1

INTRODUCTION

1.1 RESEARCH GOALS

1.1.1 *Audio Corpus Analysis*

The last decade saw the rapid growth of the digital humanities, an interdisciplinary area of research in which research topics and methods from the humanities and computing come together. The rise of digital humanities research can be explained by the unprecedented availability of tools and resources for *data-intensive* research. In linguistics, for example, it is now easier than ever to evaluate the evidence for a theory or hypothesis not just in a small selection of documents, but in a large corpus. Digital linguists, musicologists and other humanists owe this opportunity to pioneering efforts that go back to the beginning of the computing era, including digitization programs, the creation of various data formats and the developments of new infrastructures for off line and on line data.

Musicology has seen decades of digital and computational research, beginning as early as the 1960's and 1970's. In the late 1990's, building on developments in early computational musicology, digital signal processing and the web, developments in the field of music information retrieval (MIR) have given music research a digital boost from another, more consumer-oriented angle: that of music search and recommendation. MIR's continued pursuit of new data analysis methods has provided the music research community with a huge array of methods for the quantitative analysis of music, at unprecedented scale.

Over the last years, researchers have turned to these technologies to engage in ever more complex, large-scale and data-intensive music analysis—researchers with diverse backgrounds in both statistical or 'database-driven' musicology, and music information retrieval. *Corpus analysis* has been used in the search for musical 'universals' (universal properties of music) [171], to find and track trends in a sample of musical works, e.g., Western classical compositions [166] and Western popular music [42, 122, 175], or to model 'stability' of musical motives under oral transmission [197]. Corpus-level music analysis has also been used to test theories of expectation [80], or correlate features of the music with performance on various tasks, ranging from a memory test to walking [104].

A large majority of this research deals with *symbolic* data: scores, chords or manual annotations. This is not surprising given the origins of this research in computational musicology, but it contrasts sharply with the predominance, in music information retrieval, of research on *audio*, i.e., music recordings. Despite the wide availability of audio data and tools for audio content analysis, very little work has been done on the corpus analysis of audio data.

This thesis presents a number of contributions to the scientific study of music based on *audio corpus analysis*. We will begin this investigation with a closer look at what audio corpus analysis is, and how it fits into the larger context of music research, reviewing the fields of research of which it is a part, and laying out the argument for a corpus analysis based on audio, rather than symbolic data (section 1.2). But first, in section 1.1.2, we introduce the COGITCH project, the initiative behind this research, and discuss its objectives, to motivate the research goals of this thesis.

1.1.2 *The* COGITCH *project*

The COGITCH project was part of CATCH, a Netherlands-based science program for research on the intersection of cultural heritage and information technology, and financed by NWO, the Dutch organisation for scientific research. The COGITCH project was a collaboration between two heritage institutes and two universities. On the side of the heritage institutes, the Meertens Institute (MI) is involved in the research and documentation of Dutch language and culture, and the Netherlands Institute for Sound and Vision (S&V) oversees, among other things, an archive of Dutch media heritage, including music. The research groups affiliated with the universities are the department of information and computing sciences of Utrecht University and the music cognition group at the University of Amsterdam.[1]

Both MI's and S&V's collections are very rich in data: the MI's *Dutch Song Database* contains metadata for over 140,000 songs, and audio recordings for a subset of them, including 7178 unique field recordings of Dutch folk songs [98].[2] The S&V collection contains metadata and audio for over 300,000 songs which it rents out to various media institutions, and an additional, physical collection of over 50,000 vinyl records (33, 45 and 78 RPM), part of which was digitized during the COGITCH project.

Access to Digital Music Heritage

The COGITCH project focused on two main objectives. The first goal of the COGITCH project was to facilitate access to both institutions' music collections, through an integrated search infrastructure, making the collections *interoperable*. Inspired by the technologies of an earlier CATCH project, WITCHCRAFT, this interoperability should extend into not just the metadata, but also the content of the collections: the music recordings.[3]

As will be explained in chapter 5, content-based retrieval within and between these collections requires a model of *similarity* between documents in the collection, which in turn requires powerful fragment-

1 `http://www.uu.nl/organisatie/departement-informatica`
`http://mcg.uva.nl`

2 `http://www.liederenbank.nl/index.php?lan=en`

3 The goal of the WITCHCRAFT project (2006-2010) was to create a 'functional content-based retrieval system for folksong melodies' using the Meertens Institute's collection of symbolic folk song transcriptions.
`http://www.cs.uu.nl/research/projects/witchcraft/`

level audio description methods and scalable methods for the comparison of these descriptions. Apart from being a useful advancement in itself, a good model of similarity for the documents in these collections can also benefit research into the evolution of folk song music in The Netherlands, the dynamics of stability and variation in oral traditions, and the emergence of popular music in the twentieth century [196].

Hooks and Memory in Popular Music

A second, equally important goal of the COGITCH project was to establish a scientific model of *hooks*. A hook, as will be explained in chapter 7, can be defined as the part of a song that is most recognizable. With a scientific investigation into hooks and the recognizability of real-world music, the COGITCH project seeks to improve our understanding of the role of memory in popular music. Eventually, we hope this will contribute to a better understanding of music memory in general.

A model of hooks can also help create better, more perceptually and cognitively informed similarity models and retrieval methods. Part of this second goal of the COGITCH project was to use any findings on hooks and memory to improve the similarity models discussed above and inform the aforementioned investigations into stability, variation and the evolution of folk and popular music. But it can be used to support other kinds of information retrieval, too: search, recommendation and browsing systems can be improved with better models of what users may or may not remember well [69].

1.1.3 *Research Goals*

The goals of this thesis are threefold. The first two relate to audio corpus analysis. First, we aim to review and advance the available audio description methods for corpus analysis research. Second, we aim to review the corpus analysis methodology itself, and explore new methods that address some of the open issues.

The third goal relates to the objectives of the COGITCH project. Not all of these will be addressed: some of the project goals that fall outside the scope of this thesis are the analysis of stability and variation, and the application of new findings on hooks and music memory to improve music retrieval technologies. The two objectives of the project that will be discussed in this thesis are the application of new audio description methods to develop new approaches to scalable song similarity, and a computational analysis of hooks.

These goals will not be addressed in the order they are now stated. Section 1.3, the last section of this chapter, will present an outline of the remainder of this thesis, with a focus on how the above objectives will be approached.

1.2 DISCIPLINARY CONTEXT

What is corpus analysis, and where does it fit in the larger context of music research? Corpus analysis was described above as 'data-intensive music research'. A more specific definition that we shall use in this thesis is: the analysis of a music collection with the aim of gaining insights into the music. Not any collection, of course: given one or more research questions, a dataset should be selected to represent the particular music to which the questions pertain.

The above definition seems to 'exclude' a sizable segment of computational and digital music research. Indeed, not all digital musicology involves the content-based analysis of a music collection. Conversely, most MIR research does not aim at new insights into music. To see this, we need to take a closer look at what musicology is, what MIR is, and how research in musicology and MIR is done. At the end of this discussion we also introduce *empirical musicology* and *cognitive musicology* and look at how they relate to musicology and MIR. Finally, we discuss how corpus analysis research fits in.

1.2.1 *Musicology in the Twentieth Century.*

Summarising a century of research in just a few paragraphs necessarily involves oversimplifications. However, it is fair to say that music scholarship, at the beginning of the twentieth century, had a strong focus on individual musical works as represented by the score. A lot of work was done recovering, editing and analyzing music sources, and much of it was concerned with the music that was at that time seen as Europe's great works of art, identifying the particular structural, musical aspects of these compositions that made them into the masterpieces for which they were seen. For many, the aim was to expose the nature of the true and the beautiful, so as to advance the art itself. Mastership, genius and beauty were seen as absolute, and in the perspective of the romantic ideals of the time, were to be found in instrumental music above all, and particularly, in its use of harmony [36].

This 'analysis'-centered musicology continued further into the twentieth century, with a positivist approach to music analysis which assumed that, like in the study of the physical world, objective laws could be found that underlie the way art works. Proponents of this approach, like Schoenberg and Schenker, pushed musicologists towards formalisation, establishing a theory of music that emphasized structure, abstraction and rules. Similar principles would inform the first endeavours into computational music analysis, which included attempts to implement these positivist approaches to music analysis, including 'Schenkerian analysis', as a computer program (e.g., [87]).

As musicology evolved, however, scholars started to see problems with the positivist project. One way this manifested, is through a shift away from the emphasis on music as autonomous 'works' to a kind of process, in which the performer and the listener play an important role, too. The shift was partially pioneered by scholars of non-Western music, who, much before music theorists and historical musicologists, broke with the positivist traditions from before the second World War. Particularly, 'comparative musicology', which aimed to understand the causes and mechanisms that shape the evolution of music across cultures, was replaced with ethnomusicology, a new field based on the paradigms of cultural anthropology [170].

Nonetheless, a strong association persisted between the 'chronology' of the musical production chain, in which composition precedes performance and listening , and a hierarchy of prestige with composition ranking above performance and listening. A stronger paradigm shift came with the arrival of 'new musicology' around 1985 [88]. Looking back at the nineteenth century, and how little had changed up until the 1950's, musicologists became increasingly critical of the idea that there is such a thing as objective beauty or greatness. They also pointed to other perceived flaws in the positivist program, such as the assumption, in the search for a single authoritative reading for every musical work, that it was possible to figure out a composer's intentions. These intentions cannot be known, the new musicologists argued, and, more fundamentally, neither the intentions or the music can be understood onto itself. Music is a medium that influences and is influenced by feelings, desires and societal context such as power structures and taboo [36].

In this light, much of the historical writings on music were seen as justifications of the canon of a certain time, and the canon itself a product of the consensus of a cultural and political elite. The scholarship of the nineteenth century was therefore complicit in providing the justification for the cultural and political power structures of their time. Influenced by critical theory, feminism and gender studies, a new 'critical' musicology emerged, with the intention to understand music as it interacts with society and expose ideologies in music and music writing (e.g., [124]).

How did the proponents of computing respond to this? Judging by scientific output, computational musicology, as it is now referred to, all but disappeared in the 1980's [22, 201]. After the first initiatives mentioned at the very beginning of this chapter, a lot of effort was spent keeping up with ever-changing computing architectures (from mainframes to personal computers) and storage formats (from punch cards to floppies and hard drives). Meanwhile, the paradigm shift of new musicology tempered the heroic ambitions of these endeavours, and music computation, as an agenda, became unfashionable. A new wave of digital music research, however, came with the emergence of music information retrieval in the late 1990's.

1.2.2 *Music Information Retrieval*

Music information retrieval (MIR) is typically described as an interdisciplinary field of scientific research, with origins in computing, library science and musicology [47]. The roots of MIR largely overlapped with the early work in computational musicology so, like computational musicology, it all but disappeared in the 1980's.

In the 1990's, however, digital audio became more widely available, and computing power surged. Music information retrieval revived as an area of *applied* research, with a partial re-orientation to audio-based research, but most of all, a strong focus on *tasks*, in which specific kinds of information are extracted from musical data [22]. In the task of optical music recognition (OMR), for example, a computer program is given, as input, an image of a score, and outputs a digital version of the score in some symbolic format.

The kind of tasks at the forefront of MIR evolved throughout the 1990's and 2000's, to include input data such as images, symbolic data, digital audio, metadata and crowd-sourced social tags [22]. Types of output that MIR systems are engineered to produce can roughly be divided into: metadata (as in recommendation systems), classification labels (as in genre classification or key finding), and symbolic sequences (as in chord labeling and other kinds of transcription) [22]. Some of these tasks will be presented in more detail in chapter 2.

The most important MIR development that came in the 2000's, was probably the introduction of MIREX, an organisational framework for the joint evaluation of new MIR algorithms under rigorous tests conditions, using common datasets and evaluation scripts [47]. Growing rapidly since its first edition in 2005, MIREX now annually evaluates the algorithms of over 100 researchers. This helped identify and advance the state of the art across 24 MIR tasks.

Music information retrieval technologies not only have become a part of everyday life (as they are incorporated in, e.g., music streaming services), they have the potential to become powerful tools in musicology. But did the progress made by MIR address the concerns of new musicology better than the first efforts in computational music research?

An optimist's evaluation would probably point to the progress made in scholars' access to digital resources. Optical music recognition tools are a relevant example, as well as on line music libraries and music typesetting tools such as Finale and Sibelius, which already form an essential part of many musicologists professional 'work flow' [83]. In other word, access to data has further improved since the day of first digital encoding projects. In itself however, this facilitation of access doesn't fundamentally address any of the new musicologists' concerns.

The pessimist might conclude that MIR has moved away from musicology entirely: the success of MIREX has made evaluation such a central part of MIR that it is now only focused on the kinds of analysis for which a *ground truth* exists. And whether such a ground truth really exist is a subject of continued debate. In other words, the 'computability paradigm' of early computational music research, which treats empirical data as a hermetic ground truth, and the 'critical' perspectives of new musicology, in which any ground truth is necessarily based on assumptions and hypotheses, have not been truly reconciled.

Furthermore, this focus on tasks with a 'ground truth' may be hampering the development of methods for analyzing music data to discover something 'new'. Burgoyne et al., in their review of MIR for a new digital humanities companion, conclude: "As soon as computers became a part of the academic infrastructure, researchers became interested in using them to study music. Over a period of some decades, the computers have gotten better at answering research questions" [22]. What we argue here is that perhaps computers have not, in fact, worked as much on answering research questions as on 'solving tasks'. Research questions have centered largely on issues like: 'can new method X be used to increase performance on task Y?', and as a result, computers got better at Y. It may be time to put these technologies to work in answering actual research questions, not about method X or task Y, but about the music.

1.2.3 *Empirical Musicology and Music Cognition*

While music information retrieval branched off and diverged from musicology itself, empirical and quantitative methods in music research revived in another way. The revival coincided with a renewed support for empirical methods, a 'new empiricism', that was not just limited to music research. The new empiricism gave rise to a new *empirical musicology* (sometimes *systematic musicology*, an older and broader term mostly used in continental Europe), of which empiricism, formalisation and computation were an important part. And last but not least, both developments also propelled the new research area of *cognitive musicology* or *music cognition*. We now discuss this chain of developments, beginning with Huron's analysis of the new empiricism in [79].

Empirical Musicology

Huron first examines the new musicologists' resistance to empirical methods. New musicology, and postmodernism in general, tend to assume that there is no absolute truth to be known. Instead, truth is seen as a "social construction that relates to a local or partial perspective on the world". In other words, there is no privileged perspective or interpretation, and postmodernists are right to point this out.

However, empiricism and postmodernism are also very similar in a different way: both "cultivate institutionalized forms of skepticism". The kind of skepticism typically associated with the sciences involves holding scientific claims up to a standard of evidence that is focused on minimizing the number of 'false positive' claims, i.e., claims that are accepted as true even though they are not (strong emphasis on p-values, for example, reflects this focus). For many in arts and humanities research, however, a common fear is to make 'false negative' errors, to dismiss a claim that might in fact have merit—assigning claims such certainty that they provide explanatory "closure", may be regarded as a provocation, "a political act intended to usurp all other views" [79]. The difference may be due to the level of risk associated with each kind of error (a 'false positive' claim in the humanities, for example, may be considered as relatively harmless compared to a false positive in the sciences) and with the amount of data typically available to test claims. In other words, the humanities and the sciences "might diverge in their philosophical conceptions about the nature of the world, they nevertheless share deep methodological commonalities" [79].

Furthermore, "even if we accept the proposition that there is no privileged interpretation, it does not necessarily follow that all interpretations are equally valid", or else all knowledge would be impossible. Looking at it from a cognitive angle, Huron notes: while "we should recognize that human beings are cultural entities, we must recognize that humans are also biological entities with a priori instinctive and dispositional knowledge about the world, that originates in an inductive process of evolutionary adaptation" [79].

Huron concludes that not all forms of rigor and empiricism should be abandoned, provided that we have data and a strategy to appropriately balance type-I and type-II errors, and provided that we navigate the 'known potholes' associated with the methodologies of choice, such as logical and rhetorical fallacies and statistical self-deception.

Honing, in 2004, also looked at the comeback of empiricism in musicology, and observed three trends. The first trend was the emergence of a revitalized systematic musicology, that "is based on empirical observation and rigorous method, but at the same time is also aware of, and accounts for, the social and cultural context in which music functions" [68]. A reconnection was found between the empiricists and the new musicologists, much in the way Huron described in [79].

The second trend Honing includes is the growing role of formalisation and computation in musicology, discussed earlier. This is a trend that precedes Huron's 'new empiricism'. Did these computational approaches see a similar adaptation to the concerns of postmodernism (like the new empirical method), or not (like MIR)? Notably, Huron in [79] speaks of quantitative methods, but not of computation.

A classic example of formalisation in music theory is Lerdahl and Jackendoff's 'Generative Theory of Tonal Music' (GTTM), a highly formalized attempt to model the cognitive processing of Western tonal music, inspired in part by the proto-generative theory of Schenker and the generative grammars of Chomsky [105, 152]. The model was a landmark in the coming of age of music cognition and heavily influenced important interdisciplinary music research. But it has also been criticized, e.g., for not being a theory in the scientific sense of the word (i.e., "subject to testing and potentially falsification by hypothesis formation and experiment") and for treating music, like language in the work of Chomsky, as some kind of external absolute: "observing an idealised version of the phenomenon, and treating it as though it had its own existence" [37, 203]. To the new empiricist, in other words, it is neither empirical, nor is it much informed by the postmodernists' critique of music as an object. Another example given by Wiggins is Temperley's Grouper algorithm, a computational model for the segmentation of musical phrases [203]. In this case, the algorithm relies on the *ad-hoc* setting of a parameter to produce plausible segmentations for a particular style of music. To Wiggins, this makes it effectively a descriptive, rather than predictive theory. Nonetheless, descriptive models such as GTTM and Grouper can be a stepping stone to make progress towards a more explanatory, prescriptive model, and computational algorithms also get the merit of having given a greater visibility to of musicology outside the humanities [68, 203].

Music Cognition

Parallel to these new empirical and computational trends in musicology, was the 'cognitive revolution', Honing's third trend, marked by a similar interest in both rigorous empiricism, and computation [68].

The cognitive sciences are concerned with various aspects of the mind including perception, attention, memory, language, action, and emotion. Theories on how the mind works go back to Ancient Greece.

Only since the 1970's, however, 'cognitive science' was recognized as a research area of its own, with its own name and its own agenda, driven by twentieth century advances in the theory of computation (e.g., the work of Alan Turing), linguistics (e.g., Chomsky) artificial intelligence (Marvin Minsky and others) and other fields [152]. The emergence of cognitive science as a field of research created a new home for the study of a variety of aspects of music, too, including perception, expectation, music memory and emotion.

Music cognition is most distinct from twentieth century musicology, in that it regards music as fundamentally a psychological entity. As such, it enabled a use of empiricism and computing in music research that had not been seriously considered before: as scientific theories of music listening and performance. Wiggins et al., for example, argue for "a theory of music which starts from the position that music is primarily a construct of human minds" [203].

From early on, computational modeling became an important tool in the cognitive sciences. Honing recounts in 2011: "there is hardly a cognitive theory without a computational component" [70]. Computational modeling is distinct from other empirical methods in that it is not an instrument for quantitative observation. It is a methodological cycle that integrates theory and observation. True computational models are precisely formalized theories, therefore, like theories, they generate new hypotheses, which can in turn be tested empirically.

Summary

To summarize, some of the work done in musicology and computational musicology hasn't necessarily been striking the right balance between empiricism and the concerns of new musicology. However, an empirical approach to music research is possible that acknowledges these concerns, as laid out by Huron. A particularly promising place to look for a connection between an empirical methodology and a critical perspective of music is in cognitive musicology, which balances these concerns by emphasizing the need of predictive models and an approach to music from the perspective of the listener. By researching music at the corpus level, we can aspire to a research method that accounts for the both the context and cognition of listening, to acknowledge that music has a cultural and cognitive dimension.

1.2.4 *Why Audio?*

Before we move on to the next chapter, this section provides an argumentation for our focus on the analysis of *audio recordings*, rather than symbolic music data.

The motivation to use audio data in a corpus analysis study can be quite simple: sometimes no symbolic format representation of a particular music corpus exists. Assembling a symbolic format dataset of music often requires tedious transcription work, especially in the case of music that is not part of the Western art music tradition ('classical music', in the broad, colloquial sense), whereas audio data is much more readily available. However, there are also more fundamental reasons why audio music data can be the format of choice for research.

Today's digital symbolic music formats are based on Western notation [36].[4] As many musicologists have argued over the last decades, there is a limit to the range of musical ideas and expressions Western music notation can represent, and to which extent. It evolved to suit the needs of composers and musicians within a set of traditions now denoted together as European art music, therefore it has never been intended to capture the particularities of non-Western musical styles, or even popular music (e.g., electric guitar solos or rap vocals) [79]. On a more fundamental level, most music notation isn't primarily intended to describe music at all, rather, it serves as a means to convey some of the necessary instructions for a performer to make a composition or arrangement audible. Even though Western notation depicts a more direct representation of the musical 'surface' than, e.g., guitar or lute *tablature*, by representing music in a relatively instrument-independent format, it nonetheless omits a wide range of very important features

4 and for the case of MIDI, keyboard music.

of a sound.[5] Most significantly, those features that are historically tied to performance rather than composition or arrangement: expressive timing and dynamics, timbre, tuning, ornamentation, etc. [36, 123].

Of course, most researchers will acknowledge this, but might be interested in a particular aspect of music that can be represented symbolically without too much loss of nuance. Harmony in a pop song, for example, can to some extent be represented as chords, or the main melody of a romantic concerto as a single monophonic sequence of notes. Even in these cases, however, such a perspective implies a conscious or unconscious choice not just to leave the performer out of the equation, but also the listener. When this is not acknowledged, these perspectives carry with them an implicit assumption that what can be notated contains or correspond to something the human perceptual systems 're-extracts' from the acoustic signal. This is not the consensus in music cognition. How many individual voices can a typical listener discern in polyphonic music? How is our listening affected by difference in salience between notes in a performance? How many different chords can a listener without formal music training tell apart? How does a jazz fan process the fast arpeggios of a virtuosic saxophone solo? Findings by Huron on the first question indicate strong limitations of the perceptual systems, with clear differences between expert musicians and other listeners [76]. Many similar, related questions remain unaddressed.

On the other hand, of course, audio representations have limitations of their own. First, an audio recording cannot capture every aspect of a musical performance in perfect detail either (e.g., spatial effects) or may distort it (e.g., dynamics) However, it does capture a lot; essentially, everything required to reproduce it again to a listener in the same way most music has been experienced by most listeners in the last 50 years: over some kind of speaker system (e.g., headphones). This still leaves the listener out of the equation, but at least it contains the necessary ingredients to apply most of what we know about the perception of music as part of any analysis.

5 In tablature, instructions are represented in relation to the instrument, e.g., by indicating which strings and frets are to be played on a guitar, rather than which notes.

This touches on a second issue: even if audio representations contain the necessary material, we may not have the analysis methods to extract meaningful perceptual-level information that we humans have access to effortlessly. Therefore, an important part of this thesis will be to assess this issue in depth, curate and develop a set of simple but plausible representations of harmony, melody and timbre, building on perceptually justified feature extraction and informed by a listener-centered perspective on music, as an alternative to those employed in the analysis of symbolic music data. And as progress is made on both our understanding of the perceptual representations that make up the human auditory system and on the technology that is available to model it, we can expect more such efforts to be made in the future.

In short, there is a very large potential in audio data for corpus analysis, for three reasons: there is simply much more audio data than symbolic data available for research, symbolic music representations have fundamental shortcomings in the way they represent much of today's music, and audio data allow for a more listener-centered approach.

1.2.5 *Conclusion*

Having reviewed a number of recent evolutions in music research, namely new musicology, music information retrieval, empirical musicology and music cognition, we can better situate our intended approach to corpus analysis in this complex interdisciplinary field.

First, we recognize that indeed, music, or a piece of music, is not some externally defined object, that can be understood in terms of absolute truths. However, an empirical approach to music research is possible that acknowledges these concerns.

We also concluded that much of the work done in computational musicology hasn't been striking the right balance between empiricism and the concerns of new musicology, e.g., Lerdahl and Jackendoff's influential but mostly descriptive models of music theory. Similarly, a lot of progress has been made in music information retrieval, but there, too, the gap between the 'computability paradigm' and the critical perspectives that challenged it hasn't quite been bridged.

A better place to look for a connection between an empirical methodology and a critical perspective of music—as an abstract, intangible phenomenon—is in cognitive musicology, which balances these concerns by emphasizing the need of predictive models and an approach to music from the perspective of the listener.

We can now position the corpus analysis approaches we propose in this thesis between these observations. First, following Huron, we aspire to an approach to music research that is empirical, but avoids 'absolute' theories of 'absolute' music. Second, the corpus analysis approaches in this thesis will be quantitative, but not necessarily computational. Of course, computational models can be useful: tools from MIR or computational models of perception of cognition should be used if they help in providing a more relevant or cognitively valid representation of the data, but the 'computational' paradigm of ground truth reconstruction and its focus on performance should not be the main methodology. Third, our approach to corpus analysis will be informed by music perception and cognition, in that it assumes the perspective of the listener. The methodological consequences of these three positions will be discussed in chapter 3.

1.3 OUTLINE

1.3.1 *Structure*

This thesis is divided into three parts. The first part looks into related work and methodology. We give an overview of the audio description methods that have been proposed in the music information retrieval literature, concentrating on timbre, harmony and melody. We also give an overview of some of the applications for which they have been developed: music classification, structure analysis and content identification (chapter 2: Audio Description). In the next chapter, we review a selection of corpus analysis research, focusing on hypotheses, data, descriptors and statistical analysis methods. We also review, in a case study, two important audio corpus analysis studies on the evolution of popular music [122, 175]. Based on the review and case study, we formulate a number of methodological guidelines for corpus analysis of music, and musical audio in particular (chapter 3: Audio Corpus Analysis).

The second part of the thesis presents contributions to support future audio corpus analysis research. The first chapter of this part centers on a corpus analysis of song sections in the Billboard dataset [26]. We present a first selection of relevant audio corpus analysis features for the analysis of harmony, melody and timbre. We also apply, for the first time, a feature analysis of audio data based on probabilistic graphical models (chapter 4: Chorus Analysis). The next chapter expands the available feature set for audio corpus analysis by presenting a new type of cognition-inspired melody and harmony descriptors. The third and final chapter of this part details the computational aspects of the descriptors presented in Chapter 5, and presents an implementation of the features for use in corpus analysis and content-based retrieval (chapter 6: Audio Bigrams). Both chapters connect the audio description contributions to the project goal of improving efficient music similarity measures for musical heritage collections.

The final part of the thesis addresses the last remaining research goal, bringing several contributions together in a computational analysis of hooks in popular music. The first of two chapters on this experiment presents Hooked, a game designed to collect data for the analysis and implemented as *Hooked!*[6] and *HookedOnMusic*[7] (chapter 7: Hooked). The second chapter of this part presents the analysis itself, using the descriptors proposed in Chapter 4 and 5, and a set of novel 'second-order' audio features. The results provide new insight into what makes music catchy (chapter 8: Hook Analysis).

6 http://www.hookedgame.nl/
7 http://www.hookedonmusic.org.uk/

1.3.2 *How to Read This Thesis*

Readers only interested in a subset of the work presented in this thesis may prefer to skip through to the chapters of their interest. For example, readers interested in an introduction to music information retrieval may find chapter 2 useful in itself.

Please refer to figure 1 for an impression of the relationships between the chapters. Chapters in the left column focus primarily on audio description: audio features, feature evaluation and implementations. These might be most interesting for readers with a background in music information retrieval or signal processing. Chapters in the middle column focus mostly on corpus analysis. These might appeal to readers mostly interested in methodology, and in our eventual findings. Finally, readers with absolutely no interest in technical details and statistics may find chapters 3, on methodology, and 7, on data collection, most welcoming.

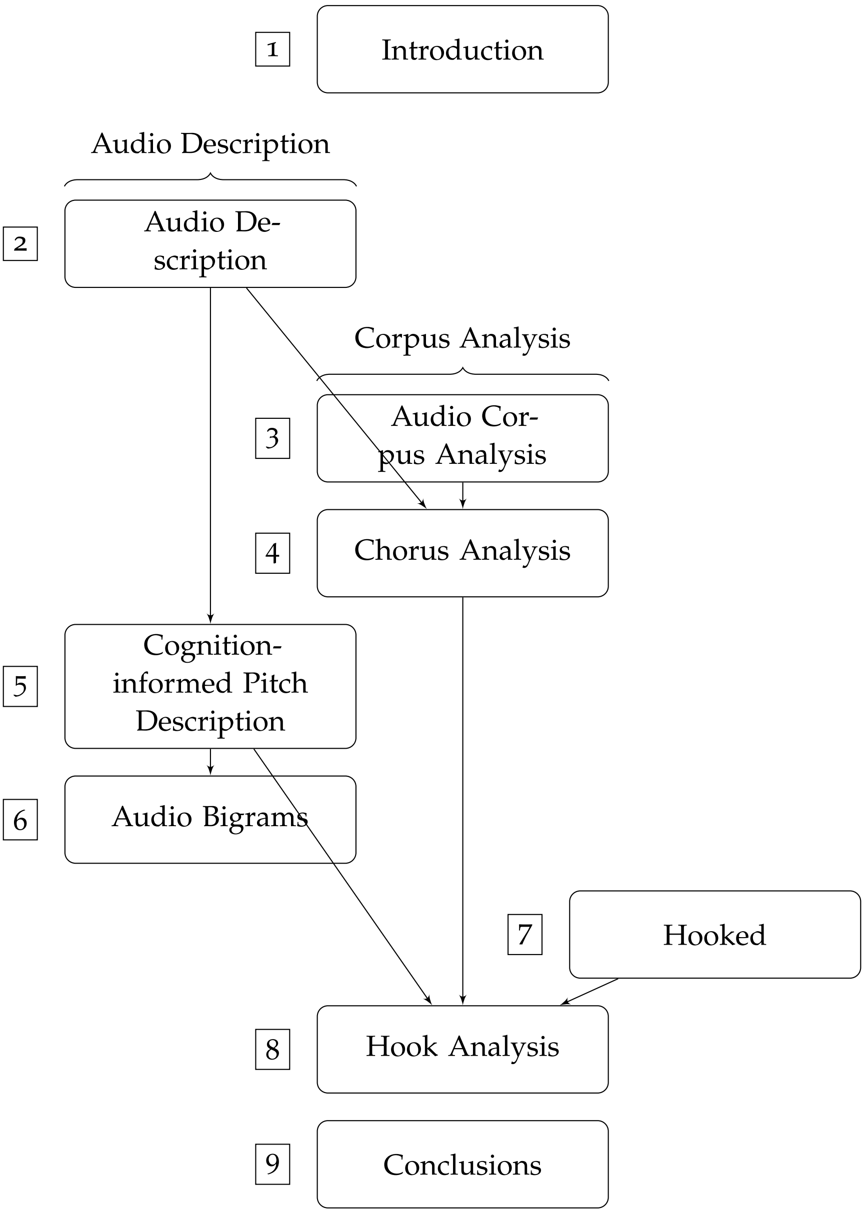

Figure 1.: Overview of chapter relations in this thesis.