Audio Based Genre Classification of Electronic Music

Priit Kirss
Master's Thesis
Music, Mind and Technology
University of Jyväskylä
June 2007

JYVÄSKYLÄN YLIOPISTO

Faculty of Humanities	Department of Music
Priit Kirss	
Audio Based Genre Classification of Electronic Music	
Music, Mind and Technology	Master's Thesis
June 2007	Number of pages: 72

This thesis aims at developing the audio based genre classification techniques combining some of the existing computational methods with models that are capable of detecting rhythm patterns. The overviews of the features and machine learning algorithms used for current approach are presented. The total 250 musical excerpts from five different electronic music genres such as deep house, techno, uplifting trance, drum and bass and ambient were used for evaluation. The methodology consists of two main steps, first, the feature data is extracted from audio excerpts, and second, the feature data is used to train the machine learning algorithms for classification. The experiments carried out using feature set composed of Rhythm Patterns, Statistical Spectrum Descriptors from RPextract and features from Marsyas gave the highest results. Training that feature set on Support Vector Machine algorithm the classification accuracy of 96.4% was reached.

Music Information Retrieval

Contents

1 Introduction

In contemporary information society, music plays a great role in different purviews and affects humans' lives in many ways being an important part of most people's everyday life. It has given reasons for numerous researches and contributed to emerging and developing of Music Information Retrieval (MIR) science. Music Information Retrieval is an interdisciplinary science that deals with retrieving information from music. During the last decade, the automatic music analysis has become one of the most important and active field of studies among others in MIR science. As Li, Ogihara and Li stated, "Efficient and accurate automatic music information processing (accessing and retrieval, in particular) will be an extremely important issue, and it has been enjoying a growing amount of attention" (2003, p. 282). As automatic music analysis was mainly based on the MIDI data earlier then during the last few years the audio based approach has become the mainstream. Among many other tackled topics, computational models of music similarity and genre classification are at central place. There are two links between those topics (Pampalk, 2006). Firstly, similarity measures can be evaluated using a genre classification scenario and, secondly, features which work well for genre classification are likely to also work well for similarity computations. Currently the main applications for automatic genre classification and music similarity are categorizing and organizing music on the Internet (e.g. iTunes Music Store), recommending similar songs and artist, generating automatic playlists (e.g. web-based radio stations such as Pandora and Last.FM) and so on. So far it has been a quite clumsy and time-consuming manual endeavour. In a word, implementing automatic genre classification and music similarity in different applications helps users to discover music. As Aucouturier and Pachet claimed, "It is only with efficient content management techniques that the millions of

music titles produced by our society can be made available to its millions of users."(2004, p. 1).

Most of the works (with a few exceptions such as (Pampalk, 2005) and (Mörchen, Ultsch, Thies, Löhken, Nöcker, Stamm, Efthymiou & Kümmerer, 2005)) dealing with music similarity and genre classification have divided music data sets approximately into 10 genres, including electronic music (often referred to as electronica). However, a major problem with this kind of approach is that many of these genres, including electronic music, can be divided in turn into many genres and subgenres that differ from each other enormously. One of the most drastic is the electronic music domain, which contains tens (or even hundreds) of hugely different genres and subgenres such as ambient, nu jazz, electronic art music, drum and bass, house, techno, electro, trance, trip-hop, intelligent dance music, and so on. Therefore they should not be classified as belonging to only one genre, and should be dissociated in order to expand the classification accuracy and broaden the music data set distribution hierarchy.

In case of electronic music, the rhythm is one of the most important features that can help distinguishing genres from each other. Therefore, extracting the rhythm patterns from electronic music, in addition to other data extracted traditionally by computational models, would provide enough information for classifying music more accurately, especially in the electronic music domain.

Thus, the aim of this thesis is to focus on genre classification of electronic music, which complements other works that have used the traditional music datasets for classification. The main idea is to combine some of the existing computational methods with models that are capable of detecting rhythm patterns and run different case studies and tests in order to determine the usefulness of this approach.

2 Previous work on genre classification

Much research has been done in music similarity and genre classification field lately in audio domain, therefore the literature containing similar topics contain somewhat different approaches. However, the broad outline for most of the works is somewhat similar and consists of a few steps. Firstly, the features are extracted from audio, then similar features are often grouped together in order to reduce amount of data (optional) and, finally, feature data is used to train machine learning algorithms for classification. In other words, music is classified using machine learning algorithms according to extracted features. The success often depends on the different algorithm variants and parameters, and features used.

The first subchapter gives the review of the papers dealing with genre classification and music similarity. In addition, there are some papers that are dealing with evaluating of different audio features for music classification. These are reviewed in the second subchapter.

2.1 Genre classification and music similarity

One of the most cited articles in the field of music genre classification and music similarity is written by Tzanetankis, Essl and Cook (2001). They claim that, although the division of music into genres is subjective, there are perceptual criteria related to the texture, instrumentation and rhythmic structure of music that can be used to characterize music. The statistics of spectral distribution over time are used to represent musical surface – characteristics of music related to texture, timbre and instrumentation – to recognize patterns. They include features such as mean of the spectral centroid, mean of the spectral rolloff, mean of the spectral flux, mean of the zero crossings, standard deviation of the spectral centroid, standard deviation of the spectral rolloff, standard deviation of the spectral flux, standard deviation of the zero crossings and low energy rate. These features are calculated over a "texture" window of 1 second consisting of 40 frames using the Short Time Fourier Transform (STFT). The calculations of features for representing the rhythmic structure of music are based on the Wavelet Transform (WT) – an alternative to STFT. The rhythmic feature set is based on detecting the most salient periodicities of the signal. Using Discrete Wavelet Transform, the signal is first decomposed into a number of octave frequency bands and time domain amplitude envelope of each band is extracted separately. Following this the envelopes of each band are summed together and autocorrelation function is computed. The peaks of the autocorrelation function correspond to the various periodicities of the signal's envelope. The performance of those feature sets has been evaluated by training statistical pattern recognition classifiers, namely Gaussian classifiers, using real world audio collections.

Li, Ogihara and Li (2003) use the same set of features as Tzanetankis, Essl and Cook (2001) but, in addition, they propose a new feature extraction method, Daubechies Wavelet Coefficient Histograms (DWCH). The authors used Marsyas (the overview is given in section 4.4) software for extracting the features. The algorithm of DWCH extraction consists of obtaining the wavelet decomposition of music signals, constructing a histogram of each subband, computing the three first moments of all histograms and, finally, computing the subband energy for each subband. Effectiveness of this feature is evaluated using machine learning algorithms such as Support Vector Machines, K-

Nearest Neighbour, Gaussian Mixture Models and Linear Discriminant Analysis. It is shown that DWCHs improve the accuracy of music genre classification significantly. On the dataset provided by Tzanetankis, Essl and Cook (2001), the classification accuracy has been increased from 65% to almost 80%.

West and Cox (2004) examine several factors that affect the automatic classification of musical audio signals. They describe and evaluate the classification performance of two different measures of spectral shape used to parameterize the audio signals, Mel-frequency filters (used to produce Mel-Frequency Cepstral Coefficient or MFCC) and Spectral Contrast feature. Genre feature extractor for Marsyas-0.1, which calculates a single feature vector piece, is also included for comparison. Next, they explore the temporal modelling of features that are calculated from the audio. The final step in the calculation of a feature classification is to reduce the covariance among the different dimensions of the feature vector. For MFCC this is performed by a Discrete Cosine Transform and for Spectral Contrast by a Karhunen-Loeve Transform. Then musical audio signals are classified into one of six genres, from which all of the test samples are drawn. The audio signals are converted into feature vectors, representing the content of the signal, which are then used to train and evaluate a number of different classifiers. The classifiers evaluated are single Gaussian models, 3 component Gaussian mixture models, Fisher's Criterion Linear Discriminant Analysis and new classifiers based on the unsupervised construction of a binary decision tree classifier with either a linear discriminant analysis or a pair of single Gaussians with Mahalanobis distance measurements used to split each node in the tree. The unsupervised construction of large decision trees for the classification of frames, from musical audio signals, is a new approach. It allows the classifier to learn and identify diverse groups of sounds that only occur in certain types of music. The results achieved by these classifiers represent a significant increase in the classification accuracy of musical audio signals.

Mohd, Doraisamy and Wirza (2005) use a similar set of features as used by Tzanetankis, Essl and Cook (2001). Marsyas software was also used to extract audio features and for classification the suite of tools available in WEKA (the overview is given in section 4.5) was used for the classification. The authors used J48 classifier that enables pre-processing, classifying, clustering, attributes selection and visualizing, for that.

Instead of traditionally used dataset of Western music, Mohd, Doraisamy and Wirza used Malay music in this paper. The results show that factors such as musical features extracted, classifiers employed, the size of dataset, sample excerpt length, excerpt location and test parameters improve classification results.

Pampalk (2006) describes different computational models of music similarity and their applications in his doctoral thesis. He combines different approaches and presents the largest evaluation of music similarity measures (features) to date. The author claims that the best combination of features performs significantly better than most of the approaches so far. A listening test is conducted to cross-check the results from the evaluation based on genre classification, which confirms that genre based evaluations are suitable to efficiently evaluate large parameter spaces. Also recommendations on the use of similarity measures are given. In addition to theoretical part three applications of similarity measures are described. The author explains that in the first application it is demonstrated how music collections can be organized and visualized so that users can control the aspect of similarity they are interested in. The second and third applications, respectively, demonstrate how music collections can be organized hierarchically, summarized with words found on web pages, and how playlists can be generated with minimum user interaction.

Lampropoulos, Lampropoulou and Tsihrintzis (2005) present a musical genre classification system based on audio features extracted from signals, which correspond to distinct musical sources. A major difference from other works is that they use first a sound source separation method to decompose the signal into a number of component signals (each corresponds to different musical instrument source). Thus timbral, rhythmic and pitch features are extracted from distinct instrument sources and used to classify a music excerpt. Next, different signals are classified into a musical dictionary of instruments sources or instrument teams. This approach attempts to mimic human listener who is able to determine a music genre and different musical instruments. The authors claim that this is a difficult task and has many limitations and shortcomings.

Lampropoulos et al. (2005) used Convolute Sparse Coding (CSC) algorithm for separating signals. In order to obtain higher perceptual quality of separated sources, the CSC algorithm uses compression. They used Marsyas software for extracting 30-

dimensional feature set proposed by Tzanetakis, Essl and Cook (2002). Lampropoulos, Lampropoulou and Tsihrintzis utilized genre classifiers based on multilayer perceptrons (a type of artificial network) for genre classification in the machine learning tool WEKA. Results show that this approach presented an improvement of 2% - 2.5% in genre classification.

E. Pampalk, A. Flexer and G. Widmer (2005) demonstrate the performance of genre classification can be improved by combining spectral similarity with complementary information. In particular, they combine spectral similarity with fluctuation patterns and derive two new descriptors thereof, namely "Focus" and "Gravity". The authors state that fluctuation patterns describe loudness fluctuations in frequency bands and that they describe characteristics, which are not described by spectral similarity measure. Fluctuation pattern is a matrix with 20 rows (frequency bands) and 60 columns (modulation frequencies) and the elements of it describe the fluctuation strength. The distance between songs is computed using Euclidean distance using matrix as a 1200-dimensional vector. According to the authors the Focus describes the distribution of energy in the fluctuation patterns and the Gravity describes the centre of gravity of the fluctuation pattern on the modulation frequency axis. Low Gravity values indicate that the excerpt might be perceived slow and also reflect effects such as vibrato and tremolo. For the classification the nearest neighbour classifier is used. They obtained an average classification performance increase of 14% but confirm the findings by Aucouturier and Pachet (2004) who averred the existence of the "glass ceiling" in genre classification.

K. West and S. Cox (2005) present an evaluation of the different audio file segmentations that have been used for music genre classification to calculate features. They include individual short frames (23 ms.), longer frames (200 ms), short sliding textural windows (1 sec) of a stream of 23 ms frames, large fixed windows (10 sec) and whole files. The authors also introduce a new segmentation based on an onset detection function, which outperforms the fixed segmentations.

2.2 Evaluation of different audio features for music classification

Pohle, Pampalk and Widmer (2005) evaluate how well a variety of combinations of feature extraction and machine learning algorithms are suited to classify music into perceptual categories such as tempo, mood, complexity, emotion, and vocal content. First, the authors calculate features that have commonly been used in the field of genre classification of a music collection, which were labelled according to the categories. Next, they convert the features into attributes that can be fed into machine learning algorithms and evaluate three different attribute sets in combination with twelve machine learning algorithms (including K-Nearest Neighbours and Support Vector Machine). Finally, confusion matrices and classification accuracies are assessed from experiments. According to the authors, the results show that most of the examined categorizations are not captured well and thus they claim that more research is needed on alternative sources of information for useful music classification.

Pohle (2005) gives a comprehensive overview of many features (including Mpeg7 set of features in addition to others) used in MIR applications and brings out the illustrations for them. Next, Pohle implements described features in T-Toolbox programmed in the Matlab which allows doing classification experiments and descriptor visualizations. For classification, the machine learning software WEKA interface is provided. Last, he gives evaluation of described methods for classification of music according to categorizations such as genre, mood, and perceived complexity. Features implemented in T-Toolbox and different machine learning algorithms are used for that. Pohle concludes that the treated features are not capable to reliably discriminate between the classes of most examined categorizations. Regardless he claims that the results could be improved by developing more elaborate techniques.

Aucouturier and Pachet (2004) give the overview of the experiments done in an attempt to improve the performance of the algorithms used frequently in music genre classification and music similarity. According to the authors this paper contributes in two ways to the current state of art. First, they report on extensive tests over many parameters and algorithmic variants which lead to an absolute improvement over existing algorithms of about 15 % R-precision. Moreover, they describe many variants that surprisingly do

not lead to any significant improvement. Experiments run by the authors suggest the existence of a “glass ceiling” at R-precision about 65 % which cannot be overcome by pursuing such variations at the same time. According to the authors the best number of Mel Frequency Cepstrum Coefficients and Gaussian Mixture Model Components is 20 and 50 respectively. Aucouturier and Pachet add that this paper does not present the absolute truth because they do not cover all the possible variants of the pattern recognition scheme. Moreover, the low-level descriptors used in MPEG7 standard and newer methods such as support vector machines are not included in the tests.

3 Musical genre

In this chapter the overview of definition of term “genre” is given and following that, the genres used for classification for this thesis are described.

3.1 Definition of genre

In general the term ‘category’ means a class, a set of objects or events, grouped according to some criteria. Many philosophers, cognitivists and semiotics agree that humans create categories in order to reduce the complexity of the empirical world and therefore in case of music the overall entropy in the musical universe. (Fabbri, 1999)

Musical genres are not just labels applied to music, rather they seem to exist both at a private level, as cognitive types, and as socialized nuclear content that is as socialized sets of instructions to detect occurrences of types. One of the attempts to define a genre is as follows: A genre is a kind of music, which is acknowledged by the community for any reason or purpose or criteria. (Fabbri, 1999)

Musical genres emerge as the names in order to define some sort of similarities, recurrences that members of a community made pertinent to identify musical events. The genre-defining rules can be related to any of the codes involved in musical event, in such a way that knowing what kind of music one will be listening to will guide and help you to choose the proper codes and tools for the participant. Therefore, the genres can be considered as accelerators that speed up the communication within a music community, as well as standardized codes that allow no margin for deviation. However, rules and

codes are made pertinent by the community, and what one sees as the most significant regularity within a certain genre may not be what the community that constituted the genre in the first place saw as its essence. As Umberto Eco stated, a hierarchy of codes always defines the ideology of a genre. (Fabbri, 1999)

The first problem that occurs is that there are no exact boundaries between different genres. Moreover, as noted before, the genres are often treated differently within different communities and therefore the boundaries between genres are also perceived very subjectively; especially nowadays, when new genres are emerging and developing faster than ever before. In addition, there are many combinational and „hybrid" genres that make different genres or subgenres often overlapping and therefore in turn makes the genre distinction often a nontrivial endeavour.

The simplest example to describe that kind of situation would be done using the term 'techno'. The term 'techno' is often unknowingly used to refer to all kinds of electronic music, whereas other people use it in order to distinguish techno music as a distinct genre of electronic music. Additional info on definition of genre can be found in (Kemp, 2005).

3.2 Genres involved

In this section the overview of genres involved in this thesis is given. The genres such as deep house (subgenre of house), uplifting trance (subgenre of trance), drum and bass, ambient and techno were chosen. The first reason why these particular genres were chosen is that they are quite well-known and widespread among different communities of (electronic) music listeners. Secondly, the cores of these genres are quite well defined and thus are distinguishable enough. Thirdly, music excerpts belonging to these genres are available and downloadable from many online record stores and therefore are easily accessible. In addition, this kind of genre collection provides both variance and similarities between genres. It means that the chosen genre taxonomy involves different distinct electronic music genres; however, three of them – house, techno and trance – are

sharing somewhat similar drumbeats and might be confusing and therefore challenging for current classification system.

3.2.1 House/deep house

Deep house is a subgenre of house music and therefore the genre of house music itself is described first. House music is a type of electronic dance music, whose common element is a prominent 4/4 drumbeat. The kick drum is pounding on every quarter note having usually the tempo of 118 – 135 beats per minute (BPM). In addition high-hats on the eight-note off-beats and snare drum or clap on beats 2 and 4 of every bar are used. In order to augment the beat, different percussion and kick fills are frequently used. However, sixteenth-note patterns are also often used, especially for percussion and/or high-hats. House music also uses a continuous, repeating, usually also electronically generated bass line. Typically added to this foundation are electronically generated sounds and samples of music such as jazz, blues and synthpop. However, there are more than 20 subgenres of house music of which probably the most well known are acid house, funky house, hard house, progressive house, tech house, tribal house, and deep house. (House music, 2006)

Deep house (Deep house, 2006) is a subgenre of house music characterized by a generally mellower, deeper sound. This deep sound is achieved through the use of atmospheric elements such as pads, keyboards, and the frequent use of deep rolling bass lines. Deep house is loosely defined; however the following characteristics distinguish it from most other forms of house music:

- relatively slow tempo (110–128 BPM – beats per minute)
- de-emphasized percussion, including:
    - simple yet syncopated drum machine programming
    - gentle transitions and fewer "build-ups"
    - less "thumpy" bass drum sound
    - less pronounced hi-hats on the off-beat

- sustained augmented/diminished key chords or other tonal elements that span multiple bars
- increased use of reverb, delay, and filter effects
- frequently, the use of vocals

Techno and trance, the two primary dance music genres that developed alongside house music, can share the basic beat infrastructure. However, techno and trance usually avoid house music's often used live music influenced feel and black or Latin music influences in favour of more synthetic sound sources and approach. (House music, 2006)

3.2.2 Trance/uplifting trance

Trance is a genre of electronic dance music which received its name from the repetitious morphing beats, and the throbbing melodies which would presumably put the listener into a trance-like state. The tempo of trance music falls usually between 130 and 160 BPM and it uses somewhat similar drumbeat to house music – kick drum is placed on every downbeat of a bar and regular high-hat is on the offbeat. Sometimes snare drum or clap is also used on beats 2 and 4. Some additional percussive elements are usually added, but, unusually in electronic dance music, tracks do not usually derive their main rhythm from the percussion. Most of the trance tracks use repeating melodic synthesizer phrases and a musical form that builds up and down throughout a track, often crescendoing[1] or featuring a breakdown[2]. Fast arpeggios[3], minor scales, and highly intermixed minor and major chords are common features in trance. Often simple sawtooth waveform based sounds are used both for short pizzicato[4] elements and for long, sweeping string and pad sounds. Sometimes vocals are also used. A lot of reverb and delay effects are often used on synthesizer sounds and vocals in trance music. That kind of approach provides the

[1] Gradually getting louder.
[2] A section where the composition is deliberately deconstructed to minimal elements.
[3] A broken chord where the notes are played or sung in succession rather than simultaneously.
[4] A technique for playing a string instrument; rather than drawing the bow across the string to make sound, the string is "plucked" with one finger

tracks with the sense of vast space that is considered to be the basis for the genre's epic quality. (Trance music, 2007)

Uplifting trance, often known as anthem trance, is a term used to describe subgenre of trance music influenced by progressive trance[5], hard trance[6], and psychedelic trance[7]/goa trance[8]. Progressive trance is characterized by extended chord progression, extended breakdowns, and relegation of arpeggiation to the background while bringing wash effects to the fore. In addition it contains melodies similar to happy hardcore[9]. The tempo of about 140 BMP is commonly used. (Uplifting trance, 2007)

2.2.3 Techno

Techno (Techno, 2007) is a form of electronic dance music with influences from Chicago House, electro, New Wave, Funk and futuristic fiction themes that were prevalent and relative to modern culture during the end of the Cold War in industrial America at that time. Techno features an abundance of percussive, synthetic sounds and studio effects used as principal instrumentation. Usually it features a constant 4/4 beat in the range of 115–160 BPM (however, it typically falls between 130 – 140 BPM). As described in (Lang, 1996), "Techno is denoted by its slavish devotion to the beat, the use of rhythm as a hypnotic tool. It is also distinguished by being primarily, and in most cases entirely, created by electronic means. It is also noted for its lack of vocals in most cases."

Techno compositions tend to have strong melodies and bass lines; however these features are not as essential to techno as they are to other electronic dance music genres. It is also quite common for techno compositions to deemphasize or omit these elements. Many dance music genres can be described in such terms; however techno has a distinct sound that aficionados can pick out very easily. In case of techno music the producers treat the electronic studio as a large, complex instrument to produce timbres that are

[5] More info can be found in http://en.wikipedia.org/wiki/Progressive_electronic_music.
[6] More info can be found in http://en.wikipedia.org/wiki/Hard_Trance.
[7] More info can be found in http://en.wikipedia.org/wiki/Psy_trance.
[8] More info can be found in http://en.wikipedia.org/wiki/Goa_trance.
[9] More info can be found in http://en.wikipedia.org/wiki/Happy_hardcore.

simultaneously familiar and alien. Machines are used to generate and complement continuous, repetitive sonic patterns also featuring unrealistic combination of sounds. (Techno, 2007)

"Techno involves sounds of which real instruments may or may not exist, and because of this provokes wholly unique thoughts and feelings, which can be played at speeds or note combinations possible only with aid of electronics, and still maintain artistic, musical quality." (Paperduck, n.d.)

However, the term "techno", which derives from "technology", is often unknowingly used to refer to all forms of electronic music.

More information on techno can be found in (Lang, 1996) and (Techno, 2007).

3.2.4 Ambient

Ambient music is a musical genre that focuses on sound and space rather than melody and form. It is music that is intentionally created to be used as both as background music and as music for active listening. It usually features slowly evolving sounds, repetition, and is relatively static. It is chiefly identifiable as having an overarching atmospheric context. However it is loosely defined and it might incorporate elements of a number of different styles - including jazz, electronic music, new age, rock and roll, modern classical music, traditional, world, and noise. The term "ambient music" was first coined by Brian Eno[10] in the late 1970s, who wanted to make music that would support reflection and space to think. (Ambient music, 2007) (Ambient music, n.d.)

Brian Eno (Music for Airports liner notes, September 1978) himself put it this way: "Ambient Music must be able to accommodate many levels of listening attention without enforcing one in particular; it must be as ignorable as it is interesting." (Ambient music, n.d.)

[10] Known as a father of modern ambient music.

3.2.5 Drum and bass

Drum and bass (abbreviated to d'n'b or drum'n'bass) also known as jungle is a genre of electronic dance music. It is characterized by fast tempo broken beat drums (not to be confused with the "broken beat" genre) generally in between 160 and 180 BPM and it uses heavy, often intricate bass lines. As the name "drum and bass" suggests, the drumbeats and bass lines are the most critical features in that genre, however drum and bass songs are not constructed solely from these elements. There have been many permutations in its style incorporating elements from dancehall, electro, funk, hip hop, house, jazz, metal, pop, reggae, rock, techno and trance. The bass lines usually originate from synthesizers or rarely from sampled sources. The complex syncopation[11] of the drumbeat is another facet of production on which producers spend a very large amount of time. The most common drumbeat samples used for drum and bass are Amen[12] break, Apache[13] break, the Funky Drummer[14] break, and others. (Drum and Bass, 2007)

There are numerous understandings of what constitutes "real" drum and bass as it has many scenes and subgenres within it. It might be anywhere between dark paranoid vocal free and relaxed singing vibes of jazzy influenced drum and bass. This genre has been compared with jazz where very different sounding music is all under the same music genre. Therefore, drum and bass is more of an approach, or a tradition, than a style. However, a drum and bass track without a fast broken beat would not be a drum and bass track and could be classified as belonging to other genres such as techno, breaks, and so forth. (Drum and Bass, 2007)

[11] A shift of accent in a composition that occurs when a normally weak beat is stressed.
[12] A drum-solo performed by Gregory Cylvester Coleman. More info can be found in http://en.wikipedia.org/wiki/Amen_break.
[13] A drumbeat sampled from "Apache" written by Jerry Lordan and recorded by The Shadows.
[14] A drum solo from "Funky Drummer" recorded by James Brown and his band. More info can be found in http://en.wikipedia.org/wiki/Funky_Drummer.

4 Methodology

In this chapter the overview of the methodology used is given. It provides the description of single features, feature sets, description of classification algorithms and overview of other tools used for this thesis.

4.1 Overview

In order to make songs comparable to each other by computers some kind of parameters or descriptors describing the audio content according to which the comparison could be carried out are needed. This process is called feature extraction – the process of generating a set of numerical descriptors that characterize the audio. One of the biggest challenges is to choose the right set of features that would reflect the perceived similarities and differences between music excerpts as good as possible. It means that songs that are perceived as being similar must be described by features that are located nearby in feature space, and contrary, in case of songs that are perceived as being different, the distance between features describing the music content must be as big as possible. The second important thing to pay attention to is that features should preserve all the important information contained in the initial data (Kosina, 2002). The successfulness of genre classification depends heavily on the chosen features and therefore is one of the most crucial parts in the chain of processes. However, it is often done by trial and error and is a difficult task (Kosina, 2002).

The second major step is using feature data in machine learning algorithms for classification. The classification, a subfield of decision theory, relies on the assumption that each observed pattern belongs to a category, which can be taken as a model for the pattern. It suggests that regardless of the differences between individual patterns, there is a set of features that are similar in patterns belonging to the same category, and different between patterns from different categories. That kind of features can be used to determine belonging to the certain class, according to the assumption that there are certain fundamental properties shared by music excerpts belonging to one genre. (Kosina, 2002)

Other possibility to understand classification is to observe it in geometrical terms. Using the feature vectors that regard to points in feature space it is possible to find decision boundaries that segment the feature space into regions that correspond to particular classes. The classification of new items is based on what region they lie in. (Kosina, 2002)

To conclude, the methodology follows the methods primarily used in other papers dealing with genre classification and it consists of two main stages:

- Extracting the features from music
- Classification using machine learning algorithms

The classification schema used for this thesis is described in section 4.8.

4.2 Features

This section gives the overview of the features used in this thesis.

- **Spectral Centroid** (first moment of the power spectrum) is defined as the average frequency, weighted by magnitude, of spectrum:

$$SpectralCentroid = \frac{\sum_{i=0}^{NF-1} f_i P(f_i)}{\sum_{i=0}^{NF-1} P(f_i)}$$

Spectral centroid is a feature adapted from psychoacoustics and music cognition. It is frequently used as an approximation for a perceptual brightness measure. The lower the spectral centroid, the more energy is located in the lower frequency components and vice versa (Tanghe et al., 2005). (Pfeiffer and Vincent, 2001) (Pfeiffer, 2002)

- **Zero-crossing rate** is defined as the number of time domain zero crossings (sign-changes) of signal per time unit (also can be measured per sample of signal). It has been used widely in both MIR and speech recognition and known as a good descriptor for genre classification. Mathematically it is defined as (Pohle, 2005):

$$ZeroCross = \frac{1}{2}\sum_{n=1}^{N} \left| sign(x(n)) - sign(x(n-1)) \right|$$

- **Spectral brightness** is defined as the amount of spectral energy corresponding to frequencies higher than a given cut-off threshold. As noted in the MIRToolbox help, typical values for the frequency cut-off threshold are 3000 Hz (Juslin 2001, p. 1802.) and 1500 Hz and 1000 Hz (Laukka, Juslin & Bresin, 2005). Brightness is a measure of the higher-frequency content of the signal. (Typke et al., 2005)

- **Spectral Spread** is defined as the differences between the indices of the highest and the lowest subband that have an amplitude above threshold, defined on logarithmically spaced frequencies (similar to bandwidth, which is defined on linearly spaced frequencies). (Pohle, 2005)

- **Spectral Skewness** is defined as the third moment of the power spectrum:

$$SpectralSkewness = \frac{\sum_{i=0}^{NF-1}(P(f_i) - \mu)^3}{N\sigma^3}$$

where μ = mean and σ = standard deviation. Spectral skewness describes the symmetry of the distribution of the amplitude spectrum values, whereas positive value means that the distribution has a tail at the higher values, negative values correspond to the tail at lower values and a value of zero shows that the distribution is symmetric. (Tanghe et al., 2005)

- **Spectral Kurtosis** is defined as the fourth moment of the power spectrum, offset by -3:

$$SpectralKurtosis = \frac{\sum_{i=0}^{NF-1}(P(f_i) - \mu)^4}{N\sigma^4} - 3,$$

where μ = mean and σ = standard deviation. Spectral kurtosis describes the size of the tails of the distribution of the amplitude spectrum values. Positive spectral kurtosis values mean that the distributions have relatively large tails, distributions with small tail have negative kurtosis, and normal distributions have zero kurtosis. (Tanghe et al., 2005)

- **Spectral Entropy** is a measure of disorganization of audio signals and can be used to measure the peakiness of distribution. Also it gives a measure of the number of bits required to represent some information.

- **Spectral Flatness** is defined as the ratio of the geometric mean to the arithmetic mean of the power spectrum:

$$SpectralFlatness = \frac{\sqrt[N]{\prod_{i=0}^{N-1} P(f_i)}}{\frac{1}{N}\sum_{i=0}^{N-1} P(f_i)}$$

It is a measure of the flatness of the spectrum, obtaining values near 1 for a flat spectrum and values near 0 for a peaky spectrum. (Tanghe et al., 2005)

- **Spectral Irregularity** is defined as a logarithm of the spectral deviation of component amplitudes from a global spectral envelope derived from a running mean of the amplitudes of three adjacent harmonics. It shows the smoothness of the spectrum. (Misdariis et al., 1998)

- **Spectral Low Energy Rate** is defined as the percentage of frames that have less than average energy within the audio excerpt. It is often used to separate speech from music. (Pfeiffer, 2002)

- **Spectral Rolloff** is defined as the lowest frequency at which the accumulated sum of all lower frequency power spectrum values reach a certain amount of the total sum (R) of the power spectrum. Mathematically it is defined as:

$$SpectralRolloff = \min\left\{ f_i \mid \sum_{i=0}^{j} P(f_i) \geq R \sum_{i=0}^{NF-1} P(f_i) \right\}$$

Usually R = 85% is used as rolloff fraction. However, as mentioned in (Pohle, 2005), R values such as 80%, 92% and 95% have been also used in different works. Spectral rolloff is a measure of spectral shape and is often used to distinguish voiced from unvoiced speech and music. (Tanghe et al., 2005)

- **Spectral Flux (also known as Delta Spectrum Magnitude)** is defined as a measure that determines changes of spectral distribution of two successive windows. Mathematically it is defined as:

$$SpectralFlux_t = \sum_{n=1}^{N} (N_t(n) - N_{t-1}(n))^2$$

where $N_t(n)$ is the normalised magnitude of the Fourier transform at window t. (Kosina, 2002)

- **Mel frequency cepstral coefficients (MFCC)** are coefficients that represent audio and have been widely used in speech recognition systems. They provide a representation of the sound spectrum that closely corresponds to distances between timbres perceived by human. In other words, the perceived similarity in timbre equals to similarity in Mel Frequency Cepstral Coefficients.

  The cepstrum is defined as the inverse Fourier transform of the log-spectrum log(S):

$$c_n = \frac{1}{2\pi} \int_{w=-\pi}^{w=\pi} \log S(w) \exp jwndw$$

If the log-spectrum is given in the perceptually defined mel-scale, then the cepstra are called MFCC. The mel scale is an approach to model the perceived pitch; 1000 mel are defined as the pitch perceived from pure sine tone with 40 dB above the hearing threshold level. Other mel frequencies are found empirically (e.g. sine tone with 2000 mel is perceived twice as high as a 1000 mel sine tone and so on). The mel-scale and Hz-scale are correlated as follows:

$$mel(f) = 2595 \cdot \log_{10}\left(1 + \frac{f}{700}\right)$$

In order to eliminate covariance between dimensions to produce MFCC-s, the discrete cosine transform is used instead of the inverse Fourier transform. When using the discrete cosine transform, the computation for mel frequency cepstral coefficients is done as described in the following steps.

First, the audio signal is converted into short (usually overlapping by one half) frames of length usually about 23 milliseconds. Then the discrete Fourier transform is calculated for each frame and the magnitude of the FFT is computed.

Next, the log base 10 is calculated from the amplitudes of the spectrum. Then the mel-scaled filterbank[15] is applied to FFT data. Finally, the discrete cosine transform is calculated and typically 12 first (most important) coefficients are used. (Aucouturier & Pachet, 2004) (Aucouturier & Pachet, 2003) (Kosina, 2002) (Pohle, 2005)

- **Linear predictive coding (LPC)** is one of the most powerful speech coding analysis techniques providing very accurate estimates of speech parameters and is known as being relatively efficient for computation at the same time. The linear prediction[16] voice model is best classified as a parametric, spectral, source-filter model, in which the short-time spectrum is decomposed into a flat excitation spectrum multiplied by a smooth spectral envelope capturing primarily vocal formants The speech signal is produced by a buzzer at the end of a tube that produces a continuous signal, which is passed through a variable model of the vocal tract and its transfer function is denoted H(z). The vocal tract can be approximated by an all-pass filter, whose z-transform is:

$$H(z) = \frac{G}{1 - \sum_{k=1}^{p} a_k z^{-k}}$$

where G is gain of filter, p order of filter, z_k the k samples delay operator, and a_k are the filter coefficients. (Pohle, 2005) (Gravier, 2005) (Smith, 2006)

- **Spectral Inharmonicity** calculated as the cumulative sum of differences of each harmonic frequency from its theoretical value. (Paiva et al., 2005)

[15] The filterbank is constructed using 13 linearly spaced filters and 27 log-spaced filters than follow a common model for human auditory perception

[16] The overview of linear prediction is given in J. Makhoul. Linear prediction: A tutorial review. Proceedings of the IEEE, 63 (5):561–580, April 1975., and also in http://en.wikipedia.org/wiki/Linear_prediction

- **Spectral histogram** represents the statistical distribution of amplitude values in waveform.

4.3 MIRtoolbox

MIR toolbox[17] is an integrated set of functions written in Matlab by Olivier Lartillot and Petri Toiviainen and is dedicated to the extraction of musical features from audio files. Among others, features related to timbre, tonality, rhythm or form can be extracted with MIRtoolbox. The toolbox also includes functions for statistical analysis, segmentation and clustering. The design of syntax offers both simplicity of use and transparent adaptiveness to a multiplicity of possible input types. All the feature extraction methods can accept audio file as an argument, or any preliminary result from previous operations. The same syntax can also be used for analyses of single audio files, bunch of files, folder full of audio files, series of audio segments, multi-channel signals, and so on. (Lartillot & Toiviainen, 2007)

4.4 Marsyas 0.2

Marsyas[18] (Music Analysis Retrieval and Synthesis for Audio Signals) is a free software framework for audio analysis, synthesis and retrieval. This software has been written by George Tzanetakis and used for a variety of both academic and industrial projects. The major underlying theme under design of Marsyas software has been to provide and efficient and flexible framework for Music Information Retrieval. It is also said to be regularly maintained by its author and there are plans to extend its functionality in the future. The Marsyas implementations include the standard temporal and spectral low-

[17] http://www.cc.jyu.fi/~lartillo/mirtoolbox/
[18] http://opihi.cs.uvic.ca/marsyas/

level features like spectral centroid, spectral rolloff, spectral flux, zero crossing rate, and mel frequency cepstral coefficients (MFCC). (McKay et al., 2005)

4.5 RPextract music feature extractor

RPextract[19] toolbox for Matlab is a feature extraction tool developed by Thomas Lidy in the Vienna University of Technology, which is based on the Music Analysis Toolbox for Matlab[20] by Elias Pampalk. Three different feature sets can be derived from content-based analysis of musical data and they reflect the rhythmical structure in the musical pieces. These three feature sets are:

- Statistical Spectrum Descriptors describe the fluctuations by statistical measures on critical frequency bands of a psycho-acoustically transformed sonogram
- Rhythm Patterns (also called Fluctuation Patterns) reflect the rhythmical structure in musical pieces by a matrix describing the amplitude of modulation on critical frequency bands for several modulation frequencies
- Rhythm Histograms aggregate the energy of modulation for 60 different modulation frequencies and therefore indicate general rhythmic in music

Since the algorithms used for RPextract Music Feature Extractor consider psychoacoustics in order to resemble human auditory system and extract suitable semantic information from music, the classification of sounds and automatic organization of music according to similarity are made possible. This system can read audio files such as au, wav and ogg as an input. These feature sets are appropriate as a basis for an unsupervised organization task, and for machine learning and classification tasks. RPextraction was submitted to the Audio Description Contest of the International

[19] http://www.ifs.tuwien.ac.at/~lidy/rp/
[20] http://www.oefai.at/~elias/

Conference on Music Information Retrieval (ISMIR 2004), winning the rhythm classification track. More detailed information is provided at (Lidy, 2005).

4.6 Weka

Weka[21] (Waikato Environment for Knowledge Analysis) is collection of state-of-the-art machine learning algorithms for different data mining tasks, which is open source software issued under the GNU General Public License[22]. It contains tools for data pre-processing, classification, regression, clustering, association rules, and visualization. The algorithms can either be applied directly to a dataset or called from the Java code. In addition it is possible to develop new machine learning schemes.

4.7 Machine learning algorithms

In this section, the description of machine learning algorithms used for classification in Weka is given.

4.7.1 K-nearest neighbours

K-nearest neighbours classification technique is a variation of nearest-neighbour (which is also known as a special case of K nearest neighbours) and considered to be one of the simplest classification methods. The idea behind this method is to simply separate the data based on the assumed similarities between different classes. A distance measure is calculated between all the points in a dataset using Euclidean or Mahalanobis distance. According to these distances, a distance matrix is constructed between all the possible

[21] http://www.cs.waikato.ac.nz/~ml/weka/index.html
[22] http://www.gnu.org/copyleft/gpl.html

pairings of points in dataset. The k-closest neighbours (data points) are then analyzed to determine which class label is the most common among the dataset. Finally, the most common class label is then assigned to the data point being analyzed. These resulting class labels are used to classify each data point in the data set. (Mower, 2003) (Teknomo, 2006)

4.7.2 Naïve Bayes (additionally with kernel)

Naïve Bayes classifier technique is a probabilistic classifier based on Bayesian theorem with strong independence assumptions (on attributes). It is a simple, yet powerful algorithm, which achieves surprisingly good results, especially when the dimensionality of the inputs is high. Naïve Bayes uses all attributes and allows them to make contributions to the decision as if they were all equally important and independent of one another, with the probability denoted by the equation:

$$p(C_i|v_1, v_2, ...v_n) = \frac{p(C_i)\Pi_{j=1}^{n} p(v_j|C_i)}{p(v_1, v_2, ...v_n)}$$

where C_i is class i, and $v_1, v_2, \ldots, v_n$ are the values of the item that is to be classified, and $p(C_i||v_1, v_2, \ldots, v_n)$ denotes the conditional probability of C_i given v_k (where $k = \{1,2\ldots,n\}$). An advantage of the Naive Bayes classifier is that it requires a small amount of training data to estimate the parameters (means and variances of the variables) necessary for classification. (Naïve Bayes Classifier, 2002) (Naïve Bayes Classifier, n.d.) (Naïve Bayes Rule Generator, 2002)

4.7.3 C 4.5

C 4.5 is a decision tree generating algorithm, based on ID3 (Iterative Dichotomiser 3) algorithm. Each branch of a tree is a decision rule depending on only one attribute (Pohle, 2005). First, the decision tree is built and then each of the instances is classified by starting with the rule at the root node[23], and moving through the tree until a leaf[24] is reached.

The ID3 tree is built recursively starting at the root node. If all the instances are of same class, then the current node becomes a leaf, which belongs to the same class and the process stops. Otherwise, the current node is expanded by choosing attribute for which information gain is maximal, and building a sub-tree for each of its possibly appearing values. The algorithm uses a greedy search, that is, it picks the best attribute and never looks back to reconsider earlier choices.

Information gain is defined as expected reduction in entropy due to sorting A, and it can be mathematically presented as:

$$gain(S, A) = entropy(S) - \sum_{v \in V} \frac{|S_v|}{S} \cdot entropy(S_v)$$

where S is the set of remaining training instances at the current node, and V is the set of attribute values from attribute A that appear in S. Entropy of S_v is defined as follows:

$$entropy(S_v) = -\sum_{c \in C} p(c|v) \cdot \log_2 p(c|v)$$

where C denotes the classes that appear in S and p(c|v) denotes the conditional probability of c given v. Entropy(S_v) can be interpreted as expected number of bits needed to encode a value of S_v. (Mitchell, 2005) (Dankel, 1997) (Pohle, 2005)

[23] A starting node in the decision tree that has only outputs and no inputs
[24] A final node in the decision tree that is not split into further nodes

4.7.4 Support vector machines

Support Vector Machine (SVM) is a supervised learning method that belongs to a family of linear classifiers used for classification and regression. However, SVM is closely related to neural networks. It is based on some relatively simple ideas but constructs models that are complex enough and it can lead to high performances in real world applications.

The basic idea behind Support Vector Machines is that it can be thought of as a linear method in a high-dimensional feature space nonlinearly related to input space, therefore in practice it does not involve any computations in that high-dimensional space. All necessary computations are performed directly in input space by the use of kernels. Therefore the complex algorithms for nonlinear pattern recognition, regression, or feature extraction can be used pretending that the simple linear algorithms are used. The implementation of SVM is called SMO in Weka.

SVM performs classification by constructing a N-dimensional hyperplane that optimally separates the data into two categories. Therefore the goal of SVM modelling is to find the optimal hyperplane that separates clusters of vector in such a way that cases with one category of the target variable are on one side of the plane and cases with the other category are on the other size of the plane. The vectors near the hyperplane are the support vectors. (Hearst, 1998) (SVM – Support Vector Machines, n.d.)

4.7.5 Adaptive boosting

Adaptive Boosting, abbreviated to AdaBoost, is a meta-algorithm, which can be used in conjunction with many other machine learning algorithms in order to improve their performance. The main idea behind the algorithm is to construct a "strong" classifier as a linear combination f(x) of "weak" classifiers $h_t(x)$ (algorithm whose performance is to be improved):

$$f(x) = \sum_{t=1}^{T} \alpha_t h_t(x)$$

In order to do that, the weak algorithm is invoked several (denoted T) times specified by the user with varying subsets of the training data, and the several obtained hypotheses are combined to one final hypothesis of higher accuracy. For each iteration, a distribution of weights D_t (initially the equal distribution) is used to select training instances with witch the weak algorithm is executed. Depending on the classification results for current training data subset, a new distribution is calculated in order to use it in the next iteration. (Matas & Šochman, 2004) (Pohle, 2005) (Boosting, 2007)

In this thesis AdaBoost is used with C 4.5 and Support Vector Machines algorithm. The pseudo code of the algorithm can be found in (Freund & Schapire, 1999) and (Boosting, 2007)

4.7.6 Classification via regression

In this context the term "regression" refers to process of estimating a numeric target variable in general (as opposed to a discrete one) and it is used to solve a classification problem with a learner that can only produce estimates for a numeric target variable. Classification via regression (CVR) is a method that uses model trees[25] (algorithm, which combines regression and tree induction for tasks where the target variable to be predicted is numeric) for modelling the conditional class probability function of each class. During training, one function learned for each class; the attribute values are used as input and with possible output values 1 and 0, indicating whether the current training instance belongs to this class or not.

In this thesis, classification via regression is evaluated with two algorithms for function approximation: M5 and linear regression. (Landwehr et al., 2004) (Pohle, 2005)

[25] More information on model trees can be found in http://www.informatik.uni-freiburg.de/~ml/papers/mljlandwehr2005.pdf

4.7.7 Linear logistic regression

Logistic regression is a well known technique for classification, which describes the relationship between a dichotomous dependent variable and a set of independent variables (continuous or discrete) and determines the percent of variance in the dependent variable explained by the independents. It can be also used to rank the relative importance of independents; to assess interaction effects; and to understand the impact of covariate control variables. The only distributional assumption with this method is that the log likelihood ratio of class distributions is linear in the observations. This way the logistic regression estimates the probability of a certain event occurring. Categorical independent variables are replaced by sets of contrast variables, each set entering and leaving the model in a single step. (Amini & Gallinari, 2002) (Friendly, 2007) (Garson, 2006)

4.7.8 Random forest

Random forest is a classification method that consists of many decision trees and outputs the class that is the most frequent value of the classes output by individual trees. For each tree in the forest, if the number of cases in the training set is N, then N cases are sampled at random with replacement from the original data. This sample is the training set for growing the tree. Next, if there are M input variables, a number m, which should be much less than M (m<<M) is specified such that at each node, m variables are selected at random out of the M and the best split on these m is used to split the node. The value of m is constant during the forest growing. Each tree is grown to the largest extent possible and not pruned (as may be done in constructing a normal tree classifier). (Random forest, 2007) (Breiman & Cutler, n.d.)

4.8 N- Fold cross validation algorithm for evaluation

N-fold Cross Validation means that the dataset is split into N equal-sized sub-sets and in N runs, each subset is once selected as a test set, while other N-1 sub-sets are used for training. Then the measures are calculated from each of the N tests and the final result is averaged.

5 Results and evaluation

In this chapter the results using the feature sets described in sections 4.2 and 4.5 and classifiers described in section 4.7 are presented. The tables representing the classification accuracies of all the combinations of features and classifiers used are presented. In addition, detailed classification accuracies by genre and confusion matrices of the best performers are provided.

5.1 Audio collection

5.1.1 Introduction

Five different genres such as house, techno, trance, drum and bass and ambient were chosen as the dataset for this thesis. Although there are many music datasets for MIR research such as In-house Small, In-House Large, Magnatune Small, Magnatune Large, In-house Extra Large and so on, none of these seemed to be suitable for current work, since their music distribution hierarchy and taxonomy is different and does not meet the requirements for that work. Therefore, a totally new music dataset, that suits the objective of this work, had to be compiled. Juno Records[26] online record store that sells mostly electronic music was used as a main source for music excerpts for house, techno, trance

[26] www.juno.co.uk

and drum and bass. For ambient music Shopsonic[27] online record store and http://www.hypnos.com[28] were used.

On Juno webpage all the music is already labelled and divided into genres and some of them also into subgenres. From genres for which the subgenres were provided, one of the subgenres was used for dataset. Whilst trance and house both were divided into subgenres on website, subgenres such as deep house and uplifting trance were chosen for dataset instead of using excerpts from house and trance in general. Unfortunately techno and drum and bass were not divided into subgenres and therefore that kind of approach could not be used for these genres.

Some of the records on Juno records website can be found under different genres; for example the same record could be found under genres such as techno, deep house, and tech house. It is mainly because the same record includes influences from all of these genres and/or contains tracks from different genres and subgenres. Usually the kind of songs that were presented under many genres were not used in order to reduce that the ambiguity of songs' genres and have more distinct classes.

5.1.2 Dataset

A total of 50 songs were chosen from each of the five genres: ambient, deep house, techno, drum and bass, and uplifting trance, therefore giving a total of 250 songs. The music on the Juno website was in stereo mp3 format having a bit rate of 64 kbps, sampling frequency of 22 kHz, and 16 bit. The initial ambient excerpts were usually of higher quality.

For the analysis all 250 songs were converted to wave format using Winamp[29] Nullsoft Disk Writer plug-in version 2.11 using the same parameters as mp3-s on Juno website had (16 bit, 22 kHz, and Stereo). That means that the quality of ambient music also got reduced in order to have the same quality as the other four genres. Many of the

[27] http://www.shopsonic.com
[28] The website for the Hypnos recordings. Also hosts a number of resources related to the ambient music.
[29] www.winamp.com

works (Dannenberg, 2005) (West & Cox, 2005) average the channels to produce a monaural signal in order to reduce the amount of data and computational time. However, as in current case the preliminary tests showed, there is no reason to convert the stereo files into mono, since the classification accuracy reduces significantly using Marsyas feature sets – in some cases almost by 8%. However, more research should be done to find the ground for that. Finally, 15 - 20 seconds long excerpts, that contained the most representative (i.e. the most instruments playing at the same time) audio information within the each of source excerpts, were cut by hand from all the 250 chosen songs. The total length of the 250 excerpts was about 80 minutes. For the full list of songs please refer to the Appendix.

5.2 Feature sets

In this section the overview of feature sets from Marsyas, RP toolbox and MIRtoolbox evaluated using Weka is given. Altogether 8 different feature sets were used, of which four were from Marsyas, one from MIRtoolbox and three from RPtoolbox. In addition a seven combinational feature sets were used. The overview of the used features is given in chapter 4.

5.2.1 Marsyas features sets

The following descriptor sets were extracted for evaluation using Marsyas:

- STFTMFCC – The feature set that consists of spectral centroid, spectral rolloff, spectral flux, zero-crossing rate, Mel-frequency Cepstral Coefficients.
- STFT – The feature set that consists of spectral centroid, spectral rolloff, spectral flux, zero-crossing rate.
- LPCC – The feature set that consists Linear Prediction Cepstral Coefficients

- MFCC – The feature set that consists of s Mel-frequency Cepstral Coefficients

5.2.2 RPextract feature sets

The following three descriptor sets were extracted for evaluation using RPextract:

- Rhythm Patterns features
- Statistical Spectrum Descriptor
- Rhythm Histogram features

5.2.3 MIRtoolbox feature set

The feature set consisting of the following features was used: spectral centroid, zero crossings rate, spectral spread, spectral skewness, spectral kurtosis, spectral flatness, spectral entropy, spectral brightness, spectral low energy, spectral irregularity, spectral rolloff, Mel-frequency Cepstral Coefficients, spectral irregularity, and spectral histogram. In the following sections it will be denoted as MIRTSet.

5.2.4 Combinational sets

Seven different combinational feature sets were constructed in order to see how these sets would work together. A short program that would concatenate two sets of features from 2 different text files into one was written in C language. It was used to compose the following combinational sets:

- Statistical Spectrum Descriptor and Rhythm Histogram features
- STFT & Statistical Spectrum Descriptor and Rhythm Histogram features

- STFTMFCC and Statistical Spectrum Descriptor
- STFTMFCC and Rhythm Histogram features
- STFTMFCC and Statistical Spectrum Descriptor and Rhythm Histogram features
- MIRTSet and Rhythm Histogram features
- MIRTSet & Rhythm Histogram features & Statistical Spectrum Descriptor

The feature sets selected for combinational ones were chosen keeping in mind not to create extensively large combinational feature sets. Therefore, Rhythm Patterns feature set, which extracts 1200 parameters for each music excerpt was not used for combinational sets. In contrast, the number of features per audio file in case of Statistical Spectrum Descriptor is 140 and in case of Rhythm Histogram 60

5.3 Classifiers

The machine learning algorithm variants used for evaluation in Weka are:

- K Nearest Neighbours for k = 1, 3, 5, 8, 10
- Naïve Bayes (additionally with kernel)
- C 4.5 (called J48 in Weka)
- Support Vector Machine (with default parameterization (c = 1 and exponent = 5) and with c = 3, exponent = 5)
- AdaBoost with C4.5 and SVM (with default parameterization (random seed = 1) and with random seed=5)
- Classification via regression (applying M5 and linear regression)
- Simple Logistic (with default parameterization (NumBoostingIterations = 1) and with NumBoostingIterations = 5)
- Random Forest (with default parameterization (numTrees = 10) and with numTrees = 20)

For evaluation 10-fold Cross Validation algorithm was used (see section 4.8 for further information).

5.4 Results

In this section the evaluation results are presented. It covers the tables that contain the classification accuracies of all the combinations of features and classifiers used. In addition the tables describing the detailed accuracy by class and confusion matrix of the best performing classifier within each feature set are provided.

In the table 1 (please see next page), the RP corresponds to the Rhythm Patterns descriptor set, the RH to the Rhythm Histogram set, the SSD to the Statistical Spectrum Descriptors set, the RS to the set combined of the RH and SSD sets, the SMR to the set combined of the STFTMFCC (described in section 5.1.1) and RH sets, the SMS to the set combined of the STFTMFCC and SSD sets, the SSR to the set combined of the STFT (described in section 5.1.1) and SSD and RH sets, and finally the SMSR to the set combined of the STFTMFCC and SSD and RH sets.

Most of the cells at the CVR2 line in the tables 1 and 2 are missing because the computational times were extensively long and these tests were skipped. For example, in case of RP the test was cancelled after a few hours of computing since the end did not seem to be near. The results using this machine learning algorithm are either no so good and therefore should be probably left out from future works.

Table 1. The overall classification accuracy

	RP	RH	SSD	RS	SMR	SMS	SSR	SMSR
1-nearest neighbours	78.4	78	63.2	79.2	83.2	70.4	83.6	82.4
3-nearest neighbours	80.8	80	64.4	78.4	81.2	74.4	84	85.2
5-nearest neighbours	79.6	79.6	66.8	78.8	82.4	74.4	83.6	84.4
8-nearest neighbours	79.2	78	68	80	84.4	70.4	85.8	82.4
10-nearest neighbours	76.4	78.4	68	79.2	84.4	71.6	81.2	84
Naïve Bayes	77.6	79.2	70.4	84	89.2	76.4	87.2	88.4
Naïve Bayes with kernel	77.6	79.6	68.4	83.6	89.2	76	85.6	88.8
C4.5	66.8	68.4	58.8	75.2	76.8	64.4	78.4	78.4
SVM	86.2	82	75.2	**91.2**	**92.8**	**87.6**	96	**95.6**
SVM	69.2	77.6	70	84.4	92	82	90.8	94
AdaBoost	81.4	81.6	72.8	82.8	85.2	74.4	85.2	87.2
AdaBoost(SVM 1)	**86.8**	82.8	73.6	**91.2**	92.4	85.6	94.4	**95.6**
AdaBoost(SVM 2)	**86.8**	80.4	74.4	90.8	92.4	85.6	**96.4**	**95.6**
CVR1	79.2	81.6	69.6	83.6	88.8	74	87.2	88
CVR2	x	x	x	x	87.2	x	80.4	x
Simple Logistic1	83.6	86	**75.6**	90	89.6	83.6	95.2	94.4
Simple Logistic2	82	**84.8**	72.4	84.4	89.6	80.8	88.4	90.8
Random Forest1	74.4	80	66.4	81.2	86	76.4	81.6	78.8
Random Forest2	78.8	82.8	69.2	82.8	88	76.8	86	83.2

In the table 2 (please see next page), the SM corresponds to the STFTMFCC, the STF to the STFT, the LPC to the LPCC, the MFC to the MFCC, the MIR to the MIRToolbox feature set, the MRH to the set combined of MIRToolbox and Rhythm Histogram feature set, and MRS to the set combined of the MIRToolbox, Rhythm Histogram and Statistical Spectrum Descriptors feature sets.

Table 2. The overall classification accuracy.

	SM	STF	LPC	MFC	MIR	MRH	MRS
1-nearest neighbours	72.8	65.2	72.8	68.8	61.6	83.2	81.6
3-nearest neighbours	74.8	68.8	70.8	65.6	63.6	82.8	82.4
5-nearest neighbours	73.2	69.6	71.6	66.4	67.2	82	82.4
8-nearest neighbours	74.8	68	74.4	68.4	65.6	80.8	82
10-nearest neighbours	75.2	68.4	73.2	70	65.6	80	81.2
Naïve bayes	76.4,	66.8	73.6,	65.6	66.4	84.8	86
Naïve bayes with kernel	72.8	68.4	72.4	65.2	68.6	82	84.8
C4.5	71.6	64	**74.8**	63.2	54.8	75.2	74.4
SVM	81.6	72.8	**74.8**	73.2	**73.2**	**91.6**	90
SVM	**84**	72.8	**74.8**	74	70	88	87.2
AdaBoost	78.8	74.4	72.8	72.8	67.6	86	85.6
AdaBoost	84.8	72.8	74	**75.6**	70.8	90.8	90
AdaBoost	84.8	72.8	74	74	70.8	90.8	90
CVR1	74	71.2	68	72.8	67.2	84.4	85.2
CVR2	x	x	x	69.6	x	x	x
Simple Logistic1	82	**76**	71.6	72.8	72.4	90.4	**91.2**
Simple Logistic2	77.2	72.6	73.6	69.6	66	84.4	84.4
Random Forest1	76.4	72	70	71.2	65.2	82.4	85.6
Random Forest2	79.6	71.6	72.8	74	65.6	86.4	87.2

As can be seen from the tables 1 and 2 the highest classification rates are produced by the feature set combined of three before mentioned distinct feature sets: STFT, Statistical Spectrum Descriptors and Rhythm Histograms (denoted SSR). Using AdaBoost with Support Vector Machine resulted in the overall classification accuracy as high as 96.4%, which is a truly striking outcome. In this case, as shown in the tables 3 and 4, ambient and drum and bass both were classified 100% correctly and the accuracy for house and trance was 98% (one music excerpt from both classes was misclassified as belonging to techno). Techno was the least correctly classified having the accuracy of 86% (the all 7 misclassified songs were classified as belonging to trance and house).

Table 3. Detailed Accuracy by Class for SSR

Precision	Recall	F-Measure	Class
0.956	0.86	0.905	Techno
0.925	0.98	0.951	House
0.942	0.98	0.961	Trance
1	1	1	Dnb
1	1	1	Ambient

Table 4. Confusion Matrix for SSR

Techno	House	Trance	Dnb	Ambient	Class
43	4	3	0	0	Techno
1	49	0	0	0	House
1	0	49	0	0	Trance
0	0	0	50	0	Dnb
0	0	0	0	50	Ambient

The set composed of STFTMFCC, Statistical Spectrum Descriptors and Rhythm Histograms (denoted SMSR) also resulted in very high values of which the highest was 95.6% produced by Support Vector Machines. As can be seen from the tables 5 and 6, in this case drum and bass and ambient, again, were classified absolutely flawlessly. However, both techno and trance were classified slightly less accurately than in case of top performing feature set SSR.

Table 5. Detailed Accuracy by Class for SMSR

Precision	Recall	F-Measure	Class
0.955	0.84	0.894	Techno
0.875	0.98	0.925	House
0.96	0.96	0.96	Trance
1	1	1	Dnb
1	1	1	Ambient

Table 6. Confusion Matrix for SMSR

Techno	House	Trance	Dnb	Ambient	Class
42	6	2	0	0	Techno
1	49	0	0	0	House
1	1	48	0	0	Trance
0	0	0	50	0	Dnb
0	0	0	0	50	Ambient

The third best descriptor set, which also gave very good results, consisted of two feature sets: STFTMFCC and Rhythm Histograms (denoted SMR). Using Support Vector Machines it resulted in overall classification accuracy as high as 92.8%, which is also relatively high. Despite the fact that techno, trance and house were classified less accurately than in case of previously described top performing feature sets, both drum and bass and ambient still had the classification accuracy of 100%. More detailed information on this case can be found from the tables 7 and 8.

Table 7. Detailed Accuracy by Class for SMR

Precision	Recall	F-Measure	Class
0.867	0.78	0.821	Techno
0.833	0.9	0.865	House
0.941	0.96	0.95	Trance
1	1	1	Dnb
1	1	1	Ambient

Table 8. Confusion Matrix for SMR

Techno	House	Trance	Dnb	Ambient	Class
39	8	3	0	0	Techno
5	45	0	0	0	House
1	1	48	0	0	Trance
0	0	0	50	0	Dnb
0	0	0	0	50	Ambient

The feature set consisting of MIRToolbox and Rhythm Histogram feature set (denoted MRH) resulted in classification accuracy of 91.6% using Support Vector Machines. As can be seen from the tables 9 and 10 drum and bass was classified flawlessly. Also uplifting trance and deep house were classified very accurately having the correctly classified instances rate of 96%. Ambient had a little lower classification accuracy comparing to other combinational feature sets; albeit accuracy of 90% is quite high percentage. The other genres had the classification accuracies between 80% – 90%.

Table 9. Detailed Accuracy by Class for MRH

Precision	Recall	F-Measure	Class
0.98	1	0.99	Dnb
0.857	0.96	0.906	House
0.884	0.76	0.817	Techno
0.889	0.96	0.923	Trance
0.978	0.9	0.938	Ambient

Table 10. Confusion Matrix for MRH

Techno	House	Trance	Dnb	Ambient	Class
50	0	0	0	0	Dnb
0	48	2	0	0	House
0	5	38	6	1	Techno
0	1	1	48	0	Trance
1	2	2	0	45	Ambient

The feature set combined of MIRToolbox, Rhythm Histogram and Statistical Spectrum Descriptors (denoted MRS) obtained the classification accuracy of 91.2% using Simple Logistics. Using this feature set none of the genres were classified 100% correctly; however uplifting trance was close to it having the accuracy of 98%. The least accurately classified genres were deep house and techno with 88%, which are not bad results at all. Detailed information can be found from the tables 11 and 12

Table 11. Detailed Accuracy by Class for MRS

Precision	Recall	F-Measure	Class
0.898	0.88	0.889	Techno
0.846	0.88	0.863	House
0.98	0.98	0.98	Trance
0.92	0.92	0.92	Dnb
0.918	0.9	0.909	Ambient

Table 12. Confusion Matrix for MRS

Techno	House	Trance	Dnb	Ambient	Class
44	4	0	1	1	Techno
3	44	1	0	2	House
0	0	49	1	0	Trance
1	2	0	46	1	Dnb
1	2	0	2	45	Ambient

RS, the feature set composed of Rhythm Histograms and Statistical Spectrum Descriptors (denoted RS) also resulted in the overall classification accuracy of 91.2% using Vector Machines and AdaBoost with Support Vector Machines. The most accurately classified genre was drum and bass having the correctly classified instances rate of 100%. Deep house, trance and ambient also obtained good results; the accuracies were as high as 96%, 92% and 88% respectively. Techno was classified the least accurately having the percentage of 80%. More detailed information can be found from the tables 13 and 14.

Table 13. Detailed Accuracy by Class for RS

Precision	Recall	F-Measure	Class
0.909	0.8	0.851	Techno
0.814	0.96	0.881	House
0.92	0.92	0.92	Trance
1	1	1	Dnb
0.936	0.88	0.907	Ambient

Table 14. Confusion Matrix for RS

Techno	House	Trance	Dnb	Ambient	Class
40	5	4	0	1	Techno
0	48	0	0	2	House
3	1	46	0	0	Trance
0	0	0	50	0	Dnb
1	5	0	0	44	Ambient

The feature set consisting of Statistical Spectrum Descriptors and STFTMFCC (denoted SMS) resulted in overall classification accuracy of 87.6% using Support Vector Machines. As can be seen from the tables 12 and 13, ambient is classified 100% correctly and the other four genres have classification rates of 80% - 90%. Additional information can be found from the tables 15 and 16.

Table 15. Detailed Accuracy by Class for SMS

Precision	Recall	F-Measure	Class
0.816	0.8	0.808	Techno
0.896	0.86	0.878	House
0.882	0.9	0.891	Trance
0.788	0.82	0.804	Dnb
1	1	1	Ambient

Table 16. Confusion Matrix for SMS

Techno	House	Trance	Dnb	Ambient	Class
40	5	4	0	1	Techno
4	43	1	2	0	House
3	1	45	4	0	Trance
5	4	0	41	0	Dnb
0	0	0	0	50	Ambient

According to the before mentioned results there is clear evidence that concatenating different widely used feature sets with descriptors that extract the rhythmical patterns

from music would provide very good overall genre classification performance. Therefore similar approaches should be taken into use in the future works.

The best single performer was Rhythm Patterns feature set from RPextract resulting in overall classification accuracy of 86.8% using AdaBoost with Support Vector Machine. As can be seen from the tables 17 and 18 the drum and bass excerpts were classified totally flawlessly and ambient, house and trance were classified a little less accurately than in the best performing combinational cases. That probably comes from the fact that drum and bass has totally different drumbeat structure than other genres. However, applying Rhythm Patterns feature set the classification accuracy of techno is only 60%, which is relatively low compared to the other genres.

Table 17. Detailed Accuracy by Class for RP

Precision	Recall	F-Measure	Class
0.769	0.6	0.674	Techno
0.772	0.88	0.822	House
0.836	0.92	0.876	Trance
0.98	1	0.99	Dnb
0.979	0.94	0.959	Ambient

Table 18. Confusion Matrix for RP

Techno	House	Trance	Dnb	Ambient	Class
30	10	9	0	1	Techno
6	44	0	0	0	House
3	1	46	0	0	Trance
0	0	0	50	0	Dnb
0	2	0	1	47	Ambient

The Rhythm Histograms feature set narrowly came off second-best having the overall classification accuracy of 86%, which is only 0,8% less than in case of Rhythm Patterns. Within this feature set the best result was achieved using Simple Logistic machine learning algorithm having the parameterization variant 2 (see section 5.3 for details). The detailed information can be found in the tables 19 and 20.

Table 19. Detailed Accuracy by Class for RH

Precision	Recall	F-Measure	Class
1	0.94	0.969	Dnb
0.792	0.84	0.816	House
0.708	0.68	0.694	Techno
0.863	0.88	0.871	Trance
0.941	0.96	0.95	Ambient

Table 20. Confusion Matrix for RH

Dnb	House	Techno	Trance	Ambient	Class
47	0	2	0	1	Dnb
0	42	7	0	1	House
0	8	34	7	1	Techno
0	1	5	44	0	Trance
0	2	0	0	48	Ambient

The third relatively good single performer was STFTMFCC with the classification accuracy of 84.4% using AdaBoost with Support Vector Machines. It was the only single feature set that classified ambient 100% correctly. This time, again, the classification performance was the least accurate in case of techno. From the tables 21 and 22 more detailed information can be seen. This is the best single performer that was not from the RHextract toolbox. Therefore it comes of no big surprise that combining STFTMFCC with feature set that contains both Rhythm Histograms and Statistical Spectrum Descriptors gives high classification results.

Table 21. Detailed Accuracy by Class for SM

Precision	Recall	F-Measure	Class
0.66	0.7	0.68	Techno
0.774	0.82	0.796	House
0.922	0.94	0.937	Trance
0.885	0.76	0.817	Dnb
1	1	1	Ambient

Table 22. Confusion Matrix for SM

Techno	House	Trance	Dnb	Ambient	Class
35	9	1	5	0	Techno
8	41	1	0	0	House
2	1	47	0	0	Trance
8	2	2	38	0	Dnb
0	0	0	0	50	Ambient

The STFT set resulted in the overall classification accuracy of 76% using Simple Logistic algorithm with parameterization nr. 1. As comparison shows, this result is significantly lower than the three first single performers. However, ambient was classified 100% correctly and therefore other genres were classified significantly less accurately. Techno, deep house and uplifting trance all had the correctly classified percentage a little less than 70%. More detailed information on this case can be found from the tables 23 and 24.

Table 23. Detailed Accuracy by Class for STF

Precision	Recall	F-Measure	Class
0.66	0.66	0.66	Techno
0.694	0.68	0.687	House
0.66	0.66	0.66	Trance
0.784	0.8	0.792	Dnb
1	1	1	Ambient

Table 24. Confusion Matrix for STF

Techno	House	Dnb	Trance	Ambient	Class
33	7	6	4	0	Techno
8	34	7	1	0	House
7	4	33	6	0	Dnb
2	4	4	40	0	Trance
0	0	0	0	50	Ambient

The fifth best single feature data set was Statistical Spectrum Descriptors set having the classification accuracy of 75.6% using Simple Logistic machine learning algorithm with

first parameterization. The most accurately classified genre was ambient, which did not contain obtrusive drumbeats, having the accuracy of 88%. Other 4 genres were classified more or less equally having the classification rates around 70%. Further information on this case can be found from the tables 25 and 26.

Table 25. Detailed Accuracy by Class for SSD

Precision	Recall	F-Measure	Class
0.673	0.7	0.686	Techno
0.725	0.74	0.733	House
0.755	0.74	0.747	Trance
0.692	0.72	0.706	Dnb
0.957	0.88	0.917	Ambient

Table 26. Confusion Matrix SSD

Techno	House	Trance	Dnb	Ambient	Class
35	8	3	3	1	Techno
5	37	1	6	1	House
1	5	37	7	0	Trance
9	0	5	36	0	Dnb
2	1	3	0	44	Ambient

The MFCC set also had the overall classification accuracy of 75.6% using AdaBoost with Support Vector Machines. Again, the classification accuracy of ambient was 88% and the other 4 genres had the correctly classified instances rate of about 70%. Detailed information can be found from the tables 27 and 28.

Table 27. Detailed Accuracy by Class for MFC

Precision	Recall	F-Measure	Class
0.673	0.7	0.686	Techno
0.725	0.74	8.733	House
0.692	0.72	0.706	Trance
0.755	0.74	0.747	Dnb
0.957	0.88	0.917	Ambient

Table 28. Confusion Matrix for MFC

Techno	House	Dnb	Trance	Ambient	Class
35	8	3	3	1	Techno
5	37	6	1	1	House
9	0	36	5	0	Dnb
1	5	7	37	0	Trance
2	1	0	3	44	Ambient

Next single performer was the LPCC set, which resulted in classification accuracy of 74.8% using C4.5 and both variants of Support Vector Machines. The tables 29 and 30 reflect the performance using Support Vector Machines. Ambient was classified the best resulting in classification accuracy of 100%. Classification of trance also showed relatively good performance having the accuracy of 90%. However, in case of techno and house, the classification accuracies were very low, 54% and 58% respectively. Many techno excerpts were classified as belonging to deep house and vice versa.

Table 29. Detailed Accuracy by Class for LPC

Precision	Recall	F-Measure	Class
0.614	0.54	0.574	Techno
0.617	0.58	0.598	House
0.667	0.72	0.692	Dnb
0.818	0.9	0.857	Trance
1	1	1	Ambient

Table 30. Confusion Matrix for LPC

Techno	House	Dnb	Trance	Ambient	Class
27	13	9	1	0	Techno
12	29	6	3	0	House
5	3	36	6	0	Dnb
0	2	3	45	0	Trance
0	0	0	0	50	Ambient

The lowest performing single performer was MIRTSet with classification rate of 73.2%. This time both ambient and deep house obtained the results of 84%, which was the best within this dataset. Again, techno was classified the least accurately having the accuracy of 56%. More detailed information on this case can be found from the tables 31 and 32.

Table 31. Detailed Accuracy by Class for MIR

Precision	Recall	F-Measure	Class
0.757	0.56	0.644	Techno
0.724	0.84	0.778	House
0.684	0.78	0.729	Dnb
0.582	0.64	0.61	Trance
0.977	0.84	0.903	Ambient

Table 32. Confusion Matrix for MIR

Techno	House	Dnb	Trance	Ambient	Class
28	9	9	4	0	Techno
3	42	4	0	1	House
3	3	32	12	0	Dnb
1	1	9	39	0	Trance
2	3	1	2	42	Ambient

6 Summary and conclusions

The objective of the thesis was to focus on genre classification of electronic music combining different traditionally used audio descriptors (features) for music classification with models that are capable of extracting rhythm patters from music and evaluate the performance of such approach. Five different electronic music genres such as deep house, uplifting trance, techno, drum and bass and ambient were used.

In general, most of the outcomes showed very good classification performance. However, there are no prior works with similar dataset distribution hierarchy to compare these results with. As these findings suggest, combining Rhythm Patterns and Rhythm Histograms feature sets from RPtoolbox with top performing feature sets from Marsyas would produce very high classification accuracies. The combined feature set consisting of feature sets such as STFT, Statistical Spectrum Descriptors and Rhythm Histograms resulted in the overall classification accuracy of 96.4%. Moreover, in this case the classification accuracy of 100% for ambient and drum and bass was obtained. In addition, combining these two sets with MIRToolbox set (which did not perform as good as the best Marsyas sets) and other sets from Marsyas, would also perform surprisingly well.

The general tendency shows that ambient and drum and bass are classified the most accurately and techno is classified the least accurately. The considerably high misclassification rate of techno is probably conditioned from the fact that the techno dataset contained excerpts from various subgenres (e.g. minimal techno, banging techno etc.) of techno and therefore the variance in sound within the genre was relatively big. The misclassified techno tracks were usually classified as belonging to uplifting trance and deep house. What it basically means is that confusions occurred mostly between 3

genres containing similar drumbeats. To illustrate this, ambient and drum and bass were not confused neither with each other nor with other 3 genres containing somewhat similar drumbeats. Therefore it can be said that this kind of approach is able to distinct different music excerpts containing different rhythmical patterns. However, depending on the feature set deep house and uplifting trance are also classified considerably accurately.

In most of the cases, the best machine learning algorithm seems to be Support Vector Machines. Simple Logistic and Naïve Bayes also performed better than most of the other classifiers. The least accurate method is C4.5, which produced a classification accuracy of 78.4% using the best performing combinational feature set.

This study also showed that using stereo files for feature extraction using Marsyas gives surprisingly significantly better results than using traditionally used mono files. In some cases the classification accuracy reduced almost by 8% using mono files. However, more research would be needed in order to find explanation for that.

6.1 Further work

While doing this master's thesis a few limitations occurred. First of all, the number of genres was limited and therefore should be increased in future works in order to provide more in-depth results. Since creating the combinational sets was quite troublesome, many of the sets that were intended to use were left out of the scope of this work. Therefore more emphasis should be put to genre classification using different new combinational feature sets not tackled here. In addition unsupervised clustering should be utilised in order to see how well that kind of approach would work. As a surprising result, the results using Marsyas feature sets extracted from the stereo files give drastically higher results; that phenomenon should also be considered to be researched.

Bibliography

Li, T., Ogihara, M., Li, Q. (2003). A comparative study on content-based music genre classification. In Proceedings of the International ACM SIGIR Conference on Research and Development in Information Retrieval, 282 – 289, Toronto, Canada.

Pampalk, E. (2006). Computational Models of Music Similarity and their Application in Music Information Retrieval. Doctoral Thesis, Vienna University of Technology, Austria.

Aucouturier, J., Pachet, F. (2004). Improving Timbre Similarity: How high is the sky? Journal of NegativeResults in Speech and Audio Sciences 1(1).

Pohle, T. (2005). Extraction of Audio Descriptors and their Evaluation in Music Classification Tasks Master's thesis, Technische Universität Kaiserslautern, Austrian Research Institute for Artificial Intelligence (ÖFAI), Deutsches Forschungszentrum für Künstliche Intelligenz (DFKI).

Tzanetakis, G., Essl, G., Cook, P. (2001). Automatic musical genre classification of audio signals. In Proceedings of The Second International Conference on Music Information Retrieval and Related Activities.

West, K., Cox, S. (2004). Features and classifiers for the automatic classification of musical audio signals. In Proceedings of the Fifth International Conference on Music Information Retrieval (ISMIR).

Mohd, N., Doraisamy, S., Wirza, R. (2005). Factors Affecting Automatic Genre Classification: an Investigation Incorporating Non-Western Musical Forms. In Proceedings of the Fifth International Conference on Music Information Retrieval.

Homburg, H., Mierswa, I., Moeller, B., Morik, K., Wurst, M. (2005). A benchmark dataset for audio classification and clustering. In Proc. ISMIR, 528-531.

Pampalk, E., Flexer, A., Widmer, G. (2005). In the Proceedings of the 6th International Conference on Music Information Retrieval, London, UK, September 11-15.

Morchen, F., Ultsch, A., Thies, M., Lohken, I., Nocker, M., Stamm, C., Efthymiou, N., Kummerer. M. (2005) MusicMiner: Visualizing timbre distances of music as topograpical maps. Technical report, CS Dept., Philipps-University Marburg, Germany.

Pohle, T., Pampalk, E., Widmer, G. (2005). Evaluation of Frequently Used Audio Features for Classification of Music into Perceptual Categories.. Proceedings of the Fourth International Workshop on Content-Based Multimedia Indexing (CBMI'05) (Riga, Latvia).

Lampropoulos, A. S., Lampropoulou, P. S., Tsihrintzis, G. A.. (2005). Musical genre classification enhanced by improved source separation techniques. In Proceedings of the Fifth International Conference on Music Information Retrieval.

West, K., Cox, S. (2005). Finding an Optimal Segmentation for Audio Genre Classification. In Proceedings of the Fifth International Conference on Music Information Retrieval.

Pampalk, E., Flexer, A., Widmer, G. (2005). Improvements of Audio-Based Music Similarity and Genre Classification. In Proceedings of the Fifth International Conference on Music Information Retrieval.

Liu, X., Wang, D. (2001). Neural Networks. Proceedings. IJCNN apos;01. International Joint Conference on Volume 2, Issue , 2001 Page(s):1083 - 1088 vol.2 Fabbri, F. (1999). Browsing music spaces: categories and the musical mind. In Proceedings of the IASPM Conference.

Schedl, M., Knees, P., Widmer**,** P. (2005). Discovering and Visualizing Prototypical Artists by Web-based Co-Occurrence Analysis. Proceedings of the 6th International Conference on Music Information Retrieval (ISMIR'05), pp. 21-28,
London, UK, September 11-15, 2005.

Kemp, C. (2005). Towards a holistic interpretation of musical genre classification. Doctoral Thesis, University of Jyväskylä, Finland.

House music. (June 2007) In Wikipedia, The Free Encyclopedia. Retrieved April 28, 2007, from http://en.wikipedia.org/wiki/House_music

Deep house. (June 2007) In Wikipedia, The Free Encyclopedia. Retrieved April 28, 2007, from http://en.wikipedia.org/wiki/Deep_house

Trance music. (June 2007) In Wikipedia, The Free Encyclopedia. Retrieved April 28, 2007, from http://en.wikipedia.org/wiki/Trance_music

Uplifting trance, (June 2007) In Wikipedia, The Free Encyclopedia. Retrieved April 28, 2007, from http://en.wikipedia.org/wiki/Uplifting_Trance

Lang, M (1996, December) Futuresound: Techno Music and Mediation. Etnomusicology Senior Project, University of Washington. Retrieved May 10, 2007, from http://music.hyperreal.org/library/fewerchur.txt

Paperduck. (n.d.). What is techno? Retrieved May 10, 2007, from http://analogik.com/article_techno_4.asp

Ambient music. (n.d.). Retrieved April 28, 2007, from http://www.synthtopia.com/Articles/ElectronicMusicStylesAmbi.html

Ambient music. (June 2007) In Wikipedia, The Free Encyclopedia. Retrieved April 28, 2007, from http://en.wikipedia.org/wiki/Ambient_music

Drum and Bass. (June 2007) In Wikipedia, The Free Encyclopedia. Retrieved April 28, 2007, from http://en.wikipedia.org/wiki/Drum_and_bass

Kosina, K. (2002). Music Genre Recognition. PhD thesis, Technical College Hagenberg.

Pfeiffer, S., Vincent, J.P. (2001). Formalization of MPEG-1 compressed domain audio features, Report No 01/96 of CSIRO Mathematical and Information Sciences, Australia, Dec.

Misdariis, N., Smith, B. K., Pressnitzer, D., Susini, P., McAdams S. (1998). Validation of a Multidimensional Distance Model for Perceptual Dissimilarities among Musical Timbres. The Journal of the Acoustical Society of America, Volume 103, Issue 5, May 1998, pp.3005-3006

Gravier, G. (March 5th 2004). Speech analysis techniques. Retrieved April 11, 2007, from http://www.irisa.fr/metiss/guig/spro/spro-4.0.1/spro_2.html

Smith III, J.O. (August 2006) Physical Audio Signal Processing,"Linear Predictive Coding of Speech" section. Retrieved May 13, 2007, from http://ccrma.stanford.edu/~jos/pasp/Linear_Predictive_Coding_Speech.html.

Tanghe, K., Degroeve, S., De Baets, B. (2005). An algorithm for detecting and labeling drum events in polyphonic music. In: Proceedings of the first Music Information Retrieval Evaluation eXchange (MIREX), London, United Kingdom, September 11-15.

Paiva, R. P., Mendes, T., Cardoso, A. (2005). On the Detection of Melody Notes in Polyphonic Audio. in Proc. of the 6th International Conference on Music Information Retrieval, ISMIR'2005, London, UK.

McKay, C., Fiebrink, R., McEnnis D., Li, B., Fujinaga, I. (2005). ACE: A Framework for Optimizing Music Classification. ISMIR 2005: 42-49

Pfeiffer, S. (2002). Maante Fundamental modules. Retrieved March 12, 2007, from http://www.cmis.csiro.au/maaate/docs/modules.html#spflux

Typke, R., Wiering, F., Veltkamp, R.C. (2005). A Survey of Music Information Retrieval Systems. In: Proceedings of the Fifth International Conference on Music Information Retrieval (ISMIR 2005), pp 153-160.

Aucouturier, J.J., Pachet, F. (2003) Representing Musical Genre: A State of the Art. In Journal of New Music Research, 32(1).

Lartillot, O., Toiviainen, P. (2007) A Matlab Toolbox for Musical Feature Extraction from Audio. Paper presented at the 8th International Conference on Music Information Retrieval, Vienna, Austria.

Mower, E. (2003). K nearest neighbor. Retrieved March 24, 2007, from http://www.cra.org/Activities/craw/dmp/awards/2003/Mower/KNN.html

Teknomo, K.(2006). What is K nearest neighbor algorithm? March 24, 2007, from http://people.revoledu.com/kardi/tutorial/KNN/What-is-K-Nearest-Neighbor-Algorithm.html

Naïve Bayes rule generator. (n.d.). Retrieved March 28, 2007, from http://grb.mnsu.edu/grbts/doc/manual/Naive_Bayes.html

Mitchell, T. M. Decision trees, MDL, Boosting. (n.d.). Retrieved May 12, 2007, from http://www.cs.cmu.edu/~guestrin/Class/10701-S05/slides/DTrees-MDL-Boosting-2-9-05.pdf

Dankel II, D. (1997). The ID3 algorithm. Retrieved May 12, 2007, from http://www.cise.ufl.edu/~ddd/cap6635/Fall-97/Short-papers/2.htm

Hearst, M.A. (1998). Support vector machines. IEEE Intelligent System. v13 i4. 18-28.

Matas & Šochman. (2004). AdaBoost. Retrieved May 12, 2007, from http://www.robots.ox.ac.uk/~az/lectures/cv/adaboost_matas.pdf

Boosting. (2007). In Wikipedia, The Free Encyclopedia. Retrieved May 12, 2007, from http://en.wikipedia.org/wiki/Boosting

Freund, Y., Schapire, R. E. (1999). A short introduction to boosting**.** Journal of Japanese Society for Artificial Intelligence, 14(5):771-780, September, 1999. (Appearing in Japanese, translation by Naoki Abe.)

Landwehr, N., Hall, M., Frank, E. (2003). Logistic model trees. In Proc 14th European Conference on Machine Learning (pp. 241-252). Springer-Verlag.

Amini, M., Gallinari, P. (2002). Semi Supervised Logistic Regression. In F. van Harmelen (ed.): ECAI2002, Proceedings of the 15th European Conference on Artificial Intelligence, IOS Press, Amsterdam, 2002, pp.390-394.

Friendly, M. (2007). Logistic regression. Retrieved May 13, 2007, from http://www.math.yorku.ca/SCS/Courses/grcat/grc6.html

Garson, G. G. (2006). Logistic regression. Retrieved May 13, 2007, from http://www2.chass.ncsu.edu/garson/PA765/logistic.htm

Breiman, L., Cutler, A. Random forests. Retrieved May 13, 2007, from http://www.math.usu.edu/~adele/forests/cc_home.htm

Lidy, T. (2005). Rhythm patterns. Retrieved March 5, 2007, from http://www.ifs.tuwien.ac.at/~lidy/rp/

Naïve Bayes classifier. In Statsoft Electronic statistics textbook. Retrieved March 28, 2007, from http://www.statsoft.com/textbook/stnaiveb.html

Naïve Bayes. In GRB tool shed manual. Retrieved March 28, 2007, from http://grb.mnsu.edu/grbts/doc/manual/Naive_Bayes.html

SVM – Support vector machines. In DTREG software homepage. Retrieved March 30, 2007, from http://www.dtreg.com/svm.htm

Dannenberg, ``Toward Automated Holistic Beat Tracking, Music Analysis, and Understanding," in ISMIR 2005 6th International Conference on Music Information Retrieval Proceedings, London: Queen Mary, University of London, (2005), pp. 366-373

K. West and S. Cox, “Finding an optimal segmentation for audio genre classification,” in Proceedings of 6th International Conference on Music Information Retrieval (ISMIR '05), pp. 680–685, London, UK, September 2005.

Appendix

Tracklist

Deep House

01. Andreas Bender-Untitled
02. Arch Typ-Shades of Blue
03. Arch Typ-Love in Slow Motion
04. Atnarko-Don't Ya Know
05. Black Fuse-Siuation Green
06. Catalan Fc & Sven Love-Real Love
07. D Trueitt & Ric-Stormy Day
08. John Daly-Sky Dive
09. Digital Minds-Be Yourself
10. Filsonik-Evolution
11. Fish Go Deep-ESL
12. Craig Hamilton-Average Day
13. Hanna-Sanctuary
14. Lee Jones-There Comes a Time
15. Karu-Desire
16. Kevin Yost-Like a Dream to Me
17. Shawn Ward-Jazzy Dream
18. Mikee Deep-Baby
19. Shawn Ward-Time Machine

20. Marathon Men-Bye Bye Babe
21. W Beeza-Feel My Lovin
22. jay Tripwire-Call&Answer
23. Soul System-Desperate Measures
24. Jay Tripwire-Denman Place
25. Jay West-Power to Create
26. Jay Tripwire-Harmony & Peace
27. Kevin Yost- Untitled
28. William Flynn-Sian
29. Boundzound-Louder
30. Carlos Sanchez-Body Motion
31. Soul Buddha-Realize
32. Slowly & Alison Crocket-Black Sun
33. Richard Les Crees-Deep Thought
34. Karu-Perfect Love
35. Mark O'Sullivan-Prayers
36. Powel Kobak-Always be Around
37. From P60-I'm Not the Same
38. Jay Tripwire-Brothers&Sisters
39. Priit Kirss-Break Away
40. Priit Kirss-Sounds Of Autumn
41. Ananda Project-Many Starred Sky
42 Dj Replee-I Love the Way
43. Larry Heard-Changes
44. Gawron Paris-Workaholic Man
45. Solar House-Everything Changes
46 L'Renee-Say My Same
47. Powel Kobak-Always Be Around
48. Tom&Joyce-Vai Minha Tristeza
49. The Sound Diggers-Dave Snare
50. Palm Skin Productions-Wrecked

Techno

01 Convexton-Miranda
02. Messenger-Wanderer
03. Aardvarck-Cult Copy
04. Rino Cerrone-Burnt It
05. Zoxfeld-Devon
06. Kali-Tribetech
07. Fusiphorm-Childhood
08. Andreas Kremer-Polarlicht
09. Andreas Kremer-Weltenbummler
10. Mark Broom-Highs & Lows
11. Basic Implant-Disharmony
12. Valentino Kanzyani-Summer in Slovenia
13. Safety Scissors-Where Is Germany & How Do I Get There
14. Detroit Grand Pubahs-Skyddive from Venus
15. Funk D'Void & Phil Kieran-Black as You Like
16. Christian Fisher-Undisturbed
17. Pascal Feos-Ausklang
18. Carl Falk-Entry
19. Carl Falk-Plast
20. Vitiello maurizio-Just a Click
21. Uto Karem-Different Shapes
22. Marko Furstenberg-Untitled
23. Elliot Dodge-City Lights
24. Elliot Dodge-Stalker
25. Dejan Milicevic-Sort of a Flower
26. Mindhole-Clown's Pit
27. Cari Lekebusch-Level of Reality
28. Echoplex-Close Up
29. Audio53- Unknown

30. Alexi Delano-I'm Tired
31. Methodology-Path of Least Resistance
32. Raul Mezcolanza-Fried Eggs
33. Dejan Milicevic-Spectrum Of Sound
34. A Paul-Genration
35. Loner9-Minimal
36. Grimes Adhesif-Educated Derelicts
37. Grimes Adhesif-Locked Minds
38. Co Fusion-Pixy Zap
39. Co Fusion Als
40. Ben Klock-Similar Colors
41. Joel Mull-Begun The End Has
42. Monoplex-Zipzap
43. Monoplex-Sounds Of Time
44. A Paul-Awareness
45. Glenn Wilson-Northen Rise
46. Glenn Wilson-Sub Wave
47. A Paul & Industriaizer-Whatever
48. David Moleon-Episodio
49. Dito Masat-El Rio
50. Bassdrum-Ugoluna

Uplifting Trance

01. Denga&Manus-Firefly
02. Denga&Manus-Firefly
03 Chemistry-Prophecy
04. Gabriel Batz-Inner Touch
05. Cressida-The Secred Inredient
06. Cressida-Laika
07. Leon Bolier-Lyra

08. Above & Beyond-Can't Sleep

09. Cern-The Message

10 Enmass-Beyond Horizon

11. Cern-Go Fly

12. Super8 & DJ Tab-Needs To Feel

13. Tronic-Inside Outside

14. Elevation-Blinding truth

15. Carl B-Optimum

16. Emotional Horizons-Autumn

17. Nitrous Oxide-Frozen Dreams

18. Daniel Kandi-Breathe

19. Motionchild & Armenian Sun-GodSend

20 Northen Comfort-Don't Look Back

21. Andre Visior-Skyline

22. Activa-Airflow

23 4Fach Zoom-Pixel One

24. Sean Tyas-Lift

25. Nunrg-Kosmosy

26 Denga&Manus&Mque-Loosing Senses

27. Beetseekers-Reflections 07

28. Beetseekers-Synthesize

29. Activa&Tom Colontinio-Enlighten

30. Aly & Fila-Ankh

31. Aly & Fila-Ureus

32. Matt Abbot-Illusions

33. AB Project-Eternal Optimism

34. Temple One-Forever Searching

35. BBE-Seven Days &One Week

36. Lawrence Palmer-Streamline

37. Stuart C-Airborn

38. Octagen & Arizona-Starburst

39. Mr Sam-Smeya
40. Markus Schulz-First Time
41. Niklas Harding Presents Arcane-Blue Circles
42. Clear & Present-Elevate
43. John O'Callaghan-Split Decision
44. Armin Van Buuren-Blue Fear
45. Davide Bomben-So Real
46. James Wood presents WANDII-Kinetic Caper
47 Sonic Division-Painting The Scilence
48. Tiesto-Bright Morningstar
49. Tiesto-Elements Of Blue
50. Armin Van Buuren-4 Elements

Drum and Bass

01. 2529 & Contour-Xotic
02. Adam Form-Down Inside
03. Agent Alvin- Unknown
04. Alter Ego-Infection
05. Arp XP-Night Train
06. ASC&MAV-Too Deep For Ya
07. ASC&MAV-Sceptical
08. Assonance & Jazz Thieves- Unknown
09. Atlantic Connection-Let It Burn
10. NHS118- Unknown
11. Atlantic Connection- Unknown
12. Autumn- Unknown
13. Motion-Elements Of Truth
14. Nookie-Get Down
15. Nocturnal-Been So Long
16. Kryptic Minds and Leon Switch-After Life

17. Kryptic Minds and Leon Switch- The Forgotten
18. Dan Marshall-Side Step
19. Klute-Flight
20. Heist Jazz time
21. Klute-Freedom Come
22. Grand Masterz- Unknown
23. J Cut & Electrosoul System-Come Around
24. Beta2 & Zero Tolerance-The Beaten Track
25. Big Bud-Rice & Beans
26. Grand Masterz-Unknown
27. Blame- Unknown
28. Break-Not Enough
29. FX909-The Request
30. Fellowship-Unknown
31. Ez Rollers-Lost & Found
32. Cubist-Live & Let Die
33. Brkag-I'NI
34. DK Foyer & Jeber-Rhytual
35. Commix-Talk to Frank
36. Contour-Masquerade
37. Dizplay - Freakwave
38. SKC-Vandalism
39. Mistical-Time to Fly
40. Moving Fusion-Radiance
41. Current Value &Infamy-Trail Of Tears
42. Greg Packer-Pheety Funk
43. Loxy & Ink-Killing Season
44. Technical Itch-Retribution
45. Gyromite & Subsonic-Unknown
46. Big Bud-Red Snapper
47. Bingo62-Unknown

48. Dose-Words Of Wisdom
49. Booty-Scenario
50. The Chosen-Superhuman

Ambient

01. Steve Roach-Circular Ceremony
02. Kiln-Unknown
03. Cyber Chump- Unknown
04. Steve Roach-After the Dream
05. Cyber Chump- Unknown
06. Unknown-Woomera
07. Diatonis-Between Fenceposts
08. Oöphoi-Beyon These Skies
09. Unknown-Breathe
10. Alpha Wave Movement- Drifted Into Deeper Land
11. Thought Guild-Silicon Alchemist
12. Diatonis-Tall Shadows
13. Farfield-Sun Across My Eyes
14. Diatonis-Night Drive
15. Danny Kreutzfeldt-Road
16. Seofon-Zeropoint
17. Unknown-The Seventh Portal
18. Hector Zazou-Unknown
19. Michae Bentley-Parsec
20. Steve Roach-Oracle
21. Unknown-Nomansland
22. Biosphere- Mestigoth
23. NID-Tower Of Babel
24. Nerthus-The Inharmonic Heater

25. Cyber Chump-First Transmission
26. Erebus and Terror- Komgawa
27. Marconi Union- Unknown
28. Diatonis-Flatland
29. Steve Roach & Roger King-Gone West
30. Steve Roach & Roger King-Ghost Train
31. Erebus and Terror-Frozen Ship
32. Oöphoi-Fragile Beauty
33. Unknown-Foresight
34. Diatonis-Fountains of Hycinth
35. Kiln- Unknown
36. Kiln-Unknown
37. Diatonis-Lucid Dreaming
38. Thought Guild-Semiotic Sequence
39. Gianfranco Grilli-Organic
40. Diatonis-Glass Of Starlight
41. Danny Kreutzfeldt-Lair
42. Lien-Mirablau
43. Mischgewebe-That Witch Swallows
44. Roel Meelkop-Thanatos Springs
45 Scott Gibbons & Socetas Rafaello Sanzio-Unknown
46. Chris Zippel-Space Dock
47. Thin Films-Unknown
48. Al Haca Soundsystem-Untitled & Farda P
49. Unknown -Unknown
50. Diatonis -Winding Road