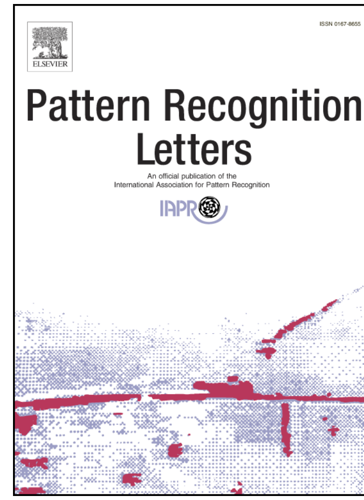


Computer analysis of similarities between albums in popular music

Joe George, Lior Shamir

PII: S0167-8655(14)00071-3
DOI: http://dx.doi.org/10.1016/j.patrec.2014.02.021
Reference: PATREC 5960

To appear in: *Pattern Recognition Letters*

Computer analysis of similarities between albums in popular music

Joe George [a] and Lior Shamir [a,*]

[a] *Lawrence Technological University*

21000 W Ten Mile Rd., Southfield, MI 48075

Abstract

Analysis of musical styles is a complex cognitive task normally performed by music fans and critics, and due to the multi-dimensional nature of music data can be considered a challenging task for computing machines. Here we propose an automatic quantitative method that can analyze similarities between the sound of popular music albums in an unsupervised fashion. The method works by first converting the music samples into two-dimensional spectrograms, and then extracting a large set of 2883 2D numerical content descriptors from the raw spectrograms as well as 2D transforms and compound transforms of the spectrograms. The similarity between each pair of samples is computed using a variation of the Weighted K-Nearest Neighbor scheme, and a phylogeny is then used to visualize the differences between the albums. Experimental results show that the method was able to automatically organize the albums of The Beatles by their chronological order, and also unsupervisely arranged albums of musicians such as U2, Queen, ABBA, and Tears for Fears in a fashion that is largely in agreement with their chronological order and musical styles.

Key words: Music, machine perception, music information retrieval.

1 Introduction

The application of pattern recognition and machine learning to automatic analysis of music enabled numerous useful tasks. Due to the paramount effect of information technology on music consumption, production, and marketing culture, these research efforts are expected to continue.

One of the most common pattern recognition problems in automatic music analysis is music classification (Tzanetakis & Cook, 2002; Guo & Li, 2003). Classification of music can be done by genre (Li et al., 2003; Bagci & Erzin, 2007), emotions (Yand et al., 2008), and musical instruments (Zlatintsi & Maragos, 2013). Due to the intensive research efforts in the field, frameworks for basic music classification were proposed and developed (McKay, 2010). Other directions of machine learning research in music include automatic music recommendation (McFee, Barrington & Lanckriet, 2012), cover song detection (Serra et al., 2012), query by humming (Rocamora et al., 2013; Tsai et al., 2013), sound quality prediction (Manders, Simpson & Bell, 2012), and also tasks such as analysis of traditional Irish music (Duggan, 2009) and detection of difficult chords (Matthias & Dixon, 2010).

An important task in content-based music retrieval is the ability to search music databases for the most similar musical pieces based on an input music sample (Downie, 2008; Casey et al., 2008). Methods for music-based content retrieval include shape similarity (Urbano et al., 2011), editing distances (Mongeau & Sankoff, 1990), alignment (Hanna et al., 2008), n-grams (Uitdenbogerd & Zobel, 1999; Bainbridge, Dewsnip & Witten, 2005), minimum area between polynomial chains (Typke, Veltkamp & Wiering, 2004; Clifford

* Corresponding author: Tel: (248) 204-3512 Fax: (248) 204-3518

Email: lshamir@mtu.edu (Lior Shamir)

et al., 2006), and hybrid modeling of music descriptors (Zhang et al., 2009). In addition to fully automatic methods, semi-automatic approaches for music similarity using textual descriptors labeled manually were used to generate playlists automatically (Pachet & Cazaly, 2002).

Here we describe an unsupervised machine learning method for automatic quantitative analysis of similarities between albums in popular music. The analysis reflects the change in the musical style of the artists, as albums are often considered milestones in the historical perspective of the artist's musical style. The method is based on comprehensive morphological analysis of the audio content as reflected by its 2D spectrogram, and the morphological descriptors are used for determining and quantifying similarities between albums. The primary application of the algorithm is analysis of music in a quantitative fashion for music research and music critic purposes, as well as analyzing and visualizing similarities between musical styles for content-based browsing and music discoverability.

2 Music data

The music data sets include all studio albums of several influential and well-discussed popular music artists such as The Beatles, Queen, ABBA, Tears for Fears, and U2. The computer analysis performed in this study is based on the assumption that an album is a musical unit that reflects a certain musical style or perspective. While this assumption is not always completely true as can be evident by frequent debates between band members about what songs should be included in a certain album, albums are widely considered milestones in the musical development of musicians, and the primary unit by which music is criticized and discussed in present and historic perspective.

The data included all tracks of all studio albums, and excluded albums from live shows or collections. The reason for excluding non-studio albums is that in these albums the artists had limited control over the sound (in the case of live shows) or content (in the case of collections), and therefore albums that were not recorded in the studio cannot be assumed to reflect the sound and music exactly as designed and created by the musicians.

The audio files were originally FLAC (Free Lossless Audio Codec) files, converted to mono WAV files. To normalize for the length of each musical piece, a 60-second long segment was trimmed from each track using the Sound Exchange (SOX) open source software (Sox, 2013). These segments do not include the entire track, but are sufficiently long to perceive the sound. The 60-second segments do not start from the beginning of the track, but from 30 seconds from the beginning of the track, so that each segment used in the experiment is a 60-second long segment from 00:30 to 01:30 of the original track. The reason for not using the first 30 seconds is that in many cases the song can have an intro played by a single instrument or an instrumental part that does not reflect the full sound of the song (e.g., the flute solo intro of the Beatles' "Strawberry Fields Forever").

For the experiment we used all 13 studio albums of the Beatles (released in the UK), 14 studio albums of Queen, the 11 albums of U2, the eight albums of ABBA, and the six studio albums of Tears for Fears.

Each of the 60-second music samples was converted into a 800×512 2D spectrogram, which is a visual representation of the audio and provides precise information of the recording (Altes, 1980). The vertical dimension of the spectrogram corresponds to frequency or pitch, usually measured in Hertz or kilohertz. The time is represented by the horizontal axis. The 2D spectrogram visualizes the sound such that edges and textures can be noticed to the eye, and therefore numerical content descriptors that reflect edges and textures can be informative for the analysis of the spectrograms, as well as other 2D descriptors such as 2D polynomial decomposition and statistical distribution of the pixel intensities.

3 Music analysis method

3.1 Feature set

The analysis of the spectrograms was based on the Wndchrm feature set (Shamir, 2008; Shamir et al., 2008a; Orlov et al., 2008; Shamir et al., 2009a), which is a comprehensive set of features that quantifies very many aspects of the visual content. The motivation for the analysis is the observation that visual features of the spectrograms such as textures reflect the audio content in an informative fashion (Deshpande, Singh & Nam, 2001; Holzapfel & Stylianou, 2008), and the low-level image features of these spectrograms can be used effectively for music classification by genre (Costa et al., 2011). The Wndchrm scheme was originally developed for bioinformatics research (Shamir et al., 2008a), and was found effective in analyzing 2D image morphology in fields such as microscopy and radiology (Shamir et al., 2008b, 2010a), astronomy(Shamir, 2009) as well as quantitative morphological analysis of visual art (Shamir et al., 2010b; Shamir, 2012; Shamir & Tarakhovsky, 2012).

In summary, Wndchrm uses a large set of 2883 2D numerical content descriptors (Shamir et al., 2008a; Shamir, 2008; Shamir et al., 2010b). These include the Haralick (Haralick, Shanmugam & Dinstein, 1973) and Tamura (Tamura, Mori & Yamavaki, 1978) texture features, Radon transform features (Lim, 1990), Gabor filters (Gabor, 1946) with a Gaussian harmonic function (Gregorescu, Petkov & Kruizinga, 2002), Fractal features (Wu, Chen & Hsieh, 1992), Chebyshev statistics features (Gradshtein & Ryzhik, 1994), multi-scale histograms (Hadjidementriou, Grossberg & Nayar, 2001), first 4 moments (Shamir et al., 2008a), Prewitt gradient (Prewitt, 1970) edge features, statistics of the high-contrast 8-connected Otsu binary mask objects (Otsu, 1979), the Euler number (Gray, 1971), Zernike features (Teague, 1979), and Chebyshev-Fourier features (Orlov et al., 2008). A detailed description of the numerical descriptors is available in (Shamir, 2008; Shamir et al., 2008a; Orlov et al., 2008; Shamir et al., 2010a,b). A block diagram of the scheme is available in (Shamir et al., 2009c, 2010a).

These content descriptors are extracted not just from the raw values, but also from the two-dimensional transforms and combinations of multi-order transforms. The transforms that are used are Fourier transform, Chebyshev transform, Wavelet (symlet 5, level 1) transform, and edge magnitude transform. A detailed description and performance analysis of the image features extracted from image transforms and multi-order transforms can be found in (Shamir, 2008; Shamir et al., 2008a; Orlov et al., 2008; Shamir et al., 2009a; Shamir, Orlov & Goldberg, 2009b; Shamir et al., 2010b).

For the feature extraction, each spectrogram is divided into 16 equal-sized tiles (200×128 pixels) such that the feature set (Shamir, 2008; Shamir et al., 2008a, 2010b) is computed separately for each tile (Shamir et al., 2008a, 2010a). Obviously, when a certain track is allocated to training or test sets, all tiles associated with it are assigned to the same set so that tiles of the same song cannot exist both in the training and test sets, and therefore tiles of the same spectrogram cannot be compared to each other.

Music and sound are complex multi-dimensional types of data, and therefore effective quantitative representation of music often requires multiple descriptors. However, since the set of 2D numerical content descriptors computed from each spectrogram is large and comprehensive, not all descriptors are expected to be equally informative for the purpose of music analysis. To weigh the 2D content descriptors by their informativeness, each feature is assigned with a Fisher discriminant score (Bishop, 2006), described by Equation 1,

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

where W_f is the Fisher discriminant score of feature f, N is the number of albums, $\overline{T_f}$ is the mean of the values of descriptor f in the entire training set, and $\overline{T_{f,c}}$ and $\sigma_{f,c}^2$ are the mean and variance of the values of feature f among all training spectrograms of album c. All variances used in the equation are computed after the values of descriptor f are normalized to the interval $[0, 1]$. After Fisher scores are assigned to the descriptors, the weakest 65% of the descriptors (with the lowest Fisher discriminant scores) are rejected, resulting in a set of 1009 2D numerical content descriptors. The threshold of 65% of the features was determined empirically as will be described in Section 4.

After computing the 2D numerical content descriptors, the distance $d_{x,c}$ between a song x and a certain album c is measured by Equation 2.

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

where T_c is the training set of album c, t is a feature vector from T_c, $|x|$ is the length of the feature vector x, x_f is the value of numerical descriptor f in the vector x, t_f is the value of feature f of training sample t, 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, which is set to -5. The -5 value has been determined empirically, and is thoroughly discussed with experimental results by (Orlov et al., 2008). The distance between a feature vector of a certain spectrogram in the test set and a certain album is the mean of its weighted distances (to the power of p) to all vectors of songs that belong in that album.

After the distances between all songs to all other songs are determined, the computed distance $M_{A,Z}$ between albums A and album Z is determined by the average distance of all songs in album A to all songs in album Z, as described in Equation 3

$$M_{A,Z} = \frac{\sum_{s \in A} D_{s,Z}}{|A|} \tag{3}$$

Where $|A|$ is the number of songs in the album A. Repeating the task for all albums in the dataset provides a matrix of all distances between all pairs of albums. That is, the cell n, m is the distance between album n to album m. The distance matrix is inverted into a similarity matrix, and the values are normalized such that the computed similarity of an album to all other albums is divided by the computed similarity of the album to itself (so that the similarity of an album to itself is set to 1). The distance matrix can be visualized by phylogenies (evolutionary trees) using the Phylib package (Felsenstein, 2004). Phylip was originally developed to visualize genomic similarities between organisms, but in this study it is used to visualize the similarities between music albums based on the distance matrix.

In all experiments described in Section 4, two songs from each album were used for testing, and the remaining songs were used for training such that each album was trained with M-2 songs, where M is the number of songs

ACCEPTED MANUSCRIPT

in the album with the lowest number of songs. Each experiment described in Section 4 was repeated 40 times such that in each run different songs were randomly allocated for training and test sets.

A disadvantage of the method is its slow response time. While converting the audio files to spectrograms is nearly immediate, computing the 2D numerical content descriptor from a single spectrogram takes ∼9 minutes using a single Intel core-i7 processor, and therefore requires a substantial multi-core computing facility. The work described in this paper was done with a 32-core cluster of Intel core-i7 and a 320-core cluster of AMD Opteron processors.

4 Experimental results

The method described in Section 3 was applied to the music datasets described in Section 2. In the first experiment, the similarity analysis of music was focused on the albums of The Beatles, which had a clear and noticeable change in their sound and musical style, and due to the influence of the band on popular music the history of their sound and albums has been thoroughly studied by music critics. As described in Sections 2 and 3, nine songs from each album were used for training and two for testing, and the experiment was repeated 40 times such that in each run different albums were randomly allocated to training and test sets. The classification accuracy of a Beatles song to the correct album was just ∼30.6%, but this accuracy is significantly higher than ∼7.7% of random guessing, indicating that the analysis can identify differences between the albums. Although the purpose of this study is not to classify the albums but to profile the similarities between them, the classification accuracy was used for determining the optimal number of features that were selected from the total of 2883 features computed for each spectrogram.

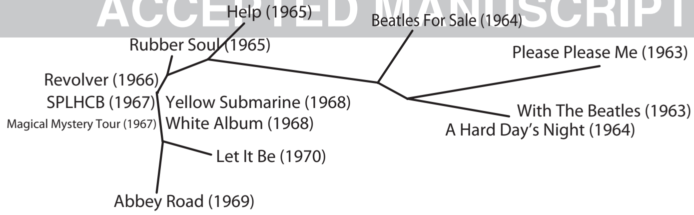


Fig. 1. Phylogeny of the studio albums of the Beatles.

For instance, when 15% of the features were used the classification accuracy of songs to albums was ∼25.5%, increased to ∼29.1% when using 25% of the features, and was ∼26.8% when 45% of the features are used. The classification accuracy was also used to evaluate the effect of the duration of the audio sample from each song. With 20 second audio samples the classification accuracy was reduced to ∼24.9%, while with 40 second samples it was comparable to the classification accuracy using the 60 second samples.

Figure 1 shows the phylogeny generated automatically for the 13 studio albums of the Beatles released in the UK. The similarities between the albums are reflected in the phylogeny by the length of the arcs, such that a shorter path between two nodes (albums) means higher similarity between them. Longer lines or longer paths mean that the albums are determined by the algorithm as less similar to each other.

The phylogeny shows that the algorithm was able to organize the albums of the Beatles in a chronological order, showing the continuous change in their musical style during these years. The early rock and roll albums "Please, please me" (1963), "With the Beatles" (1963) and "A hard days night" (1964) are positioned in the right part of the phylogeny, followed by the mid-60s pop rock albums "Beatles for sale" (1964), "Help" (1965), and "Rubber soul" (1965). Then the algorithm clustered the psychedelic rock albums "Revolver" (1966), "Sergeant Pepper's lonely hearts club band" (1967), "Magical mystery tour" (1967), and "Yellow submarine" (1968). These are followed by the later albums of The Beatles - the white album (1968), "Abbey Road" (1969), and "Let it be" (1970), which were rock albums with blues and R&B influence. The album "Let it be" (1970) was released after "Abbey road", but it included tracks that the band recorded before "Abbey road".

The content descriptors with the highest Fisher discriminant scores, which are the descriptors that had the most substantial effect, are the Zernike polynomial descriptors extracted from the Fourier transform, Wavelet transform of the Fourier transform, and the Fourier transform of the edge transform. The Haralick texture features were also informative when extracted from the raw spectrograms, from the edge transform of the spectrograms, and from the wavelet transform of the edge transform. Other informative features were the edge descriptors extracted from the raw spectrograms, and fractal features extracted from the edge transform of the spectrograms.

In another experiment, the studio albums of Queen were studied to detect and profile the possible change in the band's musical style. While the number of studio albums of Queen is comparable to that of The Beatles (15 of Queen compared to 13 of The Beatles), the band released their albums during a much longer period from 1973 to 1995, during which they changed their sound and musical style. Figure 2 shows the phylogeny created by the computer using the studio albums of Queen, except for "Flash Gordon", which is officially considered a Queen studio album but is actually a movie soundtrack and includes mostly instrumental pieces.

The phylogeny shows that the method was able to accurately organize Queen albums in an order close to the chronological order by which these albums

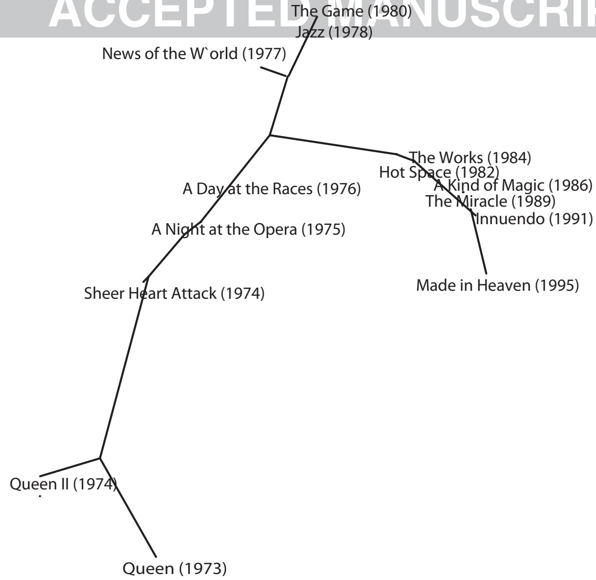


Fig. 2. Phylogeny of the studio albums of Queen. The phylogeny shows a chronological order of the albums. The branch starts with the album "The Works" shows the sonic departure of the band from their signature 70's sound.

were recorded, starting from "Queen" (1973) and "Queen II" (1974), through "Sheer heart attack" (1974), "A night at the opera" (1975), and "A day at the races" (1976).

After "A day at the races" the algorithm placed "News of the world" (1977), "Jazz" (1978), and "The game" (1980) in chronological order. However, the phylogeny also features another branch starting with "The works" (1984) and "Hot space" (1982), which were recorded after "The game". These albums are not positioned on the same line, but on a different branch forking before "News of the world". That violation of the chronological order can be explained by the strong style change starting with "Hot space", in which the band adopted 80s sound that was significantly different from the sound the band was recognized with through the 70s (Miccio, 2011). The next albums on that branch are "A kind of magic" (1986), "The miracle" (1989), "Innuendo" (1991), and "Made in heaven" (1995), positioned relatively far from the previous albums, which is also expected since "Made in heaven" is a compilation of Queen songs

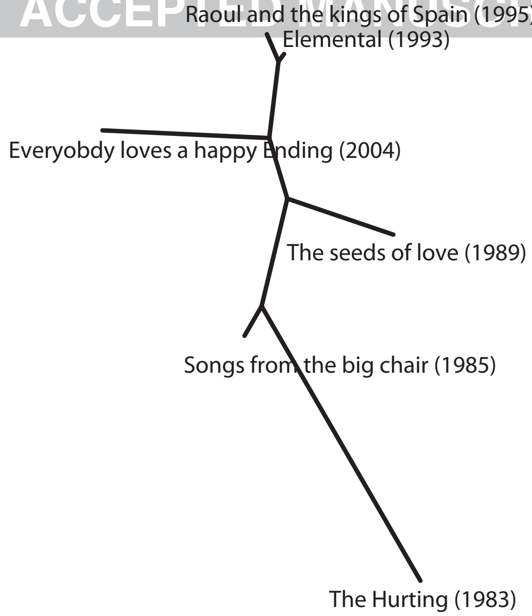


Fig. 3. Phylogeny of the studio albums of Tears for Fears. "Everybody loves a happy ending" was released after "Elemental" and "Raoul and the kings of Spain", but was produced by the original founders of the band who reunited, and consider the album a continuation of "The seeds of love".

featuring Freddie Mercury's vocals recorded before his death in 1991, but the songs were produced after his death. The classification accuracy of songs to albums in that experiment was ∼24%.

Another experiment that provided results that were largely in agreement with the chronological and historical perspective of the musical style is the computer analysis of the studio albums of Tears for Fears. The band released six studio albums between 1982 and 2004. The relatively long periods between albums and the carefully crafted album sound (Thrills, 1990) provides a noticeable change of the musical style. Figure 3 displays the phylogeny that was generated from the official studio albums of Tears for Fears

The first album, "The hurting", was positioned by the algorithm at the bottom of the phylogeny, with significant distance from the second albums "Songs from the big chair". The reason for the distance between the two albums can be explained by the change in the sound of the band, shifted from the New Wave synthpop sound of "The hurting" to the more sophisticated big sound that became its signature musical style in "Songs from the big chair" (Swihart, 1985) and consequent albums. Upper in the phylogeny the algorithm placed "The seeds of love", which was released in 1989 and featured a warmer and more spacious sound (Holden, 1989). The album is followed by two albums clustered close to each other: "Elemental" (1993), and "Raoul and the kings of Spain" (1995). These two albums are considered official Tears for Fears albums, but were the sole work of Roland Orzabal after the band split in 1991, and were directed towards a smaller, more sophisticated audience (Sinclair, 1990). Interestingly, the algorithm placed the album "Everybody loves a happy ending" between "The seeds of love" and "Elemental". The album was recorded in 2004, after "Elemental" and "Raoul and the kings of Spain", but was the work of the original band founders who reunited in 2000. The band members see the album as the continuation of their previous collaborative work in "The seeds of love" (Reynolds, 2004; O'hara, 2004). The classification accuracy of songs to albums was ∼34%, which is clearly higher than random guessing accuracy of ∼16.7%.

In the next experiment, the method was applied to the albums of U2. The method accurately assigned a song to its album in ∼29% of the cases. The phylogeny that was generated by the computer is displayed in Figure 4.

The phylogeny placed U2's early 80s post-punk albums "October" (1981), "Boy" (1980), and "The unforgettable fire" (1984) at the bottom of the tree. Then, the algorithm clustered the rock albums of the late 80s "The Joshua tree" (1987), and "Rattle and hum" (1988), followed by the albums "Zooropa" (1993), "Achtung baby" (1991), and "Pop" (1997), which had alternative rock style influence (Eno, 1991).

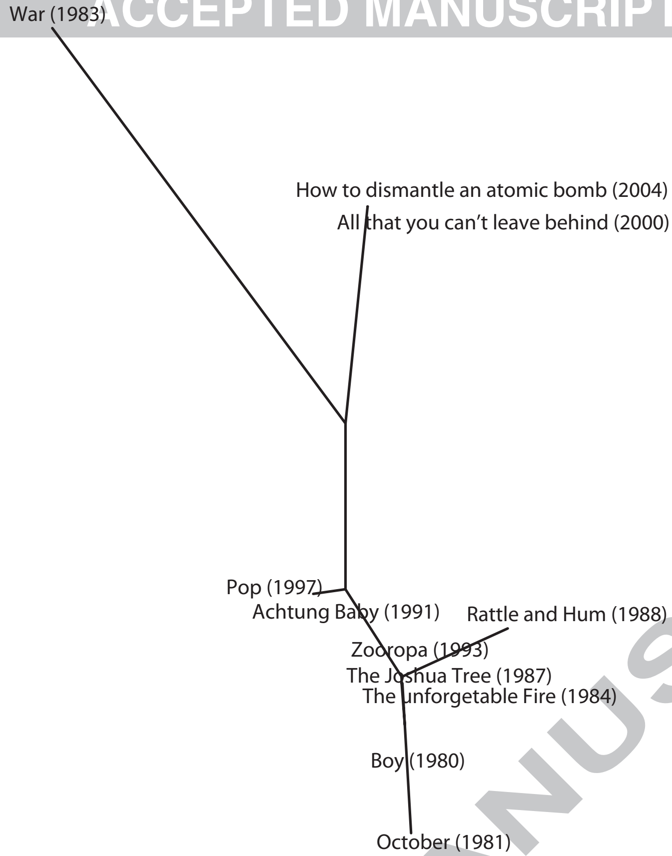


Fig. 4. Phylogeny of the albums of U2.

The cluster is followed by the next albums "All that you can't leave behind" (2000), and "How to dismantle and atomic bomb" (2004), which were considered by the band as rock albums, and a significant change from the alternative sound of their albums during the 90's (Sheffield, 2004). The album "War" (1983) is placed by the algorithm far from the other albums, although its post-punk musical style is not noticeably different from the albums recorded in the mid 80's.

We also tested the music of ABBA, and the phylogeny is displayed in Figure 5. As the phylogeny shows, the algorithm automatically positioned all albums on a clear line and by their chronological order, from the band's first album "Ring Ring" to their last album "The visitors". The classification accuracy of songs to albums was ∼43%.

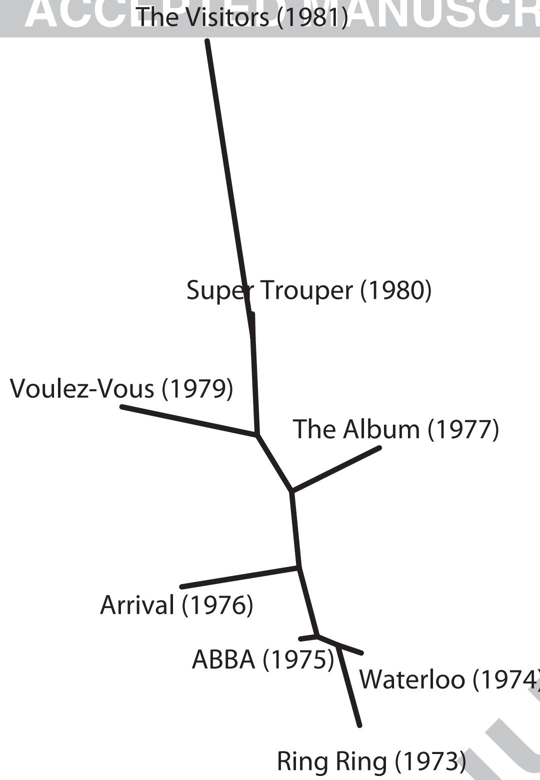


Fig. 5. Phylogeny of the studio albums of ABBA.

4.1 Comparison to results using audio features

To test the efficacy of using the spectrograms for unsupervised analysis of music, we compared the results described in this paper to results produced by using audio features computed directly from the audio. The features were extracted using the jAudio open source tool, which is part of the jMIR open source music and audio analysis package (McKay, 2010). jAudio extracts audio content descriptors that reflect various aspects of the audio such as 1D and 2D moments, area moments, spectral and harmonic spectral properties (flux, centroid, smoothness), beat histograms, zero crossing, Mel-Frequency Cepstral Coefficients (MFCC) and more, as described in (McKay, 2010) and in http://jaudio.sourceforge.net/jaudio10/features/feature.html. jAudio provides a total of 78 numerical audio content descriptors.

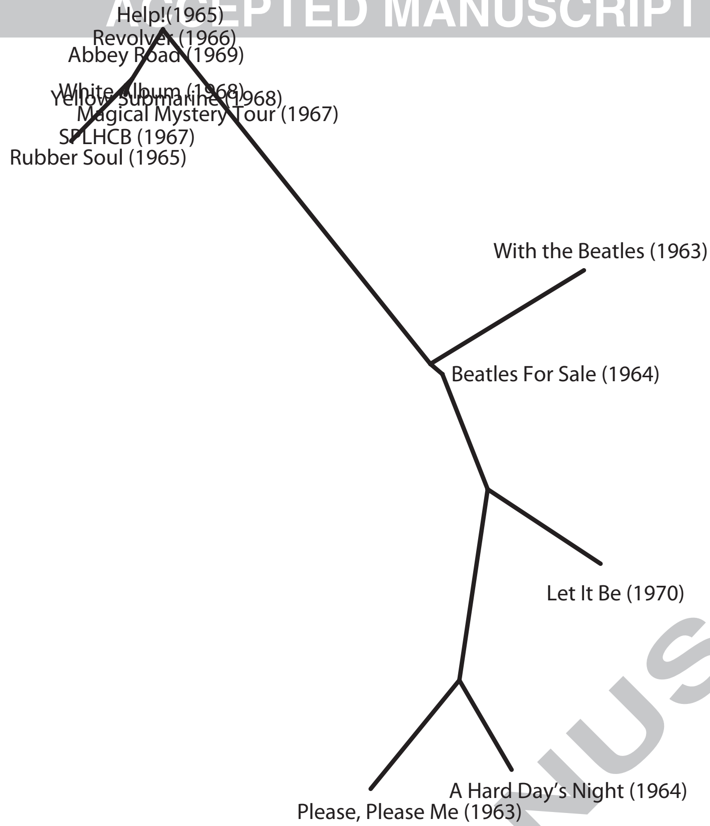


Fig. 6. Phylogeny of the Beatles UK studio albums with features extracted using jAudio.

Figure 6 and 7 shows the phylogeny of the studio albums of The Beatles and ABBA, respectively, analyzed using audio features extracted by jAudio.

As Figure 6 shows, the jAudio features were sufficiently informative to allow separation between the first four albums and the rest of the Beatles albums. However, inside the two groups there is no particular chronological order of the albums. It is also noticeable that the album “Let It Be”, recorded during and after 1968 and released in 1970, is positioned among the early albums of the band despite the fact that the rock style of the album and musical instruments that were used are fundamentally different from the musical style and instruments the band used in 1964. Figure 7 shows no chronological order in the albums of ABBA, which is in contrast to using the analysis of the

ACCEPTED MANUSCRIPT

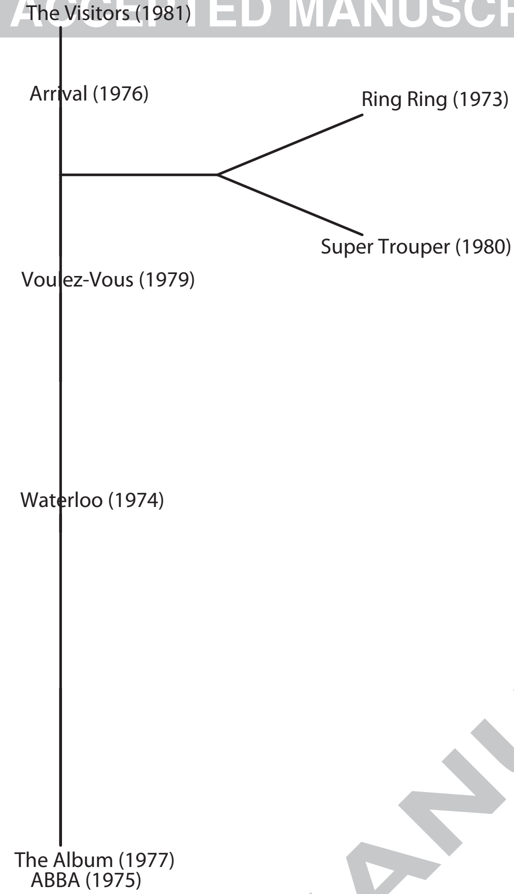


Fig. 7. Phylogeny of the albums of ABBA with audio features extracted using jAudio.

spectrograms. The classification accuracy of the Beatles songs to albums was ~20.2%, and was ~16% in the case of ABBA.

5 Conclusions

Music is highly complex and multi-dimensional data that introduce a challenge when automatically analyzed by computing machines. Here we describe a method that can use comprehensive morphological analysis of the spectrograms of songs to profile the similarities between albums. The results are largely in agreement with the chronological order of the albums, as well as the perception of the change in the musical style.

The number of descriptors extracted from the spectrograms is larger than some other studies in which the features are extracted directly from the audio files. These features reflect the textures of the spectrogram, polynomial decomposition, statistical distribution of the pixel values, etc', leading to a larger set of features required to effectively reflect the spectrograms. Due to their large number, the features are weighted by their informativeness. Unsupervised analysis of the albums with audio features extracted directly from the audio files provided partial or no chronological order of the albums. These results show that morphological analysis of the spectrograms can be used effectively to analyze sound and music.

The similarity between two albums is determined in this study by the average distance between all songs of one albums and all songs of the second album. Other distances can also be used, such as the minimum distance between two songs of the two albums.

An analysis that visualizes similarities between musical styles can be used for music discoverability and content-based navigation in large music databases. Such methods are required to satisfy the growing need to organize and manage music data, which is currently one of the most popular and most consumed types of digital data.

6 Acknowledgments

This work was supported in part by grant 1157162 of the National Science Foundation.

References

Altes, R.A., 1980) Detection, estimation, and classification with spectrograms. J. Acoust. Soc. Am., 67, 1232–1246.

Bagci, U., Erzin, E., 2007. Automatic Classification of Musical Genres Using Inter-Genre Similarity. IEEE Signal Processing Letters, 14, 521–524.

Bainbridge, D., Dewsnip, M., Witten, I.H., 2005. Searching digital music libraries. Information Processing and Management, 41, 41–56.

Bishop, C. M., 2006. Pattern Recognition and Machine Learning, Springer Press.

Casey, M., Veltkamp, R., Goto, M., Leman, M., Rhodes, C., Slaney, M., 2008. Content-based music information retrieval: Current directions and future challenges. Proceedings of the IEEE, 96, 668–695.

Clifford, R., Christodoulakis, M., Crawford, T., Meredith, D., Wiggins, G., 2006. A fast, randomised, maximal subset matching algorithm for document-level music retrieval. In: International Conference on Music Information Retrieval, 150–155.

Costa, Y.M.G., Oliveira, L.S., Koericb, A.L., Gouyon, F., 2011. Music genre recognition using spectrograms. In: 18th International Conference on Systems, Signals and Image Processing. 1–4.

Deshpande, H., Singh, R., Nam, U., 2001. Classification of music signals in the visual domain. In:Proceedings of the COST G-6 Conference on Digital Audio Effects (DAFX-01).

Downie, S., 2008. The music information retrieval evaluation exchange (2005-2007): A window into music information retrieval research. Acoustical Science and Technology, 29, 247–255.

Duggan, D., 2009. Machine annotation of traditional Irish dance music. PhD thesis. Dublin Institute of Technology.

Eno, B., 1991. Bringing up baby. Rolling Stone, November 18, 1991. Retrieved from http://www.atu2.com/news/bringing-up-baby.html

Felsenstein, J., 2004. PHYLIP Phylogeny Inference Package, Version 36.

Gabor, D., 1946. Theory of communication. Journal of IEEE, 93, 429–457.

Gradshtein, I., Ryzhik, I., 1994. Table of integrals, series and products. 5 ed. Academic Press, p. 1054.

Gray, S.B., 1971. Local properties of binary images in two dimensions. IEEE Trans. on Computers, 20, 551–561.

Gregorescu, C., Petkov, N., Kruizinga, P., 2002. Comparison of texture features based on Gabor filters. IEEE Trans. on Image Proc., 11, 1160–1167.

Guo, G., & Li., S. Z., 2003. Content-based audio classification and retrieval by support vector machines. IEEE Trans. on Neural Networks, 14, 209–215.

Hadjidementriou, E., Grossberg, M., Nayar, S., 2001. Spatial information in multiresolution histograms. In: IEEE Conf. on Computer Vision and Pattern Recognition, 1, p. 702.

Hanna, P., Ferraro, P., Robine, M., 2007. On optimizing the editing algorithms for evaluating similarity between monophonic musical sequences, Journal of New Music Research, 36, 267–279.

Hanna, P., Robine, M., Ferraro, P., Allali, J., 2008. Improvements of alignment algorithms for polyphonic music retrieval. In: International Symposium on Computer Music Modeling and Retrieval, 244–251.

Haralick, R.M., Shanmugam, K., Dinstein, I., 1973. Textural features for image classification. IEEE Trans. on Syst., Man and Cyber., 6, 269–285.

Holden S., 1990. Disciples of the Beatles. New York Times, February 21, 1990.

Holzapfel, A., Stylianou, Y., 2008. Musical genre classification using non-negative matrix factorization-based features. IEEE Transactions on Audio, Speech, and Language Processing, 16, 424–434.

Li, T., Ogihara, M., Li, Q., 2003. A comparative study on content-based music genre classification. In: SIGIR03 (July 2003), 282–289.

Lim, J.S., 1990. Two-Dimensional signal and image processing. Prentice Hall, 42–45.

Manders, A.J., Simpson, D.M., & Bell, S.L., 2012. Objective prediction of the sound quality of music processed by an adaptive feedback canceller. IEEE Trans. on Audio, Speech, and Language Processing, 20, 1734–1745.

Mauch, M., Dixon, S., 2010. Approximate note transcription for the improved identification of difficult chords. In: Proceedings of the 11th International Society for Music Information Retrieval Conference (ISMIR 2010).

McFee, B., Barrington, L., Lanckriet, G.R.G., 2012. Learning content similarity for music recommendation. IEEE Transactions on Audio, Speech, and Language Processing, 20, 2207–2218.

McKay, C., 2010. Automatic music classification with jMIR. PhD dissertation, McGill University.

Miccio, A., 2011. Queen Hot Space. Stylus Magazine May 31, 2011.

Mongeau, M., & Sankoff, D., 1990. Comparison of musical sequences. Computers and the Humanities, 24, 161–175.

O'hara, K., 2004. Houston Chronicle, September 19, 2004, p. 5.

Orlov, N., Shamir, L., Macura, T., Johnston, J., Goldberg, I., 2008. WND-CHARM: Multi-purpose image classification using compound image transforms. Pattern Recognition Letters, 29, 1684–1693.

Otsu, N., 1979. A threshold selection method from gray level histograms. IEEE Trans. on Syst., Man and Cyber., 9, 62–66.

Pachet, F., Aucouturier, J.J., 2002. Music similarity measures: Whats the use? In: International Symposium/Conference on Music Information Retrieval.

Prewitt, J.M., 1970. Object enhancement and extraction. Picture processing and psychopictoris. B. S. Lipkin and A. Rosenfeld, Eds. New York, NY: Academic Press, 75–149.

Reynolds, R., 2004. Album Review: Tears for Fears - Everybody loves a happy ending. City Monthly Magazine, March 2004.

Rocamora, M, Cancela P, Pardo, A., 2013. Query by humming: Automatically building the database from music recordings. Pattern Recognition Letters, In Press.

Serr, Y., Kantz, H., Serra, X., Andrzejak, R.G., 2012. Predictability of music descriptor time series and its application to cover song detection. IEEE Transactions on Audio, Speech, and Language Processing, 20, 514–525.

Shamir, L., 2008. Evaluation of face datasets as tools for assessing the performance of face recognition methods. International Journal of Computer Vision, 79, 225–230.

Shamir, L., Orlov, N., Eckley, D.M., Macura, T., Johnston, J., Goldberg, I., 2008a. Wndchrm an open source utility for biological image analysis. Source Code for Biology and Medicine, 3, article 13.

Shamir, L., Orlov, N., Eckley, D.M., Macura, T., Goldberg, I., 2008b. IICBU 2008 - A proposed benchmark suite for biological image analysis. Source Code for Biology and Medicine, 46, 943–947.

Shamir, L., 2009. Automatic morphological classification of galaxy images. Monthly Notices of the Royal Astronomical Society, 399, 1367–1372.

Shamir, L., Ling, S., Scott, W., Boss, A., Orlov, N., Macura, T., Eckley, D.M., Ferrucci, L., Goldberg, I., 2009a. Knee X-ray image analysis method for automated detection of Osteoarthritis. IEEE Transactions on Biomedical Engineering, 56, 407–415.

Shamir, L., Orlov, N., Goldberg, I., 2009b. Evaluation of the informativeness of multi-order image transforms. In: International Conference on Image Processing Computer Vision and Pattern Recognition, 37-42. Las Vegas, NV. 2009b.

Shamir, L., Ling, S., Scott, W., Hochberg, M., Ferrucci, L., Goldberg, I. 2009c.

Early detection of radiographic knee osteoarthritis using computer-aided analysis. Osteoarthritis and Cartilage, 17, 1307–1312.

Shamir, L., Macura, T., Orlov, N., Eckley, M., Goldberg, I., 2010a. Impressionism, expressionism, surrealism: Automated recognition of painters and schools of art. ACM Transactions on Applied Perception, 7, article 8.

Shamir, L., Delaney, J., Orlov, N., Eckley, M., Goldberg, I., 2010b. Pattern recognition software and techniques for biological image analysis. PLoS Computational Biology, 6, e1000974.

Shamir, L., 2012. Computer analysis reveals similarities between the artistic styles of Van Gogh and Pollock. Leonardo, 45, 149–154.

Shamir, L., Tarakhovsky, J., 2012. Computer analysis of art. ACM Journal on Computing and Cultural Heritage, 5, no. 2, article 7.

Sheffield, R., 2004. U2 how to dismantle an atomic bomb. Rolling Stone Reviews, December 9, 2004.

Sinclair, T., 1990. Raoul and the kings of Spain. Entertainment Weekly, October 13, 1995. Retrieved from http://www.ew.com/ew/article/0,,299113,00.html

SOX, 2013. Sox: Sound Exchange. SourceForge. http://sox.sourceforge.net.

Swihart, S., 1985. AllMusic. http://www.allmusic.com/album/songs-from-the-big-chair-mw0000038029

Tamura, H., Mori, S., Yamavaki, T., 1978. Textural features corresponding to visual perception. IEEE Trans. on Syst., Man and Cyber., 8, 460–472.

Teague, M.R., 1979. Image analysis via the general theory of moments. Journal of the Optical Society of America, 70, 920–920.

Thrills, A., 1990. Tears for Fears: The seeds of love. UK: Virgin Books, ISBN-13: 978-0863693298.

Tsai, W.H., Tu, Y.M., Ma, C.H., 2013. An FFT-based fast melody comparison method for query-by-singing/humming systems, Pattern Recognition

Letters, 33, 2285–2291.

Typke, R., Veltkamp, R.C., Wiering, F., 2004. Searching notated polyphonic music using transportation distances. In: ACM International Conference on Multimedia, 128–135.

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

Uitdenbogerd, A., Zobel, J., 1999. Melodic matching techniques for large music databases. In: IEEE ACM International Conference on Multimedia, 57–66.

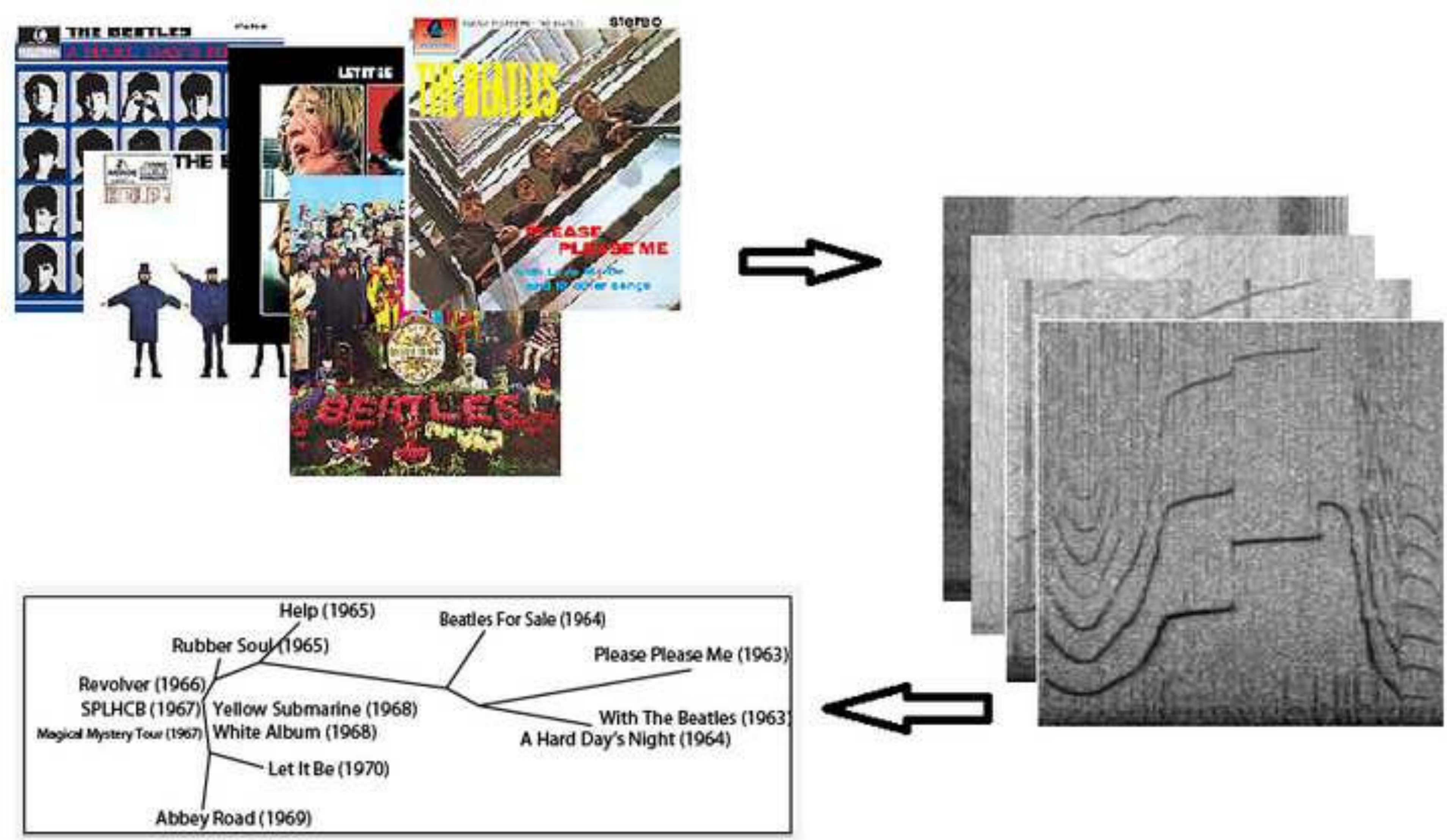
Urbano, J., Llorns, J., Morato, J., Snchez-Cuadrado, S., 2011. Melodic similarity through shape similarity. Lecture Notes in Computer Science, 6684, 338–355.

Wu, C.M., Chen, Y.C., Hsieh, K. S., 1992. Texture features for classification of ultrasonic liver images. IEEE Trans. Med. Imag., 11, 141–152.

Yang, Y.H., Lin, Y.C., Cheng, H.T., Liao, I.B., Ho, Y.C., Chen, H.H., 2008. Toward multi-modal music emotion classification. In: Proceedings of the 9th Pacific Rim Conference on Multimedia: Advances in Multimedia Information Processing. 70–79.

Zhang, B., Shen, J., Xiang, Q., Wang, Y., 2009. CompositeMap: a novel framework for music similarity measure. In: SIGIR09 (July 2009), 403–410.

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


THE BEATLES
LET IT BE
THE BEATLES
PLEASE PLEASE ME
STEREO
BEATLES
Help (1965)
Beatles For Sale (1964)
Rubber Soul (1965)
Please Please Me (1963)
Revolver (1966)
SPLHCB (1967)
Yellow Submarine (1968)
With The Beatles (1963)
Magical Mystery Tour (1967)
White Album (1968)
A Hard Day's Night (1964)
Let It Be (1970)
Abbey Road (1969)

Unsupervised analysis of albums in popular music is presented.

The analysis is done by a first step of transforming the audio files of the songs to 2D spectrograms

The method was able to sort the albums of bands in an order that is very close to their chronological order.

The spectrogram analysis provided more informative analysis than audio features extracted directly from the audio files.