees using random bootstrapped data.^[95] Generating the phylogeny with the 23 nodes of Figure 3 takes just several seconds, but the response time increases quickly as the size of the tree grows. Creating a phylogeny with 100 nodes requires almost two hours of computing using a single worker, but in the case of a large number of nodes the task can

be completed in a much shorter time using multiple workers.^[95]

5 Conclusions

While perceiving music is a complex cognitive task that introduces a challenge for computing machines, automatic analysis of music has been becoming an increasingly important need for music-related information systems. While machines have demonstrated good performance in supervised machine learning tasks such as automatic association of a song with its musical genre, in this paper we addressed the problem of unsupervised machine learning of music, and showed that machines can identify musical genres automatically based on the audio content alone, and without using any pre-defined annotation of the musical pieces. The analysis is unsupervised in the sense that the system does not have any information on the association of the musicians to genres, and in fact does not have information about the existence of genres in the dataset. The method computes the similarities between the musicians, and then associates the musicians by the computed similarities without using any knowledge about the genre each of the musicians is associated with.

The selection of artists who have strong association with a certain musical genre allows to evaluate the performance of the system. On the other hand, the same selection also allowed the algorithm to make a clear separation between the genres, and did not analyze cross-genre musicians that could challenge the algorithm by analyzing musical pieces that do not have a strong and clear association to a specific musical

genre.

The results show that the very large set of numerical content descriptors extracted from the spectrogram transformation of the audio files is more informative for the purpose of unsupervised analysis of musical genres compared to using numerical audio descriptors extracted directly from the audio files. Clearly, the experiments performed in this study are data-driven, and is based on the contention that the spectrogram contains the same information as the original audio file, but structured in a different fashion.

The spectrogram analysis, however, is far more computationally intensive, and requires significant computing resources. Fortunately, computing the numerical content descriptors of the 2D spectrograms can be parallelized effectively,^[66] and each core can process a different spectrogram file so that a medium-sized computing cluster can process the entire audio dataset in less than two hours.

The analysis of unsupervised musical styles can be used for organizing and managing large music databases, as well as music discoverability and browsing of musical databases based on their audio content and musical styles. Such methods are required to satisfy the growing need to manage and navigate efficiently inside large databases of music, which is currently one of the most popular and most consumed types of digital data.

Acknowledgements

This work was supported in part by grant 1157162 of the National Science Foundation. We would also like to thank Dr. Franco Delogu for the fruitful discussions.

References

- [1] Casey MA, Veltkamp R, Goto M, et al. Content-based music information retrieval: Current directions and future challenges. *Proceedings of the IEEE*. 2008; 96(4): 668-696. <http://dx.doi.org/10.1109/JPROC.2008.916370>
- [2] Pachet F, Cazaly D. A taxonomy of musical genres. In: 6th International Conference on Computer-Assisted Information Retrieval, 2000. p. 1238-1245.
- [3] Rebelo A, Fujinaga I, Paszkiewicz F, et al. Optical music recognition: state-of-the-art and open issues. *International Journal of Multimedia Information Retrieval*. 2012; 1(3): 173-190. <http://dx.doi.org/10.1007/s13735-012-0004-6>
- [4] Knees P, Schedl M. A survey of music similarity and recommendation from music context data. *ACM Trans Multimedia Comput Commun Appl*. 2013; 10(1): 1-2, 21. <http://dx.doi.org/10.1145/2542205.2542206>
- [5] Pampalk E, Rauber A, Merkl D. Content-based organization and visualization of music archives. In: *Proceedings of the tenth ACM International Conference on Multimedia*. 2002. p. 570-579. <http://dx.doi.org/10.1145/641007.641121>
- [6] Downie JS. Music information retrieval. *Annual review of information science and technology*. 2003; 37(1): 295-340. <http://dx.doi.org/10.1002/aris.1440370108>
- [7] Knees P, Schedl M, Celma O. Hybrid music information retrieval. *International Journal of Multimedia Information Retrieval*. 2013; 2(1): 1-2. <http://dx.doi.org/10.1007/s13735-013-0033-9>
- [8] Wold E, Blum T, Keislar D, et al. Content-based classification, search, and retrieval of audio. *IEEE Transactions on Multimedia*. 1996; 3(3): 27-36. <http://dx.doi.org/10.1109/93.556537>
- [9] de Nooijer J, Wiering F, Volk A, et al. Cognition-based segmentation for music information retrieval systems. Master's thesis, Utrecht University. 2007.
- [10] Alphonse BH. Music analysis by computer: A field for theory formation. *Computer Music Journal*. 1980: 26-35.
- [11] Li T, Ogihara M. Toward intelligent music information retrieval. *IEEE Transactions on Multimedia*. 2006; 8(3): 564-574. <http://dx.doi.org/10.1109/TMM.2006.870730>
- [12] Jensen JH, Christensen MG, Ellis DP, et al. A tempo-insensitive distance measure for cover song identification based on chroma features. In: *IEEE international conference on acoustics, speech and signal processing*. 2008; 2209-2212. <http://dx.doi.org/10.1109/ICASSP.2008.4518083>
- [13] Salamon J, Serra J, Gomez E. Tonal representations for music retrieval: from version identification to query-by-humming. *International Journal of Multimedia Information Retrieval*. 2013; 2(1): 45-58. <http://dx.doi.org/10.1007/s13735-012-0026-0>

- [14] Celma O. Music recommendation and discovery: The long tail, long fail, and long play in the digital music space. Springer New York, NY; 2010.
- [15] Braunhofer M, Kaminskas M, Ricci F. Location-aware music recommendation. *International Journal of Multimedia Information Retrieval*. 2013; 2(1): 31–44. <http://dx.doi.org/10.1007/s13735-012-0032-2>
- [16] Manders AJ, Simpson DM, Bell SL. Objective prediction of the sound quality of music processed by an adaptive feedback canceller. *IEEE Transactions on Audio, Speech, and Language Processing*. 2012; 20(6): 1734–45. <http://dx.doi.org/10.1109/TASL.2012.2188513>
- [17] Mongeau M, Sankoff D. Comparison of musical sequences. *Computers and the Humanities*. 1990; 24(3): 161–175. <http://dx.doi.org/10.1007/BF00117340>
- [18] Hanna P, Robine M, Ferraro P, et al. Improvements of alignment algorithms for polyphonic music retrieval. In: *Proceedings of the International Computer Music Modeling and Retrieval Conference (CMMR)*; 2008. p. 244–251.
- [19] Typke R, Veltkamp RC, Wiering F. Searching notated polyphonic music using transportation distances. In: *Proceedings of the 12th annual ACM International Conference on Multimedia*. 2004. p. 128–135.
- [20] Clifford R, Christodoulakis M, Crawford T, et al. A fast, randomised, maximal subset matching algorithm for document-level music retrieval. In: *International Conference on Music Information Retrieval*; 2006. p. 150–155.
- [21] Uitdenbogerd A, Zobel J. Melodic matching techniques for large music databases. In: *IEEE/ACM International Conference on Multimedia*; 1999. p. 57–66.
- [22] Bainbridge D, Dewsnip M, Witten IH. Searching digital music libraries. *Information Processing & Management*. 2005; 41(1): 41–56. http://dx.doi.org/10.1007/3-540-36227-4_13
- [23] Zhang B, Shen J, Xiang Q, et al. Compositemap: a novel framework for music similarity measure. In: *Proceedings of the 32nd International Conference on Research and Development in Information Retrieval*; 2009. p. 403–410.
- [24] Aucouturier JJ, Pachet F. Music similarity measures: What's the use? In: *International Symposium on Music Information Retrieval*; 2002.
- [25] Fu Z, Lu G, Ting KM, et al. On feature combination for music classification. In: *Structural, syntactic, and statistical pattern recognition*. Springer; 2010. p. 453–462.
- [26] Bergstra J, Casagrande N, Erhan D, et al. Aggregate features and adaboost for music classification. *Machine Learning*. 2006; 65(2-3): 473–484. <http://dx.doi.org/10.1007/s10994-006-9019-7>
- [27] Tzanetakis G, Cook P. Musical genre classification of audio signals. *IEEE transactions on Speech and Audio Processing*. 2002; 10(5): 293–302. <http://dx.doi.org/10.1109/TSA.2002.800560>
- [28] Xu C, Maddage M, Shao X. Automatic music classification and summarization. *IEEE Transactions on Speech and Audio Processing*. 2005; 13(3): 441–450. <http://dx.doi.org/10.1109/TSA.2004.840939>
- [29] Fu Z, Lu G, Ting KM, et al. A survey of audio-based music classification and annotation. *IEEE Transactions on Multimedia*. 2011; 13(2): 303–319. <http://dx.doi.org/10.1109/TMM.2010.2098858>
- [30] Weihs C, Ligges U, Morchen F, et al. Classification in music research. *Advances in Data Analysis and Classification*. 2007; 1(3): 255–291. <http://dx.doi.org/10.1007/s11634-007-0016-x>
- [31] Berenzweig A, Logan B, Ellis DP, et al. A large-scale evaluation of acoustic and subjective music-similarity measures. *Computer Music Journal*. 2004; 28(2): 63–76. <http://dx.doi.org/10.1162/014892604323112257>
- [32] Li T, Ogihara M, Li Q. A comparative study on content-based music genre classification. In: *Proceedings of the 26th Annual International Conference on Research and Development in Information Retrieval*. ACM; 2003. p. 282–289.
- [33] Barreira L, Cavaco S, da Silva JF. Unsupervised music genre classification with a model-based approach. In: *Progress in artificial intelligence*. Springer; 2011. p. 268–281. http://dx.doi.org/10.1007/978-3-642-24769-9_20
- [34] Cataltepe Z, Yaslan Y, Sonmez A. Music genre classification using MIDI and audio features. *EURASIP Journal on Applied Signal Processing*. 2007; 2007(1): 150–150. <http://dx.doi.org/10.1155/2007/36409>
- [35] Mostafa MM, Billor N. Recognition of western style musical genres using machine learning techniques. *Expert Systems with Applications*. 2009; 36(8): 11378–11389. <http://dx.doi.org/doi:10.1016/j.eswa.2009.03.050>
- [36] Huang YF, Lin SM, Wu HY, et al. Music genre classification based on local feature selection using a self-adaptive harmony search algorithm. *Data & Knowledge Engineering*. 2014; 92: 60–76. <http://dx.doi.org/10.1016/j.datak.2014.07.005>
- [37] Yang YH, Lin YC, Cheng HT, et al. Toward multi-modal music emotion classification. In: *Advances in Multimedia Information Processing 2008*. Springer; 2008. p. 70–79.
- [38] Mokhsin MB, Rosli NB, Adnan WAW, et al. Automatic music emotion classification using artificial neural network based on vocal and instrumental sound timbres. *Frontiers in Artificial Intelligence and Applications*. 2014; 10(12): 2584–2592. <http://dx.doi.org/10.3233/978-1-61499-434-3-3>
- [39] Herrera-Boyer P, Peeters G, Dubnov S. Automatic classification of musical instrument sounds. *Journal of New Music Research*. 2003; 32(1): 3–21. <http://dx.doi.org/10.1076/jnmr.32.1.3.16798>
- [40] Cataltepe Z, Sonmez A, Adali E. Music classification using kolmogorov distance. In: *Representation in Music/Musical Representation Congress*; 2005.
- [41] Liem CC, Larson M, Hanjalic A. When music makes a scene. *International Journal of Multimedia Information Retrieval*. 2013; 2(1): 15–30. <http://dx.doi.org/10.1007/s13735-012-0031-3>
- [42] Rao P, Ross JC, Ganguli KK, et al. Classification of melodic motifs in raga music with time-series matching. *Journal of New Music Research*. 2014; 43(1): 115–131. <http://dx.doi.org/10.1080/09298215.2013.873470>
- [43] Kostek B. Perception-based data processing in acoustics: applications to music information retrieval and psychophysiology of hearing. Springer; 2005;3.
- [44] McKay C. Automatic music classification with jMIR. [Ph.D. thesis]. McGill University; 2010.
- [45] Duggan B. Machine annotation of traditional Irish dance music. [Ph.D. thesis]. Dublin Institute of Technology; 2009.
- [46] Mauch M, Dixon S. Approximate note transcription for the improved identification of difficult chords. In: *Int. Conf. on Music Information Retrieval (ISMIR)*; 2010. p. 135–140.
- [47] Novello A, McKinney MF, Kohlrausch A. Perceptual evaluation of music similarity. In: *Int. Conf. on Music Information Retrieval (ISMIR)*; 2006. p. 246–249.
- [48] Bigand E, Vieillard S, Madurell F, et al. Multidimensional scaling of emotional responses to music: The effect of musical expertise and of the duration of the excerpts. *Cognition & Emotion*. 2005; 19(8): 1113–1139. <http://dx.doi.org/10.1080/0269930500204250>
- [49] Samson S, Zatorre RJ, Ramsay JO. Deficits of musical timbre perception after unilateral temporal-lobe lesion revealed with multidimensional scaling. *Brain*. 2002; 125(3): 511–523. <http://dx.doi.org/10.1093/brain/awf051>
- [50] Wedin L. A multidimensional study of perceptual-emotional qualities in music. *Scandinavian Journal of Psychology*. 1972; 13(1): 241–257. <http://dx.doi.org/10.1111/j.1467-9450.1972.tb00072.x>
- [51] Grey JM. Multidimensional perceptual scaling of musical timbres. *The Journal of the Acoustical Society of America*. 1977; 61(5): 1270–1277. <http://dx.doi.org/10.1121/1.381428>
- [52] Spreckelmeyer KN, Altenmüller E, Colonius H, et al. Preattentive processing of emotional musical tones: a multidimensional scaling

- and ERP study. *Frontiers in psychology*. 2013; 4. <http://dx.doi.org/10.3389/fpsyg.2013.00656>
- [53] Saari P, Barthet M, Fazekas G, et al. Semantic models of musical mood: Comparison between crowd-sourced and curated editorial tags. In: *IEEE International Conference on Multimedia and Expo Workshops (ICMEW)*. IEEE; 2013. p. 1–6.
- [54] Downie J, West K, Ehmann A, et al. The 2005 music information retrieval evaluation exchange (MIREX 2005): Preliminary overview. In: *6th Int. Conf. on Music Information Retrieval (ISMIR)*; 2005. p. 320–323.
- [55] Downie JS, Byrd D, Crawford T. Ten years of ISMIR: Reflections on challenges and opportunities. In: *Conference of the International Society for Music Information Retrieval*; 2009. p. 13–18.
- [56] Goto M, Hashiguchi H, Nishimura T, et al. RWC music database: Popular, classical and jazz music databases. In: *Conference of the International Society for Music Information Retrieval*. 2002; 2: 287–288.
- [57] Scaringella N, Zoia G, Mlynek D. Automatic genre classification of music content: a survey. *Signal Processing Magazine*. 2006; 23(2): 133–141. <http://dx.doi.org/10.1109/MSP.2006.1598089>
- [58] Hu X, Downie JS, Laurier C, et al. The 2007 MIREX audio mood classification task: Lessons learned. In: *International Society for Music Information Retrieval*; 2008. p. 462–467.
- [59] Bagwell C. Sox-sound exchange. 2010.
- [60] Altes RA. Detection, estimation, and classification with spectrograms. *The Journal of the Acoustical Society of America*. 1980; 67(4): 1232–1246. <http://dx.doi.org/10.1121/1.384165>
- [61] Deshpande H, Singh R, Nam U. Classification of music signals in the visual domain. In: *Proceedings of the Conference on Digital Audio Effects*. SN; 2001. p. 1–4.
- [62] Nakano M, Le Roux J, Kameoka H, et al. Nonnegative matrix factorization with markov-chained bases for modeling time-varying patterns in music spectrograms. In: *Latent variable analysis and signal separation*. Springer; 2010. p. 149–156.
- [63] Costa YM, Oliveira LS, Koerich A, et al. Music genre recognition using spectrograms. In: *18th International Conference on Systems, Signals and Image Processing (IWSSIP)*; 2011. p. 1–4.
- [64] Chou CH, Liao BJ. Music genre classification by analyzing the subband spectrogram. In: *international conference on Information Science, Electronics and Electrical Engineering (ISEEE)*. 2014; 3: 1677–1680.
- [65] Shamir L. Evaluation of face datasets as tools for assessing the performance of face recognition methods. *International Journal of Computer Vision*. 2008; 79(3): 225–230. <http://dx.doi.org/10.1007/s11263-008-0143-7>
- [66] Shamir L, Orlov N, Eckley DM, et al. Wndchrm an open source utility for biological image analysis. *Source Code for Biology and Medicine*. 2008; 3: 13. <http://dx.doi.org/10.1186/1751-0473-3-13>
- [67] Shamir L, Ling SM, Scott WW, et al. Knee x-ray image analysis method for automated detection of osteoarthritis. *IEEE Transactions on Biomedical Engineering*. 2009; 56(2): 407–415. <http://dx.doi.org/10.1109/TBME.2008.2006025>
- [68] Shamir L, Yerby C, Simpson R, et al. Classification of large acoustic datasets using machine learning and crowdsourcing: Application to whale calls. *The Journal of the Acoustical Society of America*. 2014; 135(2): 953–962. <http://dx.doi.org/10.1121/1.4861348>
- [69] Shamir L, Orlov N, Eckley DM, et al. IICBU 2008: a proposed benchmark suite for biological image analysis. *Medical & biological engineering & computing*. 2008; 46(9): 943–947. <http://dx.doi.org/10.1007/s11517-008-0380-5>
- [70] Shamir L, Macura T, Orlov N, et al. Impressionism, expressionism, surrealism: Automated recognition of painters and schools of art. *ACM Transactions on Applied Perception (TAP)*. 2010; 7(2): 8. <http://dx.doi.org/10.1145/1670671.1670672>
- [71] Shamir L, Delaney JD, Orlov N, et al. Pattern recognition software and techniques for biological image analysis. *PLoS Computational Biology*. 2010; 6(11): e1000974. <http://dx.doi.org/10.1371/journal.pcbi.1000974>
- [72] Shamir L. Computer analysis reveals similarities between the artistic styles of Van Gogh and Pollock. *Leonardo*. 2012; 45(2): 149–154. http://dx.doi.org/10.1162/LEON_a_00281
- [73] Shamir L, Tarakhovskiy JA. Computer analysis of art. *ACM Journal on Computing and Cultural Heritage (JOCCH)*. 2012; 5(2): 7. <http://dx.doi.org/10.1145/2307723.2307726>
- [74] George J, Shamir L. Computer analysis of similarities between albums in popular music. *Pattern Recognition Letters*. 2014; 45: 78–84. <http://dx.doi.org/10.1016/j.patrec.2014.02.021>
- [75] Haralick RM, Shanmugam K, Dinstein IH. Textural features for image classification. *IEEE Transactions on Systems, Man and Cybernetics*. 1973; (6): 610–621.
- [76] Tamura H, Mori S, Yamawaki T. Textural features corresponding to visual perception. *IEEE Transactions on Systems, Man and Cybernetics*. 1978; 8(6): 460–473. <http://dx.doi.org/10.1109/TSMC.1978.4309999>
- [77] Lim JS. *Two-dimensional signal and image processing*. Englewood Cliffs, NJ, Prentice Hall; 1990. 710 p. .
- [78] Gabor D. Theory of communication. *Journal of the IEEE*. 1946; 93(26): 429–441.
- [79] Grigorescu SE, Petkov N, Kruizinga P. Comparison of texture features based on Gabor filters. *IEEE Transactions on Image Processing*. 2002; 11(10): 1160–1167. <http://dx.doi.org/10.1109/TIP.2002.804262>
- [80] Wu CM, Chen YC, Hsieh KS. Texture features for classification of ultrasonic liver images. *IEEE Transactions on Medical Imaging*. 1992; 11(2): 141–152. <http://dx.doi.org/10.1109/42.141636>
- [81] Gradsteyn I, Ryzhik IM. *Alan jeffrey: Table of integrals, series and products*. 1994.
- [82] Hadjidemetriou E, Grossberg MD, Nayar SK. Spatial information in multiresolution histograms. In: *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2001; 1: 1–702. <http://dx.doi.org/10.1109/CVPR.2001.990544>
- [83] Prewitt JM. Object enhancement and extraction. *Picture processing and Psychopictorics*. 1970; 10(1): 15–19.
- [84] Otsu N. A threshold selection method from gray-level histograms. *Automatica*. 1975; 11(285-296): 23–27.
- [85] Gray SB. Local properties of binary images in two dimensions. *IEEE Transactions on Computers*. 1971; 100(5): 551–561. <http://dx.doi.org/10.1109/T-C.1971.223289>
- [86] Teague MR. Image analysis via the general theory of moments. *Journal of the Optical Society of America*. 1980; 70(8): 920–930. <http://dx.doi.org/10.1364/JOSA.70.000920>
- [87] Orlov N, Shamir L, Macura T, et al. Wndcharm: Multi-purpose image classification using compound image transforms. *Pattern Recognition Letters*. 2008; 29(11): 1684–1693. <http://dx.doi.org/10.1016/j.patrec.2008.04.013>
- [88] Shamir L, Orlov N, Goldberg IG. Evaluation of the informativeness of multi-order image transforms. In: *International Conference on Image Processing Computer Vision and Pattern Recognition*; 2009. p. 37–42.
- [89] Bishop CM. *Pattern recognition and machine learning*. Springer New York, NY; 2006.
- [90] Shamir L. Assessing the efficacy of low-level image content descriptors for computer-based fluorescence microscopy image analysis. *Journal of Microscopy*. 2011; 243(3): 284–292. <http://dx.doi.org/10.1111/j.1365-2818.2011.03502.x>
- [91] Felsenstein J. *Phylip-phylogeny inference package (version 3.2)*. 1989.
- [92] Felsenstein J. *Phylip: (phylogeny inference package) version 3.6 a3*. 2002.
- [93] Thompson D. *Third ear: Alternative rock*. Miller Freeman, San Francisco, CA; 2000.
- [94] McKay C, Fujinaga I, Depalle P. *jAudio: A feature extraction library*. In: *Int. conf. on Music Information Retrieval (ISMIR)*; 2005. p. 600–3.
- [95] Abdennadher N, Boesch R. Porting phylip phylogenetic package on the desktop grid platform xtremweb-ch. *Studies in Health Technology and Informatics*. 2007; 126: 55. PMID:17476048.