Automatic Classification of Heavy Metal Music

D.G.J. Mulder

July 18, 2014

Thesis for BSc Mathematics and BSc Computer Science

Supervisor: Dr John Ashley Burgoyne

Korteweg-de Vries Instituut voor Wiskunde
Instituut voor Informatica
Faculteit der Natuurwetenschappen, Wiskunde en Informatica

Universiteit van Amsterdam

Abstract

In this thesis, I construct an automatic classification system for heavy metal. The aim of this system is correctly categorizing digital audio files, containing heavy metal music, into their respective subgenres. First, a history of heavy metal subgenres is discussed. Then, as a new contribution to this research field, two features are proposed: the *vertical* and *horizontal interval features*. These features are based on the concept of musical intervals and are built around the computation of *chroma vectors*, first introduced by Fujishima in 1999 [1]. After this, I discuss several distance functions and classifiers, namely a classifier based on the *Mahalanobis distance*, and the *k-nearest neighbor* classifier with the *Aitchison distance*. This classification system is evaluated on a manually assembled collection of heavy metal audio files. With their respective best-performing classifiers, we observe an average accuracy of .28 for the vertical and an accuracy of .21 for the horizontal interval feature, as compared to a chance rate of .06. However, their performances are closer together with the construction of a *confusion cost matrix*, where each classification is assigned a cost based on the severity of their particular subgenre confusion. These results are discussed and put into a musical context.

Title: Automatic Classification of Heavy Metal Music

Authors: D.G.J. Mulder, unarmedlad@gmail.com
Supervisor: Dr John Ashley Burgoyne
Second signatory: Dr Gerard Helminck

Date: July 18, 2014

Faculteit der Natuurwetenschappen, Wiskunde en Informatica
Universiteit van Amsterdam
Science Park 904, 1098 XH Amsterdam
`http://www.science.uva.nl/`

Contents

1. Introduction

Motivation

Heavy metal is a controversial genre. Outsiders often see it as unsophisticated, low-brow and 'heavy-for-the-sake-of-being-heavy'. In contrast, there exist legions of metal fans who applaud it for its intensity, unconventional song structures, lyrical references to mythology and theology, and sometimes even virtuosic musicianship. Most people have little idea that the concept 'heavy metal' functions as an umbrella term that comprises many different (sub-)subgenres, some of which are sonically and culturally incomparable. Ever since I was introduced to heavy metal music, I have been interested in these distinctions: what makes this band closer to 'folk metal' than to 'black metal', is that band closer to 'thrash metal' or to 'power metal', and why can most fans of 'death metal' not stand 'nu metal'? Two people that both claim to listen to metal can have entirely different tastes.

Another thing that has always interested me is the way computers are utilized in the recognition of music similarity. It fascinates me how a service like Spotify can, when given an artist, create a whole radio station with music it claims to be similar. Some services, like Last.fm[1], have a large user base whose listening habits are monitored and analyzed in order to find such similarity, while others, like MusicIP Mixer[2] and Pandora[3], base their similarity algorithms purely on characteristics obtained from musical analysis.

No matter which method is used, people who design such systems often have little idea of the classificational subtleties of metal music, and so do not take them into account. This can lead to metal fans (colloquially called 'metalheads') having a much worse experience with these music services than the average music listener would. As such, there is a need for a method that can more accurately distinguish among these different subgenres of metal. In this thesis, I will attempt to use algorithmic music analysis and pattern recognition techniques to construct a method that will aid in automatic subgenre classification of heavy metal music.

Related work

The research area to which this thesis belongs is called *automatic music classification* or *music genre recognition*. My academic introduction to this subject was an article by Tzanetakis and Cook, written in 2002 [2]. The field has come a long way since its writing. While the algorithms used by the commercial services discussed above are usually trade

[1] `http://www.last.fm/`
[2] `https://musicbrainz.org/doc/MusicIP_Mixer`
[3] `http://www.pandora.com/`

secrets, a state-of-the-art system that is available in academic literature is AdaBFFs [3], which assembles votes of classifiers through the use of decision trees. Another is SRCAM [4], which uses overcomplete dictionaries of auditory features for sparse representation classification. A final system is MAPsCAT [5], which chooses classes based on minimum expected risk in a framework of Bayesian statistics. In this thesis, we will build a method from the ground up, rather than utilize these pre-existing systems. Because of this, our system will not be as comprehensive or as complete as them.

A comprehensive overview of related articles is given by Sturm [6], along with a fundamental criticism of the research methods used in the field: results might be inflated due to the presence of irrelevant confounding factors, and so, systems with high reported accuracies might nevertheless be unable to recognize genres. I did not seek out to revolutionize the way this research is done and many of Sturm's criticisms still apply to this thesis. Nonetheless, I take them into account by providing a musical contextualization of the results in addition to the results themselves. As far as I know, no one has researched automatic classification specifically for heavy metal subgenres before.

2. Background

2.1. Heavy metal subgenres

What needs to be settled first is which subgenres we will use for classifying the music. What the exact subgenres of heavy metal are is a hotly debated topic. While most metal listeners will agree about broad distinctions, the existence or non-existence of many smaller genres is not agreed upon at all. Note that in the metal community, there is the opinion that most, if not all, music under the umbrella terms 'metal' or 'heavy metal' can, in fact, be categorized under one of its subgenres (as evidenced, for example, by the fact that the RateYourMusic[1] user base rejects the use of the 'Metal' genre label for releases, in favor of any of its subgenres). Therefore, the genre 'heavy metal' itself will not be used for classification.

There can also arise some confusion over the use of the term 'heavy metal' itself. It has both been used as a catch-all term for all music in its subgenres (synonymously to 'metal'), as well as to refer to the traditional style of heavy metal. In this thesis, we will restrict ourselves to the former usage. The traditional style will be called *classic metal*.

To give some context to this thesis, it is appropriate to briefly discuss the history of heavy metal and its subgenres. Readers without an interest in heavy metal music can skip to section 2.1.2 without problems, although some knowledge is presumed in the discussion chapter. Unlike some authors, we will make a distinction between *movements* and *genres* in this thesis, a *movement* being used to describe a collection of artists appearing around the same time that share a similar locality, aim, attitude or aesthetic, while *genres* group music together based on purely musical characteristics, regardless of locality and time.

2.1.1. History of heavy metal and its subgenres

As the story is often told, heavy metal started in England, 1969, with a band named Black Sabbath. Inspired by a horror film of the same name, they wrote an eponymous song that featured an ominous riff based around the musical tritone interval. Taken together with other sonic elements such as the distorted sound of Tony Iommi's guitar, this is considered by many historians to be the birth of what we nowadays call 'heavy metal' [7, 8]. Black Sabbath would continue to be a successful band throughout the 1970s and 1980s, and in fact still performs today. In the 1970s, several bands such as Motörhead and Judas Priest fused the innovations of Black Sabbath with the hard rock techniques and tempo of bands like The Jimi Hendrix Experience, Led Zeppelin

[1] `https://rateyourmusic.com/`

and Deep Purple to create the traditional heavy metal sound [7, 9], henceforth referred to as *classic metal* (in analogy to *classic rock*). Simultaneously, other bands such as Pentagram and Saint Vitus wanted to focus more exclusively on the Black Sabbath sound. They sought to recreate and expand on the brooding slowness and tritone usage of Black Sabbath's famous riff, and thus forged the subgenre of *doom metal* [7].

In the late 1970s and early 1980s, there was a movement of mostly classic metal bands in England, dubbed the *New Wave of British Heavy Metal* (or *NWOBHM*). The bands in this movement shared a do-it-yourself attitude and a desire to set heavy metal as a genre apart from hard rock by speeding up and utilizing aggressive imagery. Typical examples of NWOBHM bands are Iron Maiden and Venom [7, 10]. At the same time, heavy metal had begun to take hold in North America. While proper metal bands could still be found, the most popular 'heavy metal' in the United States from that time nowadays is often regarded as being hard rock utilizing some elements and imagery from classic metal bands [7] (Van Halen, Y&T, Kiss). In the underground, bands like Metallica and Exodus rejected this 'false metal', taking influence from the NWOBHM and the aggression of hardcore punk to create a genre called *thrash metal*. Thrash metal often features palm-muted rhythm guitar parts, virtuosic guitar solos and even faster tempos than NWOBHM bands. Sometimes, vocals are shouted, in contrast to the melodic singing that is common in classic metal [7, 11].

In the ever-burning desire for higher speeds and more extremity, in the early-to-mid 1980s thrash metal bands delivered a new string of adrenaline-fueled albums that each attempted to be faster, heavier and more brutal than what came before, culminating in Possessed's *Seven Churches* in 1985 and Slayer's *Reign in Blood* in 1986. Featuring low-register tremolo-picked riffs, chaotic chromatic guitar solos and lyrical themes of serial killers and Satanism, they are often cited as a template for the next degree of extremity in metal subgenres: *death metal* [12]. Death metal bands Death and Morbid Angel took this template and added an inhuman-sounding vocal style, often referred to as 'growling' or 'grunting', and extremely fast blasting drums ('blast beats') to solidify this new style [7, 8, 13]. Seemingly paradoxically, this extremity sometimes gave room to more atmospheric elements in the music, in the form of slower, more melancholic riffs, giving rise to fusion genres such as *death/doom*. (Not considered as a separate genre in our taxonomy. We would like to be able to describe such bands as being on the edge of doom metal and death metal.)

In the late 1980s, metal had seen the chance to proliferate all over the world. In Germany, a band called Helloween was influenced by the fantasy elements, theatrical grandeur and operatic singing of hard rock and classic metal bands such as Rainbow, Iron Maiden, Judas Priest and Dio. They fused the classic metal style with the more technical riffing techniques and speed of thrash metal, and added guitar solos that were inspired by classical music. Helloween is often seen as the first *power metal* band [14]. Power metal is a genre that is obsessed with high fantasy atmosphere, alluding to works like J.R.R. Tolkien's *The Lord of the Rings*. Bands employ grand, epic melodies and instruments like acoustic guitars, keyboards and sometimes instruments associated with folk music like flutes to achieve this. They eschewed the dissonance and chromaticism of styles like thrash and death metal that were popular in metal at the time.

In Sweden, a one-man band called Bathory ran with the tremolo picking and satanic themes of early thrash and death metal bands like Venom, Possessed and Celtic Frost. With amplifier distortion and the low-fidelity production that was dictated by a low budget, the tremolo phrases were smeared together, making the guitar sound reminiscent of buzzing bees. Again there was room for atmospherics: sound samples like the sea and seagulls flying by were deployed, evoking visions of the Scandinavian landscape. Indeed, Bathory later on took a thematic influence from Nordic folklore. In the early 1990s in Norway, bands in the underground metal scene were revolting against the perceived commercialization of American death metal. Bands such as Mayhem, Immortal, Emperor and Darkthrone looked at Bathory as their mentor and together cemented *black metal* as a separate subgenre. Black metal often features lo-fi production, high-register tremolo-picked riffs, a high-pitched 'shrieking' version of the death metal grunt and occasionally atmospheric elements like acoustic guitars and synthesizers. Bands usually use minor-key tonality and dissonance to achieve a sinister or 'evil' sound [7, 8, 13].

Progressive metal started in the United States in the late 1980s and early 1990s when Queensrÿche, Fates Warning and Dream Theater combined the epic songwriting, unusual time signatures and virtuosic musicianship of Rush, King Crimson and Pink Floyd with metal in the vein of Iron Maiden and Judas Priest. The genre has since expanded to include bands with similar musical aims but very different tonal and melodic repertoires such as Opeth and Amorphis. Wagner [15] makes a distinction between the capitalized 'Progressive metal' and 'progressive metal', the former being the genre of bands such as Dream Theater and the latter being the movement of all metal bands which are progressive by challenging and expanding existing genre definitions, also including for example the experimental thrash metal bands Celtic Frost and Voivod. We will not use this capitalization distinction in this thesis, but note that we are aiming to distinguish between genres, not movements. Related genres and movements, mostly focusing on rhythmic complexity and polyrhythms, are found under names like *math metal*, *technical metal* and *djent*.

The subgenres discussed above are arguably the most historically significant and well-documented metal subgenres. That is certainly not to say that they are the only ones. However, a study of all documented metal subgenres would be a complete thesis on itself, so for the rest we will suffice with short descriptions:

Melodic death metal (sometimes shortened to *melodeath*) is the results of bands in the Swedish and British death metal scenes wanting to reintroduce the rich melodicism of classic metal like Iron Maiden and power metal to the then-popular death metal sound. It therefore combines melodic, but downtuned guitars with growled vocals (usually slightly higher pitched than in death metal), blast beats and/or keyboards. It has evolved into its very own sound, bearing only little resemblance to death metal and instead sounding rather like a more extreme power metal or a more melodic thrash metal, and is therefore noted here as a separate subgenre.

Groove metal (or *post-thrash*) is a largely American style that was established in the early 1990s by bands such as Exhorder, Pantera and Sepultura. It descended from thrash metal, but stripped away the flashy solos and lead guitar to focus largely on rhythmic syncopation and a downtuned 'chugging' guitar sound. In the case of Sepultura and

Soulfly, elements of Latin and tribal percussion were added.

Taking elements of groove metal as well as post-hardcore, alternative rock, grunge and even funk, *alternative metal* bands put down an aggressive, modern sound with unconventional song structures and a tendency towards experimentation. Relatedly, *nu metal* is a more commercial variant of alternative metal that takes additional influence from hip-hop and turntablism. Bands often feature a DJ in their lineup. Some bands took the experimentation and outside influences so far beyond any reference points in known subgenres, that they are often grouped under their own category, *experimental metal* or *avant-garde metal*. However, this is a very loosely defined subgenre.

Artists from various regions of the metal spectrum combined their music with a heavy influence from industrial dance music and other electronic music genres to form *industrial metal*. Alternative metal is sometimes used as an umbrella genre for nu, experimental and industrial metal and related styles, and some people argue that due to the outside influences and attitude of these bands, alternative metal is not a proper subgenre of metal.

Stoner metal combines the influence of early Black Sabbath, fuzzy guitar distortion and the feedback-laced jams of early psychedelic and acid rock into a bluesy, almost lazy sound. It shares many characteristics with stoner rock, but stoner metal is usually reserved for the more Sabbathian bands. *Gothic metal* resulted of bands in the death/doom scene, who were inspired by gothic rock to take doom metal in a more dark and melancholic direction, adding violins, ethereal synthesizer sounds and exotic scales. It often emphasizes the contrast between dark and light, for example by juxtaposing low, grunted male vocals with high and fragile female vocals, sometimes called 'beauty-and-the-beast' vocals. Meanwhile, *sludge metal* bands such as Melvins and Crowbar took doom metal in a very different direction, adding the aggression and vocal style of hardcore punk. Some notable overlap with stoner metal exists. There is also a more atmospheric variant influenced by post-rock, played by Neurosis and Isis.

Symphonic metal is a subgenre that consists of metal bands that employ orchestral sounds (sometimes synthesized) and elements of classical music, opera and/or film scores in their songwriting, featuring these characteristics more prominently than any characteristic that would tie the band to a different subgenre. In the case that there *are* more prominent elements of other subgenres in addition to the symphonic elements, genre nomenclatures are often combined like in *symphonic black metal* or *symphonic power metal*. Symphonic metal is often confused with gothic metal, but unlike in that subgenre there are no inherent connections with doom metal and it is mostly attempted to sound grandiose, epic and larger-than-life, in contrast to the more claustrophobic, intimate sound of gothic metal.

Folk metal is the result of bands combining features of metal with native folk instruments, folk music melodies and sometimes drinking and party songs, again at the expense of other subgenre characteristics. Some bands take a folkloric, mythological or pagan approach. Of course, there are many regional variants, given names such as *Celtic metal* or *Oriental metal*. Sometimes, the style derived from (post-black metal) Bathory known as *Viking metal*, that features Nordic folk elements, a sorrowful mood and mostly medium tempos, is also categorized under folk metal. This should not be

confused with bands from other subgenres that happen to use Viking imagery, such as the folk/symphonic metal band Turisas.

For some more typical examples of the selected subgenres, see the selected albums in appendix A.

Note on the absence of some subgenres

We will shortly list some subgenres of metal that have been documented but are nonetheless not considered to be proper subgenres in this thesis, along with reasons why the author thinks the exclusion is justified. Note that this is all open for debate and very much reflective of one metalhead's opinion.

Speed metal is a transitional subgenre of bands that played faster than classic metal bands but had not yet reached the aggressiveness of thrash metal. Bands described as speed metal can practically always also be described as classic metal, thrash metal or power metal.

Grindcore and *metalcore* are two (very different) subgenres that are as much a part of the extended hardcore punk universe as of a metal one, and thus are not included here, although there might be overlap with death and thrash metal, among others.

Neoclassical metal is not a coherent subgenre but rather a descriptor of very different artists who happen to share a heavy influence from classical music. *War metal* is a and relatively new and small subgenre of bands that live on the overlap of intentionally primitive forms of black, death and thrash metal. *Glam metal* or *hair metal* (or even *pop metal*) is a mainly American concept of bands with a very 'glammy' image playing pop rock, hard rock or classic metal.

The author very much appreciates the metalhead custom of combining subgenre names, like in *blackened death metal* and *progressive thrash metal*, and uses it himself. However, we will not use these fusion genres in the remainder of this thesis. Rather, we will classify tracks into one primary subgenre. As remarked elsewhere, in a more sophisticated system we might also consider secondary genres.

The existence of many distinguishable sub-subgenres like *slam death metal* and *traditional doom metal*, whether fusion genre or not, is also recognized by the author. However, they are mostly ignored in favor of their parent subgenres, although exceptions are made, as discussed in the next subsection. The exclusion of movements such as the NWOBHM, Second Wave of Black Metal and NWOAHM is sufficiently discussed above.

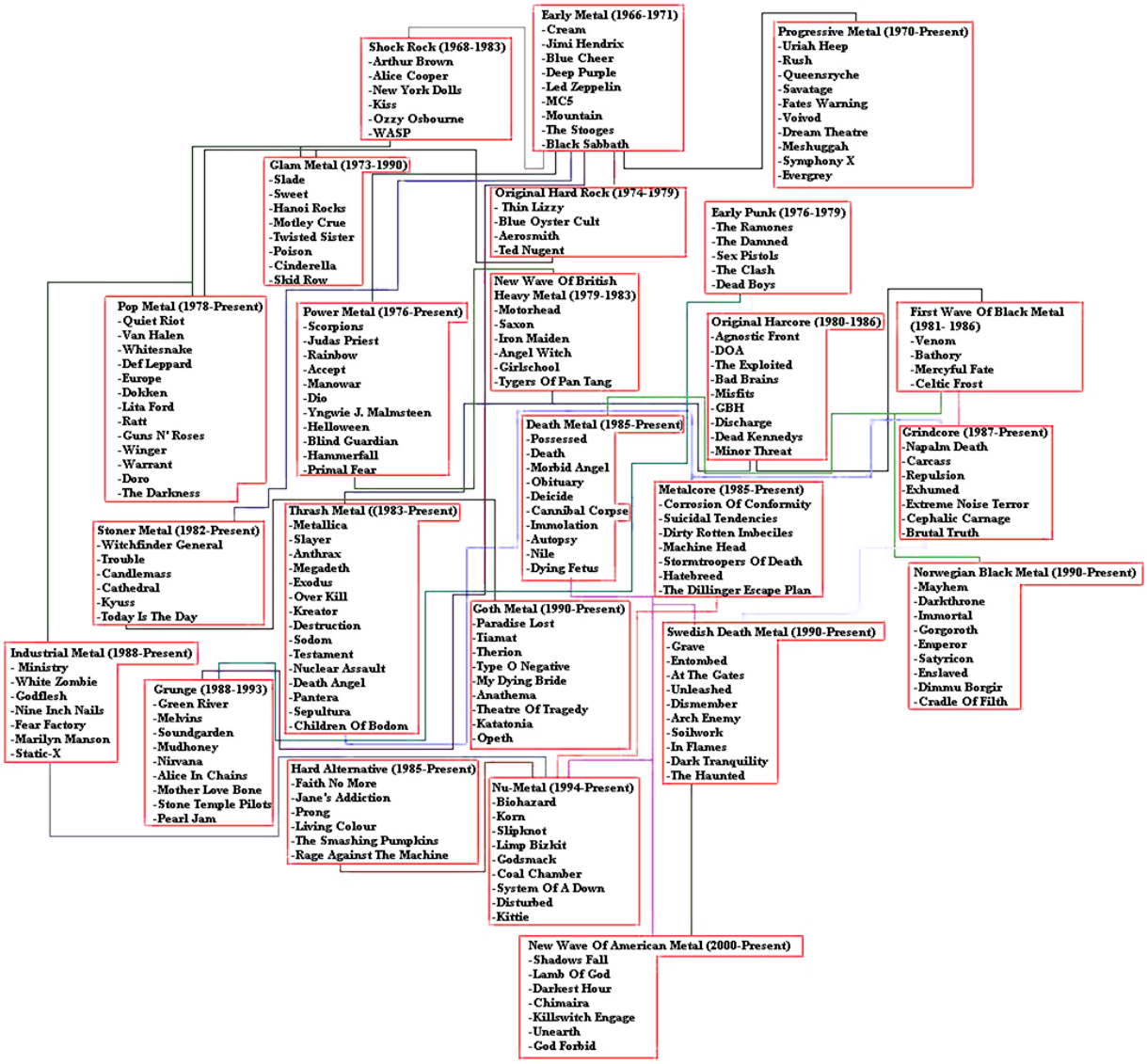

Figure 2.1.: Anthropologist-metalhead Sam Dunn's famous 'metal family tree' [8]. Notice that it is technically not a tree, seems too complex for our purposes and contains several nodes that we do not consider to be proper genres in this thesis (such as *NWOBHM*) and nodes that the author considers to be a genre but not a subgenre of metal (such as *grunge*). Source: `http://commons.wikimedia.org/wiki/File:Metal_Genealogy.jpg`

2.1.2. Subgenre taxonomy

We need to make a selection of subgenres for classification purposes. Therefore, we need to select a taxonomy of metal, or 'metal subgenre tree', which we would like to be as representative as possible of the encountered metal subgenres, without being redundant.

Let us take a look at figure 2.1, where we see anthropologist-metalhead Sam Dunn's famous 'metal family tree' [8], which traces influence between metal subgenres through time. This results in a complex directed graph (directed through time), which contains several nodes for labels not considered to be a metal subgenre in this thesis, while omitting some important ones. This makes it unrepresentative as well as redundant, and using it for classification would implicate that every class can have multiple parent classes, both in width and depth. This would severely complicate our classification algorithm. Therefore, we reject this taxonomy, as well as others constructed in a similar way. For our purpose, it is desirable to select a metal subgenre tree that is as shallow as possible, without sacrificing representativeness.

The author personally uses a tree with depth three for manual classification (see figure 2.2). The nodes in this tree largely match the subgenres treated in section 2.1.1. While this taxonomy is far less complex, it still poses the problem of multiple degrees of subclassification. Therefore, a flattened version was constructed for this project, as seen in figure 2.3. Some of the leaf nodes of depth three were omitted entirely. The sub-subgenres under folk metal were omitted because they comprise a relatively small collection of artists and it is not always clear where the borders between them and their parent genre lie. Technical and experimental metal were omitted because their existence and definition are controversial. As such, it was difficult to produce archetypical examples for them. All the other depth-three nodes of figure 2.2 reappear in figure 2.3 at depth two, and will be treated as proper subgenres of metal for the remainder of this thesis. This makes for a total of 17 metal subgenres. In abbreviated form, the collection of possible categories for our observations is therefore: $\{\mathbf{A}, \mathbf{B}, \mathbf{C}, \mathbf{De}, \mathbf{Do}, \mathbf{F}, \mathbf{Go}, \mathbf{Gr}, \mathbf{I}, \mathbf{M}, \mathbf{N}, \mathbf{Po}, \mathbf{Pr}, \mathbf{Sl}, \mathbf{St}, \mathbf{Sy}, \mathbf{T}\}$. While the selected taxonomy was based on the author's personal idea of heavy metal categorization, a more objective approach seems impossible; there is no canonical heavy metal taxonomy and every decision is bound to be controversial.

2.2. Feature extraction

For the construction of a classification algorithm, it is necessary to do *feature extraction*: we need to somehow convert the raw data describing the musical waveforms to meaningful and comparable numbers that correspond to musical properties. Examples of such properties can range from tempo (how many beats per minute) and song length to the more sophisticated, like the number of different instrument voices that are present. It is not always readily apparent how to obtain this information from an audio file.

For this project, we will focus on one musical property: musical intervals. Existing automatic classification systems use features such as timbre and tempo, but for something as precise as metal subgenre classification, it seems desirable to also have a feature that can tell us something about the music's tonality, or the different kind of melodies and harmonies that are being used. With our method, we will attempt to find some correlation between musical interval information extracted from an audio file and the metal subgenre to which it belongs. Note that by limiting ourselves to musical intervals, we

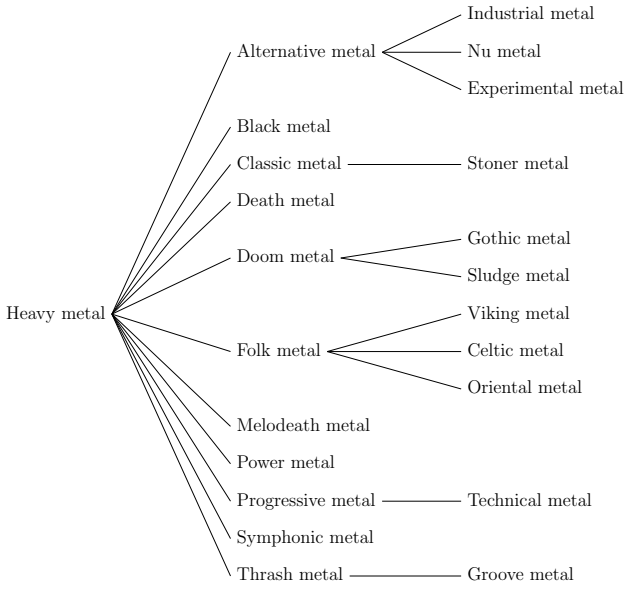

Figure 2.2.: A metal genre taxonomy that the author personally uses for manual classification. Notice that the tree has depth three. For automatic classification purposes, it is simpler to use a shallower tree.

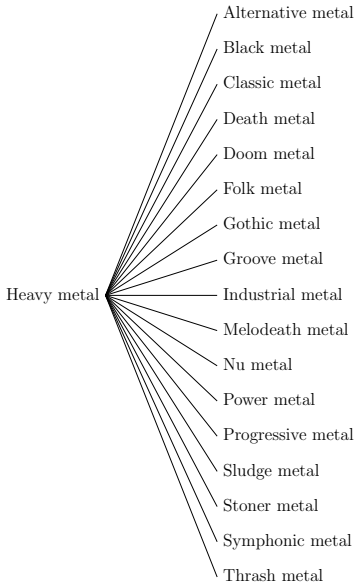

Figure 2.3.: The metal genre taxonomy that we will use in this thesis. It is a flattened version of figure 2.2 with some omitted nodes. Notice that for different nodes of depth two (the sub-subgenres) in that figure, different choices have been made in whether to omit them or to include them as a node of depth one in this figure. The umbrella genre 'heavy metal' is not to be used for classification; for our purpose we will assume all observations to be heavy metal and attempt to categorize them in a metal *subgenre*.

are not constructing a complete musical classification system. If we obtain good results, our features could be combined with existing features to achieve an even stronger metal subgenre classification system. For examples of previously researched musical features, see [2].

2.3. Pitch space and pitch class space

Sound consists of small fluctuations in air pressure. When these fluctuations are periodic over a period of time t, we can speak of the *frequency* $\frac{1}{t}$. Our ears pick up sound within a certain range of these frequencies, to be interpreted by our auditory systems. Our auditory systems are very sensitive to change in sound frequency and are especially sensitive to the ratios between different occurring frequencies. Musicians exploit this fact by arranging sounds with different frequencies in enjoyable patterns, while making use of repetition and rhythm, to create music. In music, we can often distinguish sounds with a set duration and frequency, called *notes*. The perceived frequency of a note is called the *pitch*.

All possible pitches comprise a one-dimensional continuous linear space called *pitch space*. We can label pitches in pitch space on a logarithmic scale using the following formula:

$$p = 69 + 12 \cdot \log_2 \left(\frac{f}{440} \right), \tag{2.1}$$

where f is the fundamental frequency of the note in hertz and p is a real number representing the corresponding pitch. In this scale, the pitch 69 arbitrarily corresponds to the frequency of 440 Hz. This is the MIDI Tuning Standard [16]. An increase of one corresponds with a multiplication of the frequency by a factor $\sqrt[12]{2}$. We can now use the metric $d(p, q) = |p - q|$ to denote the distance between two pitches p and q. This makes sense, because as noted before, we are most receptive to the ratio between frequencies rather than to the difference. The unit distance is also called a *semitone*.

An increase of 12 semitones corresponds with a doubling of the frequency: for any p and q in pitch space,

$$p - q = 12 \left(\log_2 \left(\frac{f_p}{440} \right) - \log_2 \left(\frac{f_q}{440} \right) \right) = 12 \cdot \log_2 \left(\frac{f_p}{f_q} \right) = 12 \tag{2.2}$$

if and only if

$$\log_2 \left(\frac{f_p}{f_q} \right) = 1, \tag{2.3}$$

which of course only holds when $f_p = 2 \cdot f_q$. A distance of 12 semitones is also called an *octave*. Notes which are spaced an octave apart tend to sound very similar to us. This phenomenon is called *octave equivalence*. We can define an equivalence relation $\sim$ by setting $p \sim q \iff d(p, q) = 12$, or $p \sim q \iff p \equiv q \pmod{12}$. The equivalence class of a pitch is called a *pitch class*. The quotient space of pitch space by $\sim$ is called *pitch class space*, which can be represented by a circle (see figure 2.4). We define distances in

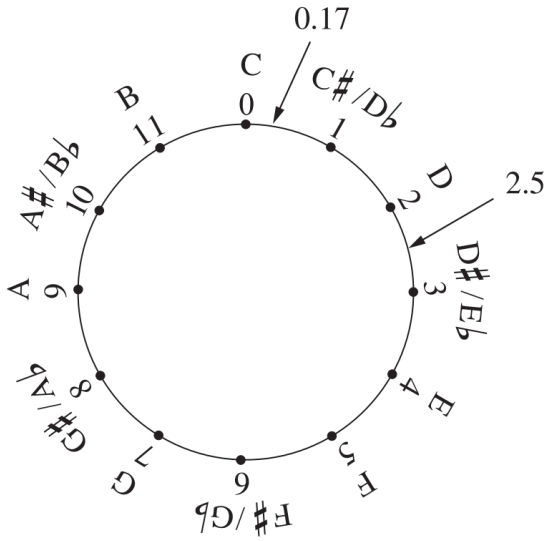

Figure 2.4.: Pitch class space, represented as a circle. For distances between pitch classes, we take the shortest distance between their points on the circle. Source: [17]

pitch class space by $d_\sim([p],[q]) = \min\{d(x,y) \mid x \sim p \wedge y \sim q\}$ for equivalence classes $[p]$ and $[q]$ of pitches p and q, respectively. (I will not prove that pitch class space with $d_\sim$ is a metric space, as we will not use any results for metric spaces. Note, however, that the given distance function is for this particular space equivalent to the usual definition of a metric on a quotient metric space, as the space is circular.)

Most (but not all) of the music we encounter today is based around a discretization of pitch space, where only the whole numbers of the scale are used. However, musicians traditionally do not use these numbers for pitches. Instead, they are relabeled as follows: pitch 60 is labeled C_4, pitch 61 is labeled $\mathrm{C}\sharp_4$ and pitch 62 becomes D_4. This goes on until pitch 71, using the letters, flat symbol $\flat$ and sharp symbol $\sharp$ in an irregular pattern, making the 12-tuple $(60, 61, \ldots, 71)$ correspond to $(\mathrm{C}_4, \mathrm{C}\sharp_4, \mathrm{D}_4, \mathrm{E}\flat_4, \mathrm{E}_4, \mathrm{F}_4, \mathrm{F}\sharp_4, \mathrm{G}_4, \mathrm{A}\flat_4, \mathrm{A}_4, \mathrm{B}\flat_4, \mathrm{B}_4)$. After that, pitch 72 becomes C_5 and the pattern is repeated again, increasing the index number every octave (and likewise decreasing it when going down). This system is called *scientific pitch notation* [17]. The pitches with sharp or flat symbols correspond to the black keys on a piano, while the ones without correspond to the white keys.

We now consider the corresponding discretization of pitch class space. Here we only maintain twelve distinct elements. We write $\mathrm{C} := [\mathrm{C}_4]$, so that C is the equivalence class of $\ldots, \mathrm{C}_3, \mathrm{C}_4, \mathrm{C}_5 \ldots$, and likewise for every pitch in discrete pitch space. Doing this, the set of elements of discrete pitch class space becomes $\{\mathrm{C}, \mathrm{C}\sharp, \mathrm{D}, \mathrm{E}\flat, \mathrm{E}, \mathrm{F}, \mathrm{F}\sharp, \mathrm{G}, \mathrm{A}\flat, \mathrm{A}, \mathrm{B}\flat, \mathrm{B}\}$. We noted earlier that pitches in the same pitch class sound similar. We call the perceived shared quality of all pitches belonging to the same pitch class, the *chroma* of the pitches [17].

2.4. Musical intervals

Besides the octave, we can distinguish more distances in pitch space. The music-theoretical name for a distance in pitch space is a *musical interval*. Every whole-number distance less than or equal to twelve is associated with a particular name, as can be seen in table 2.1 (octave omitted). Observe that the numerals in the interval names and their short forms do not directly correspond to their associated distances. Instead, a distance of zero is called the *unison* (short: P1) and after that the numeral increases only weakly. The origin of this naming oddity is in the concept of *diatonic scales* in the European musical tradition, whereby a composer would mostly use a seven-element subset of pitch class space within a composition, choosing (relative to a reference pitch or *tonic*) either the *minor* or the *major second*, either the *minor* or *major third*, etc. (but never the *tritone*) [18, 19]. In contrast, all twelve pitch classes are used in music employing *chromaticism*.

Distance in semitones	Interval name	Short form
0	Unison	P1
1	Minor second	m2
2	Major second	M2
3	Minor third	m3
4	Major third	M3
5	Perfect fourth	P4
6	Tritone	TT
7	Perfect fifth	P5
8	Minor sixth	m6
9	Major sixth	M6
10	Minor seventh	m7
11	Major seventh	M7

Table 2.1.: The different musical intervals up to the major seventh. Intervals for distances $n > 12$ are usually treated as the interval corresponding to $n \bmod 12$. Notice the discrepancy between distances and numerals in the interval names. Assuming inversion equivalence, we are only able to distinguish the first seven of these (unison through tritone).

Closely associated with intervals is the concept of *consonance* and *dissonance*. Consider two nearby pitches in a musical composition. Now consider the ratio between their associated frequencies, displayed as a simple fraction. If the numerator and denominator are small integers, the two pitches will generally sound easier on the ears together than when they are large integers. Combinations of notes that are considered to sound pleasant together are called consonant, while those that are considered unpleasant are called dissonant. P1, P4, P5 and the octave are often considered particularly consonant intervals, while m2, M7 and especially the tritone are considered dissonant intervals [18, 19].

The tritone interval has a special connection to heavy metal. Its very dissonant sound gave it the name of *diabolus in musica*, 'the devil in music', and the Catholic church forbade its use in medieval times [7, 8]. It is this interval that made Black Sabbath's eponymous song sound so ominous, and I expect the use of the tritone to be prevalent in metal in general, and especially in the subgenres closely connected with Black Sabbath, such as doom metal. Another interval with a special connection to heavy metal is the perfect fifth. Two simultaneous notes separated by the interval of a perfect fifth are called a *power chord*. Power chords are often used as the building block for riffs in rock, punk and metal.

If we assume octave equivalence, we can distinguish only seven intervals. For any interval in table 2.1 of distance n between 1 and 11, the interval of distance $12 - n$ is called its inversion. For instance, M7 is the inversion of m2, and TT is its own inversion. The distance between C_4 and D_4 is $d(\mathrm{C}_4, \mathrm{D}_4) = 2$, the major second, while the distance between C_4 and D_3 is $d(\mathrm{C}_4, \mathrm{D}_3) = 10$, the minor seventh or the inversion of M2. In the case of octave equivalence, we cannot distinguish between D_3 and D_4, so we get $d_\sim(\mathrm{C}, \mathrm{D}) = 2$, which we will also call M2. This illustrates that we cannot distinguish between interval inversions in pitch class space. This phenomenon is called *inversion equivalence* [17].

3. Method

3.1. Data set

For each of the subgenres in figure 2.3, five representative albums were selected by the author. Only in the case of doom metal, six albums were selected to compensate for the low number of tracks, since doom metal albums tend to feature longer, but fewer, songs. Each album was selected for being an archetypical example of the subgenre. Albums for which the subgenre seemed ambiguous were avoided. For an example of the considerations that went into this selection, note that Nightwish is considered to be a typical example of symphonic metal. However, their early albums are additionally associated with the subgenre of power metal. Therefore, the album where they started to distance themselves from their power-metal influences was selected.

Intro, outro and interlude tracks and other tracks that did not fit in with the style of the surrounding tracks in an obvious manner were omitted. The remaining tracks, ripped from CD to digital audio files in lossless format, will serve as our data set. As a measure for the robustness of the created data set, it was ensured that for each release, its subgenre label (or one of the underlying sub-subgenres) was among the primary genre labels as voted by the RateYourMusic[1] user base at the time of writing. Possible weaknesses of this construction is that it assumes that subgenre remains consistent over an album (excepting the manually purged tracks) and that the assumption that a song belongs to a single subgenre might be inherently wrong (for a further discussion of this, see section 2.1, section 5.3 and [6]).

3.2. Fourier transform

A digital audio file (we will use *track* synonymously from now on) describes a sampled audio signal as a function of time in quantized space. In our case, we use a format with a 44 100 Hz sampling rate and 16 bits per sample, in accordance to the Red Book audio CD standard [20]. In order to obtain frequency information from these files, we need to perform a discrete Fourier transform (DFT). A DFT converts a sampled function with a time domain to the frequency domain.

Definition 3.1. Let $x = (x_0, x_1, \ldots, x_{N-1})$ be a discrete signal of length N with sam-

[1] `https://rateyourmusic.com/`. Retrieved June 2014.

pling frequency f_s. The *discrete Fourier transform* of x is defined by:

$$\mathbf{DFT}(x) = (X_k)_{k=0}^{N-1}, \text{ where } X_k = \sum_{n=0}^{N-1} x_n e^{-\frac{2\pi i}{N} kn}. \tag{3.1}$$

Here $|X_k|$ represents the amplitude of the frequency $f_s \cdot \frac{k}{N}$, and $\arg(X_k)$ represents its phase at the start of the signal.

Calculating the transform of the signal of an entire track would give the average amplitudes of the present frequencies, but this is not what we want. To account for the fact that the signal changes in time, we divide the signal into smaller frames. Then for each frame, we multiply the signal with a window function which is nonzero only at that frame. This is known as the *short-time Fourier transform* (STFT).

Definition 3.2. Let $x = (x_0, x_1, \ldots, x_{N-1})$ be a discrete signal and w be a discrete window function. Define the *windowed signal* at sample m by $(x_w(m))_n = x_n w_{n-m}$. The *short-time Fourier transform* at sample m is defined by:

$$\mathbf{STFT}_{w,m}(x) = \mathbf{DFT}(x_w(m)). \tag{3.2}$$

Evaluating a DFT directly has a time complexity of $\mathcal{O}(N^2)$. There exist optimized algorithms that can compute it with $\mathcal{O}(N \log_2 N)$ operations. Such algorithms are called *fast Fourier transforms* (FFT). A well-known example is the Cooley–Tukey algorithm [21]. This algorithm is available in the FFTW library[2].

3.3. Chroma vectors

We are now able to analyze the pitches contained in an arbitrary frame of a digital audio file. However, they are represented as a vector that maps frequencies to their amplitude and phase, which does not correspond to the way we perceive music. We would like to be able to analyze music in the terms introduced in chapter 2. To do this, Fujishima [1] introduced the *pitch class profile* in 1999, also known as a *chroma vector*. A chroma vector is a 12-dimensional vector in which each component represents the contribution of the corresponding pitch class in a signal.

First, we need a mapping from the frequency indices of a DFT to indices of the elements of discrete pitch class space. With the frame length N and the sampling frequency f_s of the signal known, we define:

$$M(k) = \begin{cases} \operatorname{round}\left(12 \log_2\left(f_s \cdot \frac{k}{N}\right)\right) \bmod 12 & \text{if } k \in \left\{\left\lceil \frac{100N}{f_s} \right\rceil, \left\lceil \frac{100N}{f_s} \right\rceil + 1, \ldots, \left\lfloor \frac{6400N}{f_s} \right\rfloor\right\}, \\ -1 & \text{otherwise.} \end{cases} \tag{3.3}$$

By selecting these values of k, we consider only the frequencies between 100 Hz and 6400 Hz, six octaves that contain the majority of tonal information ([22] suggests 100–5000 Hz, but we slightly extend that to a whole number of octaves to avoid covering

[2] `http://www.fftw.org/`. Retrieved June 2014.

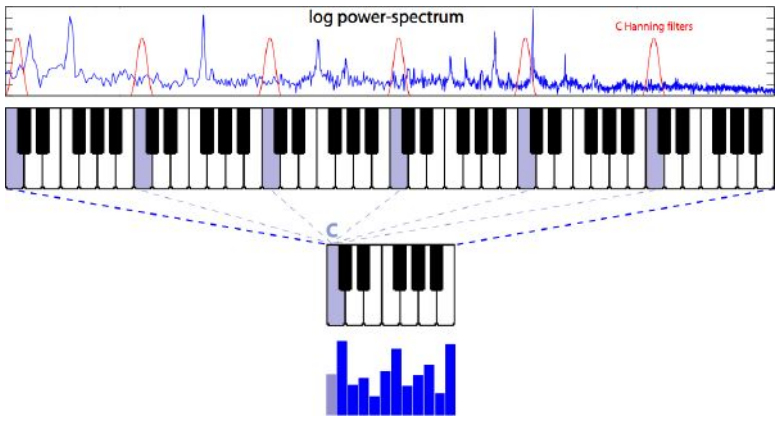

Figure 3.1.: The process of constructing a chroma vector can be envisioned in the following way: the Fourier transform is computed over a frame of a digital audio file. Then, the amplitudes of the transform are mapped to pitches of discrete pitch space, corresponding to keys of a piano in the image. Finally, in accordance with octave equivalence, values of pitches belonging to the same pitch class are added up to obtain the vector (in practice, the last two steps are done at once). Source: [23]

some pitch classes more than others). Now we can define the chroma vector of a signal x as follows.

Definition 3.3. Let $\mathbf{STFT}_{w,m}(x) = (X_k)_{k=0}^{N-1}$ be the STFT of a signal x. We define the *chroma vector* of the signal to be:

$$(v(x))_n = \sum \{|X_k|^2 \mid M(k) = n\}, \tag{3.4}$$

for $n \in \{0, 1, \ldots, 11\}$.

Note that we are not normalizing the individual chroma vectors as other authors tend to do [1]. Chroma vectors are regularly used in chord detection algorithms. Loudness differences between frames are not particularly meaningful for that purpose, but we wish to retain this information for interval analysis.

We compute a sequence of chroma vectors for each digital audio file F in the data set, using a frame length of $N = 200$ ms ($= 8820$ samples) with a 50% overlap per frame to compensate for the window function having smaller values at the sides of its domain. For a window function we use the *Hamming function*:

$$w_n = 0.54 - 0.46 \cos\left(\frac{2\pi n}{N-1}\right), \tag{3.5}$$

for $n \in \{0, 1, \ldots, N-1\}$ (and equal to zero elsewhere). We call the sequence $C_F = (v_0, v_1, \ldots, v_n)$ of chroma vectors we obtain this way a *chromagram*, so that every file in the data set is now represented by a chromagram. The Matlab MIRtoolbox 1.5 was utilized in these computations.[3]

[3] `https://www.jyu.fi/hum/laitokset/musiikki/en/research/coe/materials/mirtoolbox`. Retrieved June 2014.

3.4. Interval features

As noted in the previous chapter, we want to extract features from our audio files, because the chromagrams we have now are too complex to serve as features in their own right. Therefore, a second computational step is needed.

We wish to construct features that represent musical interval information. However, 'musical intervals' is a multifaceted notion, allowing no simple definition of the information we would like to extract. In particular, we can distinguish two dimensions over which intervals can be considered: horizontal and vertical. Horizontal intervals are the intervals between pitches that occur subsequently in time, while vertical intervals are intervals between pitches that occur at the same point in time. These correspond to the concepts of *melody* and *harmony*. However, we do not use these names so as not to imply that we are looking at the melodies and harmonies as intended by the composer or observed by a listener, as automatic music transcription is an open problem in signal processing with a different aim from ours. Our constructions are rather based purely on the chroma vectors extracted from the audio.

3.4.1. Vertical interval feature

We need to extract information on the intervals between pitches that occur at same points in time. The closest that we have to this in our representation is one chroma vector. Therefore, we need to look at the intervals that occur between pitches in one chroma vector. We will construct a 7-dimensional *interval vector*, each coordinate corresponding to one of the intervals given by inversion equivalence. Intuitively, if the values corresponding to the pitch classes A and E are high, we would like the corresponding interval, P5, to have a high value as well. *A fortiori*, if A and E are the only nonzero elements in our vector, we would like P5 to be the only nonzero component of the interval vector. However, this quickly becomes more complicated: if we have nonzero pitch classes A, C and E, we want nonzero components for the intervals of each of the pairs, namely m3 (for A and C), M3 (C and E) and P5. In general, to n different pitch classes correspond $\binom{n}{2}$ intervals, which may or may not be distinct. To capture all this information in our feature, we propose the following method.

Definition 3.4. For any chroma vector v, define the *interval matrix* M_v to be $v \cdot v^T$.

M_v is a 12-by-12 matrix that corresponds to the intuition that $(M_v)_{i,j}$ is high when v_i and v_j are high (see figure 3.2). However, because we desire *transpositional equivalence* (a song is still the same song when it is played n semitones higher or lower, for any reasonable n), the exact positions of all 144 values are not important. What is important is the interval to which they correspond. Therefore, we propose the following reduction of the interval matrix. First we define the squared summation over the entries parallel to the diagonal:

$$\sigma_n := \sum_{m \in \{1,2,\ldots,12-n\}} ((M_v)_{m,m+n})^2 + \sum_{m \in \{n,n+1,\ldots,12\}} ((M_v)_{m+n,m})^2, \tag{3.6}$$

for $n \in \{0, 1, \ldots, 11\}$. Note that for each choice of n, all of the entries over which we sum correspond to the same musical interval. Since we assume inversion equivalence in our representation, we then need to combine the summations for equivalent intervals. This becomes clear from figure 3.2. Therefore, our interval vector is defined as follows.

Definition 3.5. For any chroma vector v, let the *vertical interval vector* $I_V(v)$ be defined by:

$$(I_V(v))_n = \sqrt{\sigma_n + \sigma_{12-n}}, \tag{3.7}$$

for $n \in \{0, 1, \ldots, 6\}$.

Note that in this way, we sum over the main diagonal and the 'sixth parallel diagonal' twice. This is exactly what we want. We have half as many entries corresponding to the unison and tritone intervals, respectively, so we compensate by counting them twice.

By taking the root of the sum of squares instead of the raw sum, we are emphasizing the contribution of the largest values over the smaller values. When a certain interval in the interval matrix is represented by one very high value and several low values (as expected when there is one very prominent interval in the sample), we want this interval to contribute more to the feature than an interval that is represented by only medium values (as one would expect with noise, for example), even if their raw sums would be the same.

We now define our first feature to be the normed sum of all interval vectors:

Definition 3.6. Let F be a digital audio file with corresponding chromagram $C_F = (v_1, v_2, \ldots, v_N)$. The *vertical interval feature* f_V over F is defined by:

$$f_V(F) = \frac{\sum_{i=1}^{N} I_V(v_i)}{||\sum_{i=1}^{N} I_V(v_i)||_1}. \tag{3.8}$$

By normalizing only at the end of the computation, differences in loudness between frames are retained. In this manner, soft passages in the music contribute less to the feature than loud passages.

The vertical interval feature represents a distribution over the seven different intervals, where each component represents the relative contribution of the corresponding interval to the musical harmony of the audio file. As noted before, we would expect $(f_V)_5$, which corresponds to the P4 and P5 intervals, to be high throughout the metal genre in general because of the power chord. However, subgenres that seek to break away from genre conventions, like progressive metal and alternative metal, could be expected to have a relatively lower usage of the power chord in favor of more dissonant chords, which would also be reflected in the feature. I expect that a subgenre like power metal, which favors happy-sounding, consonant harmony would be easily distinguishable from the more difficult and dissonant harmony of doom metal. Furthermore, there might be a large difference between subgenres that use simple 'guitar–bass–drums–vocals' instrumentation (e.g., classic and thrash metal) vs. those that tend to favor a more layered approach (e.g., progressive and symphonic metal).

	C	C♯	D	E♭	E	F	F♯	G	A♭	A	B♭	B
C	P1	m2	M2	m3	M3	P4	TT	P4	M3	m3	M2	m2
C♯	m2	P1	m2	M2	m3	M3	P4	TT	P4	M3	m3	M2
D	M2	m2	P1	m2	M2	m3	M3	P4	TT	P4	M3	m3
E♭	m3	M2	m2	P1	m2	M2	m3	M3	P4	TT	P4	M3
E	M3	m3	M2	m2	P1	m2	M2	m3	M3	P4	TT	P4
F	P4	M3	m3	M2	m2	P1	m2	M2	m3	M3	P4	TT
F♯	TT	P4	M3	m3	M2	m2	P1	m2	M2	m3	M3	P4
G	P4	TT	P4	M3	m3	M2	m2	P1	m2	M2	m3	M3
A♭	M3	P4	TT	P4	M3	m3	M2	m2	P1	m2	M2	m3
A	m3	M3	P4	TT	P4	M3	m3	M2	m2	P1	m2	M2
B♭	M2	m3	M3	P4	TT	P4	M3	m3	M2	m2	P1	m2
B	m2	M2	m3	M3	P4	TT	P4	M3	m3	M2	m2	P1

Figure 3.2.: The interval matrix. Entries are marked with the musical interval they correspond to. Additionally, every entry also corresponds to the inversion of the noted interval, since we consider these as equivalent in our representations. Note that there are half as many entries corresponding to the unison and the tritone than to the other intervals. Therefore, we sum over these entries twice to obtain our interval vector.

It is important to stress again that chromagrams (and music recordings in general) do not necessarily correspond directly to the composition as intended by the artist. Many things happen between the composition of a piece and the pressing of the CD, like the recording of the music itself. A gifted listener may be able to reconstruct the composition upon listening to this recording. Our algorithms, however, cannot do that yet, nor do they need to. The vertical interval feature, for instance, may be very susceptible to noise. In this case, noise is non-tonal data in the audio, which is inherent in every recording. The presence of noise is not necessarily unintentional, however, as elements such as guitar distortion, cymbal crashes and growled vocals are also instances of non-tonal contributions to the recording. These would show up as a more equalized distribution over the intervals in our feature, which could be used as a discriminatory property. In other words, a high value for a certain interval in our feature does not unquestionably mean that this is a harmonic interval that occurs often in the composition (although this is still largely why we expect the feature to work in general), but may indicate the presence of high distortion or grunts. This differentiates the song from songs that do not feature those elements. A high noise ratio is expected in the more extreme metal subgenres, like death, melodeath and black metal and also somewhat in thrash, groove and sludge metal.

3.4.2. Horizontal interval feature

For our horizontal interval feature, we want to capture information on the musical intervals between different chroma vectors. But what points to choose? Intervals between subsequent vectors might seem like a logical choice, but we find that subsequent vectors are often very similar to each other, due to their relatively high frame rate. A horizontal interval feature based on interval matrices between subsequent chroma vectors would be too similar to our vertical interval vector. Ideally, we would like to identify stable regions in a chromagram, computing interval matrices between vectors in subsequent regions. However, the identification of these musical transitions belongs to the field of *harmonic change recognition*, which is still actively studied (by the development team of Chordify,[4] for example) and is outside the scope of this project. Therefore, I propose a much simpler feature.

Definition 3.7. Let $i_v \in \{1, 2, \ldots, 12\}$ denote the index of the highest component of chroma vector v. (If there are multiple highest components, we may choose the lowest index, but this occurs with extremely low probability, given the precision of floating-point representations in modern computing hardware.) For two chroma vectors v and v', define the *horizontal interval vector* $I_H(v, v')$ to be:

$$(I_H(v, v'))_n = \begin{cases} 1 & \text{if } n = |i_v - i_{v'}| \text{ or } 12 - n = |i_v - i_{v'}| \\ 0 & \text{otherwise,} \end{cases} \tag{3.9}$$

for $n \in \{0, 1, \ldots, 6\}$.

[4] `http://chordify.net/`

Inspired by [24], we convolve our chromagram C_F with a Gaussian filter with $\sigma = 8$ per row to obtain the filtered chromagram C'_F. This is to reduce the influence of transients. (*Transients* are sounds of short duration and high amplitude that occur at the onset of a note but do not necessarily have the same pitch as the initiated note, e.g., the plucking of a guitar string vs. the actual sound of the vibrating string.)

Definition 3.8. Let F be a digital audio file with filtered chromagram $C'_F = (v'_1, v'_2, \ldots, v'_N)$. The *horizontal interval feature* f_H over F is defined by:

$$f_H(F) = \frac{\sum_{i=1}^{N-1} I_H(v'_i, v'_{i+1})}{||\sum_{i=1}^{N-1} I_H(v'_i, v'_{i+1})||_1}. \tag{3.10}$$

The horizontal interval feature represents a distribution over the seven different intervals, where each component represents the relative occurrence of the corresponding interval to the melody of the leading voice (where the leading voice is simply the highest component of each vector in the chromagram, rather than the leading voice in the composition). Since we only consider intervals between the highest components of subsequent chroma vectors, and chroma vectors are often similar to the next one, this approach will cause us to count the unison interval very often, namely each time that the index of the highest component stays the same over subsequent vectors. Therefore, we could interpret the value of the unison component to represent the inverse of the dynamism in melody: a high value indicates that notes in the leading melody are being held for long periods of time, while a low value corresponds with a very swiftly changing melody. Hence, in the slow subgenre of doom metal we would expect a high value of the unison in the horizontal interval feature, while in fast and melodic subgenres like melodeath and power metal we would expect a low value.

Other intuitive expectations for this feature include a comparatively uniform interval distribution for music employing chromaticism, common in death metal, and a high value for M3/m6 for music that favors a major-key tonality, like power metal. In contrast, music with a dark atmosphere, such as doom, gothic and black metal, is expected to have a high m3/M6 value and perhaps an avoidance of the M3/m6 intervals. As hinted at before, I expect the tritone to be more prevalent in metal than in other genres, but particularly for bands that are especially influenced by early Black Sabbath, as is common in doom and stoner metal. In addition, m2/M7 and TT would be relatively high for bands with a high utilization of dissonance in their songwriting, associated with black, death and some strains of progressive and alternative metal.

3.5. Classification

Now that we have defined our features, we will turn our attention towards categorizing our digital audio files. Each file in the data set, as represented by one of our features, is henceforth called an *observation*. The task of grouping observations together based purely on their similarity is called *clustering*. While seeing what clusters we would obtain in our data set when applying clustering algorithms to the features we have calculated

is surely interesting on its own accord, we already have some expectation of what we want our groups to look like: we want them to match the subgenres we have chosen in figure 2.3. It is there that we enter the realm of *classification*. In classification, we have a set of observations of which it is already known in what category they belong, called the *training set*. We can use information obtained from the training set to classify each observation in our other set: the *test set*. In the terminology of the field of machine learning, clustering is also known as *unsupervised learning* and classification is *supervised learning*.

Given n observations, let each observation be represented by a D-dimensional vector x_i, for $i \in \{1, 2, \ldots, n\}$ (for instance by one of our features). The vector space $\mathbb{R}^D$ in which the observations lie is also called the *feature space*. The category θ_i to which an observation x_i belongs is called its *class*. We can view observations belonging to a certain class as realizations of stochastic variables, distributed by a probability distribution associated with that class [25]. If there are M different possible classes, we represent each class with a number in the range $1, 2, \ldots, M$. Knowing the training set $T = \{x_i \mid i \in \{1, 2, \ldots, N-1\}\}$ of observations with known classes θ_i, we must estimate the class θ of a new observation x. The function $g : \mathbb{R}^D \to \{1, 2, \ldots, M\}$ that we use to map an observation to its estimated class is called a *classifier*.

Definition 3.9. Let $x \in \mathbb{R}^D$ be a stochastic variable of class θ. The *error probability* of a classifier g is:

$$E_g = P(g(x) \neq \theta). \tag{3.11}$$

Of course, in any practical application of classification, we seek to minimize this E_g. We would like to obtain information on the minimal possible error rate given a set of observations. To express this, we first define $P(c_i)$ as the *prior probability* of class i and $P(x \mid c_i)$ as the *class likelihood*, which is the conditional probability density of x given that it belongs to class i [26]. Now, according to Bayes' theorem, the *posterior probability* of class i given observation x is given by:

$$P(c_i \mid x) = \frac{P(x \mid c_i)P(c_i)}{\sum_{j=1}^{M} P(x \mid c_j)P(c_j)}. \tag{3.12}$$

Definition 3.10. Let $x \in \mathbb{R}^D$ be a stochastic variable. The *Bayes error* is defined by:

$$E_{\text{Bayes}} = \inf_{g:\mathbb{R}^D \to \{1,2,\ldots,M\}} \{E_g\}. \tag{3.13}$$

Per definition, for any classifier g:

$$E_{\text{Bayes}} \leq E_g \tag{3.14}$$

For a given feature space, the Bayes error represents a lower bound on the error probability of any classifier acting on that space.

Theorem 3.11. *The Bayes error is equal to:*

$$E_{\text{Bayes}} = 1 - \sum_{i=1}^{M} \int_{C_i} P(c_i)P(x|c_i)dx, \tag{3.15}$$

where C_i is a partition of the feature space $\mathbb{R}^D$ given by

$$\{x \in \mathbb{R}^D \mid P(c_i)P(x \mid c_i) > \max_{\substack{1 \le j \le M \\ j \neq i}} \{P(c_j)P(x \mid c_j)\}\} \tag{3.16}$$

This is proven in [25]. The values of the prior probability and the class likelihood are in general not explicitly known for a given feature space, however. Therefore, we must rely on estimates for information on the Bayes error.

Classifiers are most often based on differences between observations. Therefore, we need to define a metric on our feature space. We do not want to use the Euclidean distance because we do not want distances to be invariant under orientation in our feature space. For instance, a difference of 0.1 in the value for the unison interval does not mean the same as a difference of 0.1 for the tritone. Furthermore, we are interested in the occurrence of intervals only with respect to the occurrence of other intervals (our features represent distributions, not absolute values), so in our features only the ratios between the vector components are important, not their values in themselves. Therefore, we want to use metrics with the property of *scale invariance*.

3.5.1. Mahalanobis distance and classification

The first metric that we define on our feature space is the *Mahalanobis distance*. The Mahalanobis distance is not a traditional metric, in the sense that it is not defined between two observations, but between an observation and a set of observations.

Definition 3.12. The *Mahalanobis distance* of an observation x from a set of observations X with mean μ_X and covariance matrix S_X is defined as:

$$d_M(x, X) = \sqrt{(x - \mu)^T {S_X}^{-1} (x - \mu_X)}. \tag{3.17}$$

The Mahalanobis distance is scale-invariant and accounts for the variance in each dimension [27]. When the covariance matrix is diagonal, it is equal to the Euclidean distance to the mean. We can define a classifier using the Mahalanobis distance.

Definition 3.13. Let x be an observation from the test set and for each $m \in \{1, 2, \ldots, M\}$, let A_m be defined as follows.

$$A_m = \{x_i \in T \mid \theta_i = m\}. \tag{3.18}$$

The *Mahalanobis classifier* is then expressed as:

$$g_{\text{Mahal}}(x) = m \text{ if } d_M(x, A_m) = \min_{m' \in \{1,2,\ldots,M\}} \{d_M(x, A_{m'})\}. \tag{3.19}$$

If this is not uniquely defined, we select one of the possible minima at random. However, this occurs almost never when all observations are continuously distributed in $\mathbb{R}^D$ and with extremely small probability in our quantized vector space. The Mahalanobis classifier works best when for each class, observations can be expected to be distributed around the class mean in a unimodal manner.

3.5.2. Aitchison distance

The *Aitchison distance* is another useful distance function for our feature space. It was defined by Aitchison in 1986 for distances between compositional data (data representing exclusively relative information, for example percentages) [28]. This makes it well-suited for our feature space.

Definition 3.14. The *Aitchison distance* between two vectors x and y with dimension D is defined as:

$$d_A(x, y) = \sqrt{\frac{1}{D} \sum_{i=1}^{D} \sum_{j=i+1}^{D} \left(\log \frac{x_i}{x_j} - \log \frac{y_i}{y_j} \right)}. \tag{3.20}$$

(Note that here, x_i and x_j denote vector components and not separate observations.)

The Aitchison distance is scale-invariant. Note that it is only defined on the subset of $\mathbb{R}^D$ where all of the vector components are nonzero. Indeed, ratios between vector components that include zeros are meaningless. This presents us with a practical problem: what if there are zeros in our features? We could conceive of zero components as components that actually were present in the data, but did not contribute enough to be counted in our quantized representation. In that view, as we have only defined non-negative features, we can solve this by adding a very small amount ϵ to each component in our features and renormalizing. (A Bayesian interpretation of this strategy is discussed in [29].) We choose $\epsilon = \frac{.0001}{7}$.

3.5.3. k-nearest neighbor classification

The *k-nearest neighbor classifier* (or k-NN) was first proposed by Fix and Hodges in 1951 [30]. It is a very simple classifier, but it often works well because it does not assume anything about the distribution of the observations. k-NN simply assumes that observations in a given class are closer to observations of the same class than to those of other classes. A new observation is assigned the class that is most frequent among its k closest neighbors. We use the Aitchison distance for this.

Definition 3.15. Let x be an observation from the test set and let $B_k \subset T$ be the set of $k \in \mathbb{N}$ nearest neighbors of x in the training set. Let $x_j \in B_k$ be the absolute nearest neighbor of x. Define $N_{k,m} = |\{x_i \in B_k \mid y_i = m\}|$. The k-NN classifier is expressed as:

$$g_k(x) = \begin{cases} m & \text{if } N_{k,m} > N_{k,m'} \text{ for all } m' \in \{1, 2, \ldots, M\} \setminus \{m\} \\ \theta_j & \text{if there is no } m \text{ s.t. } N_{k,m} > N_{k,m'} \text{ for all } m' \in \{1, 2, \ldots, M\} \setminus \{m\}. \end{cases} \tag{3.21}$$

As can be seen, in the case of a tie we select the class of the nearest neighbor. Some sources alternatively propose decreasing k until there is a plurality [31], but we will not use this. For small values of k the difference will be small.

Bounds on the Bayes error can be given in terms of the error probability of 1-NN [32].

Theorem 3.16. *Define $E_{\mathrm{NN}} = E_{g_1}$. The following lower bound on the Bayes error holds:*

$$\frac{M-1}{M}\left(1-\sqrt{1-\frac{M}{M-1}E_{\mathrm{NN}}}\right) \leq E_{\mathrm{Bayes}}. \tag{3.22}$$

As the size of the observation set converges to infinity, the error rate converges to the error probability. The obtained mean error rates can serve as estimations for the asymptotic error probabilities and will be used to estimate the Bayes error using the bounds given above [26].

3.6. Confusion cost

Not all subgenres are experienced as equidistant. Metalheads listening to a radio station that plays sludge metal but suddenly features a stoner metal track will be less upset than metalheads listening to a classic metal station suddenly hearing black metal. Therefore, classification errors should not be all weighted equally. I propose the weights displayed in the symmetric matrix in table 3.1. These correspond roughly to the author's answer to the question *"How bad would it be if a person expecting subgenre A got subgenre B instead?"* on a scale from 0 to 3.

	A	**B**	**C**	**De**	**Do**	**F**	**Go**	**Gr**	**I**	**M**	**N**	**Po**	**Pr**	**Sl**	**St**	**Sy**	**T**
A	0	3	3	2	2	3	1	1	1	2	1	3	2	2	1	3	2
B	3	0	3	1	2	2	2	3	2	1	3	3	2	2	3	2	1
C	3	3	0	3	1	2	2	2	3	2	3	1	2	2	1	2	1
De	2	1	3	0	2	3	3	1	2	1	3	3	2	2	3	3	1
Do	2	2	1	2	0	2	1	2	3	2	2	3	2	1	1	3	2
F	3	2	2	3	2	0	2	3	3	2	3	1	2	3	3	2	2
Go	1	2	2	3	1	2	0	2	1	2	2	2	2	2	3	1	3
Gr	1	3	2	1	2	3	2	0	1	2	1	3	2	1	1	3	1
I	1	2	3	2	3	3	1	1	0	2	1	3	2	3	3	2	2
M	2	1	2	1	2	2	2	2	2	0	3	1	2	2	3	2	1
N	1	3	3	3	2	3	2	1	1	3	0	3	2	2	2	3	2
Po	3	3	1	3	3	1	2	3	3	1	3	0	1	3	3	1	2
Pr	2	2	2	2	2	2	2	2	2	2	2	1	0	2	2	1	2
Sl	2	2	2	2	1	3	2	1	3	2	2	3	2	0	1	3	1
St	1	3	1	3	1	3	3	1	3	3	2	3	2	1	0	3	2
Sy	3	2	2	3	3	2	1	3	2	2	3	1	1	3	3	0	3
T	2	1	1	1	2	2	3	1	2	1	2	2	2	1	2	3	0

Table 3.1.: Cost matrix. With a and b being the row index label and column index label, respectively, each entry represents a cost for the classification of an observation with label a as label b. Note that the matrix is diagonal, and that the only zero entries are on the diagonal, since these are the only correct classifications so should not cost anything.

4. Results

The performance of the two features was tested. There are 38 representatives of doom metal, which is the least represented subgenre in the data set. To make the results more comparable between labels, 38 tracks were chosen at random (without replacement) from each subgenre, for a total of 646 tracks. Both features were computed over each track. Then, to construct training and test sets, 10-fold cross-validation was used: from the remaining set of tracks, ten disjoint partitions were made of roughly equal size. For every fold, a different partition was used as training set, while the remaining partitions were used as test set. For each feature, all observations in the test set were classified by g_{Mahal} and g_k for $k \in \{1, 3, \ldots, 19\}$. The estimated classes of the observations were then compared to their original labels. The ratio of correctly classified tracks is called the *accuracy*, while the ratio of incorrect classifications is the *error rate* (these can serve as estimates of the classifier's error probability, as noted in section 3.5.3).

In this chapter, a selection of the most relevant results is presented. For both features, we first look at a box plot of the accuracies of the ten folds for each classifier (see figures 4.1 and 4.2). Like in a regular box plot, the central mark in each box is the median and the edges are the 25th and 75th percentile of the data. The ends of the whiskers are the lowest and highest data points, respectively. Unlike in a regular box plot, a big dot was added for each classifier's mean accuracy. For comparison, a solid horizontal line was added to represent the expected accuracy of a uniformly random classifier, which is $\frac{1}{17}$ or approximately .06. After that, the mean error rates of 1-NN and the classifier with the best accuracy are used to estimate bounds on the Bayes error for the respective feature spaces, using equations 3.14 and 3.22. This is shown in tables 4.1 and 4.4. These bounds are also displayed as dotted horizontal lines in the aforementioned box plot.

Then, the classifications of the best-performing classifier, accumulated over the ten folds, are layed out in a *confusion matrix* (tables 4.2 and 4.5). A confusion matrix sets out the actual labels against the estimated labels. The rows contain the actual classes, while the columns contain the classes that were estimated by the classifier. In this way, every entry contains the number of observations labeled with its row header that were estimated to have the class of its column header. The diagonal contains all correctly classified observations. The last column is reserved for the *recall* per label, or ratio of correctly classified observations with this actual label (diagonal entry divided by 38) and the last row is the *precision* per label, or ratio of correctly classified observations with this estimated label (diagonal entry divided by column sum). The bottom right value is the total accuracy. Because there is an equal numbers of representants for each class, this accuracy is equal to the mean of the recalls.

Finally, we look at the *confusion cost matrix* of the best-performing classifier (tables 4.3 and 4.6). This reads in a similar way to the confusion matrix, except that the

entries are now the error cost of the classifications labeled with the row header that were classified as the column header. This is obtained by pointwise multiplication of the confusion matrix with the cost matrix of table 3.1. In contrast to the regular confusion matrix, here, lower values indicate a better performance. The last column contains the *recall cost* per label, or cost per observation with this actual label (row sum divided by 38) and the last row is the *precision cost* per label, or cost per observation with this estimated label (column sum divided by corresponding column sum of confusion matrix). The bottom right value is the cost per observation (matrix sum divided by 646). Because there is an equal numbers of representants for each class, this is equal to the mean of the recall costs.

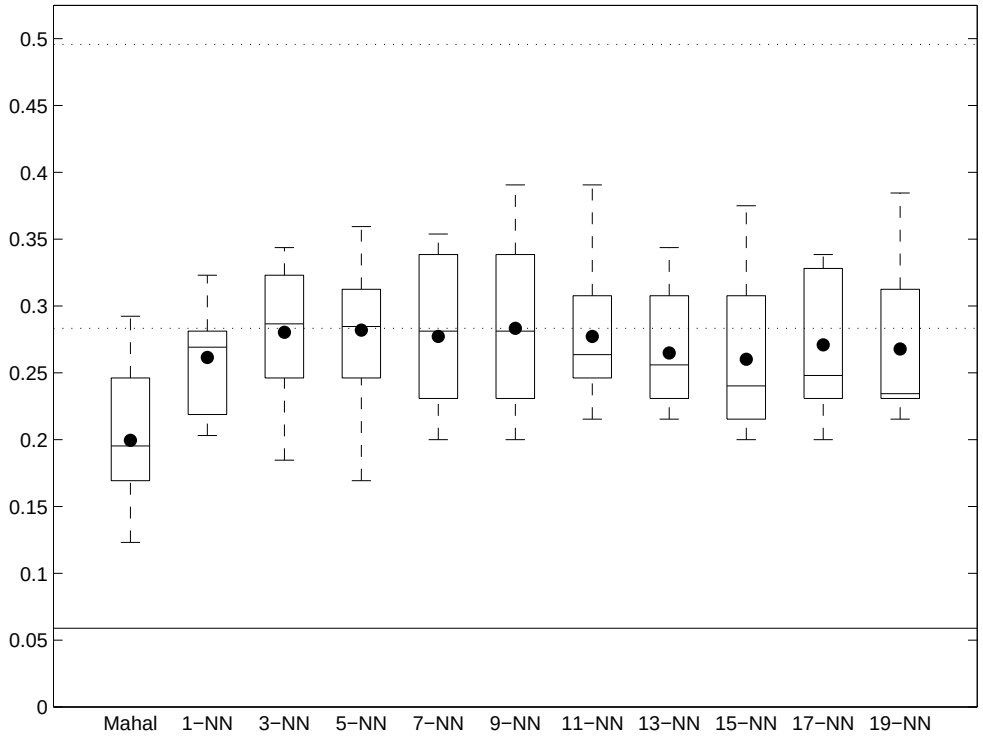

Figure 4.1.: Box plot of the accuracies of the vertical interval feature for the tested classifiers. The big dots represent the mean accuracy for each classifier. The solid horizontal line is the expected accuracy $\frac{1}{17}$ of a uniformly random classifier. The dotted lines are bounds on the accuracy of the optimal classifier on this feature space, given by the estimated bounds on the Bayes error of table 4.1.

Mean error rate 1-NN	Min. mean error rate (9-NN)	Bayes error estimation
.74	.72	$.50 \leq E_{\text{Bayes}} \leq .72$

Table 4.1.: The mean error rates of 1-NN and 9-NN are used to estimate bounds on the Bayes error for the vertical interval feature.

	A	**B**	**C**	**De**	**Do**	**F**	**Go**	**Gr**	**I**	**M**	**N**	**Po**	**Pr**	**Sl**	**St**	**Sy**	**T**	*R*
A	5	0	0	3	1	7	1	0	2	1	1	1	4	5	1	3	3	*.13*
B	2	12	1	4	1	0	0	5	1	4	2	1	0	1	1	0	3	*.32*
C	1	0	13	6	1	1	2	1	3	1	1	1	0	2	3	1	1	*.34*
De	1	3	1	12	0	0	0	2	4	1	0	0	0	3	0	0	11	*.32*
Do	1	0	1	0	5	2	2	2	1	0	0	2	5	8	2	7	0	*.13*
F	2	0	2	2	0	8	2	2	0	2	2	2	8	3	0	3	0	*.21*
Go	1	0	0	0	0	2	10	0	2	1	0	6	11	2	0	3	0	*.26*
Gr	1	1	0	3	3	1	0	11	2	5	1	2	0	0	0	0	8	*.29*
I	1	2	3	7	1	1	1	2	4	5	1	4	1	0	1	3	1	*.11*
M	1	4	1	0	0	1	0	1	1	25	1	2	0	0	0	1	0	*.66*
N	4	3	0	0	0	1	0	2	1	2	6	7	1	0	4	1	6	*.16*
Po	1	1	0	0	1	0	3	1	3	1	2	12	5	2	1	3	2	*.32*
Pr	1	0	0	0	2	2	7	0	0	1	0	5	13	0	1	6	0	*.34*
Sl	2	0	2	1	5	1	1	1	0	3	0	1	1	11	6	1	2	*.29*
St	3	0	2	2	2	1	1	3	1	2	1	3	2	4	9	0	2	*.24*
Sy	1	0	0	0	2	1	3	0	2	1	0	3	6	3	1	14	1	*.37*
T	2	3	1	8	1	1	0	1	2	0	1	2	0	1	1	1	13	*.34*
P	*.17*	*.41*	*.48*	*.25*	*.20*	*.27*	*.30*	*.32*	*.14*	*.45*	*.32*	*.22*	*.23*	*.24*	*.29*	*.30*	*.25*	*.28*

Table 4.2.: Confusion matrix for the vertical interval feature with the 9-NN classifier, accumulated over ten folds. The row headers are the true labels of observations in the test set and the column headers are the labels estimated by the classifier. The last row contains the precision per label and the last column contains the recall per label. The bottom right value is the total accuracy.

	A	**B**	**C**	**De**	**Do**	**F**	**Go**	**Gr**	**I**	**M**	**N**	**Po**	**Pr**	**Sl**	**St**	**Sy**	**T**	μ
A	0	0	0	6	2	21	1	0	2	2	1	3	8	10	1	9	6	*1.9*
B	6	0	3	4	2	0	0	15	2	4	6	3	0	2	3	0	3	*1.4*
C	3	0	0	18	1	2	4	2	9	2	3	1	0	4	3	2	1	*1.4*
De	2	3	3	0	0	0	0	2	8	1	0	0	0	6	0	0	11	*0.9*
Do	2	0	1	0	0	4	2	4	3	0	0	6	10	8	2	21	0	*1.7*
Fo	6	0	4	6	0	0	4	6	0	4	6	2	16	9	0	6	0	*1.8*
Go	1	0	0	0	0	4	0	0	2	2	0	12	22	4	0	3	0	*1.3*
Gr	1	3	0	3	6	3	0	0	2	10	1	6	0	0	0	0	8	*1.1*
I	1	4	9	14	3	3	1	2	0	10	1	12	2	0	3	6	2	*1.9*
M	2	4	2	0	0	2	0	2	2	0	3	2	0	0	0	2	0	*0.6*
N	4	9	0	0	0	3	0	2	1	6	0	21	2	0	8	3	12	*1.9*
Po	3	3	0	0	3	0	6	3	9	1	6	0	5	6	3	3	4	*1.4*
Pr	2	0	0	0	4	4	14	0	0	2	0	5	0	0	2	6	0	*1.0*
Sl	4	0	4	2	5	3	2	1	0	6	0	3	2	0	6	3	2	*1.1*
St	3	0	2	6	2	3	3	3	3	6	2	9	4	4	0	0	4	*1.4*
Sy	3	0	0	0	6	2	3	0	4	2	0	3	6	9	3	0	3	*1.2*
T	4	3	1	8	2	2	0	1	4	0	2	4	0	1	2	3	0	*1.0*
μ	*1.6*	*1.0*	*1.1*	*1.4*	*1.4*	*1.9*	*1.2*	*1.3*	*1.8*	*1.1*	*1.6*	*1.7*	*1.4*	*1.4*	*1.2*	*1.4*	*1.1*	*1.4*

Table 4.3.: Confusion cost matrix for the vertical interval feature with the 9-NN classifier, accumulated over ten folds. Every entry is the error cost of the corresponding classifications in table 4.2. For each label, the value in the last row is the *precision cost* and the value in the last column is its *recall cost*. The bottom right value is the average cost per observation.

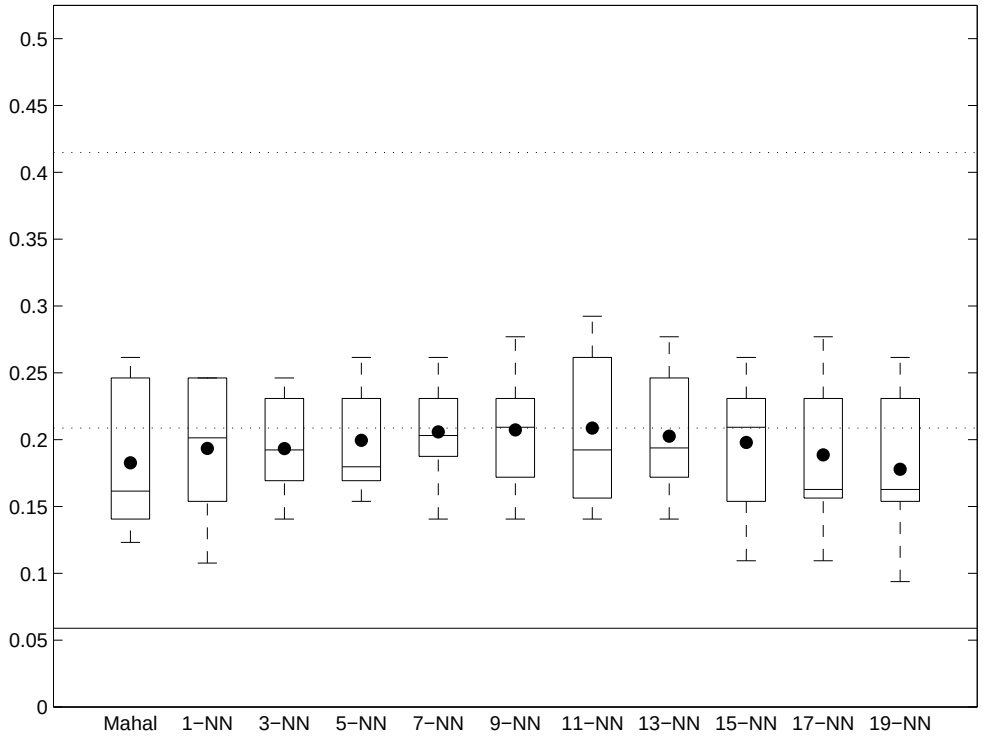

Figure 4.2.: Box plot of the accuracies of the horizontal interval feature for the tested classifiers. The big dots represent the mean accuracy for each classifier. The solid horizontal line is the expected accuracy $\frac{1}{17}$ of a uniformly random classifier. The dotted lines are bounds on the accuracy of the optimal classifier on this feature space, given by the estimated bounds on the Bayes error of table 4.4.

Mean error rate 1-NN	Min. mean error rate (11-NN)	Bayes error estimation
.81	.79	$.59 \leq E_{\text{Bayes}} \leq .79$

Table 4.4.: The mean error rates of 1-NN and 11-NN are used to estimate bounds on the Bayes error for the horizontal interval feature.

	A	**B**	**C**	**De**	**Do**	**F**	**Go**	**Gr**	**I**	**M**	**N**	**Po**	**Pr**	**Sl**	**St**	**Sy**	**T**	*R*
A	9	0	0	4	0	1	3	2	1	3	4	3	1	0	3	3	1	*.24*
B	1	8	0	3	6	0	2	0	1	0	0	3	0	2	1	1	10	*.21*
C	1	0	8	3	1	2	7	2	0	3	2	3	3	0	0	3	0	*.21*
De	0	0	0	13	2	0	0	5	2	1	4	0	0	3	0	0	8	*.34*
Do	1	4	1	2	8	0	3	3	0	1	0	3	2	1	0	6	3	*.21*
F	0	2	4	0	3	4	3	0	2	3	0	1	7	0	2	5	2	*.11*
Go	3	0	1	0	2	1	3	0	0	4	2	4	11	0	0	7	0	*.08*
Gr	0	0	1	5	4	0	0	11	0	5	1	1	0	1	5	0	4	*.29*
I	1	2	1	7	5	1	0	1	1	5	3	0	3	1	3	1	3	*.03*
M	1	0	0	0	0	2	2	3	2	15	1	2	5	0	2	2	1	*.39*
N	3	2	0	5	0	1	3	3	2	3	8	4	0	0	1	0	3	*.21*
Po	1	0	2	0	3	3	2	0	1	3	0	9	5	0	1	8	0	*.24*
Pr	0	1	2	0	0	2	7	1	0	0	0	6	14	0	0	4	1	*.37*
Sl	2	1	0	5	6	1	1	2	2	1	2	0	2	9	1	1	2	*.24*
St	4	2	3	4	0	4	2	4	0	2	5	0	1	2	0	2	3	*.00*
Sy	1	1	1	0	4	1	6	1	0	1	2	4	10	0	0	4	2	*.11*
T	1	3	0	10	3	0	3	3	0	0	0	1	2	0	0	1	11	*.29*
P	*.31*	*.31*	*.33*	*.21*	*.17*	*.17*	*.06*	*.27*	*.07*	*.30*	*.24*	*.20*	*.21*	*.47*	*.00*	*.08*	*.20*	*.21*

Table 4.5.: Confusion matrix for the horizontal interval feature with the 11-NN classifier, accumulated over ten folds. The row headers are the true labels of observations in the test set and the column headers are the labels estimated by the classifier. The last row contains the precision per label and the last column contains the recall per label. The bottom right value is the total accuracy.

	A	**B**	**C**	**De**	**Do**	**F**	**Go**	**Gr**	**I**	**M**	**N**	**Po**	**Pr**	**Sl**	**St**	**Sy**	**T**	μ
A	0	0	0	8	0	3	3	2	1	6	4	9	2	0	3	9	2	*1.4*
B	3	0	0	3	12	0	4	0	2	0	0	9	0	4	3	2	10	*1.4*
C	3	0	0	9	1	4	14	4	0	6	6	3	6	0	0	6	0	*1.6*
De	0	0	0	0	4	0	0	5	4	1	12	0	0	6	0	0	8	*1.1*
Do	2	8	1	4	0	0	3	6	0	2	0	9	4	1	0	18	6	*1.7*
F	0	4	8	0	6	0	6	0	6	6	0	1	14	0	6	10	4	*1.9*
Go	3	0	2	0	2	2	0	0	0	8	4	8	22	0	0	7	0	*1.5*
Gr	0	0	2	5	8	0	0	0	0	10	1	3	0	1	5	0	4	*1.0*
I	1	4	3	14	15	3	0	1	0	10	3	0	6	3	9	2	6	*2.1*
M	2	0	0	0	0	4	4	6	4	0	3	2	10	0	6	4	1	*1.2*
N	3	6	0	15	0	3	6	3	2	9	0	12	0	0	2	0	6	*1.8*
Po	3	0	2	0	9	3	4	0	3	3	0	0	5	0	3	8	0	*1.1*
Pr	0	2	4	0	0	4	14	2	0	0	0	6	0	0	0	4	2	*1.0*
Sl	4	2	0	10	6	3	2	2	6	2	4	0	4	0	1	3	2	*1.3*
St	4	6	3	12	0	12	6	4	0	6	10	0	2	2	0	6	6	*2.1*
Sy	3	2	2	0	12	2	6	3	0	2	6	4	10	0	0	0	6	*1.5*
T	2	3	0	10	6	0	9	3	0	0	0	2	4	0	0	3	0	*1.1*
μ	*1.1*	*1.4*	*1.1*	*1.5*	*1.7*	*1.9*	*1.7*	*1.0*	*2.0*	*1.4*	*1.6*	*1.5*	*1.3*	*0.9*	*2.0*	*1.7*	*1.2*	*1.5*

Table 4.6.: Confusion cost matrix for the horizontal interval feature with the 11-NN classifier, accumulated over ten folds. Every entry is the error cost of the corresponding classifications in table 4.5. For each label, the value in the last row is the *precision cost* and the value in the last column is its *recall cost*. The bottom right value is the average cost per observation.

5. Discussion

We will discuss the performance of the two features and then discuss possible future work. Keep in mind that in general with music classification systems, bad precision is more undesirable than bad recall. For an illustration of this, imagine again metalheads listening to a radio station of a particular subgenre. If only a small fraction of the played tracks actually belonged to that subgenre (bad precision), they will be much unhappier than if a fraction of the tracks of the subgenre would never play (bad recall).

5.1. Vertical interval feature

For the vertical interval feature, all classifiers perform significantly better than chance. All tested k-NN classifiers have a mean accuracy within the range of .25 to .30, consistently better than the Mahalanobis classifier with a mean accuracy of .20. This shows that observations are likely not centered around the mean of their class in a unimodal manner. Instead, the class distributions may be multimodal in our feature space. Indeed, there is no reason to think that subgenres would be centered around just one distribution of the musical intervals; it is quite likely that there are multiple different preferences for harmonic intervals within one subgenre. These different preferences may correspond to sub-subgenres or even artist style. Differentiation of style within metal subgenres is common in the metal community. For example, fans of sludge metal differentiate between 'atmospheric sludge metal' and regular sludge metal.

It proved not to be possible to obtain tight bounds on the Bayes error. However, the lower bound indicates that this feature will never perform better than roughly 50 percent, for any possible classifier. If results higher than 50% are desired, the answer must not be sought in different classifiers, but in different features [26]. This is in line with expectation, as this feature alone was not meant to constitute a complete metal subgenre recognition system, but is only a first step towards it. Another interpretation of the Bayes error bounds is that between 28% and 50% of the information relevant to metal subgenre classification has been extracted with our method [32].

Looking at table 4.2, we observe that the performance of the vertical interval feature–9-NN combination is very different between subgenres. The recall is relatively good for classic, melodeath and symphonic metal, and relatively bad for alternative, doom, industrial and nu metal. The precision is relatively good for black, classic and melodeath metal, and relatively bad for alternative, doom and industrial metal.

It was noted before that some subgenres of metal are considered to be closer together than others, which was the reason for the introduction of the confusion cost matrix. The utility of this cost matrix is illustrated by some of the results. Death and thrash metal are

frequently confused for each other. This may not be so surprising. While the distinction remains important for metal listeners, death metal was developed directly from thrash metal and as such features a lot of the same elements. To a new listener, the most readily apparent distinctions may be vocal technique (growled vs. shouted/clean) and guitar tone register (low-register/downtuned vs. standard tuning). The vocal technique might provide some distinction within the feature space, but the guitar tone register is not an element that we would expect the vertical interval feature to pick up. Indeed, it appears that the harmonic content of these two genres, when abstracted from octave information and transposition, is very similar.

Another high confusion rate both ways is between doom and sludge metal. Again, these are closely related subgenres, more so than thrash and death. As can be seen in figure 2.2, the author initially even considered sludge metal to be a subgenre of doom metal. Likewise, groove metal is often confused for thrash metal. However, the other sub-subgenres of figure 2.2 do not display a higher confusion rate than average with their parents. This validates their existence as separate classes in our system. All three confusion cases are leniently weighted in the confusion cost matrix, resulting in better average confusion costs for these labels than might be expected from their precisions and recalls.

In contrast, we observe other, more unexpected, confusions. Reasons for this could be that subgenres may be closer together in terms of harmonic tonality than expected, or that our feature is not accurate enough to pick up the harmonic distinctions. Nu metal is frequently confused for power metal. While this is certainly an undesirable result, as the average listener will find these subgenres to be nothing alike, this may perhaps be understood by considering that both genres tend to feature grand-sounding choruses as an important part of their compositions (on the side of nu metal, mostly with Ill Niño and Linkin Park, less so with Slipknot). It may be that the vertical interval feature tends to work better, i.e. produces more distinguishable information, when these choruses are present, and the harmonies that are used to make these choruses sound 'grand' are similar between these two subgenres. Even with this in consideration, it is still quite an unexpected confusion, and results in a high recall cost (2.0) for nu metal.

Also of note is the confusion of alternative metal for folk metal. These subgenres have very little in common, and this can only be taken to mean that our feature is not representing useful information with one of these genres. To help distinguish these, I suggest a classification system that also features a timbre feature. Towards the end of this chapter I will comment on the effectiveness of interval features with modern metal genres, like alternative metal, in particular. Finally, we see frequent confusions between progressive metal on one side, and gothic, symphonic and folk metal on the other side. Tonally, progressive metal is very unpredictable, so it was expected to be a difficult label for this system.

The vertical interval feature performs best for a subgenre with a very particular tonality: melodeath metal. We see exceptional recall (more than 60%) and recall cost and near-best precision and precision cost. Melodeath is quite an oddity in terms of metal tonality (as noted in section 2.1.1), employing a combination of the more conventional tonality of subgenres like classic and power metal and the more chromatic tonality and

higher expected noise ratio of extreme metal subgenres like death metal (caused by growled vocals, high guitar distortion and fast percussion, see section 3.4.1). This appears to make it easily distinguishable from either, having very low confusion rates with the aforementioned subgenres.

Apart from the specific cases we just discussed, bad confusion results are mostly caused by the accumulation of multiple, but individually relatively low, confusion rates. These may be unavoidable in a system using just one feature. Several suggestions regarding this are made in section 5.3.

5.2. Horizontal interval feature

From figure 4.2, it becomes clear that the horizontal interval feature performs worse than the vertical interval feature, but produces more consistent results, having less variance in the accuracies both between classifiers and within each classifier. The Mahalanobis classifier does not perform notably worse than the k-NN classifiers. From the Bayes error bounds we assess that between 21% and 41% of the relevant information for subgenre classification has been extracted. However, looking at the confusion scores of tables 4.3 and 4.6 rather than the unweighted accuracies, the two features seem to have a more similar level of performance (1.5 vs. 1.4). This means that while the horizontal interval feature errs more often, the errors it makes are somewhat less costly overall than those of the vertical interval feature.

Looking at table 4.5, it seems that there is some overlap of well-performing and badly-performing classes between the two features. We see good recall for death, melodeath and progressive metal, and bad for gothic, industrial and stoner (with zero correctly classified observations). Precision is relatively good for alternative, black, classic and sludge metal, and bad for gothic, industrial, stoner and symphonic metal. This overlap indicates a correlation between harmonic composition choices and melodic composition choices over metal subgenres.

In fact, when we compare the pairwise confusions, the results are surprisingly similar. Many of the same confusions that we noted for the vertical interval feature reappear here, like death and thrash metal and the confusion of folk, symphonic and gothic for progressive metal. The interpretation of this is largely the same as before: a preference for certain intervals in the melody seems to be associated with a certain preference in harmonic intervals. A few places where it does better are with the confusion of alternative for folk metal, and doom for sludge metal. However, in their place are new confusions, such as black for thrash metal and classic for gothic metal. In particular, it is interesting to note the high confusion in both directions between gothic and symphonic metal. These genres are often confused by metalheads too, as mentioned in section 2.1.1. Interestingly, we did not see this particular confusion much with the vertical interval feature. This is fitting, as symphonic metal songs are often backed by orchestra, which would produce a very different harmonic pattern than with gothic metal, although in the melodic department they might be more similar.

In spite of a few exceptions, the feature mainly performs worse because of multiple,

fairly equally distributed confusions instead of large outliers. These seemingly random confusions hinder the musical interpretation of its results. The feature seems to throw away relevant information by considering only the leading voice. This likely causes large overlap within the class distributions in this feature space. Compared to the vertical interval feature, the horizontal interval feature just does not seem worthwhile to add to a metal classification system in its current form. The few cases where it seems to make a better distinction could also be handled by the addition of more features. However, despite the observed connection between the two features, in theory they do still represent different information. Perhaps a more sophisticated version of the horizontal interval feature could make it worthwhile.

5.3. Future work

The system we constructed shows potential for growth in many different areas. The horizontal interval feature could be made more sophisticated by taking into account more than just the most prominent pitch class in each chroma vector. However, the problem of counting many of the same intervals as the vertical interval feature (because adjacent chroma vectors are usually similar given their relatively high frame rate) would have to be worked around in order to keep the two features separate. An answer could be sought in chord-change detection algorithms, counting intervals between chroma vectors only when there is a chord change.

More traditional subgenres like classic, power, death, black and melodeath each make very distinctive tonal choices, and this shows in the recall and precision rates of the two features. In contrast, the more modern, 'alternative-related' genres like industrial, alternative, and nu metal perform worse than average. It will not be a controversial observation that newer metal subgenres distinguish themselves mostly in other areas than tonality, like in rhythm and instrumentation, and this might be true also for the other subgenres where performance was bad. We will need more information than we extracted to obtain an accurate metal classification system.

As noted before, the best performance would most likely be achieved by combining our features with existing features. A tempo feature could help with classifying subgenres that distinguish themselves by (lack of) speed, such as thrash metal and doom metal. A feature with information on rhythm and meter could help with subgenres that focus on different kinds of rhythmic complexity, such as progressive and groove metal. Most significantly, I suspect a timbre feature could help clear up a lot of the confusions we witnessed with the interval features. Thrash metal would be discernable from the lows of death metal, the high overtones of the instruments used in folk metal would stand out from the more guitar-centric alternative metal, and low scores for industrial metal would certainly be avoided since industrial metal is the only metal subgenre consistently utilizing electronic music to a very large extent.

For the purpose of this research, I attempted to compile a labeled data set that avoids ambiguity by selecting archetypical example albums of our subgenres. However, in real life, such ambiguity is unavoidable. Many metal artists combine characteristics of several

subgenres, and often there is no consensus on the correct label of their music (and on whether such a 'correct label' even exists). This is reflected in the fact that users of Last.fm or RateYourMusic can vote for more than one label for artists and albums. In fact, RateYourMusic implements a system where users can vote on a primary and secondary genre for each release, and even within those, more than one can be voted for. A suggestion to make results more comparable to such data (and as such, the system more suited for real-life situations) would be to use classifiers that output a distribution over class estimations rather than only one class. Towards this purpose, the k-NN classifier could be expanded to consider all k nearest neighbors instead of just a plurality.

The vertical interval feature also appeared to be very sufficient in detecting musical outliers. For example, within the classic metal subset of the data set, *Planet Caravan* by Black Sabbath has a higher Mahalanobis distance to the rest of the subset than any other classic metal track (62.5 vs. an average of 9.60), in this feature space. This corresponds exactly to the author's intuition, as he considers *Planet Caravan* to be a notable outlier in the selected classic metal tracks: it is essentially a psychedelic rock song on a classic metal album. Relatedly, preliminary investigative research shows promising results for the vertical interval feature in a more traditional case, namely a classification system of general genres. It would be interesting to test our method on GTZAN, the data set constructed by Tzanetakis and Cook [2]. In addition, it would be interesting to see how the state-of-the-art classification systems mentioned in the introduction would perform on our data set.

6. Conclusion

We set out to create an automatic classification system for heavy metal subgenres. For this purpose, we constructed two new features and explored several classifiers. We constructed the *vertical interval feature*, which gives information on the intervals within the harmony of the music, and the *horizontal interval feature*, which gives information on the intervals in the melody of the leading voice. We then tested the performance of these two features on a manually assembled and labeled data set of heavy metal music. The results showed us that the *Mahalanobis classifier* was likely not a good choice for our system, while the performance of the *k-NN classifiers* was better. The best obtained accuracies were .28 for the vertical interval feature with the 9-NN classifier, and .21 for the horizontal feature with 11-NN, relative to a chance rate of .06. When combined with a subjective cost for each possible confusion case, this results in average confusion costs of 1.4 and 1.5, respectively, out of a maximum of 2.9 (progressive metal did not have possible confusions with a cost of 3.0). We observed that the features performed better with some subgenres than others. For example, both features had excellent recall and precision for melodeath, while both struggled with industrial metal, frequently confusing it with other subgenres.

With estimated bounds on the *Bayes error*, we assessed that the error rate of the theoretical optimal classifier was in the range of .50 to .72 for the vertical and .59 to .79 for the horizontal interval feature. While this is better than pure chance, it proves that it is not possible to construct a system viable for real-life applications (where we would certainly want an error rate of less than 50%) based on just one of these features. This is not unexpected, seeing as intervals are but one aspect of musical composition. If we would combine our features with features representing very different aspects, such as rhythm and timbre, we may therefore expect beter results. Nevertheless, with this system and through the introduced cost matrices, I hope to have built a comparative groundwork for research of metal subgenre classification systems in the future.

7. Populaire samenvatting

Als je naar de radio van Spotify luistert, weet je dat dit niet werkt zoals een ouderwets radiostation. Hiervoor zitten geen mensen als Giel Beelen in een studio nummers bij elkaar te zoeken. Nee, hier zitten slimme algoritmes achter die de muziek analyseren en nagaan of twee nummers bij elkaar passen. Het onderzoeksgebied dat deze algoritmes bestudeert heet *automatische muziekclassificatie*. Het blijkt dat de classificatie van *heavy metal* in het bijzonder moeilijk is. Als het voor sommige buitenstaanders al moeilijk is om te leren dat het ene gebrul het andere niet is, hoe moeten we dat dan ooit aan een computer leren? Voor mensen die wel van metal houden is het juist heel belangrijk of ze op dit moment naar 'thrash metal' of naar 'power metal' aan het luisteren zijn. In deze scriptie bekijken we hoe we zelf zo'n classificatie-algoritme kunnen bouwen, die we vervolgens zullen testen voor metalsubgenres.

De mp3's, flacs en wavs op je computer (en op de servers van Spotify) beschrijven de geluidsgolven van een muziekopname, zodat je speakers deze later weer kunnen reproduceren. Je kunt die geluidsgolven ook zien als een optelling van sinus- en cosinusgolven. Met een bepaalde formule, de *Fouriertransformatie*, kun je precies nagaan welke golven dit zijn, en welke frequentie (toonhoogte) en amplitude (luidheid) ze hebben. Op deze manier kunnen we bijvoorbeeld ook een geluidsbestand in kleine stukjes hakken en van elk stukje precies beschrijven welke tonen er in voorkomen. Als je dan ook nog de tonen die een *octaaf* van van elkaar wegliggen (op de piano tellen we twaalf toetsen verder; deze tonen hebben dezelfde naam: allebei 'A' of allebei 'F♯') bij elkaar optelt, krijg je een *chromavector* (zie figuur 3.1). Met een chromavector kunnen we dus van elk stukje van het muziekbestand zeggen in welke mate de twaalf verschillende toonhoogtes er in voorkomen.

Vervolgens gaan we de chromavectoren gebruiken om iets over de muzikale eigenschappen van een nummer te zeggen. Als we zo'n eigenschap kunnen uitdrukken in een getal noemen we dat een *kenmerk*. We willen kijken naar de *intervallen* die in het nummer voorkomen. In de muziek zijn intervallen de afstanden die tussen opeenvolgende of gelijktijdige noten voorkomen. Voor twee noten kunnen we bijvoorbeeld tellen hoeveel pianotoetsen we moeten opschuiven om van de ene noot naar de andere te komen. Alleen hebben we het octaaf-interval al weggedaan in onze chromavectoren, dus inclusief de verschuiving van nul houden we nog zeven intervallen over om te onderscheiden (ga maar na: tussen de middelste C en G op een piano zit misschien een verschuiving van zeven toetsen, maar als we de C een octaaf hoger nemen zitten er nog maar vijf toetsen tussen, en we kijken alleen naar de *kleinste* afstand).

Op deze manier gaan we de intervallen tellen die voorkomen binnen onze verzameling chromavectoren. We kijken zowel horizontaal als verticaal. Voor de horizontale methode tellen we tussen twee opeenvolgende chromavectoren steeds het interval tussen de tonen

die horen bij de hoogste waardes van de vectoren. Met de getallen die we zo krijgen proberen we iets te zeggen over de *melodie* van het nummer. Voor de verticale methode kijken we naar de intervallen die voorkomen tussen de twaalf waardes van één chromavector. (Dit zijn er heel veel! Het is nodig om alle 144 paarsgewijze vermenigvuldigingen in een tabel te zetten en dan diagonaal op te tellen. Kijk maar naar figuur 3.2.) Dit doen we voor elke chromavector van een bestand en de resultaten tellen we kwadratisch op. Met de getallen die we zo krijgen proberen we iets te zeggen over de *harmonie* van het nummer. De twee waardes per nummer die we zo krijgen (eigenlijk twee keer *zeven* waardes, want we hebben het voorkomen van ieder interval apart geteld), noemen we het *horizontale* en het *verticale intervalkenmerk*.

Hierna gaan we de kenmerken gebruiken om metalnummers te groeperen. We gaan ons algoritme trainen door middel van een *trainingsverzameling*. De trainingsverzameling is een verzameling nummers waarvan het algoritme al van tevoren mag weten in welk subgenre (welke klasse) ze horen. Ons algoritme moet vervolgens voor een stel andere nummers, samen de *testverzameling* genoemd, zeggen in welke klasse ze horen. Het stukje van ons algoritme dat van elk nummer in de testverzameling probeert te raden in welke klasse het hoort heet de *classificator*. De classificator mag zijn gok dus alleen baseren op de kenmerken en de trainingsverzameling. Een simpele classificator is bijvoorbeeld de *k-naaste-buren-classificator* (Engels: *k-nearest neighbor classifier*): we kiezen voor k een getal, zeg 5 of 11 (of één!). Voor een nummer in de testverzameling berekenen we het kenmerk, en dan gaan we in de trainingsverzameling zoeken van welke k nummers hun kenmerken het meest lijken op het eerste kenmerk (of *de kleinste afstand heeft* hiertoe, als we een afstandsfunctie tussen de kenmerken maken). Vervolgens kijken we tot welke klassen deze k nummers hoorden en laten we simpelweg de meerderheid hiervan beslissen tot welke klasse het eerste nummer hoort. We willen natuurlijk een classificator die zo vaak mogelijk goed raad en niet stelselmatig nummers uit het ene subgenre aanziet voor het andere subgenre. Dit is echter naast de gekozen classificator ook sterk afhankelijk van de informatie die hij krijgt van het kenmerk. Als dit geen relevante informatie is, zal geen enkele classificator goed zijn werk kunnen doen!

Als laatste testen we de prestaties van de twee kenmerken in combinatie met verschillende classificators. We zien dan bijvoorbeeld dat voor iedere geteste classificator, het verticale intervalkenmerk het beter doet dan het horizontale intervalkenmerk. Ook zien we dat ze allebei bijna nooit thrashmetalnummers verkeerd classificeren als power metal (één of twee keer van de 38), maar bijvoorbeeld wel thrash metal en death metal vaak met elkaar verwarren. Deze informatie kunnen we gebruiken om in de toekomst een beter algoritme te maken. Bijvoorbeeld door onze kenmerken te combineren met andere kenmerken, die het verschil tussen deze subgenres wel duidelijk maken.

Afterword and acknowledgements

So let it be written;
So let it be done.

Metallica – *Creeping Death*

It *has* been written. It *has* been done. The metal has been classified and the final words are on the page (almost). It is now time for Sabbath. But first, I would like to take the opportunity for some reflection and to thank the countless people whose input was invaluable during the making of this thesis.

Looking back at the process, it proved to be challenging to combine the thesis writing with other study commitments. Several things were left behind: an optimization of the k-NN classifier using neighbors in Gabriel graphs, and an exploration of several possible ways to combine the two features at the end. Nonetheless, this thesis touches upon a wide variety of different subjects, many of which I would not have originally expected to be involved.

Another challenge was writing for both a computer scientist audience and a mathematical audience. The problem I studied was an applied one, but I tried to frame the presented methods and constructions in an exact mathematical style to avoid the ambiguities that sometimes arise when this is done in a more descriptive style in computer science articles. This has the added benefit that they are now more easily generalizable to other applications, so that readers who wish to apply these methods to different problems can make their use of them. On top of that, I hope to have been able to transfer to the reader some insight, and perhaps even some interest, in the classificational subtleties and difficulties of heavy metal.

First of all, I would like to thank my supervisor, Dr John Ashley Burgoyne. Every week, I got lost inside the maze of my own thoughts and ideas, and every week, when walking out of his office, my thoughts were recollected and my objectives were clear again, thanks to his excellent guidance and advise. I hope your expertise, experience and enthusiasm shine through in this thesis. Next, I would like to thank Dr Gerard Helminck, who despite being unfamiliar with the subject, was enthusiastic and had enough faith in me to act as the second signatory for my thesis. Without you this would not have been possible!

My sincere thanks also goes to Aletta Smits, who helped me with some last-minute corrections and who has supported and encouraged me for as long as I can remember. (Which is unfortunately not very long; you know I have a brain like a sieve...) My good friend Wessel Broekhuis, who is a reviewer for Metalfan.nl[1] and author of the book

[1] `http://metalfan.nl/`

Alleen met mijn wereld – hoe ik leerde leven met autisme (Nieuwezijds B.V., 2010), is also my metal partner-in-crime and an endless, encyclopedic resource on metal. We can spend hours debating the correct classification of Gojira and the position of Swedish death metal in death metal history. As such, his words and opinions were invaluable to me while making this thesis.

My special gratitude goes out to Djera Khan, my beloved girlfriend of six years. Next to her mental support and excellent proofreading, I would like to thank her for putting up with me and my obsession with music for all these years. She endured all of my endless metal stories and loved me anyway. I could not have done this without you.

Finally, I would like to extend my thanks to the rest of my friends and family who, I am very lucky to say, are too many to list. They all contributed to this thesis, however directly or indirectly. You know who you are, thank you all very much!

Bibliography

[1] T. Fujishima, "Realtime chord recognition of musical sound: a system using common Lisp music," in *Proceedings of the International Computer Music Association*, pp. 464–467, 1999.

[2] G. Tzanetakis and P. Cook, "Musical genre classification of audio signals," *IEEE Transactions on Speech and Audio Processing*, vol. 10, July 2002.

[3] J. Bergstra, N. Casagrande, D. Erhan, D. Eck, and B. Kégl, "Aggregate features and ADABOOST for music classification," *Mach. Learn.*, vol. 65, pp. 473–484, Dec. 2006.

[4] Y. Panagakis, C. Kotropoulos, and G. R. Arce, "Music genre classification via sparse representations of auditory temporal modulation," in *Proceedings of the 17th European Signal Processing Conference*, pp. 1–5, Aug. 2009.

[5] J. Andén and S. Mallat, "Multiscale scattering for audio classification," in *Proceedings of the 12th International Society for Music Information Retrieval Conference*, pp. 657–662, Oct. 2011.

[6] B. L. Sturm, "Classification accuracy is not enough - on the evaluation of music genre recognition systems.," *Journal of Intelligent Information Systems*, vol. 41, no. 3, pp. 371–406, 2013.

[7] I. Christe, *Sound of the Beast: The Complete Headbanging History of Heavy Metal*. HarperCollins, 2003.

[8] S. Dunn (director), "Metal: A Headbanger's Journey." Seville Pictures, 2005.

[9] S. Dunn (director), "Metal Evolution – Early Metal Part 2: UK Division." Banger Films, Inc., 2011.

[10] S. Dunn (director), "Metal Evolution – New Wave of British Heavy Metal." Banger Films, Inc., 2011.

[11] S. Dunn (director), "Metal Evolution – Thrash Metal." Banger Films, Inc., 2011.

[12] S. Huey, "Reign in Blood – Slayer." `http://www.allmusic.com/album/reign-in-blood-mw0000191741`. Retrieved: June 2014.

[13] S. Dunn (director), "Metal Evolution – Extreme Metal: The Lost Episode." Banger Films, Inc., 2014.

[14] S. Dunn (director), "Metal Evolution – Power Metal." Banger Films, Inc., 2011.

[15] J. Wagner, *Mean Deviation: Four Decades of Progressive Heavy Metal*. Bazillion Points, 2010.

[16] MIDI Manufacturers Association, "The complete MIDI 1.0 detailed specification: incorporating all recommended practices," 1996.

[17] D. Tymoczko, *A Geometry of Music: Harmony and Counterpoint in the Extended Common Practice*. Oxford University Press, 2011.

[18] W. Piston, *Harmony: Fifth Edition*. W. W. Norton & Company, 1987.

[19] R. Kamien, *Music: An Appreciation*. McGraw-Hill, 2011.

[20] International Electrotechnical Commission, "IEC 60908, Red Book standard," *Audio Recording-Compact Disc Digital Audio System*, 1987.

[21] J. W. Cooley and J. W. Tukey, "An algorithm for the machine calculation of complex Fourier series," *Mathematics of Computation*, vol. 19, pp. 297–301, 1965.

[22] E. Gómez, *Tonal Description of Music Audio Signals*. PhD thesis, Universitat Pompeu Fabra, 2006.

[23] T. Jehan, *Creating Music by Listening*. PhD thesis, Massachusetts Institute of Technology, 2005.

[24] C. Harte, M. Sandler, and M. Gasser, "Detecting harmonic change in musical audio," in *Proceedings of the 1st ACM Workshop on Audio and Music Computing Multimedia*, pp. 21–26, ACM, 2006.

[25] L. Devroye, L. Györfi, and G. Lugosi, *A Probabilistic Theory of Pattern Recognition*. Springer, 1997.

[26] K. Tumer and J. Ghosh, "Bayes error rate estimation using classifier ensembles," *International Journal of Smart Engineering System Design*, vol. 5, no. 2, pp. 95–110, 2003.

[27] P. C. Mahalanobis, "On the generalised distance in statistics," in *Proceedings of the National Institute of Science, India*, vol. 2, pp. 49–55, Apr. 1936.

[28] J. Aitchison, *The Statistical Analysis of Compositional Data*. Springer, 2011.

[29] J. A. Martin-Fernandez, J. Palarea-Albaladejo, and R. A. Olea, "Dealing with zeros," in *Compositional Data Analysis* (V. Pawlowsky-Glahn and A. Buccianti, eds.), pp. 43–58, John Wiley & Sons, Ltd, 2011.

[30] E. Fix and J. L. Hodges, "Discriminatory Analysis: Nonparametric Discrimination: Consistency Properties," Tech. Rep. Project 21-49-004, Report Number 4, USAF School of Aviation Medicine, Randolf Field, Texas, 1951.

[31] "K-Nearest Neighbor." http://www.cra.org/Activities/craw_archive/dmp/awards/2003/Mower/KNN.html. Retrieved: June 2014.

[32] T. Cover and P. Hart, "Nearest neighbor pattern classification," *IEEE Transactions on Information Theory*, vol. 13, pp. 21–27, Sept. 2006.

A. Selected albums

This is an overview of the albums that were selected to construct the data set of this project. To the interested reader, they may serve as introductions to their respective subgenres.

Alternative metal:

- Deftones – White Pony
- Disturbed – Indestructible
- Faith No More – Angel Dust
- Helmet – Meantime
- System of a Down – Toxicity

Black metal:

- Bathory – Under the Sign of the Black Mark
- Darkthrone – Transilvanian Hunger
- Emperor – In the Nightside Eclipse
- Immortal – Pure Holocaust
- Mayhem – De Mysteriis Dom Sathanas

Classic metal:

- Black Sabbath – Paranoid
- Iron Maiden – The Number of the Beast
- Judas Priest – British Steel
- Motörhead – Ace of Spades
- W.A.S.P. – The Headless Children

Death metal:

- Bloodbath – Nightmares Made Flesh

- Carcass – Necroticism: Descanting the Insalubrious
- Death – Scream Bloody Gore
- Morbid Angel – Blessed Are the Sick
- Nile – Annihilation of the Wicked

Doom metal:

- Candlemass – Epicus Doomicus Metallicus
- Cathedral – Forest of Equilibrium
- Katatonia – Dance of December Souls
- My Dying Bride – The Angel and the Dark River
- Shape of Despair – Angels of Distress
- Trouble – Psalm 9

Folk metal:

- Ensiferum – Ensiferum
- Korpiklaani – Tales Along This Road
- Moonsorrow – Voimasta ja Kunniasta
- Primordial – To the Nameless Dead
- Skyclad – Prince of the Poverty Line

Gothic metal:

- Lacuna Coil – Comalies
- Moonspell – Irreligious
- Paradise Lost – Draconian Times
- Tiamat – Wildhoney
- Type O Negative – October Rust

Groove metal:

- DevilDriver – The Fury of Our Makers Hand
- Lamb of God – Ashes of the Wake
- Machine Head – Through the Ashes of Empires

- Pantera – Vulgar Display of Power
- Sepultura – Chaos A.D.

Industrial metal:

- Rammstein – Herzeleid
- Fear Factory – Demanufacture
- Ministry – ΚΕΦΑΛΗΞΘ
- The Kovenant – Animatronic
- Godflesh – Streetcleaner

Melodeath metal:

- Amon Amarth – With Oden on Our Side
- Arch Enemy – Wages of Sin
- At the Gates – Slaughter of the Soul
- Children of Bodom – Follow the Reaper
- In Flames – The Jester Race

Nu metal:

- Ill Niño – Revolution Revolución
- KoRn – KoRn
- Limp Bizkit – Significant Other
- Linkin Park – Hybrid Theory
- Slipknot – Slipknot

Power metal:

- Blind Guardian – Imaginations from the Other Side
- DragonForce – Valley of the Damned
- Helloween – Keeper of the Seven Keys, part I
- Sabaton – Primo Victoria
- Sonata Arctica – Ecliptica

Progressive metal:

- Amorphis – Skyforger
- Dream Theater – Images and Words
- Fates Warning – Awaken the Guardian
- Pain of Salvation – Remedy Lane
- Queensrÿche – Rage for Order

Sludge metal:

- Acid Bath – When the Kite String Pops
- Crowbar – Broken Glass
- Eyehategod – Take as Needed for Pain
- Melvins – Houdini
- Neurosis – Through Silver in Blood

Stoner metal:

- Corrosion of Conformity – Deliverance
- Down – NOLA
- High on Fire – Snakes for the Divine
- Kyuss – Blues for the Red Sun
- Orange Goblin – Frequencies from Planet Ten

Symphonic metal:

- Epica – The Divine Conspiracy
- Haggard – Eppur Si Muove
- Nightwish – Once
- Therion – Theli
- Within Temptation – Mother Earth

Thrash metal:

- Anthrax – Spreading the Disease
- Kreator – Pleasure to Kill
- Megadeth – Rust in Peace
- Metallica – Master of Puppets
- Slayer – Reign in Blood