

Genre-oriented Organization of Music Collections Using the SOMeJB System: An Analysis of Rhythm Patterns and Other Features

Thomas Lidy, Andreas Rauber

Department of Software Technology, Vienna University of Technology,
Favoritenstr. 9 - 11 / 188, A-1040 Wien, Austria,

<http://www.ifs.tuwien.ac.at>

Abstract:

With the advent of larger electronic music repositories, the automatic organization of music into different genre categories is receiving increased attention. The creation of such genre hierarchies, as well as ways for providing useful interfaces to these, poses an interesting challenge. With the *SOM-enhanced JukeBox (SOMeJB)* system we developed an approach for automatically organizing pieces of music in raw audio format according to their perceived sound similarity based on Rhythm Patterns. This paper reviews the SOMeJB system and presents performance evaluations of the resulting hierarchical genre clusters. We furthermore compare the Rhythm Pattern based clustering with cluster structures obtained from other sets of features employed for genre-based music analysis. We demonstrate results based on a collection of about 24 hours of music.

Keywords: Music Analysis, MP3, Genre Detection, Feature Extraction, Clustering, Neural Networks

1. Introduction

Recent development has seen an increased interest in systems supporting the automatic organization of large music collections according to different genres. There are several application scenarios for such systems, such as providing new means of distribution for the music industry, offering platforms for the legal sharing of copyright-free pieces of music, supporting the organization of large private collections of music, or simply providing an interface for mobile player devices. All of these benefit from intuitive interfaces providing flexible and convenient organization.

Existing music repositories mostly limit access to their content to query-based retrieval based on textual meta-information. A significant amount of research has been conducted recently in the area of content-based music retrieval, c.f. [4]. Specifically genre based organization and detection has gained significant interest recently. One of the first works to incorporate psychoacoustic modeling into the feature extraction process and utilizing the *SOM* for organizing audio data is reported in [3]. A system performing trajectory matching using *SOMs* and *MFCCs* is presented in [16]. Specifically addressing the classification of sounds into different categories, Wold et al. [21] use loudness, pitch, brightness, bandwidth, and harmonicity features to train classifiers. A wide range of musical surface features is used by the Marsyas system [19] to organize music into different genre categories using a selection of classification algorithms. A subset of these will be used for the experiments reported in this paper.

What we would like to have, additionally, is a way to facilitate exploration of collections of music. We would like to automatically organize music according to its sound characteristics in such a way that we find similar pieces of music grouped together, allowing us to find a classical section, or a hard-rock section etc. in a music repository. The *SOMeJB* Music Digital Library Project, as first outlined in [10], and described in more detail in [13,14], aims at creating such a browsable music archive by combining a variety of technologies from the fields of audio processing, neural networks, and information visualization, to create maps of music archives. It has its roots in the *SOMLib* Digital Library for text archives [12]. It is based on the *Self-Organizing Map (SOM)* [5], a popular unsupervised neural network, and its extension, the *Growing Hierarchical Self-Organizing Map (GHSOM)* [1]. These networks organize pieces of music available as, e.g., mp3 files, according to their musical sound characteristics, creating a kind of genre-based organization. The resulting maps of the music archive can be explored, and new, unknown pieces of music similar to ones personal likings can be discovered, with Islands of Music and Weather Charts [8] providing an intuitive interface to the system. To obtain an organization of music according to perceptual sound similarity, a novel set of features capturing Rhythm Patterns in a set of frequency bands was devised, incorporating psycho-acoustically motivated transformations [13]. In this paper we present an evaluation of the obtained music clusters, confronting them with a manually pre-defined genre hierarchy. We furthermore compare the resulting structure with one obtained by using a different feature set used in genre classification.

2. The SOMeJB System

The *SOMeJB* system uses the topology-preserving capabilities of the *SOM*, as well as its extended model, the *GHSOM*, to create a map of a music archive, where similar pieces of music are located next to each other. Similarity is defined with respect to different feature spaces, two of which are described in Sections 2.1 and 2.2. The resulting feature vectors are presented to the neural network, which in the course of a training process, learns an appropriate mapping. In a nutshell the training process can be described as follows: Input signals are presented to the map, consisting of a grid of units with n -dimensional weight vectors, in random order. An activation function based on some metric (e.g. the Euclidean Distance) is used to determine the winning unit (the 'winner'). In the next step the weight vector of the winner, as well as the weight vectors of the neighboring units, are modified following some learning rate in order to represent the presented input signal more closely. After the training process, similar input patterns are mapped onto neighboring units of the *Self-Organizing Map*. As two deficiencies of the standard *SOM* model we have to note its static architecture, which has to be defined prior to the training process, as well as the impossibility to faithfully reflect the hierarchical structure inherent in data. With the *Growing Hierarchical Self-Organizing Map (GHSOM)* [1]

we proposed a novel neural network model that addresses both deficiencies. It is composed of independent *SOMs*, each of which is allowed to grow in size during the training process until a quality criterion regarding data representation is met. This growth process is further continued to form a layered architecture such that hierarchical relations between input data are further detailed at deeper layers of the neural network.

The resulting maps offer themselves as interfaces to explore a music archive. A specifically appealing visualization based on *smoothed data histograms (SDH)* are the *Islands of Music*, which use the metaphor of geographical maps, where islands resemble styles of music, to provide an intuitive interface to music archives. Furthermore, attribute aggregates are used to create *Weather charts* that help the user in understanding the sound characteristics of the various areas on the map [8].

2.1 Rhythm Patterns

The feature extraction process for the Rhythm Patterns is composed of two stages. First, the specific loudness sensation in different frequency bands is computed, which is then transformed into a time-invariant representation based on the modulation frequency. Starting from a standard *Pulse-Code-Modulated (PCM)* signal, stereo channels are combined into a mono signal, which is further down-sampled to 11kHz. Furthermore, pieces of music are cut into 6-second segments, removing the first and last two segments to eliminate lead-in and fade-out effects, and retaining only every second segment for further analysis. Using a Fast Fourier Transform (FFT), the raw audio data is further decomposed into frequency ranges using Hanning Windows with 256 samples (corresponding to 23ms) with 50% overlap, resulting in 129 frequency values (at 43Hz intervals) every 12 ms. These frequency bands are further grouped into so-called *critical bands*, also referred to by their unit *bark* [23], by summing up the values of the power spectrum between the limits of the respective critical band, resulting in 20 critical-band values. A *spreading function* is applied to account for *masking effects*, i.e. the masking of simultaneous or subsequent sounds by a given sound. The spread critical-band values are transformed into the logarithmic *decibel* scale, describing the sound pressure level in relation to the hearing threshold. Since the relationship between the dB-based sound pressure levels and our hearing sensation depends on the frequency of a tone, we calculate *loudness levels*, referred to as *phon*, using the equal-loudness contour matrix. From the loudness levels we calculate the specific loudness sensation per critical band, referred to as *sons*.

To obtain a time-invariant representation, reoccurring patterns in the individual critical bands, resembling rhythm, are extracted in the second stage of the feature extraction process. This is achieved by applying another discrete Fourier transform, resulting in amplitude modulations of the loudness in individual critical bands. These amplitude modulations have different effects on our hearing sensation depending on their frequency, the most significant of which, referred to as *fluctuation strength* [2], is most intense at 4Hz and decreasing towards 15Hz (followed by the sensation of *roughness*, and then by the sensation of three separately audible tones at around 150Hz). We thus weight the modulation amplitudes according to the fluctuation strength sensation, resulting in a time-invariant, comparable representation of the rhythmic patterns in the individual critical bands. To emphasize the differences between strongly reoccurring beats at fixed intervals a final gradient filter is applied, paired with subsequent Gaussian smoothing to diminish un-noticable variations. The resulting 1.200 dimensional feature vectors (20 critical bands times 60 amplitude modulation values) capture beat information up to 10Hz (600bpm), going significantly beyond what is conventionally considered beat structure in music. They may optionally be reduced down to about 80 dimensions using PCA. These *Rhythm Patterns* are further used for data signal comparison.

2.2 Other Genre Features

Previous experiments revealed that Rhythm Patterns provide a good discrimination between different styles of music, yet are not able to capture all characteristics in terms of sound sensation sufficiently. We thus are currently evaluating a set of additional features for genre detection and assignment, specifically those extracted by the *Marsyas* [6] system. More specifically, we use the following subset of features, recommended for genre classification [19]:

- **FFT:** This set of 9 features consists of the means and variances of the *spectral centroid*, *rolloff*, *flux* and *zerocrossings*, based on the Short Time Fourier Transform of the signal, as well as a *low energy* feature.
- **MFCCs:** The first five Mel-Frequency Cepstral Coefficients, i.e. FFT bins that are grouped and smoothed according to the Mel-frequency scaling.
- **MPitch:** This set of features represents harmonic content based on multiple pitch analysis.
- **Beat:** This set of features represents the beat structure of music calculated by a beat detection algorithm based on Discrete Wavelet Transform, analyzing beats between 40 and 200 bpm. This feature is closely related to our Rhythm Patterns described above, computing the histogram over the whole spectrum rather than individually for different frequency bands, and within a more restricted value range.

As these attributes have significantly different value ranges, attribute-wise normalization to the interval [0,1] is performed, allowing for subsequent comparison of weight vectors using Euclidean distance.

3. Experimental Results

We are currently in the progress of setting up an evaluation framework, allowing the analysis of the system's performance on a large and diverse audio collection both through usability evaluations as well as automated recall/precision testing for different types of genre hierarchies. It consists of approximately 2.000 pieces of music from a wide range of genres, divided into several sub-sets. These are individually organized into different genre categories. The individual organizations are further integrated into a single genre hierarchy for the complete data collection. The various forms of organization reflect the strongly diverse, and sometimes even contradicting, concepts of genres employed by different users for conceptually structuring a music collection by style. The selected pieces of music include both western as well as less traditional, ethnical pieces of music, contain several different interpretations of the same piece of music by the same or different artists, both vocal and instrumental versions of

several titles, etc. Here we present preliminary results from a sub-collection representing 335 pieces (approx. 24 hours) of music. Please note, that both feature sets are evaluated with a particular focus on genre-oriented clustering. Specifically, no feature weighting is performed. Thus, the potentially higher relevance of a specific attribute type for a particular genre is not considered. This balanced treatment of all features may lead to sub-optimal performance for both feature sets, yet particularly so for the second set of features which consists of different types of attributes capturing different types of information.

3.1 Data Set

The 335 pieces of music are manually filed into 14 genres within 4 main categories, as depicted in Figure 1, listing the number of titles per category. Contrary to previous settings reported in [13,14], this sub-collection does not try to constitute a balanced corpus across a large variety of genres. It rather exhibits a strong focus on two categories of music (Electronic and Pop), which are sub-divided in finer granularity. This setting also does not try to limit itself to “clean” and nicely separable genres, resulting in significant challenges during the manual organization phase, with some titles not being straight-forward to classify. The collection also incorporates a number of songs, which are interpreted in different versions, i.e. cover songs, remixes, or new interpretations of songs.

Both map hierarchies evolved to similar structures, having a 2x3 top-layer map with more fine-grained representation provided for all of the top-layer units in the second layer of the hierarchies. Parameters had to be set to slightly different values to account for the huge differences in feature space dimensionality. Due to space considerations we provide an overall evaluation as well as detailed discussion of some clusters. The complete set of experiments is available for interactive exploration at the project homepage at <http://www.ifs.tuwien.ac.at/~andi/somejb>.

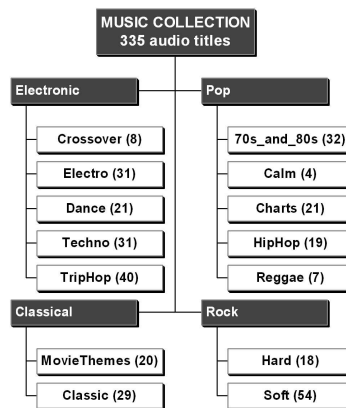


Figure 1: Music Collection

3.2 Rhythm Patterns

Figure 2 depicts the resulting top-layer map based on Rhythm Patterns, listing the percentual distribution of genres across units. The unit on the upper left contains nearly all of the Classical songs in the collection, plus a number of very calm Pop songs (e.g. songs by *Enya*, which can be regarded as very classical-like, or *Frozen* by *Madonna*), and Pop songs from the 70's and 80's genre. Also, some non-aggressive Rock and Electronic songs were classified into this unit. This led to a rather low main-genre-match of 54.7% for the Classical genre. However, considering, that also the songs from the Electronic/TripHop genre like *Lullaby* or *Gorecki* by *Lamb* consist mainly of classical instruments and a strong female voice, the grouping on the top layer can be regarded as rather good. Furthermore, only two songs manually classified as Classical were mapped onto the neighboring unit in the upper right corner, resulting in a almost perfect grouping of 96% of all classic titles within the upper left corner cluster.

The two missing classical titles, i.e. *Orbital - The Saint* and *Gladiator Soundtrack - Main Theme (Trance Mix)* - are both of the sub-category Movie Themes, and are interpretations with electronic elements. Thus, the separation of the Classical songs already on layer 1 is justified. The corresponding second-layer map evolved into 3x2 units. Here, most of the Classical songs are concentrated on the two units on the top row of the map. The two units on bottom of the map contain the rather calm Pop songs or classical-like Electronic or Rock songs, forming the transition to the neighboring second-layer maps.

The aforementioned upper right corner unit on the first layer map consists of a large amount of Rock songs (both Soft and Hard Rock), but also of many Electronic songs (mainly from the TripHop and Electro sub-categories), as well as a number of Pop songs (again from 70's and 80's sub-genre as well as Charts songs). This at first glance rather diverse unit is represented in more detail in the second layer, showing a very nice clustering of artists. The 2 titles by *Kruder & Dorfmeister* in the collection can be found together on the upper left unit. Two other units contain both the album version and the live version of *Alanis Morissette's* songs *Thank You* and *Ironic*, respectively. Also, both songs by Techno group *Dune* were grouped together. Furthermore, 5 of the 6 songs by *Air* in the music collection were grouped to a single unit, with similar collocations present for songs by *Rage against the machine*, *Coldplay*, *R.E.M.*, and others

GHSOM Layer 1 – Rhythm Patterns	
(1/1) 86 tracks CLASSICAL: 54.7 % ELECTRONIC: 8.1 % POP: 25.6 % 70's&80's: 17.4 % ROCK: 11.6 %	(1/2) 124 tracks CLASSICAL: 1.6 % ELECTRONIC: 33.9 % POP: 22.6 % ROCK: 41.9 %
(2/1) 22 tracks ELECTRONIC: 100 % Dance: 36.4 % Techno: 54.5 %	(2/2) 59 tracks ELECTRONIC: 57.6 % TripHop: 22.0 % Electro: 18.6 % POP: 28.8 % ROCK: 13.6 %
(3/1) 23 tracks ELECTRONIC: 100 % Dance: 30.4 % Techno: 56.5 %	(3/2) 21 tracks ELECTRONIC: 14.3 % POP: 76.2 % HipHop: 47.6 % Reggae: 19.0 % ROCK: 9.5 %

Figure 2: Top-Layer Map Rhythm Patterns

GHSOM Layer 1 – MARSYAS	
(1/1) 26 tracks CLASSICAL: 96.2 % POP: 3.8 %	(1/2) 40 tracks CLASSICAL: 30.0 % ELECTRONIC: 32.5 % POP: 35.0 % ROCK: 2.5 %
(2/1) 93 tracks CLASSICAL: 9.7 % ELECTRONIC: 44.1 % Electro: 16.1 % TripHop: 14.0 % POP: 35.5 % 70's&80's: 17.2 % ROCK: 10.8 %	(2/2) 68 tracks CLASSICAL: 1.5 % ELECTRONIC: 42.6 % Electro: 10.3 % TripHop: 11.8 % POP: 29.4 % ROCK: 26.5 % Soft: 17.6 %
(3/1) 73 tracks CLASSICAL: 2.7 % ELECTRONIC: 45.2 % Dance: 11.0 % Techno: 12.3 % TripHop: 12.3 % POP: 19.2 % ROCK: 32.9 %	(3/2) 35 tracks ELECTRONIC: 42.9 % Crossover: 11.4 % POP: 2.9 % ROCK: 54.3 % Soft: 40.0 % Hard: 14.3 %

Figure 3: Top-Layer Map Marsyas Features

The remainder of the layer 1 map is devoted to the Electronic genre (which is obviously due to the domination of this genre in the music collection). The organization of the Rhythm Patterns features shows a division into the Electronic sub-genres as well: the units on the left side of the map represent the songs of the Techno and Dance genres, while the right column contains TripHop. As a global criterion one can hear, that the upper units are rather calm, while the lower units are characterized by remarkable or even aggressive rhythms, following nicely the global organizational principles of the map.

The lower right unit clearly shows a focus on HipHop and Reggae. Even the four “miss-classified” songs *Fettes Brot*, *OPM*, and *Limp Bizkit*, that were manually categorized to Pop/Charts and Rock/Soft, are indeed very near at the border to HipHop or even performed by HipHop-bands.

3.3 Marsyas-based features

The map resulting from the Marsyas-based features, depicted in Figure 3, evolved to a 3x2 top-layer map, with all units being expanded in a second layer. Again, the unit on the upper left corner nearly completely consists of only classical songs - with just one exception: *Neneh Cherry - Manchild (Massive Attack Mix)*, resulting in a genre-match of 96.2% in this unit. However, only 25 out of 49 classical titles are co-located on this cluster, with the remaining titles being quite wide-spread across all but one of the remaining units, resulting in a less-consistent structure. The two movie themes separated from the strictly classical cluster in the Rhythm Pattern map are again consistently located separately in this map hierarchy.

The unit on the upper right represents, to almost equal shares, Classical, Pop and Electronic music. Besides the Classical songs, also the Pop songs are of a rather slow and calm character (songs by *Enya* and songs from the 70s_and_80s, e.g. *Bette Midler - The Rose*). However, there are also a number of songs with a rather aggressive beat, like *Moby - Go! (Remix)* and the Techno song *Kai Tracid - Your own reality* as well as some TripHop songs which have a distinct rhythm. It is not quite understandable, why these songs have been classified to this unit.

Again, the remaining units are dominated by Electronic songs, with the two units in the center containing also numerous Pop songs and the two units on the bottom containing more Rock songs. However, no clear structure of genre classification can be recognized. Investigating the sub-genres, a focus on HipHop and Reggae can be seen in the center-right unit. TripHop & Electro can be found with together 30.1% in the center-left unit, but unfortunately also in most of the other units. A slight accumulation of Techno and Dance songs can be seen in the lower left unit, but this unit, in contrast, also contains numerous Soft and Hard Rock songs. Overall, 4 of the 6 units in layer 1 contain titles from all 4 main categories. The Pop category is present in every single unit. As another example, the 12 songs of the band *Rage against the machine*, which have rather strong Rock characteristics with extensive use of electrical guitars, are divided upon 4 units: 5 songs can be found in the center-right unit, 5 songs in the lower-left unit, and 1 in the center-left and lower right units, respectively.

4. Conclusions

We have presented preliminary experiments comparing two different feature sets with respect to their suitability of providing genre-based organization of music archives. Using the *SOMEJB* system, a collection of music can be organized according to the mutual similarity of the individual pieces. The resulting maps enable the user to browse and explore the collection, to select a style of music to listen to by defining a cluster or region from which to play the individual pieces. Although the Rhythm Patterns capture only a small subset of the characteristics of music, they allow for a surprisingly good organization according to perceived acoustic similarity. This may be attributed to the large range of rhythmic information up to 600 bpm and the complex patterns present in the individual frequency bands, that are captured by the proposed model, taking it beyond mere beat information. Yet, the integration of additional features capturing aspects that go beyond rhythmic characteristics seems very promising.

While the organizational principles of the *SOMEJB* system provide a promising interface to music collections, recent experiments

have again highlighted the need for more formal evaluation of the impact of the different feature sets. Current work thus focuses on the creation of an evaluation framework using supervised learning techniques to analyze classification performance. This should, in turn, lead to an improved understanding of feature spaces and improved weighting of their individual contribution to provide a cleaner clustering and thus better browsing interfaces. It will further allow us to evaluate the characteristics of different types of genre hierarchies.

Bibliography

- 1 M. Dittenbach, A. Rauber, and D. Merkl.
Uncovering hierarchical structure in data using the growing hierarchical self-organizing map.
Neurocomputing, 48(1-4):199-216, September 2002.
- 3 B. Feiten and S. Günzel.
Automatic indexing of a sound database using self-organizing neural nets.
Computer Music Journal, 18(3):53-65, 1994.
- 4 J. Foote.
An overview of audio information retrieval.
Multimedia Systems, 7(1):2-10, 1999.
- 5 T. Kohonen.
Self-organizing maps.
Springer-Verlag, Berlin, 1995.
- 8 E. Pampalk, A. Rauber, and D. Merkl.
Content-based organization and visualization of music archives.
In *Proceedings of ACM Multimedia 2002*, pages 570-579, Juan-les-Pins, France, December 1-6 2002. ACM.
- 10 A. Rauber and M. Frühwirth.
Automatically analyzing and organizing music archives.
In *Proceedings of the 5th European Conference on Research and Advanced Technology for Digital Libraries (ECDL 2001)*, Springer Lecture Notes in Computer Science, Darmstadt, Germany, Sept. 4-8 2001. Springer.
- 12 A. Rauber and D. Merkl.
Text mining in the somlib digital library system: The representation of topics and genres.
Applied Intelligence, 18(3):271-293, May-June 2003.
- 13 A. Rauber, E. Pampalk, and D. Merkl.
Using psycho-acoustic models and self-organizing maps to create a hierarchical structuring of music by musical styles.
In *Proceedings of the 3rd International Conference on Music Information Retrieval*, Paris, France, October 2002.
- 14 A. Rauber, E. Pampalk, and D. Merkl.
The SOM-enhanced JukeBox: Organization and visualization of music collections based on perceptual models.
Journal of New Music Research, 2003.
- 16 C. Spevak and E. Favreau.
Soundspotter - a prototype system for content-based audio retrieval.
In *Proceedings of the 5. International Conference on Digital Audio Effects (DAFx-02)*, Hamburg, Germany, September 26-28 2002.
- 19 G. Tzanetakis and P. Cook.
Musical genre classification of audio signals.
IEEE Transactions on Speech and Audio Processing, 10(5), July 2002.
- 21 E. Wold, T. Blum, D. Keislar, and J. Wheaton.
Content-based classification search and retrieval of audio.
IEEE Multimedia, 3(3):27-36, Fall 1996.
- 23 E. Zwicker and H. Fastl.
Psychoacoustics, Facts and Models, volume 22 of *Series of Information Sciences*.
Springer, Berlin, 2 edition, 1999.