# Marsyas and Rhythm Patterns: Evaluation of two music genre classification systems

Thomas Lidy
Dept. of Software Technology
Vienna University of Technology
tom@student.ifs.tuwien.ac.at

## ABSTRACT

Within the last years several technologies for automatic genre- or content-based music classification have been developed. Two of them are the Rhythm Patterns, which were developed by Andreas Rauber et al. within the SOMeJB system, and the MARSYAS system by George Tzanetakis. While delivering promising results, none of them can at present be regarded as perfect. This paper describes the evaluation of both of the aforementioned systems and discusses the classification results of each of them individually as well as the differences (strengths and weaknesses) of the systems in a comparison.

## 1 INTRODUCTION

More and more people today face the challenge of organizing large music archives. While one reason for the widespread use of music collections is the still growing number of file-sharing activities, there are also professional applications: As an example, there is the film industry, which has to retrieve music and other audio titles from large archives in every day work.

Most of these archives until today are organized manually, either by using artist-based listings or by grouping the pieces manually to music genres or categories. This can be a very challenging task, as it involves listening to and comparing all songs and making sometimes very ambiguous decisions. Also, the result is often a division to several hundred categories, which makes an efficient retrieval impossible.

Thus, there is a need for an automatic genre-based music classification, which will make retrieval of songs with similar styles or characteristics much easier. That approach will relieve the user of providing information about an artist or title, when he or she is looking for certain kinds of music. While bringing advantages in grouping and organizing music archives, this technology will also be useful when searching for new music of a specific style without knowing a certain artist or title.

Within the last years much research has been done to develop systems that allow searching for audio titles without having to provide textual information. There are several research directions: There are systems, that search for audio based on a piece of music the user provides as input or reference [Welsh et al., 1999], as well as systems that allow searching for music by simply humming a melody [Ghias et al., 1995]. We concentrate on systems, that analyze and classify music based on perceived sound similarity through feature extraction.

This paper describes the evaluation and comparison of the Rhythm Patterns, which are part of the SOM enhanced Jukebox (SOMeJB) system, developed by A. Rauber and M. Frühwirth [Rauber and Frühwirth, 2001] and presented at the International Conference on Music Information Retrieval (ISMIR) [Rauber et al., 2002], and the MARSYAS system developed by George Tzanetakis [Tzanetakis and Cook, 2000]. Both systems solely rely on information which is present in the actual waveform representation of a music file, and both of them use psycho-acoustically motivated feature extraction. After feature extraction, the resulting feature vectors are passed as input to the GHSOM system (growing hierarchical self organizing map) [Dittenbach et al., 2002], which does hierarchical structuring of the provided data, and presents the results of the classification to the user.

Differences in the feature extraction process of the Rhythm Patterns and the MARSYAS system will be described in section 3 of this paper. Before that, I will give an overview of related work in this area in section 2. Section 4 will present the results of the evaluation of both of the systems and will show a comparison of them as well. Conclusions and suggestions for future research will be given in section 5.

## 2 RELATED WORK

Automatic music genre classification is a strongly investigated field. Numerous work has been done in this area, and a number of techniques are presented in an overview in [Foote, 1999]. The Rhythm Patterns feature extraction system is described in detail in [Pampalk et al., 2002]. A description of the MARSYAS system can be found in [Tzanetakis and Cook, 2000]. Further algorithms for automatic genre categorization of audio signals are described in [Tzanetakis et al., 2001]. A comparison of four classifiers for content-based audio classification

together with experiments has been conducted in [Liu and Wan, 2001].
Elias Pampalk did an extensive work about the Rhythm Patterns feature extraction in his Master Thesis [Pampalk, 2001], which also contains an evaluation of the system as well as the description of an intuitive user interface for the system.
For the evaluation of automatic genre classification, often a manual organization into genres is needed. This can lead to rather subjective and ambiguous classification. In [Pachet and Cazaly 2000] a taxonomy is presented that aims to be as much objective as possible.

# 3 IMPLEMENTATION DETAILS

Both of the systems under investigation were evaluated using MP3 compressed audio in stereo format with 16 bit, 44 kHz sampling frequency as input. In order to be able to do feature extraction, the compressed audio signal had to be decoded into the raw waveform (i.e. Pulse Code Modulation (PCM) audio format). To reduce the amount of data, in both systems the audio data is downsampled before doing feature extraction.

## 3.1 Feature Extraction with Rhythm Patterns

The Rhythm Patterns system uses an implementation in MATLAB for feature extraction. Before loading the waveform into the environment, the audio signal is downsampled to 11 kHz/mono format. This significantly reduces the data volume, while still preserving the specific characteristics of the audio title. First, the audio title is split in 6-seconds sequences, and only a subset of these sequences is used for audio analysis. For each of the remaining sequences, the raw audio data is decomposed into frequency ranges using a discrete Fourier transform. From these frequency ranges, so-called 'critical bands' are extracted. The 25 critical bands are measured in the unit 'Bark' and are based on psychoacoustic models (the human ear can distinguish certain frequencies better than others). Only 20 of the 25 critical bands are used in further processing. Spectral Masking effects (the occlusion of a quiet sound by a louder sound when simultaneously present) are reduced . The audio data is transformed in several steps into decibel, Phon and Sone. Phon defines the loudness level of a tone based on its frequency. Finally, from the loudness levels, the specific loudness sensation per critical band is calculated, which is measured in Sone and represents the audio signal closely as it is perceived by the human ear.
Afterwards, reoccurring patterns in the individual critical bands are extracted by transforming the individual frequency spectra back into the time domain using again a discrete Fourier transform. Depending on the frequency, the amplitude modulations have different effects on human sensation. The most significant one is fluctuation strength, which is most intense at a modulation frequency of around 4 Hz and decreases towards 15 Hz. At 15 Hz the sensation of roughness starts, and at about 150 Hz the sensation of hearing three separately audible tones increases. It is the fluctuation strength, i.e. the rhythm patterns up to 10 Hz (which corresponds to 600 bpm) that the Rhythm Patterns system is focussed on. For each of the 20 frequency bands, a fixed number of values for modulation frequencies between 0 and 10 Hz is obtained. Thus, the system captures significantly more data than what would conventionally be considered as pure rhythm. The values obtained form the feature space for subsequent analysis. In the conducted experiments, this was a 600-dimensional feature vector per audio title. The feature vectors of all titles were then presented to the GHSOM system for building a hierarchical classification.
More details about the Rhythm Patterns feature extraction process can be investigated in [Pampalk et al., 2002].

## 3.2 Feature Extraction with MARSYAS

Before feature extraction with MARSYAS is done, the audio files are downsampled to 22 kHz/mono. In contrast to the Rhythm Patterns system, MARSYAS supports multiple different kinds of feature extraction. The one used in my experiments is the feature extractor 'Genre'. The 'Genre' extractor uses different kinds of features for finding a proper genre classification:

FFT: This set of features is based on the Short Time Fourier Transform (STFT). They have been designed for Music/Speech classification. More specifically they are the means and variances of the spectral centroid, rolloff, flux and zerocrossings calculated every 20 ms. The centroid is the balancing point of the spectrum (the frequency where the energy of all frequencies bellow that frequency is equal to the energy of all frequencies above that frequency) and is a measure of brightness and general spectral shape. Another measure of spectral shape is the rolloff which is the 90 percentile of the power spectral distribution (the centroid would be the 50 percentile). This is a measure of the „skewness" of the spectral shape. Flux is the 2-norm of the difference between the magnitude of the Short Time Fourier Transform (STFT) spectrum evaluated at two successive sound frames. The STFT is normalized in energy. In addition another feature called low energy is used. Low energy is the percentage of 20 ms windows that have energy less than the average energy of the 1 second window.

SVFFT: Single vector version of the FFT feature extractor. Basically the mean across the whole file of the feature vectors (the mean of all the rows of the Feature Matrix).

MFCC: Mel-Frequency cepstral coefficients are perceptually motivated features used in Speech Recognition research.

MPITCH: A set of features for representing harmonic content based on multiple pitch analysis algorithm.

BEAT: This set of features is for the representation of the beat structure of music calculated using a beat detection algorithm based on the Discrete Wavelet Transform.

These different vector sets together build the feature vectors that are used for the GHSOM training process.
Note that we don't use the classifiers, that are incorporated in the MARSYAS framework itself. Details about the MARSYAS framework can be found in [Tzanetakis and Cook, 2000].

### 3.3 The GHSOM training process

The Growing Hierarchical Self Organizing Map is an extension to the Self Organizing Map (SOM). The SOM is a neural network model, that is widely used for high-dimensional feature space representation. However, with increasing amount of information, the SOM reaches some limitations: The size of the map has to be determined in advance ignoring the characteristics of the data distribution in the input data. Additionally the inherent hierarchical structure of the data is lost. The GHSOM addresses these limitations and extends the SOM. First, it is capable of identifying the required number of units during its unsupervised learning process. Second, the data set is clustered hierarchically by introducing a layered architecture. The layered architecture compromises a number of independent self-organizing maps within each layer, thus clustering the data hierarchically. Starting from a top-level map, each map grows in size in order to represent a collection of data at a particular level of detail. As soon as a certain improvement of the granularity of data representation is reached, the units are analyzed to see whether a further level of detail is necessary. Those units that have too diverse input data mapped onto them are expanded to form a new SOM at a subsequent layer. Two important parameters that are responsible for the granularity of the maps can be set as properties for the GHSOM (Tau_1 and Tau_2). It is important to mention, that the overall topology of the data is still preserved at subsequent layers. A detailed explanation of the GHSOM is given in [Dittenbach et al., 2002].

## 4 EVALUATION

### 4.1 The Music Collection

I used a music collection of 335 songs in MP3 format for the evaluation of the Rhythm Patterns and the MARSYAS feature extraction systems. The songs are spread over 14 genres within 4 main categories. This two-layered genre pre-organization was done completely manually by me, and is of course somewhat subjective. However, I made an organization that would at least fit the needs for our evaluation purpose, enabling a fast and easy overview at the first layer and investigating details at the second layer. The genre-subdivision can be seen in figure 4.1 together with the amount of music files in each genre sub-category. As you may notice, there is a modest focus on electronic songs. The category of electronic music is kind of diverse and is thought to be an interesting challenge for the music classification. Also the genre 'TripHop' is by far not a genre that is easy to classify, however, I wanted to test the two systems with songs that are rather hard to classify, too.

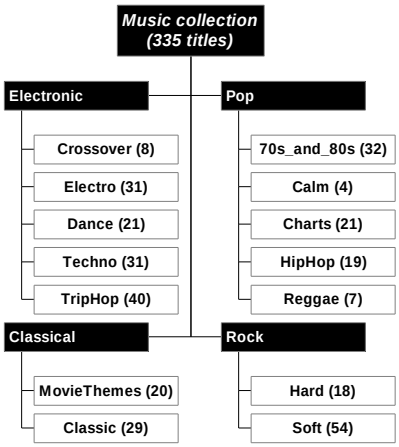


Figure 4.1: The music collection used for the experiments: 335 songs, that were manually organized in 14 genres within 4 main categories. (The numbers in brackets indicate the number of songs in each genre).

The music collection also incorporates a number of songs, which are interpreted in different versions, i.e. so-called Cover Songs, Remixes or new interpretations of songs. There is, for example, the 1980's version of the song 'Leuchtturm' by German singer 'Nena' and her new "revival" release of 'Leuchtturm' from 2003, or some live versions of songs by 'Alanis Morissette'. Table 4.5 gives a list of songs in the collection, that are included in different versions. Some of these songs have rather similar characteristics, while others are of rather different music types, which will hardly be classified together. It is interesting, if any of the classification systems will be able to recognize similarities between these versions.

As we cannot depict the complete GHSOM hierarchies here in this paper, you can investigate the evaluation results interactively on the following URL: http://student.ifs.tuwien.ac.at/~tom/sem2003/

## 4.2 Evaluation of Rhythm Patterns

In the first experiment, features of the 335 music files were extracted by the Rhythm Patterns implementation in MATLAB. Each of the songs is represented by a 600-dimensional feature vector containing the rhythmic characteristics of a number of important frequency bands (see section 3.1).

```
EXPAND_CYCLES=8
MAX_CYCLES=0
TAU_1=0.1
TAU_2=0.0035
INITIAL_LEARNRATE=0.7
INITIAL_NEIGHBOURHOOD=4
HTML_PREFIX=coll_tml_
DATAFILE_EXTENSION=.mp3
randomSeed=17
inputFile=../all_sorted.vec
descriptionFile=../somejb600.tv
savePath=.
printMQE=false
normInputVectors=NONE
saveAsHTML=true
saveAsSOMLib=true
INITIAL_X_SIZE=2
INITIAL_Y_SIZE=2
LABELS_NUM=0
LABELS_ONLY=true
LABELS_THRESHOLD=0.35
ORIENTATION=true
```

Table 4.2: The parameters of the property file for the training of the extracted Rhythm Pattern features with the GHSOM

The Rhythm Patterns vector file was trained with the GHSOM system for automatic hierarchical classification. The parameters that were used for the GHSOM can be seen in table 4.2. We will now investigate the results in layer 1 and layer 2 of the GHSOM. The layer 1 map has grown to a size of 3 x 2 units. It represents a rather coarse classification of the music collection, but nevertheless already showing a proper genre-classification. The percentage of each of the 4 main categories in each of the units in the layer 1 map can be seen in detail in figure 4.3.

**Layer 1**

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, but 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 classification which was done on the upper left unit can be regarded as rather good. Only a few songs, like e.g. 'Neneh Cherry – Manchild (Massive Attack Mix)' don't really fit into this unit. On the other hand, only a total of 2 manually "Classical" categorized songs were classified to another unit, namely to the unit on the upper right. The 2 songs – 'Orbital - The Saint' and 'Gladiator Soundtrack - Main Theme (Trance Mix)' – are both of the sub-category Movie Themes, and as the name of the latter already implies, are interpretations with electronic elements. Thus, the separation of the Classical songs on layer 1 is fully understandable.

| GHSOM Layer 1 – Rhythm Patterns | |
|---|---|
| (1/1) 86 tracks<br>**CLASSICAL**: 54,7 %<br>ELECTRONIC: 8.1 %<br>POP: 25.6 %<br>70's&80's: 17.4 %<br>ROCK: 11.6 % | (1/2) 124 tracks<br>CLASSICAL: 1.6 %<br>ELECTRONIC: 33.9 %<br>POP: 22.6 %<br>**ROCK**: 41.9 % |
| (2/1) 22 tracks<br>**ELECTRONIC**: 100 %<br>Dance: 36.4 %<br>Techno: 54.5 % | (2/2) 59 tracks<br>**ELECTRONIC**: 57.6 %<br>TripHop: 22.0 %<br>Electro: 18.6 %<br>POP: 28.8 %<br>ROCK: 13.6 % |
| (3/1) 23 tracks<br>**ELECTRONIC**: 100 %<br>Dance: 30.4 %<br>Techno: 56.5 % | (3/2) 21 tracks<br>ELECTRONIC: 14.3 %<br>**POP**: 76.2 %<br>HipHop: 47.6 %<br>Reggae: 19.0 %<br>ROCK: 9.5 % |

Figure 4.3: An overview of the classification results of the Rhythm Patterns system at GHSOM layer 1 (3 x 2 units): unit reference (row/column), number of total tracks in each unit together with percentages of the 4 main categories, with some percentages of interesting sub-genres

The upper right unit 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). Being rather mixed unit (containing 124 songs), we will go into details in this upper right unit later.

The remainder of the layer 1 map is dominated by the Electronic genre (which is obviously due to the domination of this genre in the music collection). Nevertheless, the classification of the Rhythm Patterns features shows a division into the Electronic sub-genres as well: Units 2 and 3 on the left side represent the songs of the Techno and Dance genres, while the right column focuses on TripHop. As a global criterion one can identify, that the upper units are rather calm, while the lower units are characterized by remarkable or even aggressive rhythms.

The lower right unit clearly shows a focus on HipHop and Reggae. Even the 4 songs, 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 ('Fettes Brot', 'OPM', 'Limp Bizkit'). Note that the HipHop and Reggae unit lies on the same side as the units with the TripHop songs; these genres have a specialized rhythm in common.

As a conclusion, one can say, that the Rhythm Patterns system is very good in detecting music genres with special rhythms: It separates Electronic songs into Techno&Dance and TripHop(&Electro) and it recognizes HipHop&Reggae as a separate genre. Besides, the recognition of Classical tunes was done very well.

**Layer 2**

All of the layer 1 units were further expanded to layer 2 maps. We will now have a closer look to some of these maps. When talking about a layer 2 map, I refer to its unit on layer 1 in a (row/column) notation, starting with (1/1) in the upper left corner.

As already mentioned, unit (1/1) contains all but 2 of the Classical songs in the collection. Regarding the layer 2 map of this unit, which is a 3 x 2 map, we can see, that most of the Classical songs are concentrated on the 2 units on the top of the map. The 2 units on bottom of the map contain the rather calm Pop songs or classical-like Electronic or Rock songs of unit (1/1).

The layer 2 map of unit (1/2) shows a very nice clustering of interprets. For instance, 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 (the 6th song was contained in unit (1/1)). Songs by 'Rage against the machine' can be found in 4 units – containing 6, 2, 1 and 1 song respectively. 2 songs each by 'Coldplay', by 'R.E.M.' , by 'Linkin Park' and by the 'Sugababes' are contained in one unit. On the layer 2 maps of unit (2/1) and unit (3/1) we can see a similar grouping with 'DJ Sammy feat. Carisma', 'Paul van Dyk', 'Prince Ital Joe feat. Marky Mark' and 'Paul Oakenfold'. Obviously, although many people say that especially songs from the Dance and Techno genres sound very similar, the Rhythm Pattern system manages to find differences between the interprets.

We will now investigate layer 2 of units (2/2) and (3/2). We already identified the latter as the HipHop & Reggae unit. When we have a closer look on layer 2 of unit (2/2) we see, that the lower left unit also contains 27.3 % HipHop and Reggae songs. This shows the topology preservation in the GHSOM training process on subsequent layers.

In other units on this map, we can again find a clustering of interprets: 2 songs by 'Kosheen', 2 songs by the 'Sugababes', 2 songs by 'Morcheeba' etc. can be found grouped together.

Besides these positive results, the negative aspect is, that in spite of a very nice clustering of interprets, there is still a mixture of genres on many units.

## 4.3 Evaluation of MARSYAS

The second experiment was the evaluation of the classification of the 335 music files through the MARSYAS system. Again, features of the songs were extracted independently of classification, and stored in vector files. The MARSYAS feature vectors only have 30 dimensions and thus, the vector file is much smaller than the Rhythm Patterns vector file.

This vector file was trained with the GHSOM in order to receive a hierarchical genre classification. The parameters for the GHSOM were set similar to the ones in the previous experiment (see Table 4.2). Input Vectors now have to be normalized (because of the different features MARSYAS extracts) and TAU_1 was set to a smaller value (0.03), to get a smaller layer 1 map, which ought to be comparable to the layer 1 map of the Rhythm Patterns result. The resulting map in layer 1 has 3 x 2 units, too. We will now investigate the layer 1 map.

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

Figure 4.4: An overview of the classification results of the MARSYAS system at GHSOM layer 1 (3 x 2 units): unit reference (row/column), number of total tracks in each unit together with percentages of the 4 main categories, with some percentages of interesting sub-genres

**Layer 1**

The first positive aspect is, that the unit on the upper left corner nearly completely consists of only Classical songs – with just one exception: 'Neneh Cherry – Manchild (Massive Attack Mix)'. This gives a genre-match of 96.2 % in this unit, which is a very good result. The negative aspect is, that also both the units to the right and to unit beneath the first unit contain numerous Classical songs. Additionally, 2 other units contain one resp. two classical songs, however these are MovieThemes that partly contain electronic elements: the theme from the TV series 'Babylon 5', resp. 'Orbital - The Saint' and 'Gladiator Soundtrack - Main Theme

(Trance Mix)', which were also separated from the other Classical songs in the Rhythm Patterns experiment. Nevertheless, the main genre of Classical songs is spread over 5 of the 6 units in layer 1 map.
The unit on the upper right consists nearly to a third of the categories Classical, Pop and Electronic music, each. 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. An explanation for instance for the 'Kai Tracid' song could be, that this song has a rather varying structure, and has some parts without beats.
The unit on the lower right focuses on Rock songs (54.3 %), but is also mixed together with Electronic songs (42.9 %). Also the remaining units are dominated by Electronic songs, with the 2 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 very slight accumulation of Techno & Dance songs can be seen in the lower left unit, but this unit, in contrast, also contains numerous Soft & 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! Numerous genres are widely spread over the map and a comprehensive classification cannot be recognized. As a further 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, each. Obviously, the MARSAYS system has some problems distinguishing (rather powerful) Rock songs from Techno songs or Electronic songs in general.

**Layer 2**
In the maps of layer 2 of the MARSYAS GHSOM there are, at some positions, results with a clearer topology. In the layer 2 map of the upper left unit (i.e. unit (1/1)), the only non-Classical title ('Neneh Cherry – Manchild (Massive Attack Mix)') was classified alone into a single unit (unfortunately between other Classical units). The Classical titles in unit (2/1) can be found rather on top/left of its layer 2 map, which shows, that the GHSOM training process retains the topology of data in sub-layers. Unfortunately, this cannot be said for the unit (1/2) where the Classical songs are still totally mixed with Pop songs. In unit (2/1) again, we can observe the grouping of some interprets into single units. For example, three songs by the band 'Air' are classified into one unit, two other songs by the same band into another unit (unfortunately not next to each other). Also two songs by the band 'Lamb' can be found in a unit, as well as two songs by 'Paul Oakenfold'. While the top/left of unit (2/1) focuses on Classical songs, a focus on Electronic songs can be observed at the bottom of this unit's layer 2 map. Furthermore, it is interesting that on the 4 units on the right side, there is a remarkable concentration on HipHop songs. Regarding the layer 2 of unit (2/2), here a cluster with only HipHop songs can be found at the left side (which again shows the topology preservation of the GHSOM). On unit (2/2) we can again see the clustering of several songs of certain interprets: there is a 100 % Rock cluster at the top containing 4 songs by 'Rage against the machine' and 1 song by 'R.E.M.'. There is a 100 % Electronic cluster at the lower left and 3 songs by Reggae singer 'Patrice' are grouped together in a unit on the second row from top, on the right. Unfortunately, other interprets have not been clustered in such a nice way: the songs by 'Alanis Morissette' are spread over 3 positions in unit (2/2), 2 further positions in unit (3/1) and 1 position in unit (2/1). The songs played by 'Moby' are spread in a similar way.
As we already noticed on layer 1, the lower right unit (i.e. unit (3/2)) focuses on Rock songs. At some positions on layer 2 of this unit, we can observe a separation of the main genres: There are two 100 % Rock units, and three 100 % Electronic units. Unfortunately the remaining units are still very mixed up (4 times even 50/50 %).

## 4.4 Comparison

The experiments delivered better results for the Rhythm Patterns feature extraction than for the MARSAYS system. Both systems generated a noticeable Classical cluster, however, while in the Rhythm Patterns GHSOM all but 2 Classical songs were clustered into one layer 1 unit, the Classical songs are divided over 5 of the 6 units on the MARSYAS GHSOM. Although both systems mix Rock, Electronic and Pop categories very much, we can recognize more structure in the Rhythm Patterns results. The Rhythm Patterns layer 1 map contains 2 units with 100 % Electronic music and one unit with nearly fully HipHop & Reggae songs. A few slight clusters are noticeable in the MARSYAS result, too, but the overall structure is rather messed up: 4 of the 6 units in the layer 1 map contain songs from all of

| Song Title | Genre | Song Title 2 | Genre | RP | | MA | |
|---|---|---|---|---|---|---|---|
| | | | | L1 | L2 | L1 | L2 |
| Bangles – Eternal flame | 70s_and_80s | Atomic Kitten – Eternal flame | Charts | 1.5 | - | 1 | - |
| Nena – Irgendwie, Irgendwo, Irgendwann | 70s_and_80s | Jan Delay – Irgendwie, Irgendwo, Irgendwann | HipHop | 0 | 1 | 0 | 1 |
| Nena – Leuchtturm | 70s_and_80s | Nena – Lechtturm (2003 version) | Charts | 0 | 2 | 1 | - |
| Mr. Mister - Broken Wings | 70s_and_80s | 2Pac - Until The End Of Time | HipHop | 2 | - | 1.5 | - |
| Duran Duran - Ordinary World | 70s_and_80s | Aurora - Ordinary World | Dance | 1 | - | 1 | - |
| Sister Sledge - We Are Family | 70s_and_80s | Sister Sledge Remix - We Are Family (mix) | 70s_and_80s | 0 | 0 | 1 | - |
| Neneh Cherry - Manchild | Charts | Neneh Cherry - Manchild (Massive Attack Mix) | Charts | 1 | - | 1 | - |
| DJ Sammy feat. Carisma - Sunlight | Dance | DJ Sammy feat. Carisma – Sunlight (Radio edit) | Dance | 0 | 0 | 1.5 | 2.5 |
| Cory Hart - Sunglasses At Night | Electro | Tiga & Zyntherius - Sunglasses at Night | Electro | 1 | - | 1 | b! |
| Moby - James Bond Theme | Electro | Moby - James Bond Theme (Techno Remix) | Electro | 0 | 0 | 1.5 | - |
| Gladiator Soundtrack - 17 - Now We Are Free | MovieThemes | Gladiator Soundtrack - Main Theme (Trance Mix) | MovieThemes | 1 | - | 2.5 | - |
| Alanis Morissette - Ironic | Rock/Soft | Alanis Morissette - Ironic (MTV Unplugged) | Rock/Soft | 0 | 0 | 1 | - |
| Alanis Morissette - Thank You | Rock/Soft | Alanis Morissette - Thank You (Live Woodstock99) | Rock/Soft | 0 | 0 | 1.5 | - |
| Alanis Morissette - You Oughta Know | Rock/Soft | Alanis Morissette - You Oughta Know (Acoustic live) | Rock/Soft | 1 | b! | 1 | - |
| Moby - Feeling So Real (Radio Edit) | Techno | Moby - Feeling so Real (Westbam Mix) | Techno | 1.5 | - | 1 | - |
| Moby - Go (From Hackers Soundtrack) | Techno | Moby - Go (Remix).mp3 | Techno | 1 | - | 1 | - |
| Paul van Dyk - We are alive | Techno | Paul van Dyk - We are alive (Radio Mix) | Techno | 0 | 0 | 0 | 1 |
| Safri Duo - Played-A-Live (Original Club Mix) | Techno | Safri Duo - Played-A-Live (Radio Edit) | Techno | 0 | 1 | 1 | - |
| Lamb - Sweet | TripHop | Lamb - Sweet (Soulchild radio mix) | TripHop | 0 | 0 | 1 | b! |
| Sneaker Pimps - Six Underground | TripHop | Sneaker Pimps - Six Underground (Paul Oakenfold Remix) | TripHop | 1 | - | 1.5 | - |
| Sneaker Pimps - Post-Modern Sleaze | TripHop | Sneaker Pimps - Post-Modern Sleaze (Roni Size & DJ Die Remix) | TripHop | 2 | - | 1.5 | - |
| Sneaker Pimps - Spin Spin Sugar (Moby Remix) | TripHop | Sneaker Pimps - Spin Spin Sugar (Jungle Mix) | TripHop | 0 | 1 | 2.5 | - |

Table 4.5: Songs that are contained in the music collection in different versions, plus manual genre assignment. Distance measures for Rhythm Patterns (RP) and MARSYAS (MA) on GHSOM layers 1 and 2.

the 4 genre main categories, and the Pop category is present in every unit on the map.

While, especially on the layer 2 maps, in the Rhythm Patterns GHSOM we can observe a very nice grouping of interprets, most of the interprets are widely spread over the MARSYAS GHSOM. For example, the 12 songs of the band 'Rage against the machine' are divided into 4 units in MARSYAS layer 1; in the Rhythm Patterns GHSOM, we can find all songs in units (1/2) and (2/2). It is also noticeable, that the Rhythm Patterns feature extraction is sensible to special characteristics of certain songs: It is very good in detecting similarities in songs of a certain artist and even similarities in interpretations of the same song by different artists. For instance, in the Rhythm Patterns result, the song 'Irgendwie, irgendwo, irgendwann' by 'Nena' and the new interpretation by HipHop singer 'Jan Delay' are on the same layer 1 unit and lie on units next to each other on layer 2!

This leads us to a comparison of the songs, that are contained in different versions in the music collection, which were already mentioned in section 4.1. Table 4.5 shows a list of these songs. I evaluated the results of the classification of these songs by using a distance measure in terms of units on the GHSOM maps. When 2 songs are found on the same unit, the distance measure is 0. If they are found on units next to each other, the distance is 1. A distance of 1.5 means, that the units are diagonal neighbors, etc. In the last 4 columns of Table 4.5, the distance measures for layer 1 and 2 for the Rhythm Patterns (RP) and the MARSYAS system (MA) are listed. Whenever two song versions indicate a distance of 0 on layer 1 as well as on layer 2, this means, that they were classified together into a single unit on layer 2. Songs that are to be found on different units on layer 1 already, don't have a layer 2 distance measure – with one exception: If they were found on layer 2 units, that both lay at the border to the corresponding layer 1 unit, they are marked in the layer 2 column with 'b!'. This is again a sign of the topology preservation in the GHSOM hierarchy and should be noted here.

As we see, the comparison of the classification of songs that were present in two different version, shows significantly better results for the Rhythm Patterns system. While in the Rhythm Patterns classification, 7 of the 22 'two-version songs' were classified together on a single unit on layer 2, none of them at all were found together in a unit on layer 2 in the MARSYAS classification. Moreover, in the MARSYAS classification only two of the songs were even on the same layer 1 unit! Generally, the MARSAYS classification in this comparison only shows better results on 3 positions in the given table.

# 5 CONCLUSIONS & SUMMARY

Some conclusions already have been drawn during evaluating the results in section 4. Generally, in these experiments the Rhythm Patterns system delivered much better results than the MARSYAS system.

The Rhythm patterns system, though only extracting features concerning the presence of repeating fluctuations on certain frequencies in the audio signal, delivers rather good results in dividing the audio titles into 4 main genre categories. Classical songs are very well separated from other titles, also 2 clusters with only Electronic songs have been built. Once again, I want to mention the remarkable recognition of HipHop & Reggae titles by the Rhythm Patterns. However, there are still some mixtures of Electronic with Rock and Pop songs. Some of them may be due to songs, that have varying characteristics over the time. This problem

can be addressed by evaluating several pieces of each audio title separately. But in general, the problem could be that the Rhythm Patterns system is incapable of differentiating between Electronic beats and e.g. Rock guitars when occurring at the same rhythm interval.
The MARSYAS system delivered quite unstable results. Though extracting different kinds of features (see section 3.2), a straight-forward genre separation could not be recognized in the resulting MARSYAS GHSOM. Only the Classical genre built a cluster of its own. All other genres were pretty mixed up. A rhythm (or beat) detection on multiple frequencies should deliver better results. Maybe simply more or others of the feature sets that are included in the MARSYAS system should be combined for a better genre classification.

## 6 ACKNOWLEDGEMENTS

I wish to thank Andreas Rauber for his detailed introduction to the SOMeJB system and the Rhythm Patterns and his ongoing support and motivation during the evaluation process.

## 7 REFERENCES

[Dittenbach et al., 2002]
Dittenbach, M., Rauber, A., & Merkl, D. (2002). Uncovering hierarchical structure in data using the growing hierarchical self-organizing map. Neurocomputing, 48:199-216, September 2002.

[Foote, 1999]
Foote, J. An overview of audio information retrieval. Multimedia Systems 1999. 7(1), 42-51.

[Ghias et al., 1995]
Ghias, A., Logan, J., Chamberlin, D., & Smith, B.C. (1995). Query by humming: Musical information retrieval in an audio database. In Proceedings of the Third ACM International Conference on Multimedia, pages 231-236, San Francisco, CA, November 5 - 9 1995. ACM.

[Liu and Wan, 2001]
Liu, M. & Wan, C. (2001). A study on content-based classification and retrieval of audio database. In Proceedings of the 5th International Database Engineering and Applications Symposium (IDEAS 2001), Grenoble, France, 2001. IEEE.

[Rauber and Frühwirth, 2001]
Rauber, A. & Frühwirth, M. (2001). 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.

[Rauber et al., 2002]
Rauber, A., Pampalk, E. & and Merkl, D. (2002). 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, pages 71-80, Paris, France, October 13-17 2002.

[Pachet and Cazaly 2000]
Pachet, F., Cazaly, D. A Taxonomy of Musical Genres, Contend-Based Multimedia Information Access Conference (RIAO), Paris, April 2000

[Pampalk, 2001]
Pampalk, E. Islands of Music: Analysis, Organization, and Visualization of Music Archives, Master Thesis, Department of Software Technology and Interactive Systems, Vienna University of Technology, December 2001.

[Pampalk et al., 2002]
Pampalk, E., Rauber, A., Merkl, D. Content-based Organization and Visualization of Music Archives, In: Proceedings of ACM Multimedia 2002, pp. 570-579, December 1-6, 2002, Juan-les-Pins, France.

[Tzanetakis and Cook, 2000]
Tzanetakis, G., Cook, P. MARSYAS: A framework for audio analysis, Organised Sound 4(3), Cambridge University Press, 2000

[Tzanetakis et al., 2001]
Tzanetakis, G., Essl, G. & Cook, P. (2001). Automatic musical genre classification of audio signals. In Proceedings International Symposium on Music Information Retrieval (ISMIR), Bloomington, Indiana, October 15-17 2001.

[Welsh et al., 1999]
Welsh, M., Borisov, N., Hill, J., Von Behren, R., and Woo, A. Querying large collections of music for similarity. Technical Report UCB/CSD00 -1096, U.C. Berkeley Computer Science Division, November 1999.