A Computational Approach to Content-Based Retrieval of Folk Song Melodies

Peter van Kranenburg

ISBN 978-90-393-5393-6

SIKS Dissertation Series No. 2010-43
The research reported in this thesis has been carried out under the auspices of SIKS, the Dutch Research School for Information and Knowledge Systems.

Cover design by Droppert, vld.

Photo cover: Agricultural workers in flax field, Groningen, Netherlands, beginning of twentieth century. (Private collection, Groningen, Netherlands).

Typeset with LaTeX and LyX.
Printed by Ridderprint Offsetdrukkerij BV (http://ridderprint.nl).

A Computational Approach to Content-Based Retrieval of Folk Song Melodies

Een computationele aanpak van het zoeken naar volksliederen
op melodische inhoud
(met een samenvatting in het Nederlands)

PROEFSCHRIFT

ter verkrijging van de graad van doctor
aan de Universiteit Utrecht
op gezag van de rector magnificus, prof.dr. J.C. Stoof,
ingevolge het besluit van het college voor promoties
in het openbaar te verdedigen
op maandag 4 oktober 2010 des middags te 4.15 uur

door

Pieter van Kranenburg

geboren op 9 juni 1977
te Schiedam

Promotoren: Prof.dr. R.C. Veltkamp
Prof.dr. L.P. Grijp

Co-promotor: Dr. F. Wiering

This research was supported by the Netherlands Organisation for Scientific Research (NWO) within the CATCH program (Continuous Access to Cultural Heritage) under project number 640-003 501.

רָח֖וֹק מַה־שֶּׁהָיָ֑ה וְעָמֹ֥ק ׀ עָמֹ֖ק מִ֥י יִמְצָאֶֽנּוּ׃

Contents

Preface

The research that is presented in this thesis was performed in the context of the WITCHCRAFT project, which had as its aim to develop a content-based retrieval system for folksong melodies (Wiering et al., 2009). The acronym stands for: What Is Topical In Cultural Heritage: Content-based Retrieval Among Folksong Tunes. This project was part of the CATCH program (Continuous Access to Cultural Heritage) as NWO project no. 640-003 501. The CATCH program consists of a number of projects on digitization of and access to cultural heritage. Each of these projects is a collaboration between a research institute, such as a university, and a cultural heritage institute, such as a museum or a national archive. In each project techniques from Information Science are employed to improve accessibility of data and collections from the cultural heritage institutes. The program is founded by The Netherlands Organization for Scientific Research (NWO).

The WITCHCRAFT project was a collaboration between Utrecht University and the Meertens Institute, Utrecht University as research institute and the Meertens Institute as cultural heritage institute, which hosts the collection of Dutch folk songs for which retrieval methods had to be developed. The project team was located at the Meertens Institute, which enabled close collaboration with the collection specialists.

This project offered me the opportunity to collaborate with both musicologists and computer scientists, which is exactly what I want to do. During my master projects Electrical Engineering at Delft University of Technology and Musicology at Utrecht University, I discovered the enormous potential Information Science has for Musicology. Information Science offers many techniques for knowledge discovery in large amounts of data. To apply these techniques to musical data, domain knowledge from Musicology is essential. It proves hard to realize interdisciplinary collaboration in which this approach can flourish.

Therefore, I very much appreciate the setup of the CATCH projects, which stimulates the collaboration that is necessary to successfully apply methods from Information Science to research questions from other domains.

The chapters in this thesis are based on various publications according to the following overview.

Chapter 2

- P. van Kranenburg, J. Garbers, A. Volk, F. Wiering, L.P. Grijp, R.C. Veltkamp (2007). "Towards Integration of MIR and Folk Song Research". *Proceedings of the 8th International Conference on Music Information Retrieval. Austrian Computer Society*. Vienna. pp. 505–508.

- P. van Kranenburg, J. Garbers, A. Volk, F. Wiering, L.P. Grijp, and R.C. Veltkamp (2010). "Collaboration Perspectives for Folk Song Research and Music Information Retrieval: The Indispensable Role of Computational Musicology". *Journal of Interdisciplinary Music Studies* 4, issue 1, pp. 17–43. doi: 10.4407/jims.2009.12.030.

Chapters 3 and 4

- A. Volk, P. van Kranenburg, J. Garbers, F. Wiering, R.C. Veltkamp, and L.P. Grijp (2008). "A Manual Annotation Method for Melodic Similarity and the Study of Melody Feature Sets". *Proceedings of the 9th International Conference on Music Information Retrieval. Philadelphia*. pp. 101–106.

- A.Volk, P. van Kranenburg, J. Garbers, F. Wiering, R.C. Veltkamp, and L.P. Grijp. (2008) "The Study of Melodic Similarity using Manual Annotation and Melody Feature Sets". Technical Report UU-CS-2008-013. Utrecht University. Netherlands.

Chapter 5

- P. van Kranenburg. "On Measuring Musical Style – The Case of Some Disputed Organ Fugues in the J.S. Bach (BWV) Catalogue". Computing In Musicology 15 (2007-8). Online version: http://www.ccarh.org/publications/books/cm/vol/15/cm15-07b-vankranenburg.pdf.

Chapter 6

- P. van Kranenburg, A. Volk, F. Wiering, and R.C. Veltkamp (2009). "Musical Models for Folk-Song Melody Alignment". *Proceedings of the 10th International Conference on Music Information Retrieval*. Kobe, pp. 507–512.

Chapter 7

- P. van Kranenburg and G. Tzanetakis (2010). "A Computational Approach to the Modeling and Employment of Cognitive Units of Folk Song Melodies using Audio Recordings". Accepted for presentation at the 11th International Conference of Music Perception and Cognition, Seattle, August 2010.

Chapter 1

Introduction

The main subject of this thesis is the computational modeling of similarity relations between folk song melodies for retrieval purposes. To carry out such a research project, one needs to take an interdisciplinary approach. Disciplines that are involved are Musicology (more specifically: Folk Song Research), Computer Science and Music Information Retrieval, which is an interdisciplinary research area itself. This introductory chapter sketches the broad outlines of the background of the research, while the next chapter studies more in detail the interrelationships between the involved disciplines, as well as research directions and strategies.

Apart from merely addressing the question how to relate folk song melodies automatically, this thesis also wants to contribute to Computational Musicology, which, like other research areas in computational humanities, is still in its infancy.

1.1 Context: The WITCHCRAFT Project

The research was performed at the Meertens Institute in Amsterdam, which hosts a collection of over 7000 audio recordings of Dutch folk songs, called *Onder de groene linde* (Grijp, 2008). These recordings were made from the 1950s till the 1980s, mainly by the field-workers Will Scheepers and Ate Doornbosch. They visited many people across the country with their tape recorders. Since most of the recorded songs were sung from memory, considerable variation can occur between variants of the same tune. A more in-depth discussion of the process of oral transmission and the study of it will be provided in Chapter 2. In a previous project many of these recordings were transcribed into musical notation. In the course of the WITCHCRAFT project, around 2500 of these song transcriptions, along with more than 3000 songs from written sources have been digitized and are thus available for computational processing. Because of its size, composition and quality of metadata, *Onder de groene linde* is in fact a unique resource for studying the mechanism of oral transmission of melody. And, because of the amount of melodies, it would be essential to search and order these automatically, not just using the metadata, but especially the musical content. It was the aim of the WITCHCRAFT project to design and implement methods for processing the musical content of the songs.

In order to construct such a retrieval system for melodies, methods are needed to measure the similarity of (or distance between) melodies. It is this part of the work on which this thesis focuses.

1.2 Computational Musicology

Research of music using computational methods has a history of over half a century. Hewlett & Selfridge-Field (1991) identify three main paradigms of the first decades. The 1960s were characterized by careful planning and design of global solutions to the generic problem of processing music with computers. While the available computational resources were more expensive than human resources, which made bugs and encoding errors expensive, much effort was put in designing proper data structures and methods. Actual data processing and musical results were scarce. One notable achievement was the encoding of all compositions of the Renaissance composer Josquin des Prez on punch-cards (Mendel, 1969). During the seventies the ambitious endeavor of designing universal solutions for computational processing of music was abandoned and the focus shifted towards small-scale projects that were more likely to produce practical results. In the 1980s the computer was brought to the desktop, which gave individual researchers much more freedom to experiment. Musicologists started to use standard software packages like databases and word-processors to process musical data.

During the 1990s David Huron developed the Humdrum toolkit, consisting of a formal representation of music and a collection of UNIX command line tools to analyze musical data (Huron, 1999a). Although it has a broad compass, this toolbox did not gain widespread use within Musicology. Despite all these research activities, computational studies to music were still in the margin of Musicology. In fact, they still are. In a keynote lecture at the Sixth International Conference on Music Information Retrieval, Nicholas Cook considered it necessary to urge Musicology not to miss the opportunities of computational methods (Cook, 2005). Chapter 2 will provide a more in-depth overview of computational research to folk song melodies.

Nicholas Cook was right when pointing to the potential of computational methods for music research. Within Computer Science, many abstract algorithms and data-structures have been developed during the last decades. An algorithm consists of a series of operations on formalized data. Many of these algorithms have potential for application to music. The main challenge is to formulate the musical problem at hand in terms of the algorithm and corresponding data structure. Examples of such data structures are a sequence of symbols, a vector of feature values, a parsing tree, a point set in a geometric space, and so on. The extent to which a direct relation exists between such a data-structure and the properties of music it is supposed to represent, is an important success factor for a computational study of music. Therefore, one of the basic principles of the approach in this thesis is to design computational measures of similarity that are adequate from a musical point of view and into which musical knowledge is incorporated.

1.3 Music Information Retrieval

The rise of the Internet and the world-wide-web, starting in the late 1990s, and the invention of the MP3 audio encoding gave an enormous boost to computational processing of music within the research area of Music Information Retrieval (MIR). The primary aim is not to study music as such, but to design methods to retrieve music from large data-bases using musical 'content' rather than meta-data (Downie, 2003). Typical tasks include genre classification, artist recognition, cover song retrieval, song recommendation, etc.

Concerning the representation of music, there are roughly two kinds of studies within MIR: those using audio data and those using symbolic data. Here, 'audio' means a sampled and quantized sound wave, and 'symbolic' denotes some higher level representation of music in which pitches, durations, instruments, etc. are explicitly encoded. An 'audio' representation is closer to the sound that reaches the ear, while a 'symbolic' representation is closer to how we would notate music.

It proves very difficult to infer adequate musical descriptions from the content of an audio signal. Even in the case of a single instrument, or an unaccompanied singer, detection of such basic properties as the pitch that sounds at a certain moment in time is still problematic. The problem of analyzing more complex signals representing the sound of e.g., an orchestra with many instruments or a church organ is far from being solved. Typical tasks of audio based MIR studies are beat detection, tempo tracking, pitch detection, source separation (e.g., separating the violin from the piano), etc.

For higher level representation of musical knowledge, the problems of audio processing have to be solved. For example, one cannot study harmonic progressions in audio if there is no proper chord detection algorithm available. To undertake such studies, symbolic representations are used. In most cases these have to be entered 'by hand', which has also been done within the WITCHCRAFT project. As a side note, in the case that we are dealing with scores, like in Historical Musicology, the symbolic representation is the most adequate, since it is the most direct representation of the material that is studied.

As results for audio processing get better, we can expect more and more integration of audio and symbolic approaches within MIR.

1.4 Folk Song Research

The research of 'genetic' relations between melodies in western oral culture and the desire to develop a system to classify folk song melodies have a history of over a century. In 1899, the Dutch musicologist Daniel François Scheurleer posed the question: "Welche ist die beste Methode, um Volks- und volkmässige Lieder nach ihrer melodischen (nicht textlichen) Beschaffenheit lexikalisch zu ordnen?".[1] To stimulate response, he organized a competition, which marked the starting point of a long-lasting discussion about classification systems for folksong melodies, an overview of which will be provided in Chapter 2.

Although the attention of Folk Song Research seems to have shifted away from this question in the last decades, the rise of Music Information Retrieval (MIR) has caused a renewed interest in the assessment of similarity relations between folk song melodies. MIR technology provides cultural heritage institutes that host collections of folk song recordings means for improving the accessibility

[1] What is the best method for the lexical ordering of folk and folklike tunes? (Translation by Nettl, 2005, p. 123).

to their archives for general public as well as musicological researchers. The collection *Onder de Groene Linde* is one such collection of folk song recordings.

1.5 Organization of the Thesis

The organization of this thesis is as follows. Chapter 2 presents the relevant academic background for both Folk Song Research and Music Information Retrieval. It also presents research directions and an interdisciplinary collaboration model that actually has a wider relevance than only for the computational modeling of relations between folk song melodies. In this collaboration model, Computational Musicology plays an important intermediate role.

Within Folk Song Research, the concept of *tune family* is used to model the 'genetic' relations between melodies in oral culture (Bayard, 1950). Therefore, to construct computational models of folk song similarity, the concept of tune family has to be 'deconstructed'. As an important step towards the understanding of this concept, an annotation method has been developed, which is presented in Chapter 3. This method has been used to annotate similarity relations within a corpus of 360 folk song melodies in 26 tune families. These annotations allow us to draw several conclusions about the relative importance of various dimensions of perceived melodic similarity. The songs and annotations in the Annotated Corpus are used as reference data in the experiments in the following chapters.

A common approach in pattern recognition is to represent an object by a vector of global feature values. In Chapter 4 we evaluate this approach by measuring a large amount of quantitative features in each of the songs. The relative importance of individual features and of subsets of features for classifying songs into tune families are assessed as well as the scalability: experiments are performed both for the relative small Annotated Corpus and for a large corpus of nearly 5000 songs.

Within the broader context of Computational Musicology, it is interesting to evaluate the suitability of the same approach for an entirely different kind of musicological problem. Therefore, Chapter 5 employs quantitative global features as well, in this case to study baroque fugues. For the specific problem of discerning authorship of organ fugues, quantitative features that are of the same kind as the features that were used in the previous chapter are measured in the scores. The feature values are used to assess the authorship of several disputed organ fugues in the Bach-catalog.

Sequence alignment algorithms can be employed to assess melodic similarity using local comparison of melodies. Chapter 6 presents a way to incorporate musical knowledge into alignment algorithms. A large number of variants of the algorithm are evaluated, as well as tests for scalability. Furthermore, we present the implications of the algorithmic results for the reference data (the Annotated Corpus), by studying the hard-to-classify melodies, and reconsidering their identity.

While in the preceding chapters only symbolic representations of the melodies are used, in the final chapter the audio recordings are involved as well. This chapter continues the path from global to local aspects of music. The aim is to automatically select melodic fragments that are shared exclusively by melodies from the same tune family. The challenge is to cope with melodic variation between melodies. Therefore this chapter is a first step towards employment of approximate recurrent melodic patterns to model the relations between melodies.

Chapter 2

Collaboration Perspectives for Folk Song Research and Music Information Retrieval: the Indispensable Role of Computational Musicology

Since designing computational models for the study and retrieval of folk song melodies demands a multi-disciplinary approach, in which Folk Song Research, Computational Musicology and Music Information Retrieval play a role, the relevant academic background within these disciplines as well as strategies and research directions have to be addressed.

Contribution. There is no recent study that provides an overview of the implications of the current developments in Computational Musicology and Music Information Retrieval for the research questions of Folk Song Research concerning the classification and identification of melodies. This chapter surveys relevant achievements of the disciplines and reveals a gap between Music Information Retrieval and Folk Song Research. To bridge that gap, promising directions for research are provided based on current developments, as well as a collaboration model in which Computational Musicology serves as an intermediate between Folk Song Research and Music Information Retrieval.

2.1 Introduction

Recent developments in Musicology include a growing interest in empirical approaches aiming to enrich traditional qualitative research with data-rich quantitative studies (see e.g., Clarke & Cook, 2004 or Huron, 1999b). This approach can shed new light on the objects of interest of Musicology, such as musical artifacts, their interrelations and their relations to human culture and behavior. This quantitative research is facilitated by recent improvements in computer technology enabling the use of computationally intensive methods. The current chapter explores the promises and problems of a computational approach to the study of folk song melodies.

Folk songs are sung by common people during work or social activities. One of the most important characteristics of these songs is that they are part of oral culture.[1] The melodies and the texts are learned by imitation and participation rather than from written sources such as books. In the course of this oral transmission, changes occur to the melodies, resulting in groups ('tune families') of more or less related melodies.

During the second and third quarter of the twentieth century research on this kind of music flourished in the field of Folk Song Research (FSR). Many folk songs were recorded on tape or transcribed on paper and are thus available for research. Various attempts were made to find structures and patterns in the various folk song corpora. However, after several decades, no strong theories of oral transmission, and no generally applicable classification systems have emerged.

Recent developments in Computational Musicology (CM) and Music Information Retrieval (MIR) have the potential to facilitate and support the research on folk song melodies. CM studies musicological questions with computational methods, while in the field of MIR tools are being developed to unlock large bodies of music. Providing new research methods, these developments stimulate a new interest in the questions of FSR.

There are at least two important reasons for employing MIR technology in FSR. Firstly, the musical models that have to be developed in MIR to process large amounts of folk song data are relevant to the study of the folk song melodies

[1] The definition of the term 'folk song' is not without problems. One of the most stable ingredients in the many attempts to define the concept is the process of oral transmission. "For an item to qualify as folklore it must have been in oral circulation, passing from individual to individual without aid of any written text." (Elbourne, 1975).

themselves and their histories. Secondly, MIR technology is of invaluable importance for the preservation and unlocking of large melody collections. Ethnomusicological archives contain the musical 'memory' of the world. Therefore means for maintaining and accessing these archives are necessary. Currently, many cultural heritage institutions are giving high priority to the digitization and unlocking of their collections, including musical archives.[2] Hence, the development of computational means to do so is an urgent matter.

Although attention has been paid to folk songs in the MIR community, very few studies focus on the particularities of orally transmitted melodies. In most cases folk songs were simply used because they were available as a test collection. Serious attempts to build software for processing folk song melodies should model concepts and methods that were developed in FSR. However, this is not yet standard practice. Major impediments for fruitful collaboration are the unfamiliarity of researchers in both fields with each other's methods and traditions, and the non-formalized nature of many FSR concepts and theories. Therefore we need to find an approach to bridge this gap.

This chapter provides overviews of approaches to the study of folk song melodies in Folk Song Research, Computational Musicology and Music Information Retrieval (sections 2.2 and 2.3), identifies the problems that arise when methods from Computer Science are applied to research questions of FSR (section 2.4), presents concrete directions for research (section 2.5) and describes a collaboration role model (section 2.6). Although the current focus of interest is on folk songs, most of the questions that have to be resolved to conduct a successful computational research project to folk songs do also play a role in wider scopes. Therefore, this chapter can be read as a case study in practicing and reflecting on Computational Musicology.

2.2 Some Past and Current Approaches to Melody in Folk Song Research

Since the late nineteenth century, the availability of collected folk song melodies has generated a considerable amount of musicological research. One of the primary concerns is how to deal with the specific type of melodic variation

[2] Some projects that illustrate this priority are the EthnoArc project (http://www.ethnoarc.org, accessed 1 June 2010), the DISMARC project (http://www.dismarc.org, accessed 1 June 2010) and the WITCHCRAFT project (http://www.cs.uu.nl/research/projects/witchcraft, accessed 1 June 2010).

caused by the process of oral transmission. The basic question is how to model the relationships between melodies from the same folk song culture. Therefore we will first characterize oral transmission, and then classification and identification of melodies in the context of FSR will be briefly discussed.

2.2.1 Oral Transmission and Tune Families

The capabilities of such human faculties as perception, memory, performance and creativity play an important role in the transmission of songs in an oral tradition. Performers have more or less abstract representations of songs in their memories. The only way in which others have access to a song is to listen to a performance. Research into music cognition (Peretz & Zatorre, 2005) shows that the representation of a song in human memory is not 'literal'. During performance the actual appearance of the song is reconstructed or recreated. In the process of transforming the memory representation into audible words and melody, considerable variation may occur. As long as the processes of encoding songs in, and performing songs from human memory are not sufficiently understood, we have to focus mainly on the recorded or transcribed song instances in order to infer knowledge about this kind of variation.

This approach was taken by Walter Wiora (1941), resulting in a comprehensive inventory of types of variation in German folk songs. Wiora summarizes the issue as follows: "Alles an der Beschaffenheit einer Melodie ist veränderlich". He divides the types of changes into seven categories: 1. changes in contour, 2. changes in tonality, 3. changes in rhythm, 4. insertion and deletion of parts, 5. changes in form, 6. changes in expression, and 7. demolition of the melody. He provides many examples for each of these types of change.

The difficulty of understanding the kind of melodic variation caused by oral transmission is clearly stated by Bertrand Bronson (Bronson, 1951, p. 51):

> "When we consider that there is no accessible original to impose its authority; that at every moment in its history such a tune is open to all the gusts of casual influence, subject to forgetful recollection, to individual, or local, or epochal, preferences in mode and rhythm, to willful invention or derangement, to the accidents of marriage with continually altered verbal patterns that impose their own necessities upon melodic statement, and all these operative without any counterbalancing overt external control; we can only marvel at the inner urgency with which folk tunes maintain their essential selfhood in the face of such overwhelming odds."

Together with the remark of Wiora that everything in a melody can change, this citation suggests that the melodic variation caused by oral transmission is a holistic process. Therefore, it cannot be understood by only considering a selective number of melodic features. This is confirmed by another remark of Bronson (1949, p. 169):

> "All who have worked with the problems of variation in a related body of [melodic] materials will readily acknowledge that the question of relatedness involves far more than a mere note-for-note or accent-by-accent correspondence. One very soon comes to realize that this is a problem of the utmost subtlety, in which potentially are included all or most of the elements constituting melodic identity; range, melodic and rhythmic mode, number of phrases, patterns of phrasal combinations, refrain schemes, cadence points, and so on to minuter particulars. It therefore becomes desirable to establish, at least tentatively, the relative weight to be allotted to some of these elements. It may well be that herein, with due discrimination, we shall ultimately find the basis of just distinctions between "families" and more inclusive patterns, or "types" of melody."

For a corpus of British-American folk songs Bronson (1950) obtained weights for some aspects of melodic identity. But he did not test this ordering of importance on corpora from other cultures.

We will now present some attempts from Folk Song Research to better understand the process of melodic change that is caused by oral transmission.

Klusen et al. (1978) conducted an experiment in which melodies were passed orally from one person to another. Their most general finding is in accordance with the result of Wiora: every tone in a melody can change, but some tones are more stable than others. In general, pitch was more varied than rhythm. However, as they indicate themselves, their experiment only tested relatively short-term memory (a few weeks).

The concept of tune family was developed by Samuel Bayard (1950) and defined as: "a group of melodies showing basic interrelation by means of constant melodic correspondence, and presumably owing their mutual likeness to descent from a single air that has assumed multiple forms through processes of variation, imitation, and assimilation." Bayard supposed that the entire body of Anglo-American folk songs consists of forty or fifty such families (Nettl, 2005, p. 116).

After studying traditional music of Ireland, James Cowdery (1984) proposed a "fresh" view on the concept of tune family. He criticizes Bayard's tune family model by posing that folk musicians do not compose new melodies as new instantiations of abstract archetypical airs, but relate new melodies to other concrete melodic material they know. Therefore, Cowdery does not only focus on global similarity but also on motifs that are shared among members of a tune family. Melodies may have some sections in common while other sections differ. He proposes three principles for studying relationship between melodies of which the first one corresponds to Bayard's definition: 1. the "outlining" principle: melodies correspond in their overall contour, 2. the "conjoining" principle: melodies have sections in common, while other sections differ, and 3. the "recombining" principle: melodies are composed from material from the same "pool" of melodic motifs. Thus there is no absolute need for the hypothetical historical sequence of melodies, which in virtually all cases has been lost, if existed at all. Instead, all the melodic material used to analyze melodies and to relate them by means of the tune family concept is concrete melodic material that can be observed and that is meaningful to the folk musicians. A similar approach is taken by Marcello Keller (1988), who explains the relations between Trentino folk music compositions by means of a repertoire of 'segments' that is used in the act of composing.

The approaches presented so far are based on the hypothesis that a historic link between two melodies implies some kind of melodic similarity. To show that this is not necessarily the case, Bruno Nettl (2005, p. 116) describes a way in which a song may change entirely. Starting with the structure ABCD, the first half of the song may be 'dropped', leaving CD, which may be extended with new material resulting in CDEF, which may end up in EFEF by dropping the first part again. In this case there is a historic, but not a melodic link between the first and the last song.

There is no generally accepted theory that explains melodic variation in oral cultures yet. David Rubin (1995) has developed a cognitive theory for oral transmission of texts (like epic poems or counting out rhymes) in which variation is modeled by constraint-based reconstruction of the text from memory. For example, two words with the same meaning and the same metrical characteristics may both be found at corresponding places in a set of variants of a text. The actual appearances of variants may change, but they do obey to the same constraints to a certain extent. A similar process might occur during transmission of melodies. To exploit this for retrieval purposes, the challenge is to model these constraints.

2.2.2 Identification

If two song instances are derived from the same common 'ancestor', they belong to the same tune family (Bayard, 1950) and are considered to be the same song, or, more precisely, manifestations of the same song (Nettl, 2005, p. 114).[3] The identity of a song is a complex and abstract concept. It is not obvious what constitutes the 'substance'—or, in the words of Bronson, the essential selfhood—of a song that is shared among historically derived variants. As a consequence, in folk song classification systems that are based on a limited number of features, historically linked variants may erroneously end up in entirely different classes. The possibility of interference between tune families complicates the issue even further. Because the concept of identity goes beyond individual features of song instances, it is very difficult to develop models that explain tune families.

However, identification of melodies is necessary to address a number of research questions, such as: Where do the individual songs originate from? What were the most popular melodies in a certain time or at a certain place? Which influences from abroad can be traced? How did the melodies develop over time? Because folk song collections contain only a sample of the melodies and variants that have existed, it is impossible to find all variants that are derived from a common 'ancestor' melody and thus to reconstruct the complete history of melodies from the material in the collection. However, in many cases it is feasible to find related groups of melodies within the collection, based on melodic and textual similarity and available meta-data, and to try to link them to tune families as a second step. For this a retrieval system can be an important tool.

Identification of melodies is also important for improving access to cultural heritage. For publications like folk song books, CD-boxes, etc., it is important to know which melodies are in a collection and what their relations and identities are. It is infeasible to find all relations by hand; therefore a computational approach is desirable.

2.2.3 Features for Classification

In FSR, classification systems are used to put melodies in some kind of rational order. These systems are based on such features as the number of lines, the

3 This causes an ambiguity in the term 'song', with which an individual performance can be meant, but also the tune family as a whole.

number of syllables or the cadence note sequence of a song.[4]

The purpose of the most important classification systems in FSR has been twofold. Firstly, classification is desirable for e.g. storing melodies in a card file database or for publishing a book with melodies. In those cases a one-dimensional ordering is required. Such an ordering must provide an easy way to retrieve a melody. Currently, this necessity has been overcome by using digital databases, but this requirement has been important for classification methods that are still in use. Secondly, one of the main aims of these classification systems is to group melodies that are related through the process of oral transmission together (Nettl, 2005, p. 123). Hence, a classification system can be considered a hypothesis of how melodies relate to each other in the process of oral transmission, or as a practical tool to identify melodies. However, in FSR formalized tests of classification systems with respect to their ability to group melodies from the same tune family have not been performed.

Here we give a selective overview of features used in various folk song classification systems. More complete surveys can be found in Elscheková (1966), Keller (1984) and Bohlman (1988).

Most classification systems were developed for specific corpora. One of the first was developed for Finnish songs by Ilmari Krohn (1903). In his system the number of lines and the cadence notes (ending notes of the lines, as depicted in Figure 2.2.1) are most important. Béla Bartók and Zoltán Kodály adapted his system for Hungarian folk songs. In their publications songs were hierarchically ordered by: 1. the number of lines, 2. the sequence of cadence notes, 3. the number of syllables in each line, and 4. the range (Suchoff, 1981, p. xxxiv). In later work, Bartók used another system in which he divided Hungarian songs into three classes, namely old style, new style and mixed style melodies (Suchoff, 1981, p. xlii). Subdivisions were made according to rhythmic characteristics and the number of lines. Obviously, this way of ordering the material is specifically aimed at the corpus of Hungarian songs. As Bruno Nettl (Nettl, 2005, p. 124) points out, Bartók's particular choice of features for classification could only be made by someone already familiar with the corpus for which the system was developed. This applies to folk song classification systems in general (Bohlman, 1988, p. 33).

For the British-American folk song tradition, Bertrand Bronson (1950) ordered

[4] The meaning of the term 'classification' in the context of FSR is not equivalent to the use of the term in Computer Science areas such as Pattern Recognition and Machine Learning. In Computer Science the relation between classes is not necessarily defined, while in FSR the classes are put into some kind of rational order as well, e.g., first all songs consisting of one phrase, then all songs consisting of two phrases, and so on.

Figure 2.2.1: The Dutch song *Dat gaat naar Den Bosch toe* as notated in Brandts Buys (1975, p. 81). For the first phrase, accented notes on various levels are marked.

a list of features according to importance using a punch card system: 1. final cadence, 2. mid cadence, 3. first accented note, 4. first phrase cadence, 5. first accented note of second phrase, 6. penultimate stress of second phrase, etc. Thus, a classification system based upon these features could be expected to group songs in the same tune family together to some extent.

In an important publication of The German Archive of Folk Song (Deutsches Volksliedarchiv) containing German ballads (Suppan & Stief, 1976), an ordering is used that reflects the system of Krohn. The first criterion is the number of lines and the second criterion is the cadence note sequence. If the editors judged a resulting group incoherent, subgroups were made. The exact way in which these subgroups were established, is not accounted for.

2.3 Computational Approaches to Folk Song Melodies

In this section we survey research results from Computational Musicology and Music Information Retrieval concerning the study of folk song melodies, as well as melody search engines that have large numbers of folk song melodies in their databases.

2.3.1 Folk Songs in Computational Musicology

For the overview in this section the field of Computational Musicology is taken in its broadest sense: any research on melodies that makes use of computational methods in one way or another.

2.3.1.1 The Early Days

Ordering melodies according to some specific criteria is to a certain extent a 'mechanical' activity. Therefore it is not surprising that the use of computer systems was considered soon after they became available. As early as in 1949, Bertrand H. Bronson proposed a method to represent folk songs on punch cards. Thus songs with certain desired characteristics could be retrieved using a sorting machine (Bronson, 1959). In the following decades many studies on the use of computers in folklore and folk music were published. A bibliography from 1979 on this topic lists more than 100 references (Stein, 1979).

Starting in the 1980s, an enormous digitization project was carried out under supervision of Helmuth Schaffrath. He developed The Essen Associative Code (EsAC), a music encoding for monophonic folk songs. A detailed description is provided in Schaffrath, 1997. After one and a half decade, more than 14,000 songs were digitized. The collection is still being used as test collection for melody analysis or melody retrieval. Along with the collection analytical software was implemented, which could extract numerous features from the songs, such as distributions of intervals and durations, rhythmic patterns, cadence tone sequences, pitch contour, etc. Schaffrath (1992) shows how to use these features to retrieve melodies from a database of folk song melodies.

In the 1980s Wolfram Steinbeck (1982) and Barbara Jesser (1991) did research into computer aided analysis of monophonic music. Both used a subset of the Essen folk song collection. Jesser evaluated statistics of features such as interval frequencies, duration frequencies, range, accent and cadence tone sequences and other features. In a number of example tune families, she showed that in each family common characteristics could be found, but that it was not the case that for all families the same features are important.

Wolfram Steinbeck (1982) focused primary on clustering. He used hierarchical clustering algorithms to group melodies from the Essen collection together using 13 features such as mean and standard deviation of pitches, range, size of intervals, number of direction changes, and others (Steinbeck, 1982, p. 275). With a set of 35 melodies he was able to cluster melodies into meaningful groups, such as hymns, children's songs and hunting songs. An experiment with 500 melodies also led to clusters, but in this case the clusters were more difficult to interpret musically.

2.3.1.2 Contour-Based Approaches

There are some contour-based approaches. Using tools from his Humdrum Toolkit (Huron, 1999a), David Huron (1995) confirmed the hypothesis that folk song melodies tend to show arch-like contours in single lines as well as in successions of lines. His testing material consisted of melodies from the Essen collection.

Zoltán Juhász (2000; 2002; 2004; 2006) published several studies using a large collection of digitized Hungarian folk song melodies. In his approach, a melody is represented as a contour vector, which is constructed by sampling the pitch at equal distances in time. If the sampling frequency is high enough, all details of the contour are preserved in the resulting vector. Juhász performed a principal component analysis of the contour vectors of melodic phrases.[5] The principal components can be interpreted as contours themselves. In the space spanned by the first few principal components, clusters of melodies can be found that share the same contour characteristics, namely, clusters of phrases with the same beginning and ending notes (2002). He also compared the Hungarian melodies with songs from other countries in Europe by training self organizing maps with the contour vectors (2006). By evaluating the extent to which the map of one culture is excited by contour vectors from another culture, one can evaluate the extent to which the cultures share contour types. It appears that all involved musical cultures have some contour types in common. This caused Juhász to speculate about a "common language" that reflects an archaic common origin of these European traditions.

2.3.1.3 Segmentation

Some segmentation algorithms that have been developed within CM were aimed for or tested on folk song melodies. Zoltán Juhász (2004) used the entropy of the continuation of a sequence of intervals. A high value of this conditional entropy implies that the next interval is hard to predict, which may indicate a segment boundary. This approach can be used for segmentation.

5 Juhász uses the term 'section' rather than 'phrase' or 'line'. This probably finds its origin in the terminology used by Béla Bartók.

Another data-driven approach to segmentation is presented by Rens Bod (2002). The segmentation is performed by a parser following rewrite rules that are inferred from a training set, in this case a part of the Essen collection. The rewrite rule with the highest probability is followed. He was able to reproduce 81% of the line breaks that were encoded in the Essen collection.

2.3.1.4 Other Approaches

Darrel Conklin and Christina Anagnostopoulou (2001) used the concept of viewpoint. This is a formalized way to represent a feature of a melody, by means of a sequence of symbols. Using a contour viewpoint and association rule mining Conklin (2006) could confirm the finding of Huron (1995) that melodies tend to have an arch shape. In combination with suffix trees, viewpoints can also be employed to find repeating patterns.

Bret Aarden and David Huron (2001) proposed the use of geographical information. Thus geographical variance of some features could be visualized by showing densities on a map.

The studies described show various valuable approaches to the processing of folk song melodies. With exception of those by Zoltán Juhász, none of these studies explicitly state an interest in folk songs as part of oral tradition. The particular questions of FSR, such as the understanding and modeling of tune families are barely addressed. Hence, we can state that the current available results from CS do not get us much further in addressing the problems of FSR.

2.3.2 Folk Songs in Music Information Retrieval

Folk song melodies have been used in a considerable number of MIR studies. However, in many cases folk songs were chosen because of their availability and not because of an interest in folk music as such. This applies to all 12 papers in the complete ISMIR proceedings from 2001–2009 that employ the Essen folk song collection (Schaffrath, 1995).[6] In none of these papers the implications of the choice for this data set are discussed. In most cases it is just stated that the collection is used, or a pragmatic reason is provided, e.g., the need for a large music database, or the need for a collection of monophonic songs. The

[6] Eck & Casagrande, 2005; Eerola & Toiviainen, 2004; Frieler, 2007; Hoos et al., 2001; Pearce et al., 2008; Pickens, 2000; Sapp, 2005; Singer, 2004; Skalak et al., 2008; Temperley, 2006; Toiviainen & Eerola, 2005; Ferraro et al., 2009

results of the more general questions addressed, such as meter classification, benchmark establishing or segmentation, have not been interpreted concerning their potential to contribute to folk song research.

2.3.2.1 Online Search Engines

Some online search engines allow the user to search in a large collection of folk song melodies.

The database of the **Danish Folklore Archives** contains about 10,000 instrumental melodies found in books and manuscripts with Danish folk music.[7] The collection can be queried on musical content with two kinds of accent note patterns: a so-called "incipit note sequence" (notes on the beats) and an "accent note sequence" (notes on the first beat of each bar), corresponding to the beat and the strong accents levels in Figure 2.2.1. String matching is used to evaluate the similarity of the query with the melodies in the database, with the number of permitted errors as parameter. In the explanation on the web site it is stated that the search for accent and incipit patterns in practice has proved to be the most reliable way to search a collection of melodic variants.[8] Apparently, these are supposed to be stable elements in Danish tunes by the creators of the search engine. However, a full account of the rationale behind the choice for accent and incipit note sequences is not provided.

The **Colonial Music Institute**, which promotes research in early American music and dance, offers an index for about 75,000 instrumental and vocal pieces from the period 1589–1839 (sic), including social dance tunes and songs.[9] From each melody an incipit is present in the database. There are three ways to browse these incipits: a sequence of scale degrees of all notes, a sequence of scale degrees of stressed notes, and a sequence of intervals.

Another online searchable database with folk song melodies is the **Digital Archive of Finnish Folk Tunes**.[10] The collection contains about 9,000 melodies, most of them collected in the early 20th century by Ilmari Krohn. Two types of melodic query can be used to search: gross contour (Parsons code: "`u`" (up), "`d`" (down), and "`r`" (repeat)) and a sequence of intervals in semi-

[7] http://www.dafos.dk/melodies-online.aspx (accessed 1 June 2010).

[8] http://www.dafos.dk/melodies-online/melody-codes-and-code-searching.aspx (accessed 1 June 2010).

[9] http://www.colonialdancing.org/Easmes (accessed 1 June 2010).

[10] http://esavelmat.jyu.fi (accessed 1 June 2010).

tone distance. Wildcards may be used to allow a sub-pattern search. A string matching algorithm is used for matching.

There are also some search engines that are not specifically aimed at unlocking folk song collections, but nevertheless contain a large number of folk song melodies. **Themefinder** can be used to search a large collection of themes and incipits from classical works, folk songs (the Essen collection; Schaffrath, 1995) and 16th-century Latin motets.[11] Several representations of a query melody can be used: pitch sequence (e.g., "`G G G E-`"); interval sequence (e.g., "`P1 P1 -M3`"); scale degree (e.g., "`5 5 5 3`"); Gross Contour (e.g., "`ssd`", in which "`s`" means 'same'); Refined Contour, in which steps and leaps are distinguished (e.g., "`ssD`"); Key and Meter. The user can choose whether the query should occur at the beginning, or anywhere in the theme. Internally the Humdrum Toolkit (Huron, 1999a) is used to perform the search.

The melody index **MELDEX** has been developed in the context of The New Zealand Digital Library project at the University of Waikato (McNab et al., 1997).[12] The major part of folk songs in the Meldex database consists of the Essen folk song collection together with a collection of songs from the Digital Tradition Folk-Song Database. Both text and melody queries are possible. A query melody can be generated by clicking keys on a virtual keyboard, or by whistling or humming in a microphone. An approximate string matching algorithm that is based on the alignment algorithm by Mongeau & Sankoff (1990) is used for matching (McNab et al., 1997). Unlike the previous discussed search engines, Meldex allows approximate matches without using wildcards or specifying the permitted number of errors. The user can choose whether the interval sequence or the contour of the query melody should be used. The inclusion of rhythm information in the search is optional and is switched off by default.

The database of the "Open Music Encyclopedia" **Musipedia** contains a large number of folk song melodies. Various input methods can be used to create a query: Lilypond code, contour (Parsons code), humming/whistling, or tapping. For matching, transportations distances (Typke et al., 2003) as well as string matching (for the Parsons code) are used. Among the described search engines, this is the only one in which rhythm is fully involved by default.

Only the Danish search engine and the index of the Colonial Music Institute have query methods that are explicitly motivated by knowledge or hypotheses from FSR. One can search for a sequence of accented notes, which are assumed to be more stable across variants of a tune than unaccented notes.

[11] http://www.themefinder.org (accessed 1 June 2010).

[12] http://www.nzdl.org/musiclib (accessed 1 June 2010).

2.3.2.2 An Example

As an example to show what can be accomplished by using the search engines described in the previous section, we will take a folk song melody as a query and discuss the search results.

In the Database of Dutch Songs,[13] the melody of the Dutch folk song *Dat gaat naar Den Bosch toe* as shown in Figure 2.2.1 has been identified as the song "Contre les chagrins de la vie" from the opera *Le petit matelot* by Pierre Gaveaux (1786). The first two phrases of Gaveaux' song are shown as melodies G1 and G2 in Figure 2.3.1. As indicated in Figure 2.2.1, the Dutch melody contains two different sections, A and B. Thus, the structure of the song is ABABA. This melody is our query. The search results are summarized in Figure 2.3.1.

For the search engine of the Danish Folklore Archives, we have to transform the melody into an incipit or accent note sequence. The incipit note sequence is "`13516665`". These numbers are the scale levels of the notes on the first eight beats. Querying the database with the permission of one error results in one hit, the air *Om al Verden er* as notated in a nineteenth century book with airs and dances that was property of Hans Jensen Hansen. This melody is labeled R1 in Figure 2.3.1. The B-section of this melody, which is not shown in the figure, is similar to the B-section of the Dutch song. As can be seen in the figure, this melody is encoded in 2/4 meter. Thus the accent note sequence of the 4/4 query, which is defined as the sequence of notes on the first beat of each bar, is not compatible with the accent note sequence of melody R1 in Figure 2.3.1. This complicates searching for accent note sequences. If we construct an accent note sequence of the query as if it were notated in 2/4 meter and permit one error, melody R1 is in the result list at rank three, after two unrelated hits.

The search engine of the Digital Archive of Finnish Folk Songs can be queried with an interval sequence representing each interval in half tone steps. For our query melody this would be "`+2+2+1+2+5−3`". Searching for this sequence results in melodies R2, without title, and R3, entitled *Nelosta*, at ranks 1 and 2. In both cases, the B-section, which is not shown, differs clearly from the query melody.

For Themefinder we use the scale degrees of the incipit as query: "`1 2 3 4 5 1 6`". Searching in all collections available results in two relevant hits at ranks 3 and 5, shown in Figure 2.3.1 as melodies R4 and R5. These are the songs *Loot ons noch ens drinken* and *Ueber die Beschwerden dieses Lebens* from the Essen collection. In melody R4, the B-section is absent, but melody R5 has

[13] http://www.liederenbank.nl (accessed 1 June 2010).

Figure 2.3.1: First lines of the search results. **Q** is the query. **R1** is found by the Danish engine, **R2** and **R3** by the Finnish engine, **R4** and **R5** both by Themefinder and Meldex, **R6** and **R7** by Musipedia, and **R8** by YahMuugle. All melodies are transposed to G major. The titles are: **Q** *Dat gaat naar Den Bosch toe*, **R1** *air Om al Verden er*, **R2** [without title], **R3** *Nelosta*, **R4** *Loot ons noch ens drinken*, **R5** *Ueber die Beschwerden dieses Lebens*, **R6** *Scottisch Simple de Guemene*, **R7** *I'm a little tea pot*, **R8** *Variations* by Aloys Schmitt. Two phrases from the original composition by Pierre Gaveaux, **G1** and **G2**, are added for comparison.

a B-section that resembles the B-section of the Dutch song. The title indicates that R5 is a German translation of the original French song.

Since MELDEX also searches the Essen collection, we expect the same results. Querying with the note sequence "`g a b c′ d′ g′ e′`" indeed results in these two hits at ranks 7 and 11. The other hits can be considered false positives.

When we search in the Musipedia database with the sequence "`g′4 a′4 b′4 c″4 d″2 g″2 e″4`", the first hit is a song entitled *Scottish Simple de Guemene*, which is melody R6 in Figure 2.3.1. This melody is quite distant from the query. The first line ends with a half cadence and the continuation is dissimilar. Musipedia can also "search the web". With the same query we find a nursery song called *I'm a little tea pot*, melody R7 in Figure 2.3.1. There is no B-section. The second half of the melody is like the A-section, but with a different ending.

At last, we use YahMuugle, a search engine that is not designed for folk song melodies, but searches in a collection of 476,000 RISM incipits.[14] These are incipits of classical compositions that can be found in manuscripts written before c. 1800. The transportation distance algorithm described in Typke et al., 2003 is used. The query can be created by clicking keys on a graphical keyboard. With the sequence "`g′4 a′4 b′4 c″4 d″2 g″2 e″4`" played on the keyboard as query, we find at rank 13 melody R8 in Figure 2.3.1. According to the meta data in the result list, this is a composition by Aloys Schmitt (1788–1866) called *Variations*. Since we only get an incipit, it is not possible to compare the continuation without looking up the piece in the source manuscript.

A text-based search in the Database of Dutch Songs shows that in the Dutch language at least 30 texts exist for this melody. This finding and the results for the melody searches show that this melody must have been well known in many European countries.

Some questions concerning these search results remain to be answered. The 'skeleton' of the melody is rather generic. It consists of an embellished ascending and descending movement, which is a common contour for folk song melodies (Huron, 1995). Therefore, one could argue that some of the results in Figure 2.3.1 could have been created independently instead of being derived from one original melody. A distinctive feature of this particular melody could be the quick stepwise ascension followed by a leap to the octave and a leap back to the sixth at the very beginning. The cases where a similar B-section is present (R1 and R5) are most certainly connected. For R5 the correspondence

[14] http://www.yahmuugle.cs.uu.nl (accessed 1 June 2010).

of the German and French titles is a very strong indication. Dropping the B-section is a likely simplification of the melody. Hence, the melodies without a B-section (R4 and R7) might also have historical links with the query melody. In case of a different B-section (R2, R3 and R6), we have the least certainty of a historical link.

This extensive example shows that the currently available search engines are helpful tools for studying the history and dissemination of a tune. Although this is a quite successful example, one cannot state that the current available search functionality is sufficient for research to any folk song tune. The important feature that all these variants share is the characteristic beginning. The search engines are able to match melodies using that feature. But folk songs may have variants that are similar in other aspects, e.g., a second or third phrase may be shared while the incipits differ, or the similarity may be based upon shared melodic motifs or patterns while the global features are not particularly similar. For better results, knowledge about the process of oral transmission has to be incorporated.

2.4 Problems and Challenges

Based on the previous sections we identify a number of issues that have to be addressed when taking a computational approach to study the questions of FSR.

Problem 1. There is no generally accepted theory of oral transmission of melodies. This is related to the lack of proper understanding of cognitive musical processes such as encoding songs in human memory, performing songs from memory and creating new songs. Knowledge of these processes is an important ingredient for a theory of oral transmission of melodies.

Problem 2. The diversity of classification systems (see section 2.2.3) indicates that no universally applicable system of ordering and classifying folk song melodies exists yet. There is no theory that for any collection can predict which features of melody are discriminative for tune families. Most of the existing systems have been designed for specific corpora. Once a researcher is familiar with a certain corpus, he might be able to determine some set of features that is expected to be discriminative for that corpus. This understanding of melodic variation among tunes within specific musical cultures has been obtained by studying the melodies, such as the Hungarian by Béla Bartók (1981), the Irish by James Cowdery (1990), German by Suppan and Stief (1976), Western European by Walter Wiora (1941), or the British-American by Bertrand Bronson (1959).

Problem 3. Most models and concepts from FSR are not directly implementable. For implementing a model, a very precise and unambiguous description of data structures and algorithms is required. In most cases this kind of precision is not available. As Leonard Meyer states in a more general context (1996, p. 64): "[. . .] I have no doubt about the value of employing computers in such studies [on musical style], not merely because they can save enormous amounts of time but, equally important, because their use will force us [music scholars] to define terms and traits, classes and relationships with precision—something most of us seldom do."

This remark implies that for the purpose of implementation of folk song retrieval system—or music retrieval systems in general—, new models have to be developed that are explicit in the way they evaluate similarity between melodies. In section 2.6, a general strategy for this will be presented. In the subsequent chapters of this thesis, several models will be presented and evaluated.

Problem 4. Formal testing of classification systems has not been done to a great extent. It is hard to determine whether a certain system 'works' concerning its ability to group melodies that are in the same tune family. The question is to test against what? Since the historic relationships between melodies are untraceable, it is difficult to assemble 'ground truth' data. Not only classification systems, but also theoretic overviews such as the "Systematik" of Wiora (1941) have not been tested. He provides many examples, but no formal test that is able to convince us that his overview of melodic changes lists the changes that occur in the transmission of songs indeed.

In this thesis, standard evaluation methods from pattern recognition (Chapters 4 and 5) and from information retrieval (Chapters 4 and 6) are used to test the various melodic similarity measures.

2.5 Directions for Research

Based on the overviews in the previous sections, we now provide an overview of directions for future research. This section describes research questions, while the next section addresses the question what strategies to pursue.

2.5.1 Research Question

The basic research question can be stated as follows: Given the availability of a symbolically encoded corpus of folk song melodies, given oral variation as characteristic feature of this corpus, given the already established approaches to categorize melodies from oral culture (Section 2.2.3), and given the computational studies of folk songs already undertaken (Section 2.3.1), how to design implementable models of folk song melodies that describe the interrelationships caused by the process of oral transmission? The research questions that are relevant for this task are of interest for the MIR community, since such models could be exploited in search engines, but also for FSR, for Cognitive Musicology, and for Musicology in general.

2.5.2 Cognitive Approaches

The underlying problem from a cognitive point of view is: how is a melody encoded in human memory and how is it transformed into an audible song instance during performance? This knowledge can be used to discriminate between stable and unstable elements of melodies in oral transmission. Although the understanding of cognitive processes is important, the study of these is outside the scope of the current thesis.

As we saw, folk song classification systems are to a certain extent based upon musical intuition of the researcher. Since musical intuition cannot be implemented, a proper understanding of it is needed to develop an implementable model. Therefore, it is necessary to study this musical intuition by e.g., trying to find patterns in human descriptions of musical similarity. This will be done in chapter 3, in which annotations about melodic similarity provided by domain experts are used to infer knowledge about the various ways in which related melodies are perceived as similar to each other.

The constraint-based approach proposed at the end of section 2.2.1, which will not be further elaborated in this thesis, is promising for revealing the variability in memory for melodies. The challenge is to infer the constraints that led to a particular melody or a particular group of melodies from the corpus. The set of constraints thus found shows the invariable aspects of the melodies. The aspects of the melodies that are not determined by these constraints show what can vary among the instances of a certain tune family. This can lead to hypotheses about the characteristics of human melodic memory, which may be tested using other kinds of experiments.

An interesting property of folk song corpora from the perspective of Cognitive Musicology is that the melodies are produced by common people with normal musical skills. At least, this is the case for the collection of Dutch songs that is kept by the Meertens Institute. Regardless the quality of the melodies from an artistic point of view, all these melodies are products of some human activity of musical performance. As pointed out by Isabelle Peretz (2006, section 2), most people without a formal musical training share a “common core of musical knowledge”. Although vital for understanding the nature of music, the study of this common knowledge has been underrated for a long time in cognitive research. Hypotheses about common musical skills might be tested using folk song material.

2.5.3 The Perspective of Folk Song Research

Within FSR, the research on tune families and genetic relationships of melodies seems to have been marginalized during the last two decades (Nettl, 2005, p. 130). We expect that the use of new computational methods to explore and unlock collections of melodies will result in renewed interest in this topic. The enormous increase in computational power enables the development of new kinds of algorithms that incorporate more musical knowledge and that are allowed to be computationally more demanding than the ones developed during the third quarter of the twentieth century. This enables new ways to explore the contents of the many archives of folk music. This also leads to new ways to compare different corpora of melodies from different oral traditions with each other. An example of such a comparison can be found in the work of Zoltán Juhász that has been mentioned in section 2.3.1.2. He compares the properties of the contours of songs from various folk song traditions, which leads to hypotheses about the historical relationship of these traditions.

2.5.4 Music Information Retrieval

The main question from the perspective of Music Information Retrieval in the context of this thesis is how to design computational similarity measures that can efficiently be used to retrieve melodies from a large collection.

Gained insights into the relations between folk song melodies can be used in a more general scope of MIR tasks. Knowledge about the relation between a desired melody and the way this melody is sung from memory can be helpful in processing melodic queries for a Query By Humming system. This knowledge helps to identify the most persistent aspects of a melody and it can make recognition of a melody more robust.

A common interest of FSR and MIR is the access to collections of recordings or transcriptions of folk songs. Therefore the question to be answered is: what searching or browsing functionality is needed to get access to the melodic content of a corpus of folk song melodies? More basically: what are the desired ways to get access to such a corpus from a user's point of view? How do these ways of access differ for various user groups, such as professional folk song researchers, historians, the general public, etc?

For MIR in general, the question to the needs of potential users is important. To avoid mismatches between the provided tools and the demands of various groups of users, it is necessary to get a realistic overview of possible applications of MIR technology. For the particular case of Musicology, or Folk Song Research, the availability of useful tools might also stimulate the acceptance of computational approaches as part of the generally available research methods.

Since this thesis focuses on the modeling of similarity between folk song melodies rather than the design of an end-user retrieval system, user modeling will not be discussed further.

2.5.5 Repeating Patterns and Stylistic Studies

The research on 'genetic' relationships of melodies can be considered an instance of the more general musicological problem of relating instances of music to each other. In the current context the relations occur on the level of melodic contents. One approach to relate pieces of music has been proposed by Leonard Meyer (1996). His theory of musical style is founded on the replication of musical patterns. These patterns could occur on various levels of abstraction such as actual note sequences, harmonic progressions, the structure of musical compositions, etc. Instances of music that share patterns are stylistically related. Meyer proposes a hierarchy of relations, reaching from intra-opus style, e.g., the particular style of a certain symphony, via inter-opus style, e.g., the style of the oeuvre of a certain composer, to the level of the great style periods, e.g., all Baroque works. For folk songs a similar hierarchy could be conceived: at the lowest level the style of a certain song instance, then the style of a tune family, the style of all songs from a certain oral tradition, and maybe on the highest level all singable melodies. On each level the style could be characterized with the patterns that replicate in all stylistically related instances on that level. Since Meyer's approach is based on frequencies of pattern occurrences, it offers possibilities for employment in a quantitative, computational approach.

Some recent computational approaches focus on repeating patterns in melody as well, such as the theories of Ahlbäck (2004), Lartillot (2004), or the viewpoint approach used by Conklin & Anagnostopoulou (2001). Also in Ethnomusicology there is interest in repeated patterns. Bruno Nettl (2005, p. 118) raises the question what are the basic units that are transmitted in an oral tradition. In his view these are musical motifs that in various recombinations form different songs. This is in accordance with the new conception of 'tune family' by Cowdery (1984), which focuses on motifs rather than solely on entire melodies. All together the detection of various kinds of repeating patterns is a promising approach to study the musical contents of a collection of melodies. The work presented in Chapter 7 of this thesis, is partly motivated by this approach. It studies whether tune families can be recognized by recurring, characteristic melodic patterns.

2.5.6 Global Features of Melody

A common approach in machine learning and pattern recognition studies to classify objects is to measure quantifiable features of the objects resulting in one feature vector for each object. With these feature vectors various kinds of classifiers can be trained, or clustering algorithms can be used to discover clusters of objects. This approach was taken by Steinbeck (1982) and Jesser (1991) for melodies in the Essen collection. They both defined a set consisting of features they considered relevant for the problem of classifying folk songs melodies. This approach can be extended with other melody feature sets that were not specifically designed for folk song research, such as e.g., the features that are defined by McKay (2004). In chapter 4 of this thesis, the discriminative power of single features and subsets of features from a set of 88 global features is studied.

2.5.7 Sequence Alignment

To compare two melodies, it is insightful to notate the one below the other such that the corresponding parts are aligned. Therefore, alignments of two or more songs are heavily used by folk song researchers. The overview of results in Figure 2.3.1 is an example of such an alignment. In Computer Science, algorithms that construct alignments automatically were developed some decades ago. Because these algorithms have found an application field in Molecular Biology, where they are used to find corresponding patterns in protein or nucleotide sequences, it is in that discipline that many algorithms, improvements and optimizations were developed, which have the potential to be employed for musicological research as well. For each alignment a score can be computed. The higher the score, the better the sequences could be aligned. In the computation of the score, domain knowledge can be incorporated. For aligning folk songs, musical scoring schemes are presented and evaluated in chapter 6.

2.5.8 Evaluation and Testing

Finally, an important question is how to evaluate a search engine for folk song melodies. It is usually done by manually defining an ideal ranked result list for a query and comparing the results of the algorithm to that. However, this assumes that it is possible to construct such a list. In practice, this task is to a great extent subjective and based on implicit musical skills. Therefore it is not possible to create an ideal result list that can be considered *the* result list for a particular query and that can serve as 'ground truth'. A general strategy for dealing with this problem in the context of interdisciplinary research will be presented in the next section.

2.6 Collaboration and Integration

The goal of developing useful software for folk song research and retrieval cannot be achieved without a profound collaboration between FSR, CM and MIR. However, the research in FSR summarized in the previous sections and the methods developed in CM and MIR to retrieve and to study folk song melodies do not indicate the actuality of such a collaboration. Since MIR and FSR have the least overlap, in this section we focus on strategies to achieve better collaboration between these two disciplines and we characterize the role of CM as mediator. The significance of such a collaboration model goes beyond the study of folk song melodies. Profound collaboration seems to be absent too in the relation between Musicology and MIR in general. Although both deal with music, there seems to be a gap in the ways of understanding it. In our opinion both disciplines suffer from this lack of mutual influence.

Characterizing the gap in an extreme way, we have on the one hand folk song researchers who lack a fundamental understanding of the possibilities and limitations of computational approaches, and on the other hand MIR researchers who do not have a professional musical knowledge framework, which causes a limited view on music and the way music functions in culture.

The existence of this gap and the focus on technical solutions prevents MIR often from being more than marginally relevant to FSR (or to Musicology in general), as for instance the problematic notion of ‘ground truth’ demonstrates. Sometimes it seems like MIR has a stock of so-called ‘experts’ from which truths can be drawn. Once provided by the expert, MIR does not go beyond this ground truth, thus making it a hermetic boundary between MIR and Musicology. The relevance to Musicology is determined by the answer to the question whether algorithms are being developed merely to reproduce a given ‘ground truth’, or to evaluate the theories that are behind that ‘ground truth’. The first option seems most common, while choosing the second option will obviously lead to a better understanding of music, which in turn will lead to better approaches for music retrieval.

In a recent, related paper about Computational Ethnomusicology, Tzanetakis et al. (2007) observe that in MIR existing computational techniques are frequently blindly applied to musical problems, without a clear musicological goal. Therefore their first guideline for Computational Ethnomusicology is to seek active collaboration with music scholars. “Experimental results should generally be interpreted by music scholars with an understanding of the specific music(s) involved”. They illustrate the guideline with some examples. Along the same lines, Cornelis et al. (2010) plead for a better collaboration between MIR and Ethnomusicology, which would result in more cultural independent tools for content-based MIR. In the remainder of this section we will present a more abstract model for collaboration between the involved disciplines.

Before any useful software can be developed for folk song melodies, implementable models of FSR concepts are needed. As Willard McCarty (2005, chapter 1) states in a more general discussion about the relation between Computer Science and the Humanities, the process of modeling itself is more important than the resulting models, because it is in this process that knowledge is generated about the concepts to be modeled. Therefore, the way a model fails is more interesting than the way a model succeeds, because there lies an opportunity to improve understanding. In our case, one of the most important concepts to model is tune family.

Although the modeling is more important than the models, implementations are needed for testing and for applications such as search engines. This leads to a chain of activities that will iteratively be repeated. First, the process of creating or improving the model. Second, implementing the (adapted) model.

Figure 2.6.1: Three-role model for integration.

Third, the evaluation of the implemented model, leading to improvements of the model. These activities will alternate in an iterative research process. The fourth activity is the deployment of the state-of-the-art model. This is done whenever the developed model is needed for other purposes than improving the model, such as implementing a search engine.

In the context of research to folk song melodies, we now present a possible way to overcome the observed 'gap' with the help of the three-role model that is shown in Figure 2.6.1. This model aims to identify the actors in the research process. In addition to the roles of MIR and FSR researchers, a 'man in the middle' role is needed. This role can be fulfilled by Computational Musicology. Since this is a role model, it does not necessarily imply the need for three separate persons in research teams. In exceptional cases one person might combine all three roles, but it would be more common for researchers either to combine both the MIR and CM or (probably more rarely) the FSR and CM roles. We now briefly characterize the three roles.

Folk Song Research. The data of Folk Song Research, such as recordings, transcriptions, notes, etc., are gathered during ethnomusicological fieldwork. Methods to process these data are necessary to support research on the relations between those artifacts and the cultural context they were obtained from. It is among these methods that classification and identification as described in section 2.2 can be found. For these methods a computational approach can be taken.

Since ambiguous or intuitive concepts are difficult to implement, the task for FSR is to be as precise as possible in defining concepts such that they are suited for computational modeling. Therefore, the collaboration with CM and MIR will lead to a better understanding of FSR concepts. FSR should also be involved in the evaluation of implemented models and in the process of improving these models.

In the case that the roles are not shared by the same researcher, another more practical effort is expected from folk song researchers, namely to take some time to learn how to use provided computational systems. A general understanding of the possibilities and limitations of computational methods will result in the understanding that these methods will not replace currently used methods, but open new perspectives to explore the data and to evaluate the usefulness of FSR concepts.

Music Information Retrieval. MIR designs and implements musical information systems, with the aim to improve the access to large (distributed) collections of music. Therefore, systems are generally evaluated in terms of retrieval or classification accuracy rather than by the resulting insight in the musical content. Nevertheless, as far as representations of music are used that are employable for FSR, MIR can provide numerous useful software components and user interface components that can support Folk Song Research. In our opinion MIR should employ models of music that are meaningful for FSR, and for Musicology in general. A better understanding of music will lead to more robust and more flexible music retrieval systems.

Since MIR is interested in handling large amounts of musical data, efficiency is an important constraint. Therefore, MIR could provide strategies for handling huge databases. Since CM is not primarily interested in performance issues, collaboration of MIR and CM is necessary to develop optimized algorithms that can be applied on a large scale.

Computational Musicology. In the context of this thesis, the task for the CM-role is to intermediate between MIR and FSR. The subject of interest is the same as for FSR, namely songs and their relationships. The methods, however, are from Computer Science rather than from the tradition of FSR itself. These include algorithms, data structures, evaluation strategies, etc. For the activity of modeling, the task is to 'deconstruct' FSR-concepts in order to derive implementable models (arrows 1 and 2). After the first iteration these models can be improved by providing FSR the implemented models and letting FSR examine the way in which the previous models fail (arrows 3 and 4). Another, more practical, task for CM is to provide FSR with ready-to-use software frameworks and toolboxes, which allow the combining of input, processing and output methods in various ways (Garbers, 2004). These toolboxes could consist of basic melodic transformations, feature extractors, segmentation algorithms, distance measures, clustering algorithms, classification methods, visualization tools, etc. that are relevant for evaluating musicological concepts. CM should hide implementation details that have no meaning in the musical domain. An example of such a toolbox is the Humdrum toolbox (Huron, 1999a). However, the use of this toolbox requires a level of mastering the Unix command line environment that most musicologists do not have. The search engine Themefinder is an example of a way to provide the functionality of some Humdrum tools to users that do not have the skills to handle Humdrum directly. CM could provide interfaces like this for specific tasks.

Here FSR, MIR and CM are presented as roles rather than as disciplines to stress that it is not desirable to create a need for special CM researchers to whom FSR and MIR researchers have to obey. In practice, probably all researchers within the MIR community play the role of CM to a certain extent because of their interest in music. Obviously, the quality of musical models used in music information systems affects their successfulness to a large extent. Therefore, we believe that the MIR community can gain much from pursuing the CM-role more ambitiously.

2.7 Concluding Remarks

The aim of FSR to identify melodies (i.e., to determine the tune family they belong to) seems currently too ambitious to perform automatically (see section 2.2.2), since no proper implementable model of the tune family concept is within reach, if only because of the fact that the concept is problematic within FSR itself (Cowdery, 1984). Therefore, the short-term goal for this thesis is to design computational models that support identification by finding related melodies based on evaluation of melodic similarity, leaving the decision to consider a melody as member of a specific tune family to the musicological investigator. So, for CM, in the short term, the focus has to be on designing models of melodic similarity that can be employed for classification of folk song melodies. In the long term, results of that could be used to work towards a model of oral transmission. Thus CM contributes directly to the basic questions of FSR.

From the classification approaches in section 2.2.3, we can obtain a number of features that were considered relevant by musicological researchers, such as cadence and accent note patterns, number of lines, and rhythmic characteristics. However, it will not be sufficient just to implement the models of e.g. Bartók or Bronson, since their feature sets were not assembled with the power of computational methods in mind, and they were fitted to specific corpora. The possibilities that Computer Science offers and the currently available computational power enable new kinds of models. Therefore, other features can be used, such as contours, repeating patterns, features from music cognition, features that reflect performance of untrained singers, and so on.

These new methods have to be developed in an interdisciplinary research context as described in section 2.6. Collaboration between FSR, CM and MIR provides the musical insights for computational modeling of relevant features, and for improving failing models, thus escaping the problems of ground truths that were discussed in section 2.6. We envision an iterative process of modeling and implementing that will result in an increasing understanding of the concepts of Folk Song Research, in particular the concept of tune family. This knowledge is highly valuable for both Folk Song Research and Music Information Retrieval, and might also be of interest for other disciplines, Music Cognition in particular.

As a crucial step in the process of modeling the concept of tune family, the next chapter presents an annotation method in which intuitive musical knowledge of domain experts is made explicit in order to understand the various kinds of relations between member-melodies of the same tune family.

Chapter 3

A Manual Annotation System for Folk Song Similarity

This chapter presents an annotation method that has been developed to get a better understanding of experts' evaluation of melodic similarity of melodies belonging to the same tune family. The aim of this method is to make the intuitive decisions about melodic similarity of the domain experts explicit.

Contribution 1. The newly developed annotation method is unprecedented in the detail with which aspects of melodic similarity are quantized. It has been developed in close cooperation with domain experts.

Contribution 2. A dataset has been created consisting of 360 melodies in 26 tune families with annotations according to the newly developed method. This dataset is valuable for testing models of melodic similarity and for testing Music Information Retrieval systems for melodies.

Contribution 3. Using this dataset, we show that the *kind* of melodic similarity among related melodies varies to a large extent from case to case, which has not been demonstrated clearly before.

3.1 Introduction

Similarity in the domain of music has been investigated in Ethnomusicology, Music Cognition, and Music Theory and received an intensified interest through the raise of Music Information Retrieval (MIR). The computational modeling of similarity in MIR often faces the challenge of a lack of domain knowledge about musical similarity: how do listeners perceive similarity and how can this be modeled formally? Hence, for a given data set of musical objects, it is often not obvious how these objects are linked through similarity or dissimilarity. To make that kind of information available for use in computational modeling of melodic similarity for retrieval purposes, we introduce a method to annotate folk song melodies concerning their similarity relations. In this chapter we present the method and apply it to a corpus of folk song melodies. This method is very specific concerning musical parameters that are involved in the perception of melodic similarities. To our knowledge there is no other corpus of melodies that has been annotated in such detail concerning the musical aspects of similarity.

One way to model melodic similarity computationally is to measure the values of various quantitative features of the music, such that each melody is represented by a vector of feature values. Thereupon, pattern recognition algorithms can be used to classify melodies. When taking this approach, two important questions are how many features need to be involved and whether the same features are equally important for each similarity assessment, or, in other words, are all similar melodies similar in the same way? In the current and the next chapters, we study both questions. In the current chapter we focus on dimensions of perceived melodic similarity by human domain experts, and in the next chapter we focus on low-level, easy to compute quantitative features that are more suitable for a computational approach. The question which dimensions are important will be addressed in the design of the annotation method (section 3.4). The question whether the same dimensions or features are important for different cases of melodic similarity will be addressed in two experiments that use the annotation data (sections 3.5 and 3.6), and in the next chapter. These experiments enable us to draw several conclusions about the various ways in which melodies are similar to each other.

The classification of melodies into groups of related melodies is a special case of human categorization in music. Therefore, it is relevant to relate our study to models of human categorization that are known in cognitive literature. Two different views of categorization are relevant for our study. The *classical* view on categorization, which goes back to Aristotle, defines a category as being constituted of all entities that posses a common set of features (Sutcliffe, 1993). In contrast to this, the *modern* view claims that most natural concepts are not well-defined but rather that individual exemplars may vary in the number and kind of characteristics they share with others. As a result, some exemplars may be more typical of a category than others. The most prominent models according to this view are Wittgenstein's *family resemblance* model (Wittgenstein, 1953) and Rosch's *prototype* model (Rosch, 1973). Deliège (2001) and Ziv & Eitan (2007) consider the family resemblance and the prototype model appropriate to describe the categories built in Western classical music.

We investigate similarity of folk song melodies by assessing the way in which musicological experts describe their own similarity judgments. We identify the dimensions that they indicate as being important for similarity assessment. Thus, we connect categorization in music and the modeling of melodic similarity perception, resulting in a 'deconstruction' of the intuitive musical similarity perception in its constituent dimensions.

In our annotation method, for each pair of melodies, numerical similarity ratings are given for each dimension that is considered important by the experts. These ratings are based on criteria that have been defined together with them, such that the relative importance of different musical dimensions of similarity is explicitly represented in the annotations. We directly involve domain experts in the process of designing the experimental setup and in the design of the annotation system, rather than indirectly by interpreting underlying dimensions of relevant similarity measures as, e.g. in a multi-dimensional-scaling approach (as in Müllensiefen & Frieler, 2007), or by correlating listeners ratings with retrospectively defined features according to the authors' analysis of the musical material (as in Eerola et al., 2001). The similarity assessment of the musicological experts is not the result of a spontaneous reaction, where participants listened to pairs of musical experts and gave a rating, since no time constraint was given. It is reasonable to assume that the given ratings reflect a deep understanding of the underlying similarity relations by the musicological experts.

3.2 The Context of the Experiment

We focus particularly on the corpus *Onder de groene linde*, which is described in Section 1.1. This corpus contains songs that have been transmitted within oral tradition. The domain experts of the Meertens Institute, namely Ellen van der Grijn, Mariet Kaptein and Marieke Klein,[1] classify these songs into tune families such that each tune family consists of melodies that are considered to have a common historic origin. Since the actual historic relation between

[1] All three have an academic master degree in Musicology. They were (or are) appointed as documentalists for the song collection of the Meertens Institute.

the melodies is not known from documentary evidence, the classification is based on similarity assessments. If the similarity between two melodies is high enough to assume a plausible genetic relation between them, the two melodies are classified into the same tune family. In the human process of classifying melodies into tune families the melodies that are considered the most typical of their tune family receive the status of a *prototypical* melody. Other melody candidates are then compared to this prototypical melody in order to decide whether they belong to this tune family.

The WITCHCRAFT project, that is carried out at the Meertens Institute, has at its aim to create computational methods that support folk song research.[2] In order to be able to design music information retrieval strategies to retrieve melodies belonging to the same tune family, we have to investigate whether all melodies belonging to the same tune family share the same set of common features or vary in the number and kind of characteristic features they possess.

3.3 The Annotation Method

The annotation method has been designed in an iterative way. In this chapter, we describe the results of the three final iterations of this process. Both in the first and second iteration, three experts annotated two tune families in detail, and in the third iteration they annotated 26 other tune families. After the first and second iteration the annotations were compared and the annotation method has been adapted where necessary. After the second and the third iteration we used the annotation values to draw several conclusions about the relative importance of the various dimensions, referred to as the first and second experiment in the remainder of this chapter. Table 3.1 shows a schematic overview of the process.

The procedure is as follows. For each tune family one expert determines which melody is the most prototypical. This melody receives the status of reference melody. All other melodies of the tune family are compared to this reference melody and the level of similarity for several musical dimensions is recorded.

Annotating the similarity of each pair of melodies in the corpus would be very time consuming, but by comparing all melodies to a reference melody, the comparisons have meaning for the tune family as a whole, since the reference melody is supposed to represent the tune family as a whole.

[2] http://www.cs.uu.nl/research/projects/witchcraft/ (accessed 1 June 2010).

Iteration 1:	Annotation of tune families *Frankrijk* and *Boerinnetje*.
Iteration 2:	Annotation of tune families *Meisje* and *Bergen*.
First Experiment:	Agreement among the annotators and relative importance of the dimensions.
Iteration 3:	Annotation of 26 tune families; creation of the Annotated Corpus.
Second Experiment:	Relative importance of the dimensions.

Table 3.1: Overview of iterations in designing the annotation method.

There are two drawbacks of this reduction. The first is that the extent to which the annotations are representative for the tune family as a whole depends on the extent to which the reference melody is representative for the tune family as a whole. Since the reference melodies has been used to assign the melodies to the tune families in the first place, we assume that the reference melodies are representative indeed. The second drawback is that relations between tune families are not annotated. For example, it might be possible that all melodies in a tune family are very similar concerning rhythm, but that they are also rhythmically similar to many melodies from other tune families since their rhythm is very generic. To deal with this problem, after the second iteration, additions has been made to the annotation system, which are described in section 3.6.

The annotation data consists of judgments concerning the contribution of different musical dimensions to the similarity between the melody under consideration and the reference melody. In daily practice, the experts mainly perform the similarity evaluation in an intuitive way. In order to analyze this complex and intuitive similarity evaluation, we identified the musical dimensions of the annotations in close collaboration with the experts. These dimensions are rhythm, contour, motifs, form and lyrics. They describe important factors within the decision process of assigning melodies to tune families according to the experts. In order to be used as reference data to design computational algorithms, we standardized the human evaluation such that numeric values are assigned to most of the dimensions. We distinguish three different numeric values 0, 1 and 2:

0. The two melodies are not similar according to this dimension.
1. The two melodies are somewhat similar according to this dimension.
2. The two melodies are obviously similar according to this dimension.

Differentiating more than three values proved to be an inadequate approach for the musicological experts since the exact thresholds for choosing the right value are very hard to determine. For example, when adding just one level, resulting in 0, 1, 2, and 3 as possible values, in many concrete situations, it is not clear how to discern between levels 1 and 2 exactly. The three-level approach is more appropriate since it only has the two extreme cases (not similar and obviously similar) and a mid-level in between.

In the next section explicit criteria are provided for rating the different dimensions.

3.4 Criteria for the Similarity Annotations

For each dimension we iteratively defined a number of criteria that the human decision should be based upon when assigning the numeric values that best reflect the intuitive assignments of the experts. These criteria are as concrete as necessary to enable the musicological experts to give agreeable ratings that are in accordance with their intuitive assignments. The experts were involved in defining these criteria. In the following subsections these criteria are given in detail. These are the definitions that have been established after the first iteration. After the second iteration several additions were made, as described in section 3.6, but the definitions have not been changed.

3.4.1 Rhythm

We define the following criteria for the comparison of a pair of phrases from two melodies with respect to their rhythmic similarity.

- If the two songs are notated in the same, or a comparable meter (e.g. 2/4 and 4/4), then count the number of transformations needed to transform the one rhythm into the other (see Figure 3.4.1 for an example of a transformation):
  - If the rhythms are exactly the same or contain a perceptually minor transformation: value 2.
  - If one or two perceptually major transformations needed: value 1.
  - If more than two perceptually major transformations needed: value 0.

- If the two songs are not notated in the same, or a comparable meter (e.g. 6/8 and 4/4), then the notion of transformation cannot be applied in a proper manner (it is unclear which durations correspond to each other). The notation in two very different meters indicates that the rhythmic structure is not very similar, hence a value of 2 is not appropriate.
  - If there is a relation between the rhythms to be perceived: value 1.
  - If there is no relation between the rhythms to be perceived: value 0.

Figure 3.4.1: Example of a rhythmic transformation: In the first full bar one transformation is needed to transform the rhythm of the upper melody into the rhythm of the lower melody.

In all cases the rhythmic similarity of pairs of individual phrases is annotated (local rhythm). It is not required to annotate all pairs of phrases. The annotator is free to chose the pairs of phrases to annotate.

3.4.2 Contour

The contour is an abstraction of the melody. Hence it remains a subjective decision which notes are considered important for the contour. From the comparison of the lines we cannot automatically deduct the value for the entire melody via the mean value. Therefore we also give a value for the entire melody that is based on fewer points of the melody and hence on a more abstract version of the melody than the line-wise comparison.

- For the line-wise comparison:
  - Determine begin (if the upbeat is perceptually unimportant, choose the first downbeat as begin) and end of the line and 1 or 2 turning points (extreme points) in between.

- Based on these 3 or 4 points per line determine whether the resulting contours of the lines are very similar (value 2), somewhat similar (value 1) or not similar (value 0).

- For the comparison of the global contour using the entire song:
  - Decide per line: if the pitch stays in nearly the same region choose an average pitch for this line; if not, choose one or two turning points.
  - Compare the contour of the entire song consisting of these average pitches and turning points.
  - If the melody is too long for this contour to be memorized, then choose fewer turning points that characterize the global movements of the melody.

3.4.3 Motifs

The decision to categorize a melody into a certain tune family is often based on the detection of single characteristic motifs. Hence it is possible that the two melodies are different on the whole, but they are recognized as being related due to one or more common motifs.

- If at least one very characteristic motif is being recognized: value 2.
- If motifs are shared but they are not very characteristic: value 1.
- No motifs are shared: value 0.

Characteristic in this context means that the motif serves as a basic cue to recognize a relation between the melodies.

3.4.4 Lyrics

In some cases, it is possible that the lyrics are the main reason to assign a melody to a tune family, even though the musical material does not provide decisive clues about a melodic relation. Therefore we examine whether the comparison of the lyrics of two songs suggests a relation between them or not or only vaguely. The criteria are as follows:

- If the lyrics are either literally the same, or semantically the same or the *strophic form* is characteristic and the same (or any combination of these factors): value 2.
- If parts of the lyrics are literally or semantically the same, or the strophic form is the same but not very characteristic, the combination of these factors might still indicate a significant relationship: value 2.
- If only parts of the lyrics are literally, or semantically or according to the strophic form the same (or any combination of these factors) and the partial resemblances or their combination is not very convincing: value 1.
- If none of the above cases applies: 0.

The *strophic form* is defined by the following features: number of accents per line, rhyme gender, rhyme scheme (refrain). A strophic form is characteristic if it contains uncommon patterns, such as uneven verse lengths and an irregular rhyme scheme. Usually, characteristic forms are rare, i.e., they serve just one tune family.

3.5 First experiment on similarity annotations

For an initial experiment on the similarity annotations, four tune families containing 11–16 melodies each have been selected to be annotated by three musicological experts. These are the tune families *Frankrijk buiten de poorten 1* (short: *Frankrijk*), *Daar was laatst een boerinnetje* (short: *Boerinnetje*), *Daar was laatst een meisje loos 1* (short: *Meisje*) and *Toen ik op Neerlands bergen stond* (short: *Bergen*). For each tune family one musicological expert determined the reference melody. Similarity ratings were assigned to all other melodies of the same tune family with respect to the reference melody. In a first iteration, *Frankrijk* and *Boerinnetje* were annotated, in a second iteration *Meisje* and *Bergen*. After the both iterations the results were discussed with all experts.

3.5.1 Agreement among the experts

Table 3.2 gives an overview of the agreement among the three experts for all musical dimensions using three categories. Category A counts the number of total agreement, i.e. all three experts assigned the same value. Categories PA1 and PA2 count the number of partial agreements such that two experts agreed on one value while the third expert chose a different value. In PA1 the difference between the values equals 1 (e.g. two experts assigned a 1 while one expert assigned a 2). In PA2 the difference between the values equals 2 (e.g. two experts assigned 0 while one expert assigned a 2). Category D counts the cases in which all experts disagree.

Tune Family	**A**	**PA1**	**PA2**	**D**
Frankrijk	58.7	38.1	1.6	1.6
Boerinnetje	50.8	42.6	0.5	6.1
Meisje	70.4	27.6	1	1
Bergen	77.5	18.5	1.1	2.9
Average	64.3	31.7	1.1	2.9

Table 3.2: Comparison of agreement among three experts: A for total agreement, PA1 and PA2 for partial agreement D for disagreement (see section 3.5.1 for further details). Numbers are percentages.

Both the percentage of disagreement in category D and the percentage of partial agreement PA2 containing both values for *not similar* and *very similar* are quite low. The category of total agreement A comprises the majority of the cases with 64.3%. Moreover, comparing the values obtained for *Frankrijk* and *Boerinnetje* to those for *Meisje* and *Bergen* reveals that the degree of agreement is much higher within the second iteration, after the discussion of the results of the first one. Hence, this experiment indicates that the musical dimensions have been established in such a way that there is considerable agreement among the musical experts as to how to assign the similarity values.

3.5.2 Comparing dimensions across tune families

Table 3.3 lists the distribution of the assigned values within each musical dimension and the lyrics dimension for all tune families. The dimension *lyrics* receives on average high scores for value 2 (81.2%).

Both *Frankrijk* and *Meisje* score highest for rhythm concerning the value 2, while *Boerinnetje* scores highest for motifs and *Bergen* for global contour. Hence, the importance of the different musical dimensions regarding the similarity assignment of melodies belonging to one tune family varies between the tune

	Frankrijk			*Boerinnetje*					
Value	0	1	2	0	1	2			
Rhythm/l	0	1.3	98.7	11.2	51.6	37.2			
Contour/g	0	31.7	68.3	12.8	48.7	38.5			
Contour/l	5.6	52.5	40.9	41.9	26.4	31.7			
Motifs	0	36.6	63.4	0	20.5	79.5			
Lyrics	26.7	6.6	66.7	5.1	33.3	61.5			
	Meisje			*Bergen*			Average		
Value	0	1	2	0	1	2	0	1	2
Rhythm/l	3.3	8.2	88.5	3.5	15.8	80.7	4.5	19.2	76.3
Contour/g	33.3	13.3	53.4	2.5	10.3	87.2	12.1	26	61.9
Contour/l	20.7	31.8	47.5	4.8	22.5	72.7	18.3	33.3	48.2
Motifs	13.3	16.7	70	0	17.9	82.1	3.3	22.9	73.8
Lyrics	0	3.3	96.7	0	0	100	7.9	10.8	81.2

Table 3.3: Distribution of the assigned values within each dimension per tune family as percentages.

families. Moreover, in most of the cases single dimensions are not characteristic enough to describe the similarity of the melodies belonging to one tune family.

The best musical feature of *Boerinnetje* scores 79% for value 2, the other musical dimensions score below 40%. From this perspective, the melodies of *Boerinnetje* seem to form the least coherent group of all four tune families. While *Frankrijk* receives the highest rating in a single dimension for value 2, all other dimensions score relatively low. *Bergen* scores in all dimensions above 72% for the value 2. Hence these melodies seem to be considerably similar to the reference melody across all dimensions. For *Meisje* two dimensions receive scores above 70% for value 2, on the other hand three dimensions have considerably high scores (between 13% and 33%) for the value 0. Hence this tune family contains melodies with both very similar and very dissimilar aspects.

Comparing the contribution of the musical dimensions reveals that the contour scores above 70% for value 2 for only one tune family (*Bergen*), which results in a considerably low average. Both rhythm and motifs score above 70% for value 2 in three out of four cases, resulting in an average of 76.3% for value 2 for the dimension rhythm and 73.8% for value 2 for the dimension motifs. Hence, rhythm and motifs seem to be more important than contour for the experts' perception of similarity in these experiments.

Song ID	0	1	2
70321_01	12.5	22.9	64.6
70560_01	4.2	8.3	87.5
71374_01	0	12.5	87.5
71449_01	56.3	25	18.7
71734_01	4.2	14.6	81.2
72923_01	16.4	8.3	75
73517_01	0	0	100
111465_01	0	4.2	95.8
139116_01	39.6	37.5	22.9
139121_01	43.3	41.7	15

Table 3.4: Degree of similarity of all melodies of the group *Meisje* to the reference melody 70412_01 averaged over all dimensions as percentages.

3.5.3 Similarity within tune families

By comparing the ratings for the individual songs within a tune family, we get an indication of the variation in the importance of the musical similarity dimensions.

As a measurement for the degree of similarity of a melody to the reference melody we count the number of occurrences of the three rating values and express it as percentage. For this we only use the melodic dimensions: rhythm, global contour, contour per line and motifs. The results show that the degree of similarity within the family can vary with considerable amount. For instance, in the tune family *Meisje* two melodies (73517_01 and 111465_01, see Table 3.4) score higher than 95% for value 2, while two melodies score lower than 20% for value 2 with corresponding high scores for value 0 (71449_01 and 139121_01).

The evaluation of single dimensions also shows that the degree of similarity to the reference melody varies. For instance, *Meisje* scores for the dimension rhythm on average 88.5% for value 2 (see Table 3.3). However, melody 71449_01 scores for rhythm only 42% for value 2 and 33% for value 0. It appears that there is not one characteristic (or one set of characteristics) that all melodies of a tune family share with the reference melody to the same extent.

3.5.4 Discussion

From sections 3.5.2 and 3.5.3 we conclude that both across and within the tune families the importance of the musical dimensions for perceived similarity varies.

There is not one characteristic (or one set of characteristics) that all melodies of a tune family share with the reference melody. Therefore, the category type of the tune families cannot be described according to the classical view on categorization, but rather to the modern view. The implication for the design of a retrieval system for folk song melodies is that the similarity model has to incorporate various musical dimensions in order to be able to retrieve melodies that are related in various ways.

3.6 Second experiment on similarity annotation

For the third iteration, 360 melodies (grouped into 26 tune families) have been selected by a musicological expert as a representative set out of over 6000 Dutch folk song melodies, which have been encoded at the Meertens Institute both from ethnomusicological transcriptions of field recordings and from written sources of folk songs. The size of each tune family varies between 10 and 20 melodies. These 360 songs have been selected to form a relatively small subset that is representative for the collection as a whole concerning the various similarity relations between the songs. An additional constraint was that considerable variation has to occur among the melodies that belong to the same tune family. 'Easy' tune families have not been selected. Thus, the results of studying this subset are expected to be indicative for the results that would be obtained when studying the entire corpus. We refer to this corpus as the *Annotated Corpus*. The contents of this corpus can be found in Appendix A.

The analysis of the results of the iterations as described in the previous section leads to a number of modifications concerning the annotation in the third iteration. We consider the overall agreement of the experts in the second iteration sufficient. Therefore, in the third iteration, each tune family is annotated by only one expert. Since motifs play a very important role for the classification of melodies (according to the verbal descriptions of the musicologists and the results of the numeric evaluation of dimensions within the first experiment), the location and size of characteristic motifs are annotated in the third iteration, delivering a list of motifs considered important for the classification. Moreover, the musicological experts annotate cases of doubts, in which the criterion

Melody A:

Melody B:

Global Annotations		**Local Annotations**			
		Phrase from A	Phrase from B	Rhythm	Contour
Global Contour	**1**	1	1	**2**	**1**
Global Rhythm	**1!**	2	2	**2**	**1**
Motifs	**2**	3	3	**1**	**0?**
Lyrics	**2**	4	4	**0?**	**1**

Figure 3.6.1: Example annotations for two melodies from the tune family *Daar ging een heer 1*.

as defined in section 3.4 for assigning a certain value does not correspond to the experts' intuitive rating of this dimension. Thus, the validity of the defined criteria concerning the similarity assignments of the experts is tested. Furthermore, annotation for dimensions that serves as a key for the classification has been included. If one of the dimensions has been of particularly importance for the decision to include the melody in the tune family, the annotator has a possibility to indicate that dimension. This implies that the melodies in the tune family are not generic concerning that dimension. As an example, Figure 3.6.1 shows the annotations for two melodies from the tune family *Daar ging een heer 1*.

Finally, an optional dimension has been added. In addition to the rhythmical rating of pairs of phrases, the annotators can rate the rhythmic similarity of the two entire melodies according to the same criteria (global rhythm).

	absolute values			*values in %*			*doubts in %*	*classification key in %*
	0	1	2	0	1	2		
Rhythm/global	9	62	143	4.2	29.0	66.8	3.3	22.4
Rhythm/local	87	491	1069	5.3	29.8	64.9	1.0	0
Contour/global	6	116	211	1.8	34.8	63.4	3.6	0.6
Contour/local	120	623	904	7.3	37.8	54.9	1.0	0
Motifs	8	31	293	2.4	9.3	88.3	0	9.3
Lyrics	34	10	245	11.8	3.5	84.8	0	2.4

Table 3.5: Distribution of the assigned values per dimension for 360 melodies.

Table 3.5 shows the distribution of the assigned values per dimension for all 360 melodies annotated as well as the number of doubts expressed and the number of annotated key dimensions. For all dimensions doubts have been expressed very rarely, demonstrating that the criteria defined in section 3.4 coincides in most of the cases with the intuitive assignments of the experts. Single dimensions have hardly been indicated as being the main reason for the classification of a melody, as the low values for classification key shows. In most of the cases it was the global rhythm which served as a single key for the classification. Comparing the values across the different dimensions shows that the rhythm of melodies within a tune family is still considered more similar than the contour. However, in comparison to the dimension motifs, rhythm plays a less prominent role than in the first experiment. Among the musical dimensions, the dimension motifs receives the highest scoring for value 2 with 88.3% .

Table 3.6 shows the distributions of scores among the various tune families.

Again, there are considerable differences between the tune families. This indicates that the experts' evaluation of melodic similarity is not a one-dimensional process. The various dimensions of similarity are not equally important in every case. As a consequence, it seems not possible to design an adequate (computational) model of melodic similarity of folk song melodies using only rhythm or contour (e.g., parsons contour, or a sequence of pitch intervals).

3.7 Concluding Remarks

In this chapter, we presented an annotation method to facilitate the study of similarity of melodies that are related trough the process of oral transmission. This method is based on experts' knowledge. By making implicit criteria for similarity evaluations explicit, we obtain insights in the complexity and multidimensionality of experts' evaluation of melodic similarity. It appears that various aspects of melody are important for establishing a similarity judgment: contour (both per phrase and for the entire songs), rhythm (both per phrase and for the entire songs), and motifs. In individual cases, the relative importance of these dimensions varies to a large extent. However, in general the recurrence of characteristic motifs seems most important. Motifs are local phenomena, while the other dimensions describe melodies, or individual phrases, globally. In the next chapter, we compare global features to an approach in which local comparison of melodies is possible, and in Chapter 7 we perform an experiment using characteristic motifs, which are local phenomena, to retrieve related melodies.

The Annotated Corpus that results from this study is a valuable resource for further research on melodic similarity. Similarity relations between melodies are described in detail, both quantitatively (ratings) and qualitatively (dimensions). These annotations can be used as enriched 'ground truth' to test various kinds of retrieval algorithms. The pairs of phrases that have been annotated concerning local contour and local rhythm, indicate which parts of the two melodies correspond. The large number of annotated motifs (1426 motif occurrences of 104 motif classes) can, e.g., be used to test algorithms that detect recurring patterns.

In the next chapters, we use the Annotated Corpus to evaluate several algorithmic approaches to similarity assessment of folk song melodies.

N	*Rhythm/global*			*Rhythm/local*			*Contour/global*			*Contour/local*			*Motifs*			*Lyrics*		
	0	*1*	*2*	*0*	*1*	*2*	*0*	*1*	*2*	*0*	*1*	*2*	*0*	*1*	*2*	*0*	*1*	*2*
1	*0*	*60*	*40*	*11.7*	*36.7*	*51.7*	*0*	*86.7*	*13.3*	*11.7*	*61.7*	*26.7*	*0*	*0*	*100*	*20*	*0*	*80*
2	-	-	-	*2.8*	*63.9*	*33.3*	*0*	*18.2*	*81.8*	*0*	*61.1*	*38.9*	*0*	*36.4*	*63.6*	*0*	*0*	*100*
3	*5*	*45*	*50*	*4.8*	*41.6*	*53.6*	*3.7*	*18.5*	*77.8*	*9.6*	*24*	*66.4*	*0*	*14.8*	*85.2*	*0*	*0*	*100*
4	-	-	-	*0*	*54.5*	*45.5*	*0*	*6.3*	*93.8*	*1.5*	*22.7*	*75.8*	*0*	*0*	*100*	*0*	*0*	*100*
5	-	-	-	*0*	*68.5*	*31.5*	*0*	*11.1*	*88.9*	*0*	*24.1*	*75.9*	*0*	*22.2*	*77.8*	*0*	*0*	*100*
6	*25*	*50*	*25*	*35.1*	*33.3*	*31.6*	*0*	*41.7*	*58.3*	*3.5*	*47.4*	*49.1*	*0*	*8.3*	*91.7*	*0*	*0*	*100*
7	*0*	*75*	*25*	*0*	*57.1*	*42.9*	*0*	*37.5*	*62.5*	*7.1*	*39.3*	*53.6*	*0*	*37.5*	*62.5*	*62.5*	*0*	*37.5*
8	*0*	*40*	*60*	*0*	*40*	*60*	*0*	*13.3*	*86.7*	*8.3*	*23.3*	*68.3*	*6.7*	*6.7*	*86.7*	*20*	*6.7*	*73.3*
9	-	-	-	*0*	*10*	*90*	*0*	*10*	*90*	*0*	*30*	*70*	*0*	*10*	*90*	*0*	*0*	*100*
10	*12.5*	*18.8*	*68.8*	*23.4*	*22.1*	*54.5*	*6.3*	*68.8*	*25*	*19.5*	*50.6*	*29.9*	*0*	*12.5*	*87.5*	*0*	*0*	*100*
11	-	-	-	*0*	*37.5*	*62.5*	*0*	*0*	*100*	*0*	*39.3*	*60.7*	*0*	*0*	*100*	*0*	*0*	*100*
12	*0*	*18.2*	*81.8*	*0*	*6.4*	*93.6*	*0*	*54.5*	*45.5*	*3.8*	*46.2*	*50*	*0*	*0*	*100*	-	-	-
13	-	-	-	*0*	*34.4*	*65.6*	*0*	*41.7*	*58.3*	*0*	*50*	*50*	*0*	*0*	*100*	-	-	-
14	*0*	*30*	*70*	*10*	*13.3*	*76.7*	*0*	*30*	*70*	*33.3*	*30.0*	*36.7*	*20*	*0*	*80*	*20*	*60*	*20*
15	*10*	*30*	*60*	*17.2*	*20.7*	*62.1*	*20*	*70*	*10*	*31*	*55.2*	*13.8*	*10*	*10*	*80*	*0*	*0*	*100*
16	*0*	*28.6*	*71.4*	*6.3*	*31.3*	*62.5*	*0*	*100*	*0*	*21.9*	*62.5*	*15.6*	*0*	*0*	*100*	*0*	*0*	*100*
17	-	-	-	*0*	*26.6*	*73.4*	*0*	*6.3*	*93.8*	*0*	*23.4*	*76.6*	*0*	*13.3*	*86.7*	*81.3*	*0*	*18.8*
18	*0*	*45.5*	*54.5*	*6*	*18*	*76*	*0*	*63.6*	*36.4*	*28*	*32*	*40*	*36.4*	*0*	*63.6*	*36.4*	*9.1*	*54.5*
19	*5.9*	*5.9*	*88.2*	*11.1*	*2.5*	*86.4*	*11.8*	*11.8*	*76.5*	*13.6*	*18.5*	*67.9*	*0*	*0*	*100*	*0*	*0*	*100*
20	*10*	*20*	*70*	*7.5*	*23.8*	*68.8*	*0*	*70*	*30*	*2.5*	*46.3*	*51.3*	*0*	*20*	*80*	*0*	*0*	*100*
21	-	-	-	*2.9*	*64.3*	*32.9*	*0*	*64.3*	*35.7*	*8.6*	*44.3*	*47.1*	*0*	*14.3*	*85.7*	*0*	*0*	*100*
22	*0*	*0*	*100*	*0*	*4.4*	*95.6*	*0*	*57.1*	*42.9*	*7.4*	*39.7*	*52.9*	*0*	*0*	*100*	*0*	*0*	*100*
23	*0*	*12.5*	*87.5*	*0*	*26.3*	*73.7*	*0*	*56.3*	*43.8*	*5.3*	*27.6*	*67.1*	*0*	*18.8*	*81.3*	*0*	*12.5*	*87.5*
24	-	-	-	*0*	*18.4*	*81.6*	*0*	*0*	*100*	*0*	*52*	*48*	*0*	*0*	*100*	*0*	*0*	*100*
25	*0*	*15.4*	*84.6*	*7.8*	*39.1*	*53.1*	*0*	*0*	*100*	*1.6*	*21.9*	*76.6*	*0*	*0*	*100*	*0*	*0*	*100*
26	*0*	*6.3*	*93.8*	*0*	*15.5*	*84.5*	*0*	*31.3*	*68.8*	*4.8*	*41.7*	*53.6*	*0*	*18.8*	*81.3*	*23*	*0*	*75*

Table 3.6: Distribution of the assigned values within each dimension per tune family in the second experiment

Chapter 4

The Study of Melodic Similarity using Global Melody Feature Sets

In this chapter we evaluate the usefulness of global features for recognition of melodies. First, we assess individual features to see whether we can find features that are discriminative for all tune families (section 4.4). Second, we investigate subsets of features (section 4.5). In both cases, we test the discriminative power of features for each individual tune family as well as for the entire data set. Finally, we compare the discriminative power of global features to the discriminative power of sequences of local features (section 4.6).

Contribution. This is the first study in which the discriminative power of melodic features has been evaluated for a large set of folk song melodies. It shows that none of the individual features, nor subsets of the features in the sets designed for computational folk song research by Jesser (1991) and Steinbeck (1982), and for general purpose by McKay (2004), can be used for successful classification of folk songs melodies.

4.1 Data

We use two data sets, a small set to perform feature selection, which is computationally expensive, and a large set to test for scalability. The small data set consists of the Annotated Corpus (360 songs in 26 tune families, see Chapter 3). The large set includes the small set along with 4470 songs from other tune families. We are primarily interested in the recognition of the songs in the Annotated Corpus, since this set is well understood concerning its similarity relations and it has been assembled by domain experts to be representative for the collection as a whole. In the experiments in this chapter, we evaluate the discriminative power of feature subsets twice, once with only the songs from the Annotated Corpus, and once with the other 4470 songs added.

Since the assignments of the melodies to the tune families were done in a careful process by the domain experts from the Meertens Institute, we consider the resulting partitioning of the dataset of high quality, such that it is suitable to test the discriminative power of the various classification approaches. A re-evaluation of this 'ground truth' based on algorithmic results will be perfomed in section 6.5.

4.2 Feature Set

We use the following three sets of features that are well known from literature:

- 12 features provided by Wolfram Steinbeck (1982).
- 39 features provided by Barbara Jesser (1991).
- 37 rhythm, pitch and melody features implemented in jSymbolic by Cory McKay (2004).

Steinbeck and Jesser specifically designed their feature sets to study relations between folk songs within the Essen Folk Song Collection that are connected through the process of oral transmission. Because our corpus consists of such folk song melodies, the evaluation of these two feature sets is particularly interesting. McKay's set was designed as general purpose feature set. It contains a number of features that are not in the sets by Jesser and Steinbeck.

All features for which *absolute* pitch is needed (e.g. Steinbeck's Mean Pitch) are not included because not all melodies in our corpus are in the same key.

Also the low-level, multidimensional features from the set of jSymbolic are not included because they are primarily needed to compute the values of other, higher-level features, which we do use. Furthermore, categorical features and features that have the same value for all songs have not been included. Thus, we have 88 features, which we characterize as 'global' because for each feature an entire song is represented by only one value. The complete list of features with descriptions is included in Appendix B.

Although the relationships are hard to define exactly, most of these 88 features can be considered aspects of the musical dimensions that were chosen for the manual annotations in Chapter 3. For example, features like the fraction of descending minor seconds, the size of melodic arcs and the amount of arpeggiation contribute to contour, although they do not represent the holistic phenomenon of contour exhaustively.

After computing all 88 feature values, a song is represented by a vector of 88 feature values, or, equivalently, by a point in the 88-dimensional feature space. The scaling of the values for the different features with respect to each other influences the distances between the song-representations in the feature space. Therefore, it is necessary to normalize the feature values such that they have comparable scales. For each feature we scale the values such that they have zero mean and a standard deviation of 1. This is achieved by subtracting the original mean and dividing by the original standard deviation. We do this both for the annotated set and for the full set separately.

4.3 Feature Evaluation Criterion

To determine the discriminative power of a feature, or a set of features, we need a criterion. Since we are interested in *recognition* of songs, the fraction of songs that is correctly classified into the right tune family using the feature subset under consideration, seems a good criterion. In terms of classifier evaluation, this is the leave-one-out success rate, which is computed by taking subsequently each song out, training a classifier on the other songs and classifying the song under consideration using that classifier. When this has been done for all songs, the leave-one-out success rate is computed by dividing the number of correctly classified songs by the total number of songs.

To compute the criterion value, we need a classifier. Because of the small class sizes, it is not possible to do global density estimation in the feature space. Therefore, we use the nearest neighbor classification rule to associate songs with each other: a song is classified into the tune family of the song that is closest in the feature space according to the euclidean distance. This approach only uses local densities.

Since we are interested in the differences between tune families concerning discriminative features, we want to be able to compute the criterion per tune family and for sets of tune families. Therefore we define the set of songs $\mathbf{C}$ as the set of all songs involved in the experiment, and the subset $\mathbf{S} \subseteq \mathbf{C}$ as the songs in which we are interested. In our experiments, $\mathbf{S}$ will either contain the songs from a single tune family or the songs from all 26 tune families in the Annotated Corpus, and $\mathbf{C}$ will either contain the songs of the Annotated Corpus, or the full corpus of 4830 songs. In all cases, we label the songs in $\mathbf{S}$ with their respective tune family and the other songs in $\mathbf{C}$ with 'Other'. We define $\mathbf{F}$ as the subset of features for which we want to compute the criterion.

Since we are first of all interested in subset $\mathbf{S}$, we only count the errors and successes among the songs in $\mathbf{S}$, resulting in a success rate for the songs in $\mathbf{S}$. However, when $\mathbf{C}$ contains more songs than $\mathbf{S}$, we also have to involve the false positives among the other songs in $\mathbf{C}$, since these erroneously have been assigned to (recognized as) one of the tune families in $\mathbf{S}$, and, thus, decrease the discriminative power of the feature set under consideration.

Taking the above considerations into account, we define the following criterion:

$$J(\mathbf{C}, \mathbf{S}, \mathbf{F}) = \frac{tpr(\mathbf{S}, \mathbf{F})}{1 + fpr(\mathbf{C}, \mathbf{S}, \mathbf{F})} = \frac{tp(\mathbf{S}, \mathbf{F})}{|\mathbf{S}| + fp(\mathbf{C}, \mathbf{S}, \mathbf{F})},$$

where $tpr(\mathbf{S}, \mathbf{F}) = tp(\mathbf{S}, \mathbf{F})/|\mathbf{S}|$ the true positive rate, with $tp(\mathbf{S}, \mathbf{F})$ the number of true positives and $|\mathbf{S}|$ the number of songs in $\mathbf{S}$, and $fpr(\mathbf{C}, \mathbf{S}, \mathbf{F}) = fp(\mathbf{C}, \mathbf{S}, \mathbf{F})/|\mathbf{S}|$ the false positive rate, where $fp(\mathbf{C}, \mathbf{S}, \mathbf{F})$ is the number of false positives. In this context, true positives are those songs in $\mathbf{S}$ that have another song from the same tune family as the nearest neighbor, and false positives are those songs *not* in $\mathbf{S}$ that have a song from a tune family present in $\mathbf{S}$ as nearest neighbor, but do not belong to that tune family. Since $tpr(\mathbf{S}, \mathbf{F}) \in [0, 1]$ and $fpr(\mathbf{C}, \mathbf{S}, \mathbf{F}) \geq 0$, $J(\mathbf{C}, \mathbf{S}, \mathbf{F}) \in [0, 1]$. If $\mathbf{S}$ consists of the entire data set, $fp(\mathbf{C}, \mathbf{S}, \mathbf{F})$ is zero, and thus $J(\mathbf{C}, \mathbf{S}, \mathbf{F}) = tp(\mathbf{S}, \mathbf{F})/|\mathbf{S}|$, which is the nearest neighbor leave-one-out success rate. If $\mathbf{S}$ is a proper subset of $\mathbf{C}$, J is the nearest neighbor leave-one-out success rate for the songs in $\mathbf{S}$ corrected by the false positives among the songs not in $\mathbf{S}$. This allows direct evaluation for scalability. A higher value for J indicates better class separability, since both the classes we are interested in can be separated and there is little or no interference from other classes.

Suppose, as an example, that we are interested in the discriminative power of a certain feature set **F** for the 26 tune families (360 songs) of the Annotated Corpus among the entire corpus of 4830 songs, and that we want to express the discriminative power in terms of classification results. Then, **S** consists of the 360 songs of the Annotated Corpus and **C** consists of all 4830 songs. Suppose that we have a classifier that classifies 90 songs from the Annotated Corpus (**S**) correctly and that also classifies 300 songs that are not in **S** into one of the classes (tune families) that are in **S**. Then the total number of incorrectly classified songs is 270+300=570, and, thus, the total number of correctly classified songs is 4260. Therefore, the leave-one-out success rate for the whole data set is 4260/4830=0.88. This seems a good result, but it is heavily biased by the asymmetric class sizes: the class 'Other' contains 4470 songs, while the typical size of the classes in **S** is in the order of 10 songs. In the extreme case that all songs in the data set would have been classified as 'Other', the success rate would be 4470/4830=0.93, while none of the songs we are interested in would have been classified correctly. In our example, only 90 songs in **S** have been correctly classified. Therefore a success rate of 90/360=0.25 would better reflect the performance we are interested in. Still, for the discriminative power of the feature set that was used by the classifier, this is not the right value, since 300 other songs were classified into classes that are in **S**. Therefore we correct the success rate of 0.25 using the definition of the criterion:

$$J(\mathbf{C}, \mathbf{S}, \mathbf{F}) = \frac{\frac{90}{360}}{1 + \frac{300}{360}} = 0.14.$$

An implementation-specific advantage of this criterion is the efficiency of computation. We use the implementation of the nearest neighbor classifier as is provided in the Matlab toolbox PRTools.[1] This toolbox offers a function (testk) that computes the leave-one-out success rate for an entire data set by only a few matrix operations instead of computing the error separately for each song and averaging afterward. Rewriting this function to return the value of our criterion is straightforward. Fast computation of the criterion value is especially important for finding the optimal subset of features, which has a very large solution space.

[1] http://www.prtools.org (Accessed 1 June 2010).

4.4 Individual Features

The main question of this section is: which features are discriminative for which tune families? We want to know whether there are features that are discriminative for all or many tune families, or rather for just one or a few tune families.

4.4.1 Method

For each of the 26 tune families individually, we compute for each of the 88 features the value of the criterion, both for the small dataset of 360 songs and for the large dataset of 4830 songs. Thus, **S** consists subsequently of the songs of the tune family under consideration, and **C** either of the 360 songs from the Annotated Corpus or of all 4830 songs.

We also compute the discriminative power of each individual feature for separability of the 26 classes from the Annotated Corpus as a whole. In this case, **S** contains all 360 songs from the Annotated Corpus. Again we do this both for the small and for the large data set. In the former case **C** consists of the 360 songs from the Annotated Corpus and in the latter case of all 4830 songs. Thus, for the small data set, we find the leave-one-out accuracy of the nearest neighbor classifier. The criterion value for the large data set indicates to what extent the annotated songs still can be recognized among thousands of other songs.

4.4.2 Results

Table 4.1 shows the ten features with the highest criterion values for the small data set, along with the tune families for which the features are discriminative. To compare scalability, the criterion value for the large data set is included as well. For all other combinations of features with tune families, the criterion value is less than 0.5. Clearly, the discriminative power of the features decreases dramatically for a large data set. The highest criterion value for the large data set, for individual tune families and features, is 0.2857 for tune family *Nood* and feature FractionHalfDuration (which is not shown in Table 4.1 because the performance for the small dataset is lower than 0.5). Four out of the eight songs of this tune family have been recognized along with six false positives from other tune families.

Tune Family	Feature	J for small set	J for large set
Herderinnetje	FractionEqualDurations (47)	0.8182	0.0800
Herderinnetje	FractionHalfDuration (46)	0.7692	0.0500
Maagdje	Melodic Octaves (13)	0.7273	0.0256
Maagdje	aoctave (62)	0.7273	0.0299
Meisje	Most Common Melodic Interval Prevalence (17)	0.6875	0.0488
Halewijn 2	Polyrhythms (26)	0.6667	0.1404
Lindeboom	daugfourth (69)	0.6667	0.2667
Lindeboom	Melodic Tritones (15)	0.5000	0.0800
Meisje	aminseventh (60)	0.5000	0.2273
Halewijn 2	Number of Moderate Pulses (21)	0.5000	0.1765

Table 4.1: The 10 features with the highest criterion value for the small dataset of 360 songs. The criterion value for the full dataset is also shown.

For the separability of all 26 classes, we find that for the small data set the highest value for the criterion is 0.175, which is still quite low. For the large data set, the highest criterion value is 0.027. In both cases most tune family-feature pairs yield criterion value zero. This means that none of the songs has a song from the same tune family as its nearest neighbor for the feature under consideration.

From these results, we conclude that none of the individual features is discriminative for all 26 tune families. The features that are to some extent discriminative for a single tune family are not discriminative for other tune families. The results for the Annotated Corpus are not scalable: the features that are discriminative for tune families within the Annotated Corpus are not discriminative for the same tune families within the large data set.

4.5 Feature Subsets

4.5.1 Method

To find sets of features that separate the tune families rather than individual features, we perform forward floating feature selection (Pudil et al., 1994).

Starting with an empty feature subset, this algorithm successively adds or removes a feature in order to optimize the criterion. The algorithm stops if all features are included or if the requested size of the selected subset has been reached.

Again, we do this both for each individual tune family and for all 26 tune families from the Annotated Corpus together. In the former case **S** consists of the songs from the tune family under consideration and in the latter case **S** consists of all songs from the Annotated Corpus. In both cases, during feature selection **C** consists of the 360 songs from the Annotated Corpus. We do feature selection only for the small data set, but we test the selected feature subsets on the large data set as well by computing the criterion value also for the large dataset.

4.5.2 Results

Table 4.2 shows for each tune family the indices of the selected features, and the value of the criterion for that set for both the small and the large corpus.

Although for almost all individual tune families the feature subset with the highest criterion value contains less than 10 features, 62 out of the 88 features are represented in at least one of the selected feature sets. The most common feature is STBFractionStressed (44), which occurs in only six of the 26 subsets. There are two features that occur five times, four features that occur four times, 12 features that occur three times, 17 features that occur two times, and 26 features that occur in only one of the selected subsets.

As in the previous experiments, the scalability of this global feature approach is very low. For tune families *Nood* and *Stil* the difference is even maximal. Apparently, in these cases there is a lot of interference from tune families that are not in the Annotated Corpus. The only tune family for which a moderately performing feature subset can be found for the large data set is *Meisje*.

The selection procedure for separability of all 26 classes returns a feature subset of size 60, with a criterion value of 0.8194. The criterion value for the same feature subset using the large data set is 0.3478, which shows that there is quite some confusion between tune families from the Annotated Corpus and tune families in the rest of the large corpus.

As can be seen from Figure 4.5.1, for feature subsets with more than around nine features, larger feature subsets only result in marginal improvement. The biggest improvements are reached for subsets of one, two and three features.

Tune Family	**Selected Feature Subset**	J **for small set**	J **for large set**
Heer	58 59 60 84	0.3810	0.0625
Jonkheer	4 14 20 37 38 60	0.8333	0.0769
Ruiter 2	1 6 15 28 35 44 59 67 78 86	0.7778	0.1154
Maagdje	3 13	0.8000	0
Dochtertje	37 59 71 79	0.4118	0.0962
Lindeboom	8 23 69	0.8889	0.2414
Zoeteliefjes	3 6 44 54 78 87	1.0000	0.2692
Ruiter 1	24 38 53 70 82 84	0.6667	0.3158
Herderinnetje	47 87	1.0000	0.1000
Koopman	68 70	0.6842	0.0795
Meisje	2 3 4 5 13 16 17 41 60	1.0000	0.6471
Vrouwtje	9 44 48 49 84	0.9167	0.1739
Femmes	45 51 59 81	0.7143	0.0800
Halewijn 2	9 26	0.7273	0.2162
Halewijn 4	22 28 35 37 87	0.6667	0.1905
Stavoren	15 33 45 84	0.7778	0.0769
Zomerdag	27 39 66 67	0.7895	0.0909
Driekoningenavond	9 44 57 59 68 87	0.8462	0.2500
Stad	7 13 23 36 55	1.0000	0.4000
Stil	4 23 24 34 71 75	1.0000	0
Schipper	7 12 17 47 54 58 70	0.9333	0.4000
Nood	10 17 27 46 49	1.0000	0
Soldaat	6 16 49 71	0.6111	0.2041
Bruidje	29 44 49 84	1.0000	0.0556
Verre	8 25 35 42	0.7647	0.0400
Boom	4 10 24 27 39 42 44 66 70	0.7895	0.2963

Table 4.2: The selected feature subset with the highest criterion value for the small dataset of 360 songs. The criterion value for the full dataset is also shown. Only for tune family *Ruiter 2*, more than 10 features are selected. For this tune family only the first 10 features are shown.

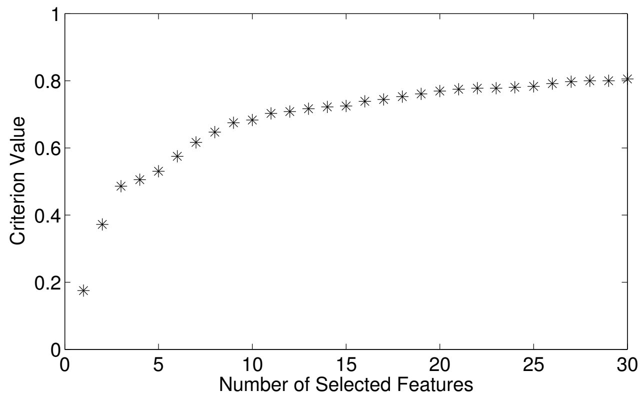


Figure 4.5.1: Criterion Values for Feature Subsets of Various Sizes for the Annotated Corpus.

The selected subset of three features contains FractionStressed (44), dminthird (66) and numlines (87). Interestingly, these three features are aspects of different dimensions of melodic similarity: rhythm and meter, pitch and length.

From these results we conclude that, although discriminative subsets of features can be found for the Annotated Corpus, no feature subset can be found that is discriminative for tune families in the large corpus. Furthermore, there is large diversity in contents of the selected features subsets for the individual tune families. There is even not one feature that is important for a substantial number of tune families.

4.6 Comparison of global and local features

In this section we compare the retrieval performances of two contrasting approaches. In the first approach, a song is represented as a vector of global feature values, where each feature value represents a characteristic of an entire song. For this, we use the feature set that is described in section 4.2 and Appendix B.

In the second approach, a song is represented as the sequence of local events that was used to compute the global feature values, or that is closely related to

Subset	Global Features	Selected Global Features
1	50–80	55 68 50 53 59 66 51 70 54 60 67 69 65 64 62 75 56 71
2	1–3 5–8 10–20 24 25 27–33 37–43 48 50–80 88	3 7 8 10 13 24 25 28 29 33 38 39 40 43 48 52 53 54 55 56 58 59 60 62 65 66 68 69 70 75 77 78 80
3	45–47 81–86	45 47 81–86
4	4 9 21–23 34 35 36 44–49 81–86	4 35 36 44 46 47 49 82 84
5	1–88	2–15 20 24–31 35–46 48–51 53–55 54 55 57–59 64–68 70–72 78 81–84 87 88

Table 4.3: Feature subsets and selected features per subset. The indices refer to the list in Appendix B.

the global features. Thus, in the second approach, the time-order of the events is preserved and local comparison of melodies is possible.

We perform the comparison between the local and the global approach separately for pitch related features and rhythm related features. The respective local features are pitch intervals and duration ratios. In the first case, a note is represented by the melodic interval between the note and the previous note. In the second case the note is represented by its duration divided by the duration of the previous note.

We use the following five global feature subsets:

1. Global interval features vs. local interval features (section 4.6.2).
2. Global interval features and other global melodic features vs. local interval features (section 4.6.3).
3. Global duration-ratio features vs. local duration-ratio features (section 4.6.4).
4. Global duration-ratio features and other global rhythmic features vs. local duration-ratio features (section 4.6.5).
5. All global features vs. a combination of local pitch and metric features.

For subsets 1 and 3, the local and global features correspond directly, while subsets 2 and 4 add less closely related global features. Table 4.3 shows the indices of the features in each subset, along with the features that were selected by the feature selection algorithm.

4.6.1 Method

4.6.1.1 Global Features

For each of the five subsets we perform forward floating feature selection (Pudil et al., 1994) in the same way as is described in section 4.5. As a measure for the retrieval performance of the global feature subset under consideration, we compute the leave-one-out success rate for the Annotated Corpus using the nearest neighbor rule. Successively all 360 songs from the Annotated Corpus are assigned to the class of the closest song in the feature space. The success rate is the number of correctly classified songs as fraction of the total number of songs. We do this both for the Annotated Corpus only and for the full corpus of 4830 songs. In both cases, we only use the 360 songs of the Annotated Corpus as 'queries'.

4.6.1.2 Local Features

For the comparison of two sequences of musical events, we use the Needleman-Wunsch-Gotoh global alignment algorithm (Needleman & Wunsch, 1990; Gotoh, 1982). A more extensive explanation can be found in chapter 6. The normalized alignment score can be used as a measure for the similarity of two songs.

The information retrieval evaluation measure that corresponds to the leave-one-out success rate is the recognition rate, which measures the fraction of queries that have a relevant item at the top-position of their respective ranking lists. A ranking list is obtained by sorting the songs according to alignment score with the query song. Again, we compute this measure for both the Annotated Corpus and the full corpus, using only the 360 annotated songs as queries in both cases.

4.6.2 Interval Features

The subset of global features that is directly related to the intervals between the notes consists of the interval features as defined by Jesser: features 50–80 (see Appendix B). Each of these features measures the occurrence-rate of a certain interval.

The subset of interval features that is most discriminative for the Annotated Corpus as obtained by the floating selection algorithm is shown in Table 4.3. The intervals corresponding to the features that were not selected are probably either too rare or too common.

For the alignment we take sequences of intervals between the successive pitches, regardless of the duration of the notes. Therefore, in this case, we have a perfect correspondence between the information used for the global features and for the alignment.

4.6.3 Melody Features

The subset of melody features consists of all interval features used in the previous section along with a number of higher-level pitch-based features from the sets of Jesser, Steinbeck and jSymbolic as shown in Table 4.3.

For the alignment, again, we take sequences of intervals between the successive pitches.

4.6.4 Global Duration-Ratio Features

The subset of duration-ratio features consists of features that relate to the relative lengths of the notes. Jesser's features relate the duration of the note to the shortest duration in the melody, while Steinbeck's features relate the duration of a note to the duration of the previous note.

For the alignment we take the sequence of duration-ratios. The duration-ratio of a note is the duration of the note as fraction of the duration of the previous note.

4.6.5 Rhythmic Features

The subset of rhythmic features contains the inter-onset-ratio features along with some other, higher-level, rhythmic features as shown in Table 4.3.

For the alignment, again, we take the sequence of duration-ratios.

4.6.6 All Features

In this case, we take all global features as well as the optimal subset that was found starting with all global features.

For the alignment, we use a combination of three types of features: pitchband, IMA and phrase-position, which are related to pitch, metric weight and phrase structure. When comparing two songs, pitch-band measures the difference in pitch between a note in the first song and a note in the second song. The larger the difference, the less similar the notes are considered. IMA (inner metric analysis) computes a metric weight for each note. The weights of a note from the one song and a note from the other song are compared resulting in a similarity score. Phrase-position compares the horizontal position of two notes within their respective phrases. A detailed explanation of this configuration and of how it is used to align two melodies is provided in chapter 6.

4.6.7 Results

Table 4.4 shows the leave-one-out success rates for each of the configurations for both the local and the global approach.

In general using pitch-related features rather than rhythm-related features result in better retrieval performance, both in the global and in the local cases. The results show clearly that the alignment approach is both more accurate and better scalable. On the small corpus, the global-feature approach achieves a good performance when the optimal subset of all involved features is used.

4.7 Conclusions

From the experiments with computational features, we draw several conclusions.

From the evaluation of individual features, we conclude that there is not a single feature that is discriminative for all tune families. Only a few features have some discriminative power for specific tune families within a small dataset.

Approach	Features	Annotated data-set	Full data-set
Global features	interval features	0.52	0.34
	selected interval features	0.59	0.34
	melody features	0.67	0.44
	selected melody features	0.74	0.46
Alignment	interval sequences	0.92	0.84
Global features	all IOR features	0.38	0.16
	selected IOR features	0.39	0.15
	all rhythmic features	0.49	0.21
	selected rhythmic features	0.55	0.26
Alignment	IOR sequences	0.74	0.46
Global features	all features	0.74	0.52
	selected features	0.82	0.55
Alignment	Pitch-band, IMA, Phrase-position	0.99	0.92

Table 4.4: Success rates for each configuration. For the global approach, the leave-one-out success rates of the nearest-neighbor classifiers are shown and for the alignment, the recognition rate.

The contents of the selected feature subsets in section 4.5 differ to a large extent. The most common feature in all selected subsets has been selected for only 6 out of 26 tune families. This indicates that each tune family is in a specific way distinct from the rest of the corpus concerning global features.

In all experiments with global features, success rates decrease dramatically for the larger data set. This indicates that the global feature approach for recognition of melodies only can be taken if the data set contains a small set of tune families.

Alignment of sequences of local features yields better classification results than comparison of global features of the same kind. The alignment approach is also better scalable.

In the analysis of the manual annotations as described in the previous chapter, we observed that recurring, characteristic motifs are important for recognizing melodies. There are many kinds of motifs: a rhythmic figure, an uncommon interval, a leap, a syncopation, and so on. Therefore it is not possible to grasp the discriminative power of motifs in only a few features. Besides that, global features are not suitable to reflect motifs, which are local phenomena. This is an important shortcoming of the approach based on global features.

As a general conclusion, we state that the global features that are known from literature are of limited use for the retrieval of related melodies from a large data base, while in small data sets good results could be obtained for some tune families only.

Chapter 5

On Measuring Musical Style—The Case of Some Disputed Organ Fugues in the J.S. Bach (BWV) Catalog

The experiments in the previous chapter show that global features of melody are not suitable for addressing the research questions of Folk Song Research concerning the ordering and classification of large corpora of melodies. To show that this does not imply that global features are inappropriate for any music classification task, the current chapter presents a study in which a well-chosen set global features is successfully used to address another musical problem, namely authorship attribution based on properties of personal styles of composers.

Contribution. This study throws new light on authorship problems of several organ fugues that are currently ascribed to J.S. Bach by comparing values of global features of musical style. In a previous publication, the authorship of the fugue in F minor (BWV 534/2) was investigated (Backer & van Kranenburg, 2005), while in the current study other fugues are involved as well: BWV 536/2, 537/2, 555/2, 557/2, 558/2, 560/2, and 565/2.

5.1 Bases for Composer Attributions

In Historical Musicology the question of authorship is important. In order to present overviews of the most important composers and their works, it is necessary to know who composed what. When preparing a critical edition of the works of a certain composer, decisions about disputed compositions have to be taken. Problems may be caused by conflicting attributions among plural sources, the lack of an authoritative source contemporary with the composer, an incomplete source, or an anonymous source which tradition holds to be by the composer. Attributions of the same work to multiple composers is a common phenomenon of European works of the fifteenth through nineteenth centuries.

Both external and internal evidence may be used to solve authorship problems (Love, 2002). In many cases, however, external evidence of a decisive nature is lacking. Here, internal evidence becomes more important. For music, stylistic evidence seems the most important kind of internal evidence. In order to assess stylistic evidence one must have a model that is able to represent musical styles in such a way that specific instances of it can be associated with a composer's personal style to a unique degree. In manual practice, proof by example is often used to support an attribution on stylistic grounds. The most pertinent feature may be a distinctive motif or chord progression that is present both in the disputed work and in an undisputed composition. Such similarities might, however, be occasional. If we want to support an attribution in a statistically sound way, we have to use events which occur frequently (such as notes and intervals).

Computer-based assessment of musical authorship was first extensively explored by Trowbridge (1982; 1985), who revealed differences in style among four Renaissance composers (Gilles Binchois, Antoine Busnois, Guillaume Dufay, and Johannes Ockeghem) by comparing the average values of 16 quantifiable features (Trowbridge, 1985). The repertory evaluated consisted of 92 Renaissance chansons, of which twothirds exist in a single manuscript copy with scribal attribution. Many of the rest are anonymous in at least one source, a few in more than one source. Many of the features are coincidentally similar to those used here. They included melodic intervals, harmonic intervals, chord types, bass progressions, root progressions, root distributions, prepared dissonances, chord durations, chord motion, texture reduction, melodic direction, rhythmic activity, average melodic range, relative melodic motion, voice crossing, and harmonic range. A good account of still earlier systems for quantitative analysis is given in Trowbridge (1982). For polyphonic music Trowbridge's thesis is by far the most thorough and comprehensive of its time. Far more prevalent today are studies that isolate and analyze musical features for differentiation of pieces by genre (e.g., McKay & Fujinaga, 2004), mood (e.g., Dannenberg et al., 1997), or idiosyncratic traits of individual composers (e.g., Cope, 1991, 1998).

5.2 A Machine Learning Approach to Stylistic Assessment

Our approach to authorship problems employs machine learning algorithms (see Duin & Tax, 2005). These algorithms learn characteristics of musical styles from representative examples, and are then able to use the obtained knowledge to classify previously unseen compositions. In an earlier publication the authorship of the fugue for organ in F minor (BWV 534/2) was evaluated (Backer & van Kranenburg, 2005). In the current chapter, another classification algorithm is used and the dataset extended with eight additional disputed fugues listed in the Schmieder (BWV) catalog (Schmieder, 1990), and with six control compositions by Johann Peter Kellner.

5.3 Modeling Musical Style

In *Style and Music* Leonard Meyer (1996) developed a theory of musical style that can be used as a starting point for studies that compare musical styles algorithmically. He defines style as a replication of patterning, whether in human behavior or in the artifacts produced by human behavior, that results from a series of choices made within some set of constraints (Meyer, 1996, p. 3). In the process of composing, a composer is subjected to certain constraints while making his choices. Meyer distinguishes three levels of constraints. Laws (1) are universal. One cannot, for example, ask a piccolo to play a contra G. Rules (2) are intra-cultural. It is in the rules that music from the Renaissance differs from music from the Baroque. Strategies (3) are constraints to which the composer subjects himself within the rules of a certain culturally established style. Thus it is in the strategies that the music of G.F. Handel differs from the music of G.Ph. Telemann.

Not all strategies reside on a conscious level. Certain patterns are ingrained during the training and development of a composer and are not replicated consciously every time during the process of composing. Meyer indicates the necessity of statistics: since all classification and all generalization about stylistic traits are based on some estimate of relative frequency, statistics are inescapable (Meyer, 1996, p. 64). It can be expected that each composer has idiomatic, countable patterns that are more often replicated in his works than in compositions by other composers. The task is to find features in which such patterns are reflected.

5.4 The Dataset

5.4.1 Selected Features

There is no well tested theory available that predicts which features have to be used to solve a particular authorship problem. Therefore, we do an 'educated guess' at features that may have discriminative power. The subset of features that can be used to solve the authorship problem in question will be selected algorithmically.

Small scale features are preferable, because the algorithms to extract them are less complicated and the results less ambiguous. It is, for example, not obvious how to quantify the extent to which a composition resembles a certain sonata form, but it is less difficult to count the number of thirds. Because in the current study we are dealing with polyphonic music (fugues), the relations between the voices are important. The composer must know, for example, whether a dissonant interval can be written between two voices, how long that interval is allowed to sound, and what can follow. It can be expected that a composer develops certain strategies to handle these situations. This can result in replicated patterns in the distances between the voices.

The following 20 features are chosen:

Features 1-9: Vertical intervals, as illustrated in Figure 5.4.1, weighed by duration. The total duration of all occurrences of each specific vertical interval is computed and at the end divided by the total duration of all intervals in all voice pairs. The intervals are folded onto one octave (e.g., a tenth is counted as a third). If the same pitch occurs in more than one voice, it is taken into account only once.

1. Seconds between parts

Figure 5.4.1: Examples of intervals between voices, illustrated in a fragment of J.S. Bach's fugue in F major BWV 540/2 (bars 25–20).

Figure 5.4.2: Some parallels between the voices.

2. Thirds between parts
3. Perfect fourths between parts
4. Augmented fourths between parts
5. Diminished fifths between parts
6. Perfect fifths between parts
7. Sixths between parts
8. Sevenths between parts
9. Octaves between parts

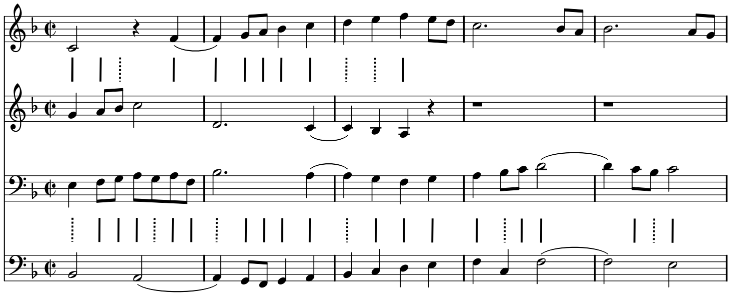


Figure 5.4.3: Examples of consonants (solid) and dissonances (dashed) between voices.

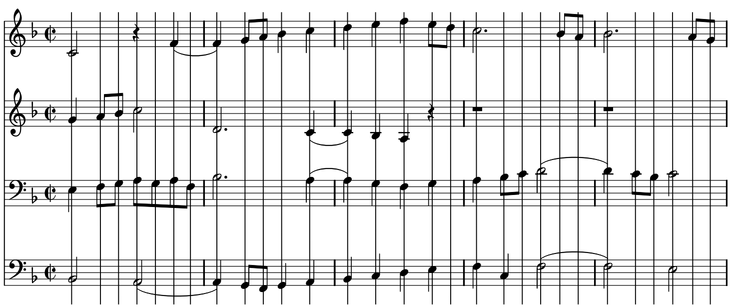


Figure 5.4.4: Time slices.

Features 10–12: Parallel motion, as illustrated in Figure 5.4.2. The quantity of parallel thirds, fourths, and sixths is computed in the same way as for Features 1-9. The total duration of all intervals involved in these parallels is computed and divided by the total duration of all intervals in all voice pairs.

10. Parallel thirds between parts

11. Parallel fourths between parts

12. Parallel sixths between parts

Features 13–15: Dissonance treatment, as illustrated in Figure 5.4.3. Perfect primes, minor and major thirds, perfect fourths, perfect fifths, and minor and major sixths are considered consonant. A fourth is considered dissonant if it is between the lowest voice and one of the upper voices. All other intervals are considered dissonant.

13. Suspension resolved stepwise in lower voice. The total duration of all intervals involved in such suspensions is computed and divided by the total duration of all intervals in all voice pairs.
14. Dissonance. The fraction of the score in which the sonorities are dissonant.
15. Bars beginning with dissonance. The percentage of bars that begin with a dissonant sonority.

Feature 16. Voice density. The average number of voices active in the composition. This is normalized for the total number of voices. Only bars that are strictly polyphonic (i.e., those in which more than one voice is active and in which each voice has not more than one note at the same time) are taken into account.

16. Voice density

Features 17–19: Entropy measures. Computed according to the concepts harmony and sonority as defined by Robert Mason (1985) and Shannon's entropy formula (1948). In the definition of Mason, a sonority is a certain type of chord, regardless of inversion, pitch, or doubling of tones. Each sonority has a unique number. The only difference between harmony and sonority is that in the case of harmony the pitch is taken into account. For example, the F-major and G-major triads are the same sonority, but different harmonies.

17. Harmony entropy. For each harmony the probability of occurrence is estimated by computing the total duration of all occurrences of the harmony and dividing that by the total duration of the piece. From these estimated probabilities, the entropy is computed.
18. Pitch entropy. Of each pitch the frequency of occurrence is estimated by computing the total duration of all occurrences and dividing that by the total duration of all pitches. From these estimated probabilities, the entropy is computed.
19. Sonority entropy. This feature is computed in the same way as Harmony entropy.

Composer	Compositions
J.S. Bach	BWV 535a/2, 535/2, 538/2, 540/2, 541/2, 542/2, 543/2, 545/2, 547/2.
J.L. Krebs	Fugue in C minor (I, 2), E major (I, 5), F minor (I, 6), G major (I, 8), F major (II, 13), F minor (II, 14), F minor (II, 15), B flat major (II, 19).
W.F. Bach	Fk 33, 36, 37, Add. 211/1, Add. 211/2.
J.P. Kellner	O08:01, O08:06, O08:07, O08:[C], O08:[F], O10:02.
Disputed fugues	BWV 534/2, 536/2, 537/2, 555/2, 557/2-560/2, 565/2.

Table 5.1: The incorporated organ fugues. The J.S. Bach numbering follows Schmieder (1990); that for Krebs, the edition of organ music by Weinberger (1985); for W.F. Bach, the catalog of Falck (1956), with additions by Peter Wollny (1993); for Kellner, the catalog by Claus (1999). Two fugues by Kellner not yet listed in Claus's catalog start with the designation "O08." In order to give them separate identities, I have added the key in square brackets. (*/2 signifies the second movement (i.e., the fugue) of a prelude-fugue pair).

Feature 20. Time slice stability. The consistency of the length of successive time slices (e.g., the time interval between two changes in the music, as illustrated in Figure 5.4.4). Stability is computed by dividing the standard deviation of the lengths of the time slices by their mean length. This normalization is necessary in order to compare pieces with different time signatures. A low value means that the music is more like a steady stream, while a larger value indicates more diversity in rhythm. Bars which are not strictly polyphonic (see Feature 16) are ignored in computation.

20. Time slice stability

5.4.2 The Compositions

Four composers are represented in the control dataset: J.S. Bach (1685–1750), his son Wilhelm Friedemann Bach (1710–84), his student Johann Ludwig Krebs (1713–80), and Johann Peter Kellner (1705–1772), who was a great admirer of J.S. Bach and played an important role in the copying and transition of Bach's organ compositions. Not many other composers among the students and contemporaries of J.S. Bach might have composed fugues comparable to those of J.S. Bach. However, an assignment of a disputed fugue to one of these four composers does not lead to an attribution without further consideration. The possibility that a composer not represented in the dataset wrote the piece should be kept open. In general, it is desirable to have external evidence that points exclusively at only a few candidates before pursuing the stylistic approach in the hope of making a definitive attribution. Because of the time consuming process of data entry, not all fugues by J.S. Bach and J.L. Krebs were encoded. To lower the probability of incorporating misattributions somewhat, only the fugues of Kellner that appeared in print are incorporated. In the case of W.F. Bach, the included five fugues are the only ones suitable for our purpose. In all, 35 works of undisputed authorship were encoded. See Table 5.1.

5.5 Data-Analysis Methods

To increase the amount of data available for control purposes, each composition was cut into overlapping segments of 30 bars, such that Segment 1 contains bars 1–30, Segment 2 contains bars 2–31, etc. (see Figure 5.5.1 for a generalized view). To produce reliable values, the minimum length of a segment has to be around 30 bars (Backer & van Kranenburg, 2005). Since there is a large degree of redundancy from one segment to the next, however, the window measurements are not independent. This must be accounted for when applying machine learning algorithms. Bars that are not strictly polyphonic are ignored in the process of splitting.

Some features may be better suited for classification than others. Choosing the 'wrong' features may even lead to more confusion. Therefore, the floating forward feature selection algorithm proposed by Pudil et al. (1994) has been applied. This algorithm successively adds or removes one or more features in order to optimize a certain criterion.

In order to get an indication of the reliability of a classification algorithm, the error rate is estimated as follows: one composition is removed from the dataset, a classifier is trained on all other compositions, and the data points of the removed composition are classified. After this has been done for all compositions, the error rates are averaged. In this way the dependency of the data points is accounted for. For convenience, I will call this error rate the leave-one-composition-out error rate (LOCO error rate).

Because we are interested in the catalog of J.S. Bach, the styles are evaluated in pairs, each consisting of J.S. Bach and one of the other composers. For each

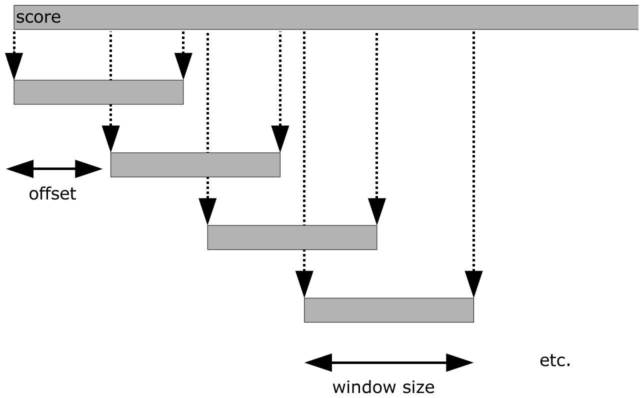


Figure 5.5.1: Schematic view of overlapping segments used in the analysis.

pair the optimal subset of features is selected using the Pudil algorithm. The criterion that is optimized is the LOCO error rate of a nearest neighbor classifier. A nearest neighbor classifier assigns the unknown object to the labeled object that is nearest in the feature space. The advantage of this classifier in the current situation is that no assumption is made about the distribution of the data points. Only local densities are used. To classify a composition, all individual segments are classified by the nearest neighbor classifier.

5.6 General Findings

For each pair, the optimal set of features that is selected by the Pudil algorithm is indicated in Table 5.2, along with the percentage of Bach-segments that has been misclassified and the percentage of segments from the other composer that has been misclassified. To give an impression of the data comparisons, scatter plots in Figures 5.6.1, 5.6.2, and 5.6.3 show for each comparison the two best musical features as selected by the Pudil algorithm.

5.6.1 J.S. Bach vs. J.L. Krebs

In the case of Krebs, the selected optimal subset consists of 12 features with an overall error rate of 1.5%, but when varying the desired size of the optimal subset, Pudil's algorithm shows that for sets with more than five features, the

Classes	Selected features	Misclassified J.S. Bach	Misclassified other
J.S. Bach vs. J.L. Krebs	Seconds between parts Thirds between parts Sevenths between parts Dissonance between parts Bars beginning with dissonance	4.5 %	2.1 %
J.S. Bach vs. J.P. Kellner	Octaves between parts Stepwise resolved suspensions Time-slice stability	0 %	3.2 %
J.S. Bach vs. W.F. Bach	Perfect fourths between parts Diminished fifths between parts Sevenths between parts Parallel fourths between parts Dissonance between parts Bars beginning with dissonance Pitch entropy Time-slice stability	1.6 %	19.3 %

Table 5.2: The selected feature subsets for each of the two-class problems with corresponding loco error rates.

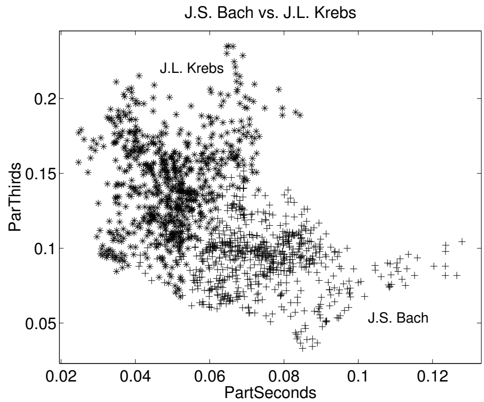


Figure 5.6.1: Projection of the segments onto the planes spanned by the two most important features for J.S. Bach (+) compared to J.L. Krebs (*).

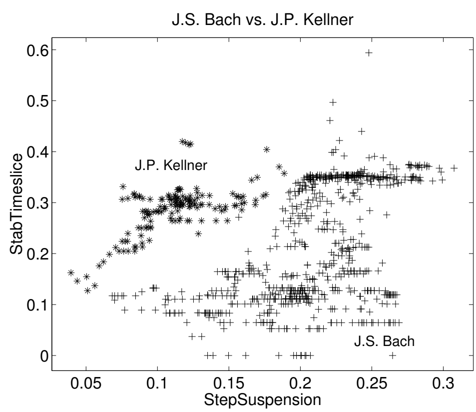


Figure 5.6.2: Projection of the segments onto the planes spanned by the two most important features for J.S. Bach (+) compared to J.P. Kellner (*)

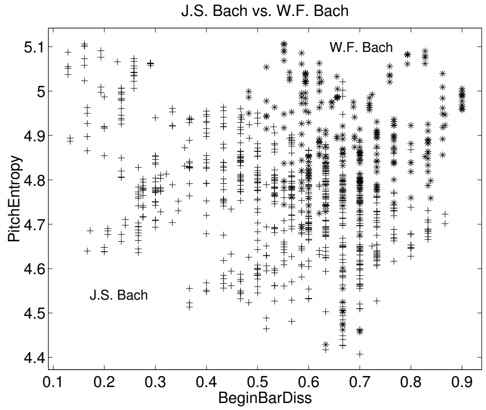


Figure 5.6.3: Projection of the segments onto the planes spanned by the two most important features for J.S. Bach (+) compared to W.F. Bach (*).

error rate decreases only marginally. The only composition that causes troubles with the set of five features is the fugue in G minor, BWV 542/2. 24 of the 80 segments are misclassified, while with the optimal set of 12 features, only one segment is misclassified. Because smaller feature sets are preferable, it is better to take the optimal subset with five features, and accept the partial misclassification of BWV 542/2. Hence, the subset that is indicated in Table 5.2 is the second best feature subset in terms of error rate, but a better subset in terms of size.

With the found subset, the differences between J.S. Bach and Krebs can be characterized as follows. Bach used more seconds and sevenths and fewer thirds than Krebs, and J.S. Bach's pieces contain more dissonances.

5.6.2 J.S. Bach vs. J.P. Kellner

For recognizing the styles of J.S. Bach and J.P. Kellner, three features proved to be sufficient. The J.S. Bach segments have more dissonances resolved by step. They also have a steadier rhythm than the Kellner segments. Kellner's O08:06 and O08:[F] utilize more octaves than the pieces by J.S. Bach.

BWV No. of work	J.S. Bach compared to	No. of segments classified as J.S. Bach	BWV No. of work	J.S. Bach compared to	No. of segments classified as J.S. Bach
534/2	Krebs	34 / 102	555/2, 557/2, 558/2, 559/2, 560/2	Krebs	5 / 84
	Kellner	54 / 102		Kellner	61 / 84
	W.F. Bach	94 / 102		W.F. Bach	84 / 84
536/2	Krebs	94 / 135	565/2	Krebs	24 / 50
	Kellner	134 / 135		Kellner	46 / 50
	W.F. Bach	135 / 135		W.F. Bach	50 / 50
537/2	Krebs	74 / 95			
	Kellner	95 / 95			
	W.F. Bach	75 / 95			

Table 5.3: Classification results for the disputed fugues. For each fugue the number of segments that are classified as J.S. Bach is shown as fraction of the total number of segments in the piece.

5.6.3 J.S. Bach vs. W.F. Bach

Eight features are needed for optimal classification. It appears that the error is mainly caused by misclassification of Fk 33 (16 out of 51 segments) and Fk add. 211/2 (27 out of 51 segments). The combination of the selected features is too complex to allow one to characterize the differences between J.S. and W.F. Bach in a few sentences. Apparently, style discrimination for this pair is more difficult than for the other two.

5.7 Classification of the Disputed Works

The classification results for the disputed fugues are shown in Table 5.3. The comparisons will now be discussed individually, since the sets of parameters which proved to be most significant varied from work to work.

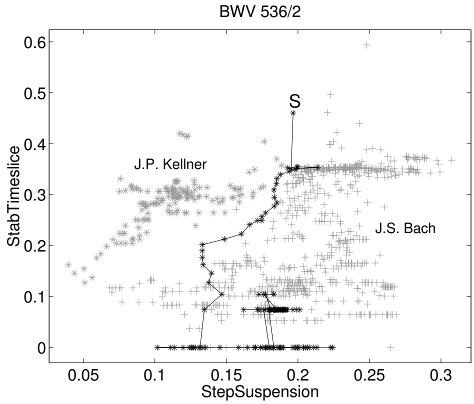


Figure 5.7.1: Projection of the trajectory of BWV 536/2 onto the plane spanned by the two most important features for the pair of composers. The first segment of the fugue is marked by 'S'.

5.7.1 BWV 534/2

Although early writers on the organ works of Bach like Philipp Spitta (1916, p. 583), Albert Schweitzer (1955, p. 238) and Hermann Keller (1950, p. 79f) did not esteem the Fugue in F Minor, BWV 534/2, as much as other fugues, the authorship was not doubted. In 1985 David Humphreys rejected this fugue as a composition by J.S. Bach. Dirksen (2000) suggested W.F. Bach as the actual composer. From Table 5.3 it is clear that the attribution to W.F. Bach is not supported. It is more difficult to adjudicate between J.S. Bach and Kellner. Classification between J.S. Bach and J.L. Krebs points strongly in the direction of Krebs, but the attribution to Krebs is not really convincing. If Krebs had composed the piece, the part of it that is misattributed (33%) is larger than for all involved undisputed fugues by Krebs. This fugue may have been composed by yet another composer.

5.7.2 BWV 536/2

The fugue in A major, BWV 536/2, has been rejected as a composition of J.S. Bach by David Humphreys (1989). In the earliest source, J.P. Kellner is the writer of the prelude, but the fugue is in a later, anonymous hand. Humphreys suggested J.P. Kellner or one of his pupils as the composer. The result of the

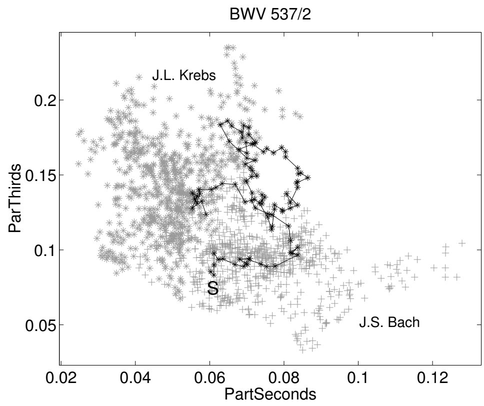


Figure 5.7.2: Projection of the trajectory of BWV 537/2 onto the plane spanned by the two most important features for the pair of composers. The first segment of the fugue is marked by ‘S’.

J.S. Bach vs. J.P. Kellner classifier does not support the authorship of Kellner. Almost all segments are assigned to the class of J.S. Bach. The trajectory of this fugue in the plane that is spanned by the most important features (regularity in rhythm and stepwise resolved dissonances) is shown in Figure 5.7.1. After the exposition, the rhythm becomes more regular than in most other pieces by J.S. Bach (a value of zero for Feature 20 means that there is no variation at all in the combined rhythm of all voices). Therefore we conclude that Kellner is in all probability not the composer of this piece, but it is also not a typical J.S. Bach fugue.

5.7.3 BWV 537/2

A very interesting hypothesis about the fugue in C minor (BWV 537/2) was posed by John O’Donnell (1989). In the earliest source the first 90 bars are written down by Johann Tobias Krebs (1690–1762) and the remaining 40 bars by his son, Johann Ludwig. This is one of the reasons for O’Donnell to suppose that the piece was left unfinished by J.S. Bach and was completed by J.L. Krebs on request of his father, who was copying the score. The classifier assigns the last 13 segments to J.L. Krebs. These correspond almost exactly with the last 40 bars. The trajectory of the piece in the plane spanned by the two most important features (seconds and parallel thirds) is interesting. The trajectory

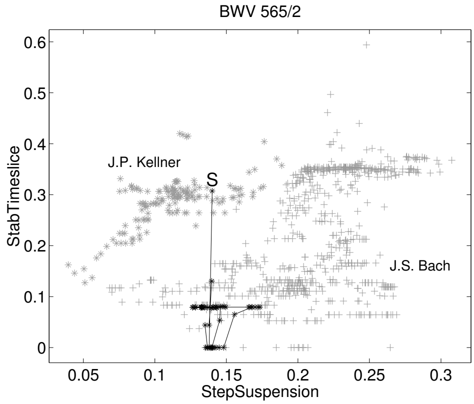


Figure 5.7.3: Projection of the trajectory of BWV 565/2 onto the plane spanned by the two most important features for the pair of composers. The first segment of the fugue is marked by 'S'.

starts in the cluster of J.S. Bach. From bar 60, a second, chromatic theme dominates the fugue. As soon as the segments contain bar 60 or higher, the trajectory goes into the cluster of Krebs, but with a relatively large number of seconds. The following part, in which the chromatic theme dominates all segments entirely, goes outside both clusters. Finally, the trajectory ends in the heart of the cluster of Krebs. A chromatic theme is rare in J.S. Bach's organ fugues. This might explain why the trajectory goes outside the J.S. Bach cluster early. Bach probably changed his strategies by writing more thirds, but Krebs was able to use his 'normal' amount of seconds and parallel thirds while composing the last 40 bars. So they treated the chromatic theme in a different way. In any case, the current results support the claim that this fugue was composed by two composers. The authorship of J.L. Krebs for the last 40 bars is likely. See Figure 5.7.2.

5.7.4 BWV 555/2, 557/2, 558/2, 559/2 and 560/2

These five fugues are part of the *Acht kleine Präludien und Fugen*. The other three fugues of this collection are too short to measure reliable features values (less than 30 bars). Because of the coherence of the group, they are treated as a single composition. The authorship of these eight little preludes and fugues has been discussed a lot. The relatively low quality has been an important reason for this. Several composers are suggested, among them J.L. Krebs (Keller, 1937, p. 67f). But, there is also a rejection of the authorship of Krebs (Tittel, 1966). The classification results in table 5.3 support the rejection of the authorship of J.S. Bach. Also W.F. Bach can be excluded. It can be concluded that from the currently involved composers, these fugues share most the characteristics of the style of J.L. Krebs. But, again, it might be very well possible that they are composed by another composer, whose style is not represented in the dataset.

5.7.5 BWV 565/2

The case of the fugue in D minor BWV 565/2 is interesting because it is part of the most famous organ work in existence, the Toccata and Fugue in D minor. Although this piece (especially its beginning) is known to almost everyone in western society as *the* organ piece by J.S. Bach, its authorship is disputed, mainly because the style of the work differs so much from all other organ works by J.S. Bach. Several theories have been posed, but it is still an unresolved question. Because the earliest source was written down by J.P. Kellner's student Johannes Ringk, Kellner might be considered a candidate. In an extensive study, Rolf Dietrich Claus (1998) concludes that Bach cannot be the composer. Neither does Claus make an attribution to Kellner. This is in accordance with the current results as shown in Table 5.3. The classification of half the piece as J.L. Krebs supports questioning the authorship of J.S. Bach, and in comparison with the style of Kellner, BWV 565 more resembles the style of J.S. Bach. The trajectory is shown in Figure 5.7.3. Apart from the first segment, the style is rather consistent under this projection. Although the proportion of dissonances that is stepwise resolved is in accordance with some pieces by Kellner, the regularity of the combined rhythm of all voices is clearly not.

5.8 Concluding Remarks

The study presented in this chapter shows that the proposed quantitative approach to the recognition of personal styles of composers results in valuable additions to existing authorship disputes (in this case about some of the disputed organ fugues in Bach's catalog). Although the current results do not offer enough evidence to draw final conclusions for these compositions, it is clear that this method is helpful in finding and testing hypotheses about differences in personal styles. Because the available data (scores) are extensively used, these hypotheses are firmly based in the scores. This is unlike many 'traditional' authorship studies, in which proof by example is the best achievable.

Chapter 6

Musical Models for Folk Song Melody Alignment

Sequence alignment algorithms offer the possibility to incorporate musical knowledge in the form of appropriate substitution scoring functions and a representation of the melody as a sequence of musical symbols. In this chapter a number of musically motivated substitution scoring functions are proposed and evaluated.

Contribution. We show that sequence alignment algorithms can successfully be employed to retrieve related folk song melodies from a large database by incorporating musical knowledge in the substitution scoring functions. Furthermore, we show how retrieval results based on the alignment scores, lead to improvements of the labeling of the songs in the Annotated Corpus.

6.1 Introduction

In this chapter we use alignment algorithms to measure the similarity of melodies. Alignment algorithms are widely used for comparison of sequences of symbols. Creating an alignment is a way to relate two sequences with each other by finding the best corresponding parts. Especially in the field of computational biology, where they are used to find corresponding patterns in protein or nucleotide sequences, many algorithms that align sequences have been developed.

Sequence alignment is also suitable for assessing musical similarity for several reasons. Firstly, music unfolds in time, therefore a model of music as a one-dimensional sequence of events seems appropriate. Secondly, manual alignments have extensively been used in folk song research to evaluate relations between melodies. Thirdly, structural alignment is a prominent model in cognitive science for human perception of similarity (Goldstone, 1994).

Most alignment algorithms use a dynamic programming approach. One of the earliest variants is the Levenshtein distance (Levenshtein, 1966), which is an edit distance: it computes how many operations are needed to transform one sequence into another. Needleman and Wunsch (1990) proposed an algorithm that finds an optimal alignment of two complete sequences. The quality of an alignment is measured by the alignment score, which is the sum of the alignment scores of the individual symbols. If we consider two sequences of symbols $\mathbf{x} : x_1, \ldots, x_i, \ldots, x_n$, and $\mathbf{y} : y_1, \ldots, y_j, \ldots, y_m$, then symbol x_i can either be aligned with a symbol from sequence $\mathbf{y}$ or with a gap. Both operations have a score, respectively the substitution score and the gap score. The gap score is mostly expressed as penalty, i.e. a negative score. The optimal alignment and its score are found by filling a matrix D recursively according to:

$$D(i,j) = \max \begin{cases} D(i-1, j-1) + S(x_i, y_j) \\ D(i-1, j) - \gamma \\ D(i, j-1) - \gamma \end{cases} , \tag{6.1.1}$$

in which $S(x_i, y_j)$ is the substitution scoring function, γ is the gap penalty, $D(0,0) = 0$, $D(i,0) = -i\gamma$, and $D(0,j) = -j\gamma$. $D(i,j)$ contains the score of the optimal alignment up to x_i and y_j and therefore, $D(m,n)$ contains the score of the optimal alignment of the complete sequences. We can obtain the alignment itself by tracing back from $D(m,n)$ to $D(0,0)$; the algorithm has both time and space complexity $O(nm)$. In our modeling, we use an extension of the algorithm proposed by Gotoh (1982), which employs an affine gap penalty function without loss of efficiency. In this approach, the extension of a gap gets a lower penalty than its opening.

Mongeau and Sankoff (1990) were among the first to adapt alignment algorithms to music. They used an extended version of the Needleman-Wunsch algorithm, in which they added two new operations: fragmentation, in which one symbol from sequence $\mathbf{x}$ is aligned with more than one contiguous symbols from sequence $\mathbf{y}$, and consolidation, which is the opposite. Their scoring function takes both pitch and duration into account. Mongeau and Sankoff's approach has been quite influential, e.g. the search algorithm implemented in the search engine MELDEX (McNab et al., 1997) is based on this algorithm.

Gómez et al. (2007) successfully tested a modified version on a MIREX dataset. In general, alignment algorithms have often been used to match short melodic phrases against a larger database (Adams et al., 2004; Gómez et al., 2007; Lemström & Ukkonen, 2000; McNab et al., 1997; Uitdenbogerd & Zobel, 1999). Typical tasks addressed with this approach are to find a tune in the database using query by humming (Adams et al., 2004), different arrangements of a piece (Uitdenbogerd & Zobel, 1999), or similar incipits given to the query (Gómez et al., 2007). We use the alignment between complete melodies in order to find folk song melodies that belong to the same tune family. The similarity relations that have to be modeled originate in the oral transmission of folk songs and differ from those in the previous tasks.

Contribution. In this chapter we model various features of music as substitution scoring functions, which we incorporate in the Needleman-Wunsch-Gotoh algorithm. Using a set of melodies that are well-described regarding their different kinds of similarity relations, we evaluate the influence of these scoring functions on the retrieval performance. Our best scoring function combines several musical features and outperforms well-known approaches from literature.

6.2 Data

6.2.1 The Data Set

The set of melodies we use is part of the collection of Dutch folk song melodies that was introduced in section 1.1.

The melodies are grouped in so-called tune families. All melodies in one tune family are considered to be historically related through the process of oral transmission, in which melodies were learned from listening and imitation rather than from written sources. Since the history of each tune family is not fully documented, it is often not known whether two melodies are historically related. Instead, musicological experts decide whether melodies belong to the same tune family by assessing their melodic and textual similarity. In order to make the experts' musical intuition behind the similarity assessments explicit, we developed an annotation system (as described in Chapter 3). For the Annotated Corpus (consisting of 360 songs in 26 tune families) several dimensions of perceived similarity (contour, rhythm, motives, lyrics) have been numerically rated by the musicologists such that the similarity between the most typical melody of a tune family (the reference melody) and all other members of the tune family has been described. The 26 tune families were chosen from the larger collection by an expert such that this set contains a representative diversity of similarity relations between members of a tune family. Comparing the annotations with the retrieval performance of alignment algorithms allows a detailed understanding of the success or failure of the models based on musicological insights.

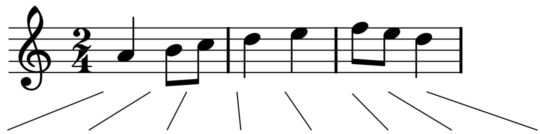
In this chapter we use the Annotated Corpus along with 4470 encoded songs that already have been classified into tune families other than the ones in the Annotated Corpus. This allows us to examine whether the annotated songs could be retrieved from a database that contains a large amount of other melodies from the same tradition. In the entire data set of 4830 songs, 1472 tune families only have one member, 300 tune families have 2 members and 78 tune families have 10 or more members. In total, the data set contains 2142 tune families.

The songs are identified by their record number in the Dutch Song Database and by the strophe number. Thus, 70078_02 denotes the second strophe of song 70078. A song can be consulted by entering the record number in the search field on the website of the Dutch Song Database (http://www.liederenbank.nl).

6.2.2 Representation of Melodies

For applying alignment algorithms, a melody has to be represented as a sequence of symbols. In our representation, each symbol represents a note. A symbol has a number of attributes, which are: pitch (in base-40 encoding, see Hewlett, 1992), duration (rational number), score time (rational number), time in bar (rational number), onset (integer), current bar number (integer), current phrase number (integer), upbeat (boolean), current meter (rational number), free meter (boolean), accented (boolean), inter-onset-interval ratio (rational number), normalized metrical weight (real number in $[0, 1]$) and the time position within phrase (real number in $[0, 1]$). These attributes are used to compute substitution scores or other attributes. Figure 6.2.1 shows an example of the representation.

Our model of metrical accents, which determines the value of the attribute accented, is simple and may need extension in the future. Based on the encoded time signature, we distinguish two metrical levels: either accented or not accented. The first note of any group of two in a double meter and the first beat



Pitch:	a^1	b^1	c^2	d^2	e^2	f^2	e^2	d^2
Duration:	1/4	1/8	1/8	1/4	1/4	1/8	1/8	1/4
Scoretime:	0	1/4	3/8	2/4	3/4	4/4	9/8	5/4
BarTime:	0	1/4	3/8	0	1/4	0	1/8	1/4
Onset:	0	2	3	4	6	8	9	10
Bar:	0	0	0	1	1	2	2	2
Phrase:	0	0	0	0	0	0	0	0
Upbeat:	false	false	false	false	false	false	false	false
Meter:	2/4	2/4	2/4	2/4	2/4	2/4	2/4	2/4
Free meter:	false	false	false	false	false	false	false	false
Accented:	true	false	false	true	false	true	false	false
IOI ratio:	[1]	0.5	1	2	1	0.5	1	2
Metr.weight:	1	0.853	0.382	0.853	1	0.853	0.382	0.853
Phrasepos:	0	0.2	0.3	0.4	0.6	0.8	0.9	1

Figure 6.2.1: Representation of a melody.

in any group of three beats in a triple meter is considered accented. All other notes are unaccented. Thus, in songs in free meter, or in songs with additive or asymmetrical meters,[1] which are very uncommon in this corpus, all notes are unaccented in this model.

Phrase boundaries have been annotated by the encoders.

6.2.3 Inconsistencies in Transcriptions

Transcription of audio recordings into musical notation depends to a large extent on the perception of the transcriber. Although explicit guidelines can be established in order to have consistent transcriptions, full systematization of the process of transcription is not possible. Therefore, we consider two types of differences between transcriptions of variants of the same song. At the one hand, there are the differences between the songs as being sung (actual differences), and at the other hand there are differences that are introduced in the process of transcription (representational differences). Not all actual differences are represented in the transcriptions. For example, a difference in tempo is in many cases not visible in the scores if the two variants are transcribed

[1] Asymmetrical meters consist of stacked groupings of dissimilar metrical groups.

in the same meter. Vice versa, not all transcriptional differences reflect actual differences in the song instances as being recorded.

The challenge in designing retrieval strategies for folk song melodies is to be robust against both types of differences, however, in different ways. The interesting type of differences from the perspective of Folk Song Research is the actual difference. The extent of actual variation should be reflected in the similarity score, but similarity measures should be invariant for representational differences. In other words, if two songs only differ representationally, they should be considered exactly the same.

In our case, the transcriptions have been made by various persons over a time span of several years. Therefore, we expect representational differences in the transcriptions. To get an idea of the types of representational differences that are typical in a collection of transcribed folk songs, we now give an overview of some common problems. It must be noted, however, that none of these problems can be fully reduced to representational differences without further consideration. In some cases there might be a reason for choosing one representation for the one variant and another representation for the other variant, while in other cases, there is no underlying actual difference. We cannot infer this from the transcriptions. Comparison with the audio recording is necessary, but then as well, more than one interpretation might be possible. For now, we consider problems in the retrieval that are caused by one of the transcriptional inconsistencies described in the remainder of this section shortcomings of the algorithm, which have to be resolved in order to get a well-performing retrieval system for folk song melodies.

6.2.3.1 Absolute Pitch

Problems Similar melodies could have been notated in various keys. Since human perception of melodic similarity is to a large extent independent of absolute pitch, our methods should be transposition-invariant.

A common procedure in the transcription of folk song melodies, is to transpose all melodies such that they have tonic or final G. These two options are not exactly the same. There are songs that do not end on the tonic, but, e.g., on the third, such as the first example in Figure 6.2.2. In the transcription of the melodies from *Onder de Groene Linde*, the choice was made to transpose all melodies to the tonality of G. This could cause problems with older, modal, melodies, which, however, are rare in this corpus. In the Annotated

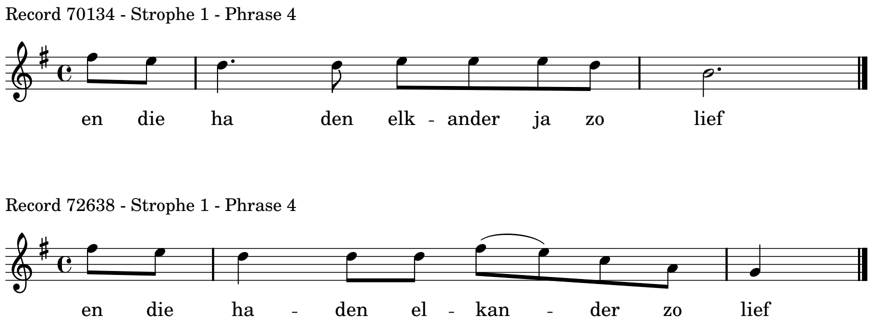


Figure 6.2.2: Different scale degrees for ending notes.

Corpus are transcriptions of songs from written sources as well. These are notated in various keys. Therefore, it is necessary to make similarity measured transposition-invariant.

Possible Solutions A common solution to this problem is to take the intervals between the notes instead of the absolute pitches. This could have severe consequences for the evaluation of similarity. Consider, for example, the two fragments in Figure 6.2.3. In absolute pitches, there are six differences, while in relative pitch (intervals) there is only one difference. The question for Folk Song researchers would be whether the transition from the one fragment to the other could be considered just one transformation in the process of oral transmission, or more transformations.

Another approach is to represent the melody with scale degrees instead of absolute pitches. For this, at any place in the melody the key must be known, which is not trivial.

Aside from using interval sequences, we also take another approach: a pitch histogram for both melodies is created that indicates for each pitch the total duration during the song. Then the shift at which the normalized histograms have maximal intersection is computed. For this procedure it is not necessary to know the absolute keys of the two melodies. Since the pitches are represented in base-40 encoding, the shift of the histogram can be interpreted as the interval with which the one melody should be transposed in order to compare it to the other.

To test this approach, we use all 2882 pairs of related melodies in the annotated corpus, and check whether the computed shift corresponds to the difference be-

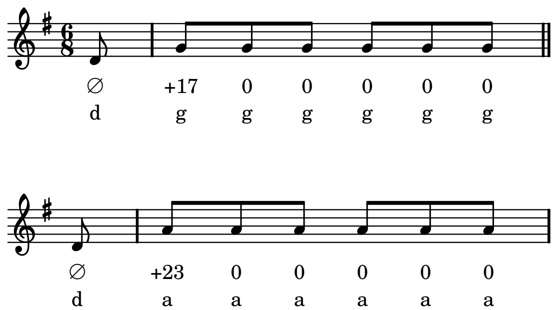


Figure 6.2.3: Absolute versus relative pitch.

tween the annotated keys of the two songs. This test yields a success rate of 96.4%. All but one of the 103 errors is a fifth or a fourth wrong, which corresponds to related keys, and for only 8 errors a reference melody is involved. Therefore, this approach seems a good solution to the problem of absolute vs. relative pitch.

In all cases in which absolute pitch is involved, we apply the transposition based on the histogram shift.

6.2.3.2 Length of cadence notes

Problems In the songs as being sung, cadence tones, which are the last stressed notes of the phrases, have very irregular durations. The pause between phrases often serves as an moment to get new breath, or to take some time to remember the next phrase. If one would try to notate the duration of these cadence tones exactly as being sung, one would need devices like meter changes and rests. Since there are several solutions to the problem of transcribing the lengths of the cadence tones, the irregularity in the recordings is reflected in the transcriptions. The choices for the notation concerning rests seems to be taken independently by different transcribers. For example, in the songs in Figure 6.2.4, there is an inconsistency in the last note of the second phrase: a dotted half note versus a quarter note and two quarter rests. The choice for a half note in 71669 extends the phrase with a third of its total duration, which is quite drastic.

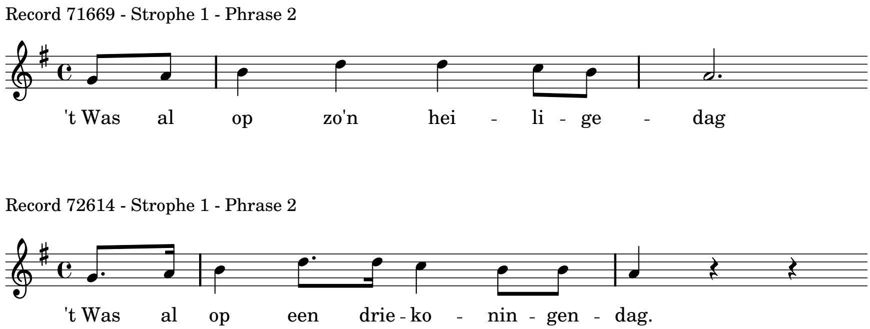


Figure 6.2.4: Inconsistencies in cadence notes.

Possible solutions The solve this problem, at least partially, we replace all rests with an extension of the preceding note. This seems a suitable procedure for a corpus of folk song melodies.

6.2.3.3 Rhythmic Unit

Problems In the process of transcription, the choice of meter is not always obvious. A common difference is that between the eight note and the quarter note as beat units. This is reflected in the notated meter as a difference between e.g., 6/8 and 6/4. Another common inconsistency is to have transcriptions both in 4/4 and 2/2. In many of these cases, the tempo of the respective recordings is not as different as the choices for the meters would suggest. An example illustrating most of the problems concerning meter is shown in Figure 6.2.5. In each of the transcriptions of four corresponding phrases, another solution is taken. There is no one-to-one correspondence between the meter and the beat unit. 73046 is notated in 6/4 and 72587 is notated in 3/2. However, in both variants the quarter note is the rhythmical unit for the syllables. Another difference is the choice for the irregular meter 5/8 in 73743, which actually seems an adequate meter for the other songs in Figure 6.2.5 as well.[2]

If the transcriber could not infer a meter from the recording at all, the song has been transcribed in 'free meter'. This applies to a substantial part of the songs from *Onder de Groene Linde*. Therefore, it is not possible to solely rely on the notated meter for finding the beat.

[2] The recordings can be consulted on the site of the Database of Dutch songs (http://www.liederenbank.nl) by entering the record number in the search field.

Figure 6.2.5: Metrical inconsistencies.

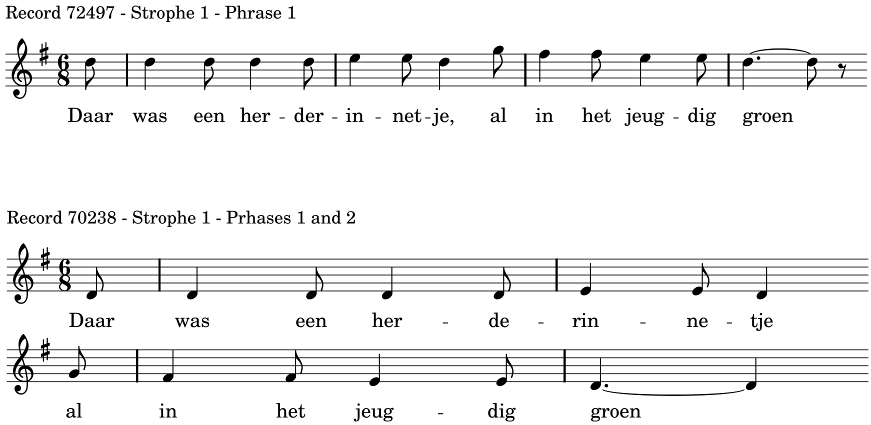


Figure 6.2.6: Relative over-segmentation.

Possible solutions The most common solution to the problem of comparing musical sequences in different tempi is to take the ratios between the durations of the notes instead of the absolute durations, i.c., the duration of each note is expressed as fraction of the duration of the previous note. This is called the duration ratio. A closely related measure for the relative duration of a note is the inter-onset-interval ratio. In this case, the difference in onset time between a note and the next note (the inter-onset interval) is taken as duration of the note, rather than the notated duration. Since we remove all rests, the ratios we use are inter-onset-interval ratios.

Another solution is to ignore duration altogether, considering just the sequence of pitches.

A third solution is to take the metric weight of the note as indicator of the importance instead of the duration. We use the Inner Metric Analysis (IMA) (Volk, 2008) to compute metric weights based solely on the onsets of the notes instead of on the notated meter. The quality of this solution depends on the quality of the metric weights as computed by the IMA algorithm. We compare all three of these solutions in the experiments.

6.2.3.4 Inconsistent Segmentations

Problems The hierarchical segmentation of the song transcription in stanzas and phrases has been performed during the process of transcription. The transcribers put each phrase on a separate line of music. It is no surprise that inconsistencies can be found here as well. A common inconsistency is a longer phrase in the transcription of one variant of the song that corresponds to two short phrases in the transcription of another variant. For example, in the second song in Figure 6.2.6, where each phrase is notated on a single staff, more phrase breaks have been entered than in the first song.

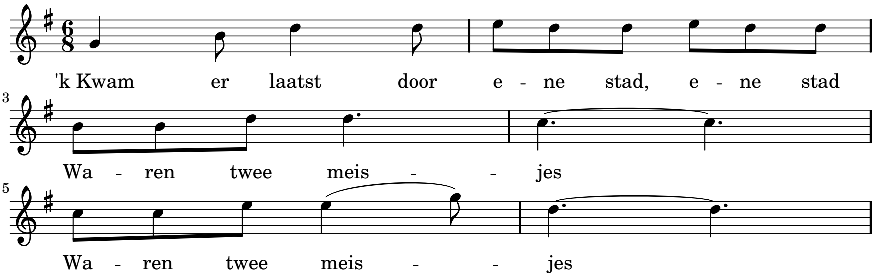
Another category of phrase ending problems is caused by the repetition of a small fragment. Does such a repetition belong to the phrase that precedes immediately, or has it to be interpreted as a phrase on its own? Figure 6.2.7 shows three different solutions to the segmentation problem caused by such a repetition. The seemingly inconsequent segmentation in 71478 is in accordance with the way the song has been sung. The singer takes a deep breath at the place where the transcriber has put the phrase break. Interestingly, the singers of the other two songs take a breath at the same place, but apparently the transcribers of these recordings did not take that as a decisive clue for segmentation.

Possible solutions The most obvious solution is not to use the phrase segmentation annotations at all. In the cases in which we do use the phrase annotations we have to be aware that we might 'lose' songs because of inconsistencies between query and database melodies. It is not possible to solve this problem in a satisfactory way in the context of the current chapter, since the segmentation of music is a complex and ambiguous process that needs further study.

6.2.4 Normalization of Alignment Scores

Since the score of an alignment depends on the length of the sequences, normalization is needed to compare different alignment scores. The alignment of two long songs results in a much higher score than the alignment of two short songs. Therefore, we divide the alignment score by the length of the shortest sequence. Thus, an exact match results in score 1, which is the maximal score. The scores are converted into distances by taking one minus the normalized score.

Record 76303 - Strophe 1



Record 72721 - Strophe 1

Record 71478 - Strophe 1

Figure 6.2.7: Segmentation of short repetitions

6.3 Substitution Scoring Functions

6.3.1 Single substitution scoring functions

In this section we introduce a number of substitution scoring functions for different musical dimensions. They result in substitution scores that are based on musicological knowledge. Each function takes two symbols of the melodic sequence as input. The output of each scoring function is in the interval $[-1, 1]$.

First, we introduce scoring functions that are based on pitch related features. The simplest scoring function determines whether two pitches are the same or not, hence the score is either maximal if the two pitches are considered exactly the same, or minimal if not:

$$S_{exactpitch}(x_i, y_j) = \begin{cases} 1 & \text{if } p(x_i) = p(y_j) \\ -1 & \text{if } p(x_i) \neq p(y_j) \end{cases}, \tag{6.3.1}$$

in which $p(x)$ is the pitch of symbol x in base-40 encoding.

In oral transmission, slight changes of pitches are likely to occur, therefore, we allow substitution with pitches that are within a pitch band with certain width:

$$S_{pitchb}(x_i, y_j) = \begin{cases} 1 - \frac{int(x_i, y_j)}{23} & \text{if } int(x_i, y_j) \leq 23 \\ -1 & \text{otherwise} \end{cases}. \tag{6.3.2}$$

We define $int(x, y) = |p(x) - p(y)| \bmod 40$. A perfect fifth is 23 in base 40 encoding. Thus, all intervals up to a perfect fifth get a positive substitution score and all larger intervals are considered a bad match.

Another way to express the distance of two pitches is by their harmonic relation. The substitution of consonances gets a higher score than the substitution of dissonances:

$$S_{harm}(x_i, y_j) = \begin{cases} 1 & \text{prime} \\ 0.5 & \text{consonance} \\ 0.5 & \text{augmented prime} \\ -1 & \text{dissonance} \end{cases}. \tag{6.3.3}$$

The intervals are taken modulo one octave. Consonances are minor and major third, perfect fourth, perfect fifth and minor and major sixth. The augmented prime gets a positive substitution score to favor alignments of songs that have both minor and major variants.

Furthermore, we define two substitution functions that are based on melodic contour, which is often considered important for melodic similarity, using either the contour of a single phrase or the contour of the entire melody:

$$S_{phrasecont}(x_i, y_j) = 1 - 2\left|p_{phr}(x_i) - p_{phr}(y_j)\right| \ , \tag{6.3.4}$$

$$S_{songcont}(x_i, y_j) = 1 - 2\left|p_{song}(x_i) - p_{song}(y_j)\right| \ . \tag{6.3.5}$$

Here $p_{phr}(x) \in [0,1]$ indicates the vertical position between the lowest and highest pitches of the phrase that x is part of, while $p_{song}(x) \in [0,1]$ indicates the vertical position between the lowest and highest pitches of the entire song. In determining the highest and lowest pitches, the notes in the upbeats of the phrases are disregarded, since these are very variable between variants of a song.

To test the interval-solution for transposition invariance, we define two scoring functions that use the interval with the previous note instead of absolute pitch:

$$S_{exactint}(x_i, y_j) = \begin{cases} 1 & \text{if } mint(x_i) = mint(y_j) \\ -1 & \text{if } mint(x_i) \neq mint(y_j) \end{cases} \ , \tag{6.3.6}$$

where $mint(x_i) = p(x_i) - p(x_{i-1})$ is the melodic interval between x_{i-1} and x_i for $i \geq 1$, and $S_{exactint}(x_0, y_j) = S_{exactint}(x_i, y_0) = 1$. The second interval-based scoring function is:

$$S_{intband}(x_i, y_j) = \begin{cases} 1 - \frac{intband(x_i, y_j)}{23} & \text{if } intband(x_i, y_j) \leq 23 \\ -1 & \text{otherwise} \end{cases} \ , \tag{6.3.7}$$

in which $intband(x_i, y_j) = |mint(x_i) - mint(y_j)| \bmod 40$. This scoring function allows a deviation of the melodic intervals up to a perfect fifth (which has value 23 in base-40 encoding).

Next, we define five scoring functions that are based on rhythmic features. In a simple approach using note durations, the score is maximal if the durations are the same, and minimal otherwise:

$$S_{exactdur}(x_i, y_j) = \begin{cases} 1 & \text{if } d(x_i) = d(y_j) \\ -1 & \text{if } d(x_i) \neq d(y_j) \end{cases} \ , \tag{6.3.8}$$

in which $d(x)$ is the duration of the symbol x.

Metric accents derived from the notated time signature describe a further aspect of the rhythmic structure of melodies. We define a substitution function that uses these metric accents in the following way:

$$S_{accent}(x_i, y_j) = \begin{cases} 1 & \text{if } a(x_i) = a(y_j) \\ -1 & \text{if } a(x_i) \neq a(y_j) \end{cases} , \tag{6.3.9}$$

in which $a(x)$ indicates whether the symbol x is accented or not (for defining accents see section 6.2.2).

A more complex notion of metric accents based on the rhythmic structure of notes instead of the time signature is provided by Inner Metric Analysis, as introduced in section 6.2.3.3. We define a scoring function that is determined by the metric weights of the notes, as computed by IMA:

$$S_{ima}(x_i, y_j) = 1 - 2\,|w(x_i) - w(y_j)| \ . \tag{6.3.10}$$

Here $w(x)$ denotes the metric weight of the symbol x scaled into the interval $[0, 1]$. For scaling, all weights are divided by the greatest weight in the song. The parameters for the IMA algorithm are the ones that are mostly used: $p = 2$, $l = 2$ (e.g. in Volk, 2008).

To cope with differences in notated durations, we define two scoring functions based on inter-onset-ratios:

$$S_{exactior}(x_i, y_j) = \begin{cases} 1 & \text{if } ior(x_i) = ior(y_j) \\ -1 & \text{if } ior(x_i) \neq ior(y_j) \end{cases} , \tag{6.3.11}$$

where $ior(x_i) = d(x_i)/d(x_{i-1})$, the ratio between the durations of x_i and x_{i-1} for $i \geq 1$, and $S_{exactior}(x_0, y_j) = S_{exactior}(x_i, y_0) = 1$, and

$$S_{iorratio}(x_i, y_j) = \begin{cases} -1 + 2\frac{ior(x_i)}{ior(y_j)} & \text{if } ior(x_i) \leq ior(y_j) \\ -1 + 2\frac{ior(y_j)}{ior(x_i)} & \text{if } ior(x_i) > ior(y_j) \end{cases} , \tag{6.3.12}$$

$S_{iorratio}(x_0, y_j) = S_{iorratio}(x_i, y_0) = 1$.

Furthermore, to use the information of phrase boundaries that is present in our data set, we introduce a scoring function based on the horizontal position of the notes within the phrase:

$$S_{phrpos}(x_i, y_j) = 1 - 2\,|phr(x_i) - phr(y_j)| \ , \tag{6.3.13}$$

in which $phr(x) \in [0,1]$ is a linear mapping of the horizontal position of symbol x between the onset of the first note and the onset of the last note of the phrase into the interval $[0,1]$. This substitution function helps to keep phrases together in alignments since the substitution of a note at the beginning of a phrase with a note at the end of a phrase gets a very low substitution score.

6.3.2 Combinations

The single substitution scoring functions defined in section 6.3.1 model isolated aspects of melodies. In order to model several aspects within one function to get closer to the multidimensionality of melodies, we combine substitution functions. We want alignments in which the aligned symbols are similar in all dimensions. Therefore, we multiply the individual scores:

$$S_{combination}(x_i, y_j) = \prod_{k=1}^{n} S_k(x_i, y_j) \ , \tag{6.3.14}$$

in which each $S_k(x_i, y_j)$ is scaled into the interval $[0,1]$, and the final score is scaled into $[-1,1]$ back again.

6.3.3 Gap Penalty Function

We use an affine gap penalty function in which the penalty for opening a gap is 0.8, and the penalty for extending a gap is 0.2. Thus, a gap opening is slightly ‘cheaper’ than a bad match. Using this gap model, variants of songs in which e.g. a phrase is repeated can be better aligned, since these penalties result in one long gap instead of many short gaps. Furthermore, the use of an affine gap penalty function prevents gaps from being scattered all over the alignment.

6.4 Evaluation of Substitution Functions

The substitution functions are evaluated by their respective retrieval performance on the Annotated Corpus as described in section 6.2. To evaluate a scoring scheme each melody from the Annotated Corpus is taken once as query and the other melodies are sorted according to the normalized score of the alignment with the query melody. At each rank the average recall and average precision over all ranking lists is computed. These values are plotted in

Scoring Function	Annotated			Full data set		
	MAP	k=1	k=10	MAP	k=1	k=10
exactpitch	0.75	0.95	0.98	0.57	0.86	0.96
pitchb	0.73	0.96	0.99	0.54	0.89	0.96
harm	0.67	0.93	0.98	0.49	0.86	0.94
phrasecont	0.61	0.90	0.97	0.44	0.81	0.92
songcont	0.63	0.91	0.98	0.45	0.84	0.94
exactint	0.60	0.92	0.97	0.42	0.84	0.93
intband	0.58	0.93	0.97	0.40	0.86	0.93
exactdur	0.45	0.77	0.92	0.20	0.47	0.74
accent	0.45	0.74	0.88	0.23	0.54	0.76
ima	0.46	0.80	0.94	0.26	0.64	0.81
exactior	0.44	0.74	0.93	0.19	0.46	0.72
iorratio	0.51	0.81	0.96	0.25	0.58	0.79
phrpos	0.56	0.84	0.95	0.33	0.69	0.89

Table 6.1: Evaluation measures for single scoring functions, showing the mean average precision (MAP), and the percentage of queries that has a relevant item among the first k ranked items for $k = 1$ and $k = 10$. Values are given both for the Annotated Corpus and for the full data set.

a diagram. Furthermore, we compute the mean average precision (MAP) by averaging the precision of all relevant items for all queries.

Identification of songs is a common task in folk song research (see Chapter 2). For this, one relevant item at the top of the ranking list can be enough. Therefore, we compute the percentage of the queries that has at least one relevant item among the first k items on the ranking list, for k is 1 (which is the recognition rate), and k is 10.

We do this both for the Annotated Corpus and for the full data set of 4830 melodies. In the latter case as well, only the 360 melodies from the Annotated Corpus are taken as queries. All precision-recall diagrams have been created using the full data set.

The criterion for relevance is the membership of the same tune family.

We test whether the differences in MAP are significant by performing a non-parametric Friedman test, with a significance level of $\alpha = .05$, using for each configuration of scoring functions the average precisions of the 360 individual queries. To determine which of the pairs of measurements differed significantly we conduct a post hoc Tukey HSD test.

Table 6.1 shows an overview of the evaluation measures for single scoring functions and Table 6.2 shows an overview of the evaluation measures for several combinations of scoring functions.

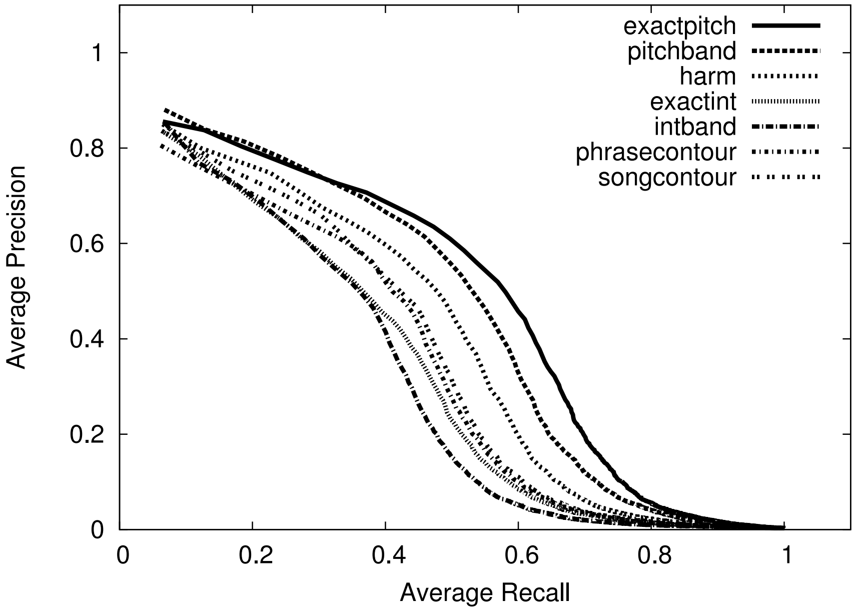


Figure 6.4.1: Retrieval performance of pitch-based substitution functions.

6.4.1 Evaluation of Single Substitution Functions

First, we study the performance of the single pitch-based substitution functions introduced in section 6.3.1. According to the Friedman-test, there are significant differences between the configurations, $\chi^2(6, N = 360) = 535.6, p < 0.001$. According to the Tukey HSD test, all pairs differ significantly except for harm and songcontour, songcontour and phrasecontour, phrasecontour and exactint, and exactint and intband.

Variation in pitch is considered an important element of oral transmission (see e.g. Klusen et al., 1978). Nevertheless, aligning melodies using the exact pitch information with $S_{exactpitch}$, which does not allow melodic deviation, results

Scoring Function	Annotated			Full data set		
	MAP	**k=1**	**k=10**	**MAP**	**k=1**	**k=10**
exactpitch & phr	0.83	0.97	0.99	0.69	0.90	0.97
pitchb & phr	0.83	0.98	1.00	0.67	0.89	0.97
harm & phr	0.78	0.96	0.99	0.63	0.87	0.98
phrasecont & phr	0.76	0.96	0.99	0.60	0.86	0.96
songcont & phr	0.78	0.96	1.00	0.61	0.89	0.96
exactint & phr	0.74	0.96	0.99	0.58	0.90	0.96
intband & phr	0.74	0.94	0.99	0.58	0.87	0.94
exactdur & phr	0.58	0.89	0.94	0.36	0.68	0.91
accent & phr	0.69	0.89	0.96	0.45	0.75	0.89
ima & phr	0.65	0.90	0.96	0.42	0.76	0.91
exactior & phr	0.60	0.88	0.97	0.33	0.65	0.87
iorratio & phr	0.66	0.90	0.97	0.41	0.73	0.89
exactpitch & accent	0.79	0.96	0.99	0.61	0.87	0.96
exactpitch & ima	0.79	0.97	0.99	0.63	0.90	0.96
exactpitch & iorratio	0.78	0.97	1.00	0.61	0.90	0.97
exactpitch & accent & phr	0.85	0.98	0.99	0.70	0.89	0.98
exactpitch & ima & phr	0.84	0.98	1.00	0.70	0.91	0.98
exactpitch & iorratio & phr	0.82	0.97	1.00	0.68	0.90	0.98
exactpitch & ima & iorratio & phr	0.80	0.98	1.00	0.65	0.89	0.98
pitchb & accent	0.77	0.94	0.98	0.60	0.86	0.95
pitchb & ima	0.75	0.96	0.99	0.58	0.87	0.96
pitchb & iorratio	0.75	0.96	0.99	0.55	0.86	0.97
pitchb & accent & phr	0.85	0.97	0.99	0.69	0.88	0.96
pitchb & ima & phr	0.84	0.99	1.00	0.68	0.92	0.98
pitchb & iorratio & phr	0.83	0.98	0.99	0.64	0.89	0.99
pitchb & ima & iorratio & phr	0.80	0.98	0.99	0.62	0.89	0.98

Table 6.2: Evaluation measures for various combinations of scoring functions, showing the mean average precision (MAP), and the percentage of queries that has a relevant item among the first k ranked items for $k = 1$ and $k = 10$. Values are given both for the Annotated Corpus and for the full data set.

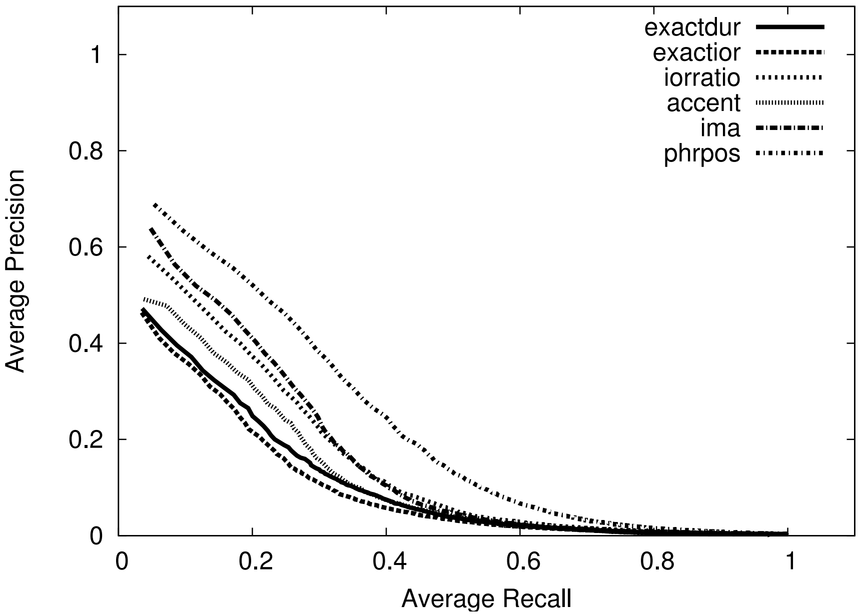


Figure 6.4.2: Retrieval performance of non-pitch-based substitution functions.

in the best performance (see Figure 6.4.1). At the top of the ranking lists, the performance of $S_{exactpitch}$ is comparable to that of $S_{pitchband}$. At lower ranks, it is even better. This is reflected in a slightly higher mean average precision for $S_{exactpitch}$ (see Table 6.1).

Both the harmonic and contour-based substitution functions perform worse than $S_{exactpitch}$. Considering the contour instead of the exact pitch sequence does not result in a better retrieval performance. Harmonic relations, which have otherwise successfully been used in models of melodic expectancy (Margulis, 2005), do not improve the alignment of melodies of a tune family in comparison to exact pitch information.

The solution for transposition invariance using intervals between the notes instead of absolute pitches, yields worse results than our histogram solution. If the two solutions would be equivalent, the performances of $S_{exactpitch}$ and $S_{exactint}$ would have been the same. Now, the interval solution performs significantly worse than the histogram method.

Figure 6.4.2 shows retrieval performance for the scoring functions that do not involve pitch information. There are significant differences between the configurations, $\chi^2(5, N = 360) = 230.2, p < 0.001$. According to the Tukey HSD test, all pairs differ significantly except for iorratio and phrpos, iorratio and ima, and accent and exactdur.

Although rhythmic features have been considered quite stable within oral transmission (see Klusen et al., 1978), all rhythm-related substitution functions perform worse than the pitch-related functions. Comparison of $S_{exactdur}$ with $S_{exactior}$ shows that the correction for representational differences by using inter-onset-ratios does not result in better performance without further improvement. Allowing some deviation using $S_{iorratio}$ does result in a better retrieval performance. S_{ima} performs better than S_{accent}, which indicates that our simple metric model that is based on the notated meter is not adequate enough. The use of the metric weights as computed by the Inner Metric Analysis leads to better performance. The best performing of the non-pitch substitution functions is S_{phrpos}. Apparently the correspondence in number and length of phrases is an important aspect of similarity in this corpus.

6.4.2 Evaluation of Combinations of Single Substitution Functions

In a next step, we combine various substitution scoring function. Since there are many possible combinations of substitution scoring functions, we do not evaluate them all, but we take a musically motivated evaluation strategy. We combine pitch, rhythmical and segmentation data. The combinations are shown in Table 6.2.

First, we combine the best of the pitch-related functions ($S_{exactpitch}$ and S_{pitchb}) with rhythmical functions as shown in Figure 6.4.3. If we involve the single substitution functions $S_{exactpitch}$ and S_{pitchb} as well, it appears that there are significant differences between the configurations, $\chi^2(7, N = 360) = 289.9, p < 0.001$. According to the Tukey HSD test, the single function pitchband performs significantly worse than all combined functions, while exactpitch performs significantly worse than all combined functions except for pitchband-iorratio. In case of the combined functions, there are no significant differences between exactpitch-accent, exactpitch-ima, exactpitch-iorratio and pitchband-accent, while both pitchband-ima and pitchband-iorratio perform significantly worse than the other configurations, but do not differ significantly from each other.

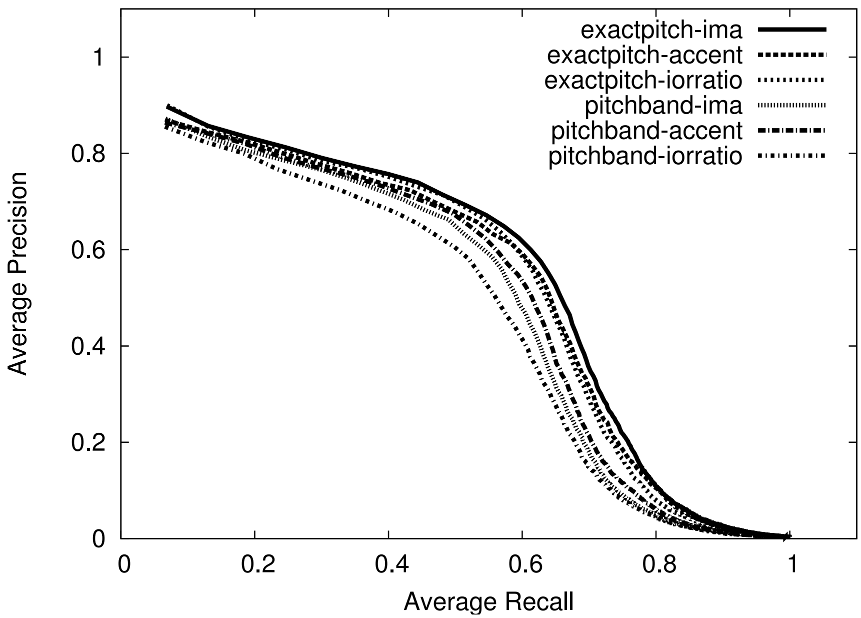


Figure 6.4.3: Retrieval performance of various combinations of substitution functions.

Figures 6.4.1, 6.4.2 and 6.4.3 show that, although the individual rhythmical substitution functions perform worse than $S_{exactpitch}$, the combined functions $S_{exactpitch-accent}$, $S_{exactpitch-ima}$, and $S_{exactpitch-iorratio}$ yield better retrieval performance than each of the single functions. Again, the combinations with S_{pitchb} perform slightly worse than the combinations with $S_{exactpitch}$.

Combination with each of the rhythmic functions ($S_{iorratio}$, S_{ima} and S_{accent}) shows comparable improvement. To find out whether these substitution functions are really exchangeable, we evaluate the differences between the rankings of the relevant items in these three cases. These rankings are obtained with the full data set and reference melodies as queries. Thus, the size of the ranking lists is 4830 items. Table 6.3 shows the results of this comparison. The average increase in rank for those songs that are higher on the ranking list of exactpitch-accent compared to that of exactpitch-ima (avgposdiff) is 75, while the average decrease in rank for those songs that are lower on the ranking list of excactpitch-accent is 112 (avgnegdiff). For the same pair of scoring functions,

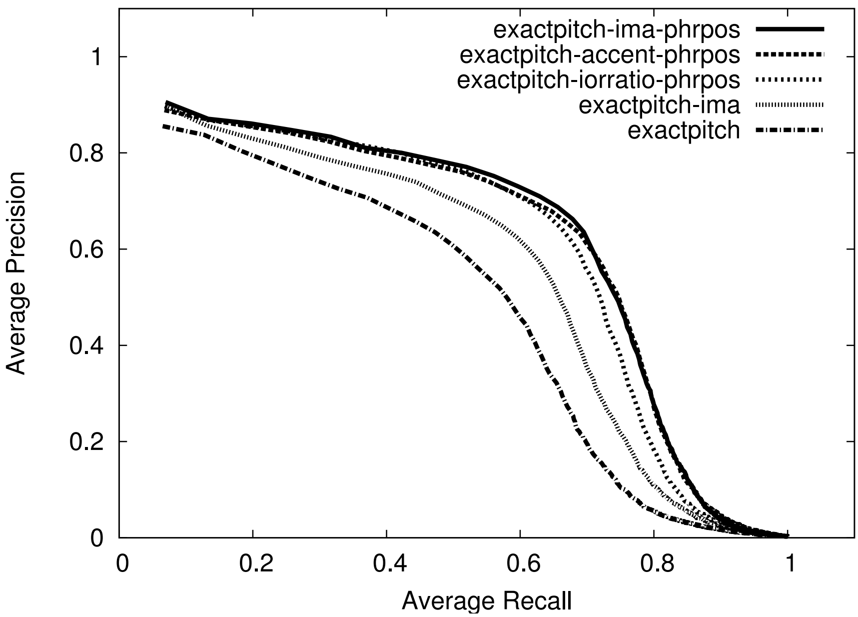


Figure 6.4.4: Retrieval performance of various combinations with phrase position. Exactpitch and exactpitch-ima are added for reference.

Scoring Functions	avgposdiff	avgnegdiff	numpos	numneg	numpostier	numnegtier	into2tier	outof2tier
exactpitch-ima $\rightarrow$ exactpitch-accent	75	112	137	150	37	47	13	16
exactpitch-ima $\rightarrow$ exactpitch-iorratio	107	95	145	140	39	44	11	17
exactpitch-accent $\rightarrow$ exactpitch-iorratio	121	77	154	130	46	42	17	20

Table 6.3: Comparisons of rankings of relevant items for exact-ima, exact- accent and exact-iorratio.

137 songs are higher on the ranking list of exactpitch-accent (numpos) while 150 songs are lower (numneg). 37 songs are substantially higher (numpostier) and 47 songs are substantially lower (numnegtier). Here, substantially means more than the size of the tune family. Of these songs, 13 'rise' into the second tier for exactpitch-accent (into2tier) and 16 songs are 'lost'. The relatively small differences between these figures, especially between into2tier and outof2tier, correspond to the similarity in overall performance. Although on a total of 360 songs, these are relatively small numbers, there are songs that are found with one substitution function, but not with the other, and vice versa. The same goes for the other two pairs of scoring functions.

Finally, we evaluate the retrieval performance when involving the phrase boundaries by combining with S_{phrpos}. There are significant differences between the configurations in Figure 6.4.4, $\chi^2(4, N = 360) = 593, p < 0.001$. According to the Tukey HSD test, the three combinations including phrpos perform significantly better than the configurations without phrpos. The two pairs that do not differ significantly are exactpitch-ima-phrpos and exactpitch-accent-phrpos, and exactpitch-ima-phrpos and exactpitch-iorratio-phrpos.

The retrieval performance of the combination $S_{exactpitch-ima-phrpos}$ shown in Figure 6.4.4 shows even better performance results than the combinations of two single substitution functions. If we average the precision of all relevant items for all queries, we get a mean average precision of 0.70 for this combination. Choosing S_{accent} or $S_{iorratio}$ instead of S_{ima} gives similar values for the mean average precision. In all three cases, for 98% of the queries, a relevant item can be found among the first 10 items on the ranking list (see Table 6.2).

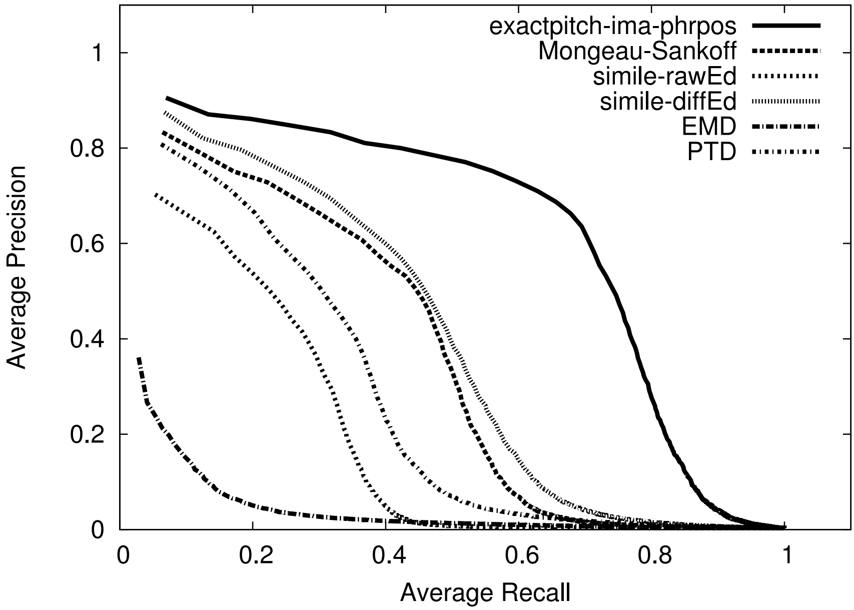


Figure 6.4.5: Comparison with related methods.

Therefore, this approach is very suitable for identification of melodies. The researcher only has to examine the top 10 of the ranking list.

6.4.3 Comparison with Related Methods

Figure 6.4.5 shows comparisons of one of our best scoring schemes with alignment methods from literature. For the method of Mongeau and Sankoff (1990) the parameters were taken as given by Mongeau and Sankoff. The normalization was done by dividing the alignment score by the sum of the durations of both sequences. DiffEd and rawEd were taken from the Simile alignment toolbox without change (Müllensiefen & Frieler, 2004). It appears that our $S_{exactpitch-ima-phrpos}$ performs best. According to the Friedman-test, there are significant differences, $\chi^2(5, N = 360) = 1070, p < 0.001$. According to the Tukey HSD test, all pairs differ significantly except for DiffEd and Mongeau-Sankoff.

Apart from string and sequence matching, another important approach that proved successful for automatic comparison of melodies is the use of transportation distances (Typke, 2007). Such a method computes the distance between two weighted point sets by calculating the amount of 'work' that is needed to transform the one into the other. An often used metaphor to explain this class of distance measures is to represent the first point set by heaps of earth, where the weight of each point correponds to the amount of earth, and the other point set by holes in the ground, where the size of the hole corresponds to the weight of the points. The distance between the two point sets is the minimal amount of work needed to fill the holes, measured by multiplying the 'transported' weight by the ground distance over which it is transported. Therefore, this distance measure bears the name Earth Movers Distance (EMD). A melody can be represented as a weighted point set by representing the notes as points in a two-dimensional plane, where the dimensions are time and pitch. The weight of the point is determined by the duration of the corresponding note (Typke, 2007). We test two variants of this distance measure. In the first variant, the EMD, the points keep their original weight, while in the second variant, the Proportional Transportation Distance (PTD), the weights of the points are normalized such that for each of the point sets the total amount of weight equals 1. From Figure 6.4.5 is it clear that the PTD performs in line with other related methods, but not very well, while the EMD fails. Indeed, the EMD works better by cutting the piece of music into short segments, and combining the segment-based matching results (Typke, 2007).

At the top of the ranking lists, all methods, except for the EMD, show reasonable to good performance. In all these cases, the easy to retrieve melodies have been found. The largest differences can be seen for the songs that are lower at the ranking lists. For example, at the rank at which on average 80% of the songs have been retrieved, the average precision for diffEd, rawEd and Mongeau-Sankoff is very low, while for $S_{exactpitch-ima-phrpos}$ the average precision is still 28%. Thus, to find 80% of the relevant songs, on average nearly four times as many items should be consulted.

6.5 Evaluation of false negatives

If we use the reference melodies as queries, a small number of songs have a persistently low ranking for all of the single substitution functions and combinations of functions we use. In this section we investigate those songs to see whether the algorithms fail or whether there are intrinsic reasons for these

Song ID	Tune Family	Median rank	Average similarity rating
15569_01	Maagdje	49.5	1.208
70078_01	Stavoren	755	1.08
71666_01	Maagdje	331	1.417
71957_03	Ruiter 1	4157	0.733
72253_01	Halewijn 4	849	0.867
72851_01	Ruiter 1	3672.5	0.55
72851_02	Ruiter 1	3433	0.8
73277_01	Boom	1249	1.267
73929_01	Jonkheer	599.5	0.917
74603_01	Halewijn 4	984.5	0
75525_01	Stil	1518.5	1.033
76271_01	Vrouwtje	72.5	1.625
76495_01	Femmes	867.5	1.333

Table 6.4: Ranking and annotations for the false negatives. The median rank is the median of the rankings of the song on the ranking lists produced by all evaluated substitution scoring functions and combinations of substitution scoring functions.

songs not to be found.

We take three steps. Firstly, we examine the manually annotated similarity ratings that were described in Chapter 3. Secondly, we use the problematic songs as queries to see whether the reference melodies are not representative enough. Thirdly, we ask domain experts to reconsider the tune family membership of these songs.

Table 6.4 shows the id's of the problematic songs. A song is 'problematic' if, when using the reference melodies as queries, the highest rank for any of the evaluated scoring functions is lower than the size of the second tier, which is twice the number of songs in the tune family. Table 6.4 also shows the median of all rankings and the average of the manual similarity ratings. This average is the weighted mean of the annotated degrees of similarity of the respective song with the reference melody, where the weight of the global ratings (global rhythm, global contour and global motifs) is the number of phrase comparisons, and the weight for the phrase comparisons (rhythm per phrase, contour per phrase) is one. The individual ratings are either 0 (not similar), 1 (somewhat similar) or 2 (similar). Thus, the maximum value for the average similarity ratings is 2.

Except for 76271_01 and 73277_01, the average similarity rating of these songs is the lowest for any of the songs in the respective tune families. This explains why these songs have low ranking when querying with the reference melodies. It indicates that these low rankings are not just a failure of the algorithm; the songs are really problematic.

It might be possible that two songs of the same tune family are more similar to each other than to the reference melody. In that case, the reference melody is not sufficiently representative for all melodies in the tune family. To examine the extent to which this effect causes the low rankings, we take each problematic song as query and inspect whether there are songs from the same tune family at the top of one of the ranking lists of three well-performing substitution scoring functions: $S_{pitchb-ima-phrpos}$, $S_{pitchb-accent-phrpos}$, and $S_{exactior-phrpos}$. For six out of the 13 songs, this appears to be the case. Thus, it would be possible to recognize these songs using the rankings lists of these three scoring functions.

We asked the domain expert at the Meertens Institute to re-evaluate the seven remaining songs, resulting in the following decisions. 75525_01 is considered quite dissimilar indeed because only the first two phrases have some melodic similarity with the other songs. Therefore a new tune family is defined for this song: *Kom laat ons nu zo stil niet zijn 2*. 74603_01 is reassigned to *Halewijn 2* instead of *Halewijn 4*. This song has four other songs from *Halewijn 2* among the first 10 ranks using $S_{pitchb-ima-phrpos}$. One of them even on the top rank. This confirms that this song is related to melodies in *Halewijn 2* indeed. 72851_01,72851_02 and 71957_03 are strongly interrelated. They were originally assigned to the tune family *Ruiter 1* mainly because of lyrical resemblance, but are considered melodically dissimilar indeed by the domain expert. Therefore, their tune family membership is considered questionable.

The two songs that remain, 73929_01 and 73277_01, are the only songs that are really hard to recognize automatically. They have no other songs from their respective tune families at the top of the rankings lists, and upon re-examination, their tune family membership is confirmed. The phrase endings in 73929_01 are inconsistent with the other songs in the tune family, which causes low substitution scores for related notes. Furthermore, this song is the only one in 4/4 meter, while all other songs in the tune family are in 6/8. Originally, the main reason for 73277_01 to be classified into tune family *Boom* is the occurrence of a melodic motif that is very characteristic for that tune family. Indeed, in the manual annotations, 73277_01 has rating 2 for motifs and rating 1 for contour. For both cases, the global alignment approach we take in this study seems inappropriate.

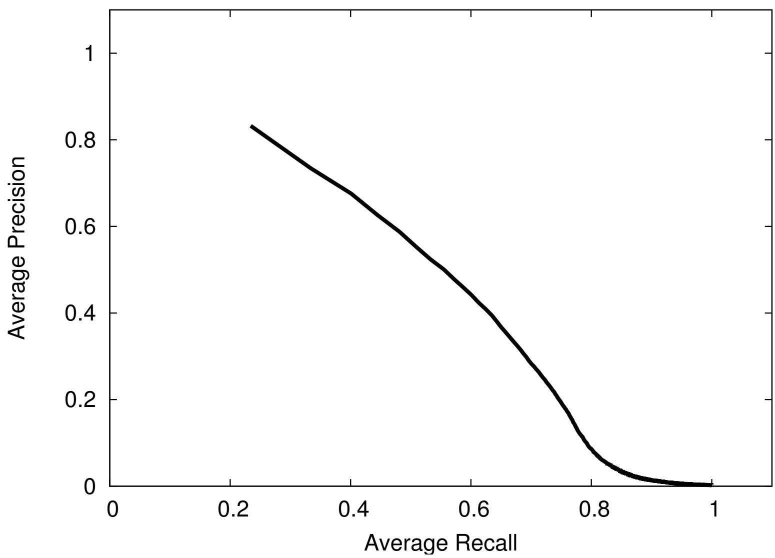


Figure 6.5.1: Average precision-recall diagram for the full corpus using all melodies with relevant items in the data set as queries.

In summary we conclude that all of these 13 problematic songs are not considered very similar by the annotators indeed. For six songs the reference melody is not representative enough for all songs in the tune family. Two songs are reassigned to another tune family, indeed. For three songs, the domain expert confirms the questionable state of tune family membership. For only two songs the global alignment approach seems inappropriate.

6.6 Evaluation of the Full Corpus

In the previous sections, we only used the 360 melodies of the Annotated Corpus as queries. In this section, we test one of the best performing score functions, $S_{exactpitch-ima-phrpos}$, using as many melodies as query as possible. 1472 tune families only have one member. These are skipped since no relevant items are expected to be retrieved at all. Each of the remaining 3358 melodies is taken as query. The resulting mean average precision is 0.66, the recognition rate is 0.83 and the fraction of queries that results in a relevant item within the

10 first items on the ranking list is 0.91, which indicates that the vast majority of the melodies can be identified using this scoring function. The average precision-recall diagram is shown in Figure 6.5.1. The average recall at the first rank of the ranking lists is 0.23, which explains why the curve starts at 0.23 for average recall. At the rank at which on average 80% of the relevant songs has been found, the precision is 8.6%. This means that to find 80% of the relevant songs, one has to consult on average approximately 12 times as many hits.

6.7 Hard to Classify Melodies

In the past few years, the musicological experts at the Meertens Institute were able to classify many songs from oral tradition 'by hand' with help of the text incipits. There are, however, 111 songs that have been left unclassified, because the experts could not find appropriate tune families to assign them to. As a test of the practical usability of the alignment-based retrieval, we use each of these 111 songs as query and present the musicologists the ranking lists according to three well performing configurations, and ask them whether the songs can be classified with help of these ranking lists. The configurations we use are $S_{pitchb-ima-phrpos}$, $S_{pitchb-accent-phrpos}$, and $S_{exactior-phrpos}$. The latter one is included to have a rhythm-only configuration as well.

The result is that 42 songs have been classified. For 32 songs, a relation with one ore more tune families has been found, but the relation is not strong enough to assign the song to the tune family. For the remaining 37 songs, no related melodies have been found at all. Another result is that two tune families have been merged into one. These two tune families were established based on the lyrics. Using the melodic ranking, it now turns out that these two lyrics are sung with the same melody.

Considering that these 111 songs were the 'hard' cases, this is a good result.

6.8 Conclusion and Future Work

This chapter shows that the inclusion of musical knowledge in alignment algorithms improves the assessment of similarity among folk song melodies. By evaluating different substitution scoring functions, we found that our pitch-related functions lead to better recognition than rhythm-related functions. The use of phrase information improved the retrieval results significantly. The best combination of functions, combining a pitch-based, a rhythm-based and a segmentation-based scoring function, outperforms related methods from literature.

Study of false negatives turns out that almost all of these songs have intrinsic reasons for not being found. Upon re-examination, the domain expert reassigned several songs to another tune family. With the results of the alignment-based retrieval, 42 out of 111 hard to classify melodies could be classified, and for 32 melodies, interesting relations with other songs in the data base could be discovered.

Thus, the results of the algorithm were useful to improve the 'ground-truth'. This is an example of interaction between the computational and the musical aspects of the research. As discussed in Chapter 2, too many computational studies within Music Information Retrieval take the musical 'ground truth' data as absolute reference. By evaluating the meaning of the algorithmic results for the ground truth data, the research gets a more interdisciplinary character, resulting in new knowledge for both the computational and the musical aspects of the research.

Based on these results, we conclude that the alignment approach as presented in this chapter is appropriate for studying relations between melodies from oral tradition.

Other findings are that the use of intervals is not an optimal solution to obtain transposition invariance—our histogram-based pitch shift outperforms the interval approach—and that the use of inter-onset-ratios does not improve rhythm based retrieval in our case.

Chapter 7

Retrieval of Folk Song Recordings using Musically Meaningful Audio Segments

The alignment approach presented in the previous chapter allows for local comparisons between melodies. However, the alignment algorithm computes the score of the alignment of two entire melodies, which is a global comparison. The results of the analysis of the annotations in Chapter 3 and the Cowdery's study (1984) of the tune family concept both strongly suggest to use recurring melodic motifs for recognition of melodies. In this chapter we take a first step towards folk song retrieval using small audio segments. The musical meaningful segmentation is obtained by splitting the recordings at silences and breathing.

Contribution. We widen the scope of automatic folk song classification by using audio recordings rather than symbolic data. To our knowledge, this is the first study in which aspects of folk song performance (breathing and pauses) are used to mark segment boundaries, and this is the first computational study that models a tune family by its most representative recurring segments.

7.1 Background

Large collections of monophonic folk song recordings are interesting from a music cognition perspective since they represent musical performances of common people. Most people share a 'common core of musical knowledge' (Peretz, 2006, section 2). Since recorded folk songs were sung from memory, knowledge about the process of remembering and reproducing melodies can be used to employ these recordings in the context of folk song research, music information retrieval or music cognition studies.

This chapter combines ideas and approaches from Ethnomusicology, Music Cognition and Computer Science. One of the research questions of Ethnomusicology is how melodies in an oral tradition relate to each other (see Chapter 2). Samuel Bayard (1950) developed the concept *tune family* to denote a group of melodies that share a common origin, which, in the most simple case, is a single tune. The idea that melodies from the same tune family are related by shared melodic motifs has a long history in folk song research. Nettl (2005, p. 117f) discusses the relative independence of shorter units of musical thought. These might 'wander' from melody to melody and from country to country (Tappert, 1890). Marcello Keller (1988) explains the relations between Trentino folk music compositions by means of a repertoire of 'segments' that is used in the act of composing. To cope with specific relations between melodies he encountered in Irish folk music, James Cowdery (1984) extended Bronson's concept of tune family by including melodies that are related by sharing melodic material from the same 'pool of motives'. Finally, one of the conclusions from Chapter 3 is that recurring motifs are more important than contour and rhythm for recognizing a song.

Understanding the way melodies change in oral transmission involves understanding of encoding of melodies in, and reproduction of melodies from human memory. Cognitive studies indicate that melodies are not reproduced note by note, but as a sequence of higher level musical units, or chunks (Miller, 1956). Much research has been done to model these chunks (e.g., Lerdahl & Jackendoff, 1983; Cambouropoulos, 1998; Namour, 1992). All mentioned approaches use a symbolic transcription of the melody in the form of a musical score and try to group notes into musically meaningful segments in a bottom-up or top-down fashion. In the current study we take as our starting point the audio recording of a song performance rather than its transcription. Thus, we can use aspects of the performance that are lost during transcription into musical notation.

The computational methods we use enable a data-rich, empirical approach to the study of segmentation and similarity of melodies (Clarke & Cook, 2004).

The two main questions in this chapter are whether recurrence of audio segments can be exploited to classify a folk song recording into the correct tune family, and whether the use of cognitively and musically meaningful audio segments yields better classification performance than the use of fixed-length audio segments.

Our classification method consists of four stages: pitch extraction, segmentation, selection of representative segments for each tune family, and classification using these representative segments. These four stages are described in the next sections. The main idea for segmentation we employ in this chapter is to take breathing and pauses during singing as segment boundaries, which can be conceived as chunk boundaries. Thus, segmentation results in musically and cognitively meaningful units. We do not assume a one-to-one relation between these breathing and pause boundaries at the one hand and chunk boundaries at the other hand, but we do assume a strong relationship.

7.2 Data-Set

We use a the Annotated Corpus that has been presented in Chapter 3. Of these 360 songs we use the 305 that are available as audio recordings.

7.3 Pitch Extraction

Since a sound recording not only contains the fundamental frequency, corresponding to the perceived pitch, but other harmonics as well, pitch extraction is necessary. We use the YIN algorithm (De Cheveigne & Kawahara, 2002) along with a newly developed post-processing filter. The YIN algorithm combines the well-known autocorrelation method with a number of modifications that prevent errors. We use time frames of 1024 samples (23.2 ms for audio sampled at 44.1 kHz) and a YIN-threshold of 0.7. Our post-processing filter uses the dependencies between subsequent time frames to correct remaining errors. For each time frame, the filter replaces the detected pitch with the median of that pitch and surrounding pitches. This has a smoothing effect on the detected pitch curve. A small scale test shows that a window of 11 pitches (using the 5 preceding and 5 following pitches) gives good results. After pitch extraction, each recording is represented as a sequence of frequencies.

A manual examination of all detected pitch curves reveals some main causes for bad pitch extraction: tape recorder hum, accompaniment, polyphonic singing, singing in octaves by male and female voices, and heavy noise in very old recordings. It seems that improvements of the pitch detection are achievable.

7.4 Segmentation

There are time frames for which the YIN algorithm cannot detect a pitch. We assume that regions with a lot of these ‘pitch-less’ frames correspond to pauses in singing or to breathing. In addition to failing to detect a pitch, another indication of pause is a low energy of the signal. Our main idea for segmentation is to use these pitch-less regions as segment boundaries. This results in melodic segments in which a continuous flow of melody is present.

Since a short sequence of pitch-less time frames could also indicate a consonant like ‘h’ or a ‘z’, we set a lower limit to the length of the pitch-less regions to be considered as segment boundaries. After a small test on some representative examples, it appears that a good value for the minimal number of adjacent pitch-less time frames is 10, and that the median of the root-mean-square values of the time frames in the candidate boundary region should be smaller than 0.012.

As stated in the introduction, we assume that segment boundaries correspond to chunk boundaries. Therefore, relatively long segments are likely to be caused by under-segmentation. For that reason, and to decrease computation time, segments longer than 360 time frames (8.4 s) are removed from the data set. This leaves a data set with 5254 segments from 260 songs in 26 tune families. The threshold of 360 is somewhat arbitrary, but the exact value is not very important. It is unlikely that we throw away too many valid segments and the remaining set contains enough useful segments for the classification experiment.

The segmented audio recordings are available at http://give-lab.cs.uu.nl/music/icmpc2010/segments.

7.5 A similarity measure for segments

To measure the similarity of two segments we use a variant of the Smith-Waterman local alignment algorithm (Smith & Waterman, 1981). This algorithm finds the longest approximate common subsequence of two sequences of symbols along with an alignment of the matching parts and a score indicating the quality of the alignment. This score is the sum of the alignment scores of the individual symbols. If we consider two sequences $\mathbf{x} : x_1, \ldots, x_i, \ldots, x_n$, and $\mathbf{y} : y_1, \ldots, y_j, \ldots, y_m$, then symbol x_i can either be aligned with a symbol from sequence $\mathbf{y}$ or with a gap. Both operations have a score, the substitution score and the gap score. The gap score is mostly expressed as penalty, i.e. a negative score. The local alignment with the highest score is found by filling a matrix D recursively according to:

$$D(i,j) = \max \begin{cases} D(i-1, j-1) + S(x_i, y_j) \\ D(i-1, j) - \gamma \\ D(i, j-1) - \gamma \\ 0 \end{cases}, \qquad (7.5.1)$$

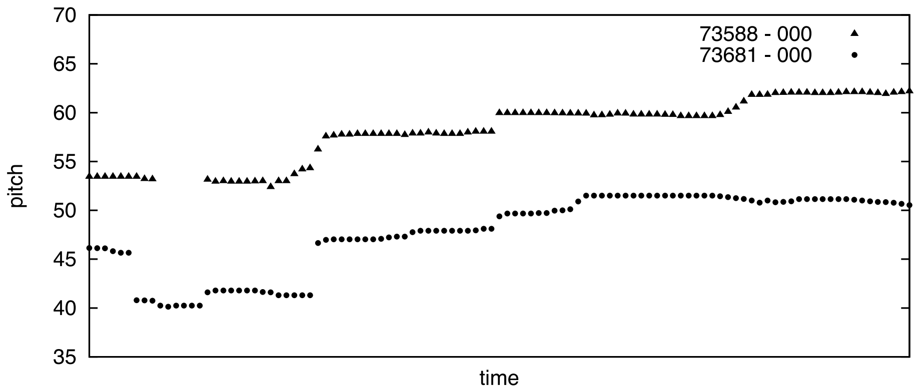
in which $S(x_i, y_j)$ is the substitution scoring function, γ is the gap penalty, $D(i, 0) = 0$ for $0 < i \leq n$, and $D(0, j) = 0$ for $0 < j \leq m$. $D(i, j)$ contains the score of the optimal local alignment up to x_i and y_j. The optimal local alignment can be found by starting at the cell with the highest value, which is the score of the alignment, and tracing back to the first cell with value zero. The standard dynamic programming algorithm has both time and space complexity $O(nm)$.

An audio segment is represented as a sequence of pitches for the consecutive time frames. The pitches are represented in continuous midi encoding, in which the middle c is represented by value 60.0, c# by 61.0, d by 62.0, and so on. By allowing fractional pitches we have a one-to-one correspondence to the frequencies, and a linear scale in the pitch domain.

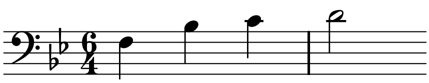
The substitution scoring function, which returns values in the interval $[-1, 1]$, is defined as:

$$S(x_i, y_j) = \begin{cases} 1 - \frac{interval(x_i, y_j)}{7.0} & \text{if } interval(x_i, y_j) \leq 7.0 \\ -1 & \text{otherwise} \end{cases}, \qquad (7.5.2)$$

where $interval(x_i, y_j) = |p(x_i) - p(y_j)| \bmod 12$, with $p(x)$ the pitch of symbol x_i. A perfect fifth has value 7 in midi-encoding. Thus all intervals up to a perfect fifth get a positive substitution score and all larger intervals are considered a bad match. This substitution score function was successful in the experiments on symbolic data that have been presented in Chapter 6. We use an extension of the algorithm proposed by Gotoh (1982), which employs an affine gap penalty function without loss of efficiency. In this approach, the extension of a gap gets a lower penalty than its opening. This prevents gaps from being scattered all over the alignment. We use 0.8 as gap opening penalty and 0.2 as gap extension penalty.



73588-000:



73681-000:

Figure 7.5.1: Example alignment of two segments (identified by: song id - segment). The alignment of the matching parts of the pitch curves as well as the symbolical transcriptions of the audio segments are shown at original pitch. 73588-000 matches with the second half of 73681-000. Apparently, a gap is needed at the beginning of 73588-000.

Since the score of an alignment depends on the length of the alignment, we normalize by dividing the alignment score by the score of the query segment with itself. Thus, an exact match that is embedded in a longer segment results in the maximal score (which is 1.0). Alignment with a short segment that has an exact match embedded in the query segment, results in a lower score. This makes our approach robust against under-segmentation as well as over-segmentation. As long as we have enough correctly detected segments, we will find related segments that are embedded in longer segments, but we will not find segments that are considerably shorter and that possibly match with many unrelated segments, so called hubs. The drawback of this normalization is that the score is not symmetrical, i.e., the score of aligning sequence $\mathbf{x}$ with $\mathbf{y}$ is not equal to the score of aligning $\mathbf{y}$ with $\mathbf{x}$.

Since the songs are sung at various pitch heights, the alignment needs to be transposition-invariant. The tentative solution we use for this is to add a constant to the pitches of one of the segments such that the means of the pitches are the same for both segments.

The normalized scores are converted to distances by taking one minus the normalized score. This results in distances within the interval $[0, 1]$. Figure 7.5.1 shows an example of an alignment.

7.6 Selecting Representative Segments

As discussed in section 7.1, shared melodic patterns are important to recognize relations between tunes. Therefore, it seems a good approach to search for similar melodic segments among the songs that belong to the same tune family. We use an automatic selection procedure. For each tune family we select the two segments that have the largest number of similar segments within the tune family, but that are not similar to each other. The selection procedure is as follows. For each segment all other segments in the dataset are ordered by distance. For a particular tune family, the segment that has the largest number of segments from the same tune family in the top 100 of the ranking list is selected as the first representative segment. To find the second representative segment the same criterion is applied with the additional constraint that the distance to or from the first selected segment is greater than 0.35. The histogram of all distances reaches its peak around 0.35. Therefore, this seems a safe value not to get a similar second segment. Thus, we find two dissimilar representative segments for each tune family. The threshold of 100 was established by inspecting the ranking lists manually. There is no segment for which the 100 nearest neighbors are all from the same tune family.

For some selected segments, the most common tune family among the 100 nearest neighbors is another tune family. These are removed from the set of representative segments. For six tune families no representative segment could be found. The numbers of songs in these families are 9, 7, 5, 4, 4, and 3. The small size of most of these families seems the cause for the failure to find representative segments. After removing these tune families, 228 recordings from 20 tune families remain. Furthermore, there are nine tune families for which only one representative segment could be found. The selected segments are shown in Table 7.1.

7.7 A Classification Experiment

In a classification experiment we use the selected representative segments to find the tune family of a query recording. The aim of the experiment is both to evaluate whether the method presented in this chapter is able to recognize a song at all and to show the improvement of using 'cognitive' segmentation over fixed-length segmentation, in which the recordings are split into segments with a fixed length of 4.3 seconds, the average length of the 'cognitive' segments.

The procedure is as follows. We take the distances from all selected representative segments to all segments of the query song. After sorting, the tune family that is most common among the first n segments (the n nearest neighbors) is the tune family that is assigned to the query recording. It appears that 3 is a good value for n.

We cannot assume that the distribution of the distances to a particular segment is the same for each segment. Especially the variation in the minimal distance is problematic. To cope with this problem, for each representative segment, the distance from the first nearest neighbor to the representative segment is subtracted from all distances. The result of this linear shift is that all segments that are close to any of the representative segments are at the top of the sorted list.

When using segments of fixed-length, 95 of the 224 (42.4%) recordings are classified into the correct tune family. This result is positively biased because the songs that contain the selected representative segments are among the classified songs. If we disregard these songs, 62 out of 191 songs (32.5%) are correctly classified. Using the same 228 recordings, with the 'cognitive' segmentation, the respective numbers of correctly classified songs are 121 out of 228 (53.1%) and 92 out of 199 (46.2%), which is considerably better. If we take into account that there are 20 tune families, these are a quite good success rates.

In Table 7.1 the results are shown per tune family. Most, but not all, tune families show an improvement in the case of 'cognitive' segments.

Tune Family	Selected Segments	Correct 'Fixed-length'	Correct 'Cognitive'
Daar ging een heer 1	74004 - 012 74227 - 007	5 / 12	6 / 12
Daar was laatstmaal een ruiter 2	73639 - 005	3 / 12	10 / 13
Daar zou er een maagdje vroeg opstaan 2	72306 - 002	4 / 5	1 / 5
Een Soudaan had een dochtertje 1	73269 - 003 74452 - 004	1 / 10	5 / 10
En er waren eens twee zoeteliefjes	74583 - 017	6 / 13	10 / 14
Er reed er eens een ruiter 1	72559 - 009 72898 - 020	5 / 16	2 / 16
Er was een koopman rijk en machtig	72441 - 014	1 / 8	3 / 8
Er was een meisje van zestien jaren 1	74336 - 000	3 / 9	7 / 10
Er woonde een vrouwtje al over het bos	74309 - 012 75771 - 014	7 / 9	0 / 9
Femmes voulez vous éprouver	76495 - 009 74470 - 019	2 / 8	2 / 8
Heer Halewijn 2	73994 - 001	2 / 7	6 / 8
Heer Halewijn 4	72253 - 014	1 / 7	0 / 8
Het was laatst op een zomerdag	73516 - 008 73516 - 001	4 / 11	2 / 12
Ik kwam laatst eens in de stad	74840 - 020 74077 - 012	6 / 14	8 / 13
Kom laat ons nu zo stil niet zijn 1	73562 - 003 73562 - 020	2 / 9	4 / 10
Lieve schipper vaar me over 1	73374 - 015	2 / 4	1 / 4
O God ik leef in nood	74038 - 021	-	1 / 3
Soldaat kwam uit de oorlog	73311 - 013 72287 - 001	6 / 15	13 / 15
Wat zag ik daar van verre 1	72688 - 027 73304 - 045	1 / 12	7 / 12
Zolang de boom zal bloeien 1	74100 - 001 73225 - 020	1 / 10	4 / 9
total:		**62 / 191 (32.5%)**	**92 / 199 (46.2%)**

Table 7.1: For each tune family the selected representative 'cognitive' segments are shown (song id - segment number), along with the number of correctly classified songs as fraction of the total number of songs in the tune family for both fixed-length segmentation and 'cognitive' segmentation. The segments can be consulted at http://give-lab.cs.uu.nl/music/icmpc2010/segments. For *O God ik leef in nood* no representative segments could be found in the fixed-length case.

7.8 Discussion and Future Work

The results clearly show that the recurrence of musically meaningful melodic patterns contributes to the identification of folk songs (i.e., to find the tune family to which they belong). We also conclude that 'cognitive' segments are more useful than fixed-length segments. This indicates a limitation of the n-gram approach that is widely used for similarity assessment or indexing of melodic material.

The segmentation we employ is entirely based on features of the audio recordings that are lost in the process of transcribing the songs into musical score. This shows that the focus of computational folk song research on symbolic musical data has to be widened. Integration of methods from both fields will lead to richer computational models of the concept of tune family.

The system is successful as a proof-of-concept. Since all phases clearly show many opportunities for improvement, we expect that the current results can be substantially improved. For example, the segmentation can be improved by using a proper breath detection algorithm instead of our simple model. The selection of representative segments could be improved by inferring the number of representative segments from the data rather than using a fixed number for all tune families. Thresholds were often defined by quick inspection. These could be determined in a more robust way. Probably these thresholds have different optimal values for different tune families.

This study offers many leads for further research that is relevant to Ethnomusicology, Computer Science and Music Cognition. It would be interesting to evaluate the musical properties of the selected representative segments. Do they have occurrences in all tunes in the tune family? Are there types of representative segments? Investigating the false positives and negatives, might reveal relations between tune families that were unnoticed before. Also, a further study of the relation between the obtained segments and cognitive models of melodic chunks seems necessary.

Finally, this research strongly indicates that musically and cognitively meaningful models are very important for Music Information Retrieval and other computational approaches to music, and therefore indicate that interdisciplinary collaboration between music scholars and computer scientists is of major importance.

Chapter 8

Conclusions and Future Work

8.1 Conclusions

In this thesis, several approaches to the computational modeling of similarity relations among folksong melodies have been studied. From Chapter 2, it follows that a multidisciplinary approach is crucial for the development of adequate algorithms. Incorporating musical knowledge into computational models and being involved in the academic study of music will raise the level of Music Information Retrieval towards musically more adequate methods. One of the most important aspects of a more musical approach is the involvement of ground-truth data as object of study, rather than accepting ground-truth data as an impenetrable barrier between the algorithm and the musical concepts that are modeled in the algorithm.

A crucial step towards musical understanding of similarity relations between folk song melodies and of the concept of tune family that has been used in Folk Song Research, is the development of the annotation method that is presented in Chapter 3. The annotations that are made with this method form a quantitative representation of the musical intuition of human experts concerning similarity of melodies. It appears that the most important dimensions of similarity relations among folk song melodies are contour, rhythm and motifs. These are basic aspects of melody indeed, and have been recognized as such in numerous studies before. Our annotation method shows how and to which extent these dimensions are important for melodic similarity assessment. An important conclusion from our study is that these dimensions are not equally important between and within tune families. There is not one dimension that 'keeps a tune family together'. It is e.g., possible that one melody is rhythmically dissimilar to other melodies in the same tune family, while for another melody in the same tune family the rhythmic similarity to other melodies has been the decisive clue to assign that melody to the tune family. Nevertheless, from the annotations we are able to state that the recurrence of characteristic melodic motifs seems the most important dimension of melodic similarity in the case of 'genetically' related folk song melodies.

One of the connecting themes in this thesis is the exploration of local versus global approaches. In the one extreme, one value represents all notes in a song, while in the other extreme, individual notes or tones with their immediate surroundings are used.

In Chapter 4 we studied sets of global quantitative features. The results confirm the findings from the annotations: there is not one feature or subset of features that is equally discriminative for all tune families. Therefore, it is not possible to develop an adequate similarity measure for folk song melodies using only few quantitative features. Another finding is that the subsets of features that are discriminative for a tune family in a small data set, are not discriminative for the same tune family among a large dataset of thousands of songs.

To show that the negative results of the previous chapter do not mean that the global feature approach is inappropriate for Computational Musicology in general, we performed a study to the authorship of baroque organ fugues using the same kinds of features (Chapter 5). We conclude that the approach is successful in the case of discerning the authorship of a set of organ fugues. In both cases, we used features that reflect the relative frequencies of basic properties of music like the occurrence of certain pitch intervals, melodically as well as between voices in the polyphonic fabric of an organ fugue. Apparently such features are able to grasp the individuality of a composer in the case of the organ fugues, but not the 'individuality' of a tune family in the case of folk song melodies. This suggest that the success of a computational approach to a musicological problem involves careful design from the musical as well as from the computational side. It is not self-evident that a certain computational technique is appropriate for the musicological problem at hand. Therefore, if the aim of the research is to gain musicological knowledge, collaboration between computer scientists and musicologists is of crucial importance.

In Chapter 6, the use of sequence alignment algorithms has been studied. Since a global alignment algorithm has been used, this approach features global as well as local comparison of melodies. It proves that the incorporation of appropriate musical knowledge in the form of substitution scoring functions, leads to a successful retrieval experiment. Our best performing configuration in terms of retrieval results uses the combination of a pitch-related, a rhythm-related and a structure-related substitution scoring function. It outperforms related methods from literature on our dataset. Contrary to the global feature approach, the alignment approach is scalable. Melodies that are related to a query melody are found in a small dataset as well as in a large dataset. The results of the alignment method have been used to reconsider the identity of melodies in the Annotated Corpus, which served as test collection. With these results, improvements could be made in the classification of several songs in the corpus. Furthermore, the ranking lists created using the alignment scores enabled the classification of more than one third of the 111 melodies that proved hard to classify 'by hand'.

The final chapter (7) introduced a motif-based approach to the retrieval of folk song melodies using the audio recordings rather than the symbolic transcriptions. All melodies have been segmented by detecting breath and pauses. Using the resulting audio segments, it proves possible for most of the tune families in the Annotated Corpus to find a characteristic segment that has approximate local matches in other melodies from the same tune family. Comparison between two ways of segmenting the recordings, the one using fixed-length segments and the other using segment boundaries according to musical properties (breath and pauses), shows an improvement in the retrieval results for the latter type of segmentation. This, again, indicates the importance of incorporating musical knowledge in the computational models.

For further study, the local approach seems very promising. The consistently high ratings for the dimension motifs in the annotations, and the good performance of the alignment approach, which employs local correspondences of melodies, strongly indicate that these local correspondences contribute much to the identity of a melody. The enormous amount of annotations concerning motif classes and motif occurrences that have been provided by the expert annotators for the melodies in the Annotated Corpus, enables the evaluation of algorithms that detect and employ recurrent melodic patterns for similarity assessment. For better understanding of oral transmission of melodies, and thus of memory for melodies, it would be interesting to study the variation in characteristic motifs. One question would be whether melodic patterns that are perceived as very characteristic for a particular tune, show less variation than parts of the melodies that are not perceived as very characteristic.

The collection of songs hosted by the Meertens Institute, and the annotations that have been made concerning similarity relations between melodies, are a rich source for future research on melodies, oral tradition, human memory for melody, and undoubtedly a much wider range of topics from Musicology and Music Cognition.

8.2 Implications and Final Remarks

As stated in section 2.7, the general aim of this thesis was to design adequate measures of melodic similarity that can be employed in a retrieval system for folk song melodies to support identification. This aim has been achieved. As part of the WITCHCRAFT project, the alignment-based similarity measures that are described in Chapter 6 have successfully been implemented in the prototype of a retrieval system for the Dutch Song Database. This system is currently being used by the specialists of the Meertens Institute. A version of this retrieval system that is publicly available will be implemented in a follow-up project.

One of the bases for the research in this thesis has been the design of musically meaningful computational methods (section 1.2). The musicological potential of methods from Computer Science is related to the extent in which the musical problem at hand can be formulated in terms of the data structures and algorithms. The further the computational setup moves away from musical concepts and musical relations, the less relevant the results are from a musicological point of view. The better the terms and traits of an algorithm are interpretable from a musicological point of view, the more the results contribute to musical knowledge. Therefore—again—collaboration between music scholars and computer scientists is crucial for the success of Computational Musicology. Obviously, this can be extrapolated to Computational Humanities in general. Computational techniques should not 'blindly' be applied to problems from humanities. Instead, computational models should be based on theories and insights from humanities.

In the computational studies in this thesis, the choices of methods and data structures have been musically motivated, except for the use of global features for folk song classification (Chapter 4). There, our motivation was that this approach is common within pattern recognition (Section 2.5.6). There do not seem to be clear musical reasons to expect that the involved features are discriminative for tune families. But, there are musical reasons to expect the specific set of similar global features that was used in Chapter 5 to be discriminative for the authorship of baroque organ fugues (Sections 5.3 and 5.4). The use of alignment algorithms has been motivated from a musical point of view as well (Section 6.1). Finally, the segment based approach from Chapter 7 was motivated by the results from the annotations (Chapter 3) in which the approximate recurrence of motifs appeared to be the most important factor in recognizing related melodies.

We did not regard the 'ground truth' data unquestionable. In fact, because of the difficulty of recognizing related melodies, the assignment of melodies to tune families at the Meertens Institute always reflects current insights. These assignments are subject to change as a result of new discoveries or better understanding of the corpus. So is the assignment of melodies to tune families within the Annotated Corpus. The results of the computational evaluation of the similarity relations between the melodies based on the alignment scores caused reconsideration of tune-family membership of several melodies, and, for some songs, even reassignment to another tune family. This is a modest example of the possible interactions between computational and musical aspects of the research. The potential of Computational Musicology lies in the exploitation of these kinds of interactions.

From the perspective of Folk Song Research, we identified four problems in Section 2.4. The first two problems are related to models and theories that have been developed within Folk Song Research, up until now. Firstly, there is no theory that describes the general characteristics of oral transmission of melodies and of the melodic variation that occurs within this transmission, and, secondly, there is no classification system for folk song melodies that can generally be applied. These problems have not been solved in this thesis. Nevertheless, a computational approach can provide new grounds for developing such theories. Especially the approach taken in Chapter 7, where segments of melody were isolated, can be considered a step in that direction. Several folk music scholars, such as Cowdery (1990) and Keller (1988), indicated the importance of small units of melody for the identity of melodies. Algorithms to discover recurrent melodic patterns in a corpus of folk songs, literal as well as approximate, can be used to model relationships between songs, tune families and song cultures, firmly based upon the melodic material.

The third problem concerns the informal way in which concepts have been defined within Folk Song Research, especially the concept of tune family. For a computational approach, a very precise description in terms of algorithms and data structures is required. In this thesis, we took an important step to overcome this problem by designing an annotation method that is able to represent musical intuition about melodic similarity in a quantitative way. Thus, we get a quantitative 'deconstruction' of the bases on which decisions about tune family membership are taken by the experts of the Meertens Institute. This is not yet a computational model of the tune family concept, but it can serve as starting point for designing such a model. The annotations show which dimensions of melody are important for tune family membership, and to what extent.

The fourth problem addressed the issue of testing. Classification systems that were developed in the past within Folk Song Research have not been tested for their ability to automatically classify melodies into tune families. In this thesis, the abundance of data and the current speed of computers enabled full evaluation of the proposed similarity measures. Thus, the conclusions about the success or failure of a method are firmly based on the data. This is a desirable consequence of a data-rich approach.

From the research that is presented in this thesis, it is clear that computational methods have a rich potential for the study of music; not as a replacement of 'traditional' methods, but as an extension of the research methods that are available to the musicologist. Pursuing computational methods would result in better understanding of musical data, precise formulation of models, and, thus, to better understanding of musicological concepts. Therefore, it would be to the benefit of Musicology not to leave these methods unused.

Bibliography

Aarden, B. & Huron, D. (2001). Mapping European Folksong: Geographical Localization of Musical Features. *Computing in Musicology*, 12, 169–183. 18

Adams, N., Bartsch, M., Shifrin, J., & Wakefield, G. (2004). Time Series Alignment for Music Information Retrieval. In *Proceedings of the 5th International Conference on Music Information Retrieval (ISMIR 2004)* Barcelona. 93

Ahlbäck, S. (2004). *Melody beyond Notes: A Study of Melody Cognition*. PhD thesis, Göteborgs Universitet. 29

Backer, E. & van Kranenburg, P. (2005). On musical stylometry—a pattern recognition approach. *Pattern Recognition Letters*, 26(3), 299–309. 71, 73, 79

Bartók, B. (1981). *The Hungarian Folk Song*. Number 13 in Béla Bartók Archives. Studies in Musicology. Albany: State University of New York Press. 24

Bayard, S. (1950). Prolegomena to a Study of the Principal Melodic Families of British-American Folk Song. *Journal of American Folklore*, 63(247), 1–44. 5, 11, 13, 124

Bod, R. (2002). Memory-Based Models of Melodic Analysis: Challenging the Gestalt Principles. *Journal of New Music Research*, 31(1), 27–37. 18

Bohlman, P. (1988). *The Study of Folk Music in the Modern World*. Bloomington: Indiana University Press. 14

Brandts Buys, M., Ed. (1975). *Gezelschapsliederen*. Leiden: A.W. Sijthoff. 15

Bronson, B. H. (1949). Mechanical help in the study of folk song. *The Journal of American Folklore*, 62(244), 81–86. 11, 16

Bronson, B. H. (1950). Some Observations About Melodic Variation in British-American Folk Tunes. *Journal of the American Musicological Society*, 3, 120–134. 11, 14

Bronson, B. H. (1951). Melodic Stability in Oral Transmission. *Journal of the International Folk Music Council*, 3, 50–55. 10

Bronson, B. H. (1959). Toward the Comparative Analysis of British-American Folk Tunes. *The Journal of American Folklore*, 87(284), 165–191. 16, 25

Cambouropoulos, E. (1998). *Towards a General Computational Theory of Musical Structure*. PhD thesis, University of Edinburgh. 124

Clarke, E. & Cook, N., Eds. (2004). *Empirical Musicology: Aims, Methods, Prospects*. Oxford: Oxford University Press. 8, 125

Claus, R. D. (1998). *Zur Echtheit von Toccata und Fuge d-moll BWV 565*. Köln-Rheinkassel: Verlag Dohr, second edition. 88

Claus, R. D. (1999). *Johann Peter Kellner. Studien zu Leben und Werk*. PhD thesis, Universität Rostock. 78

Conklin, D. (2006). Melodic analysis with segment classes. *Machine Learning*. 18

Conklin, D. & Anagnostopoulou, C. (2001). Representation and Discovery of Multiple Viewpoint Patterns. In *International Computer Music Conference, September 17-22, in Havana, Cuba*. Havana. 18, 29

Cook, N. (2005). Towards the complete musicologist. In *Proceedings of the 6th International Conference on Music Information Retrieval (ISMIR 2005)* London. 3

Cope, D. (1991). *Computers and Musical Style*. Madison, WI: A-R Editions, Inc. 73

Cope, D. (1998). Signatures and Earmarks: Computer Recognition of Patterns in Music. *Computing in Musicology*, 11, 129–138. 73

Cornelis, O., Lesaffre, M., Moelants, D., & Leman, M. (2010). Access to ethnic music: Advances and perspectives in content-based music information retrieval. *Signal Processing*, 90(4), 1008–1031. 31

Cowdery, J. (1984). A Fresh Look at the Concept of Tune Family. *Ethnomusicology*, 28(3), 495–504. 12, 29, 34, 123, 124

Cowdery, J. R. (1990). *The Melodic Tradition of Ireland*. Kent, Ohio: Kent State University Press. 25, 137

Dannenberg, R., Thom, B., & Watson, D. (1997). A Machine Learning Approach to Musical Style Recognition. In *Proceedings of the 1997 International Computer Music Conference* (pp. 344–347). 73

De Cheveigne, A. & Kawahara, H. (2002). YIN, a fundamental frequency estimator for speech and music. *Journal of the Acoustical Society of America*, 111(4), 1917–1930. 125

Deliège, I. (2001). Prototype effects in music listening: An empirical approach to the notion of imprint. *Music Perception*, 18, 371–407. 39

Dirksen, P. (2000). Het auteurschap van Praeludium en fuga in f (BWV 534). *Het Orgel*, 96(5), 5–14. 85

Downie, J. S. (2003). Music Information Retrieval. *Annual Review of Information Science and Technology*, 37, 295–340. 3

Duin, R. & Tax, D. M. (2005). Statistical Pattern Recognition. In C. Chen & P. Wang (Eds.), *Handbook of Pattern Recognition and Computer Vision* (pp. 3–24). Singapore: World Scientific Publishing. 73

Eck, D. & Casagrande, N. (2005). Finding Meter in Music Using An Autocorrelation Phase Matrix and Shannon Entropy. In *Proceedings of the 6th International Conference on Music Information Retrieval (ISMIR 2005)* London. 18

Eerola, T., Järvinen, T., Louhivuori, J., & Toiviainen, P. (2001). Statistical Features and Perceived Similarity of Folk Melodies. *Music Perception*, 18(3), 275–296. 39

Eerola, T. & Toiviainen, P. (2004). MIR in Matlab: The MIDI Toolbox. In *Proceedings of the 5th International Conference on Music Information Retrieval (ISMIR 2004)*. 18

Elbourne, R. (1975). The Question of Definition. *Yearbook of the International Folk Music Council*, 7, 9–29. 8

Elscheková, A. (1966). Methods of Classification of Folk-Tunes. *Journal of the International Folk Music Council*, XVIII, 56–76. 14

Falck, M. (1956). *Wilhelm Friedemann Bach: Sein Leben und seine Werke*. Leipzig: Kahnt. 78

Ferraro, P., Hanna, P., Imbert, L., & Izard, T. (2009). Accelerating Query-By-Humming on GPU. In *Proceedings of the 10th International Conference on Music Information Retrieval (ISMIR 2009)* Kobe. 18

Frieler, K. (2007). Visualizing Music on the Metrical Circle. In *Proceedings of the 8th International Conference on Music Information Retrieval (ISMIR 2007)*. 18

Garbers, J. (2004). *Integration von Bedien- und Programmiersprachen am Beispiel von OpenMusic, Humdrum und Rubato*. PhD thesis, Fakultät IV – Elektrotechnik und Informatik der Technischen Universität Berlin. 33

Goldstone, R. (1994). Similarity, interactive activation, and mapping. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 20, 3–28. 92

Gómez, C., Abad-Mota, S., & Ruckhaus, E. (2007). An Analysis of the Mongeau-Sankoff Algorithm for Music Information Retrieval. In *Proceedings of the 8th International Conference on Music Information Retrieval (ISMIR 2007)* Vienna. 93

Gotoh, O. (1982). An Improved Algorithm for Matching Biological Sequences. *Journal of Molecular Biology*, 162, 705–708. 66, 92, 127

Grijp, L. P. (2008). Introduction. In L. P. Grijp & I. van Beersum (Eds.), *Under the Green Linden — 163 Dutch Ballads from the Oral Tradition* (pp. 18–27). Amsterdam: Meertens Institute + Music & Words. 1

Hewlett, W. B. (1992). A Base-40 Number-line Representation of Musical Pitch. *Musikometrika*, 4, 1–14. 94

Hewlett, W. B. & Selfridge-Field, E. (1991). Computing in Musicology, 1966-91. *Computers and the Humanities*, 25, 381–392. 2

Hoos, H. H., Renz, K., & Görg, M. (2001). GUIDO/MIR—An Experimental Musical Information Retrieval System based on GUIDO Music Notation. In *Proceedings of the 2nd International International Symposium on Music Information Retrieval (ISMIR 2001)*. 18

Humphreys, D. (1985). Did J.S. Bach Compose the F minor Prelude and Fugue BWV 534? In P. Williams (Ed.), *Bach, Handel, Scarlatti – Tercentenary Essays* Cambridge: Cambridge University Press. 85

Humphreys, D. (1989). A Bach polyglot – the A major Prelude & Fugue BWV 536. *The Organ Yearbook*, XX, 72–87. 85

Huron, D. (1995). The Melodic Arch in Western Folksongs. *Computing in Musicology*, 10, 3–23. 17, 18, 23

Huron, D. (1999a). Music Research Using Humdrum: A User's Guide. Accessed 1 June 2010 from http://csml.som.ohio-state.edu/Humdrum/guide.toc.html. 2, 17, 20, 33

Huron, D. (1999b). The New Empiricism: Systematic Musicology in a Post-modern Age. Accessed 1 June 2010 from http://musiccog.ohio-state.edu/Music220/Bloch.lectures/3.Methodology.html. 8

Jesser, B. (1991). *Interaktive Melodieanalyse*, volume 12 of *Studien zur Volksliedforschung*. Bern: Peter Lang. 16, 29, 55, 56, 157

Juhász, Z. (2000). Contour Analysis of Hungarian Folk Music in a Multidimensional Metric-Space. *Journal of New Music Research*, 29(1), 71–83. 17

Juhász, Z. (2002). The Structure of an Oral Tradition - Mapping of Hungarian Folk Music to a Metric Space. *Journal of New Music Research*, 31(4), 295–310. 17

Juhász, Z. (2004). Segmentation of Hungarian Folk Songs Using an Entropy-Based Learning System. *Journal of New Music Research*, 33(1), 5–15. 17

Juhász, Z. (2006). A Systematic Comparison of Different European Folk Music Traditions Using Self-Organizing Maps. *Journal of New Music Research*, 35(2), 95–112. 17

Keller, H. (1937). Unechte Orgelwerke Bachs. *Bach-Jahrbuch*, 34, 59–82. 88

Keller, H. ([1950]). *Die Orgelwerke Bachs*. Leipzig: Peters. 85

Keller, M. S. (1984). The Problem of Classification in Folksong Research: A Short History. *Folklore*, XCV(1), 100–104. 14

Keller, M. S. (1988). Segmental Procedures in the Transcription of Folk Songs in Trentino. *Sonus*, 8(2), 37–45. 12, 124, 137

Klusen, E., Moog, H., & Piel, W. (1978). Experimente zur mündlichen Tradition von Melodien. *Jahrbuch für Volksliedforschung*, 23, 11–32. 11, 109, 112

Krohn, I. (1903). Welche ist die beste Methode, um Volks- und volksmässige Lieder nach ihrer melodischen (nicht textlichen) Beschaffenheit lexikalisch zu ordnen? *Sammelbände der internationalen Musikgesellschaft*, 4(4), 643–60. 14

Lartillot, O. (2004). A Musical Pattern Discovery System Founded on a Modeling of Listening Strategies. *Computer Music Journal*, 28(3), 53–67. 29

Lemström, K. & Ukkonen, E. (2000). Including Interval Encoding into Edit Distance Based Music Comparison and Retrieval. In *Proceedings of the AISB'00 Symposium on Creative & Cultural Aspects and Applications of AI & Cognitive Science* (pp. 53–60). Birmingham. 93

Lerdahl, F. & Jackendoff, R. (1983). *A Generative Theory of Tonal Music*. Cambridge, Mass.: MIT Press. 124

Levenshtein, V. (1966). Binary Codes Capable of Correcting Deletions, Insertions and Reversals. *Soviet Physics Doklady*, 10(8), 707–710. 92

Love, H. (2002). *Attributing Authorship: An Introduction*. Cambridge: Cambridge University Press. 72

Margulis, E. (2005). A Model of Melodic Expectation. *Music Perception*, 20(4), 663–714. 111

Mason, R. M. (1985). *Modern Methods of Music Analysis using Computers*. Peterborough, NH: Schoolhouse Press. 77

McCarty, W. (2005). *Humanities Computing*. Basingstoke: Palgrave Macmillan. 31

McKay, C. (2004). Automatic genre classification of MIDI recordings. Master's thesis, McGill University, Canada. 29, 55, 56, 153, 154, 155, 156

McKay, C. & Fujinaga, I. (2004). Automatic genre classification using large high-level musical feature sets. In *Proceedings of the 5th International Conference on Music Information Retrieval (ISMIR 2004)* (pp. 525–530). 73

McNab, R. J., Smith, L. A., David, B., & H., W. I. (1997). The New Zealand Digital Library MELody inDEX. *D-Lib magazine*, 3(5). 20, 92, 93

Mendel, A. (1969). Some Preliminary Attempts at Computer-Assisted Style Analysis in Music. *Computers and the Humanities*, 4(1), 41–52. 2

Meyer, L. (1996). *Style and Music—Theory, History, and Ideology*. Chicago: University of Chicago Press. 25, 28, 73, 74

Miller, G. (1956). The Magical Number Seven, Plus or Minus Two: Some Limits on our Capacity for Processing Information. *Psychological Review*, 63, 81–97. 124

Mongeau, M. & Sankoff, D. (1990). Comparison of Musical Sequences. *Computers and the Humanities*, 24, 161–175. 20, 92, 116

Müllensiefen, D. & Frieler, K. (2004). Cognitive Adequacy in the Measurement of Melodic Similarity: Algorithmic vs. Human Judgements. *Computing in Musicology*, 15, 147–177. 116

Müllensiefen, D. & Frieler, K. (2007). Modelling experts' notions of melodic similarity. *Musicae Scientiae*, Discussion Forum 41, 183–210. Institute of Musicology, University of Hamburg. 39

Namour, E. (1992). *The Analysis and Cognition of Basic Melodic Structures*. Chicago: University of Chicago Press. 124

Needleman, S. B. & Wunsch, C. D. (1990). A general method applicable to the search for similarities in the amino acid sequence of two proteins. *Journal of Molecular Biology*, 48(3), 443–453. 66, 92

Nettl, B. (2005). *The Study of Ethnomusicology: Thirty-one Issues and Concepts*. Urbana and Chicago: University of Illinois Press, 2 edition. 4, 11, 12, 13, 14, 27, 29, 124

O'Donnell, J. (1989). Mattheson, Bach, Krebs and the Fantasia & Fugue in C minor BWV 537. *The Organ Yearbook*, XX, 88–95. 86

Pearce, M., Müllensiefen, D., & Wiggins, G. A. (2008). A Comparison of Statistical and Rule-Based Methods of Melodic Segmentation. In *Proceedings of the 9th International Conference on Music Information Retrieval (ISMIR 2008)* Philadelphia, PA, USA. 18

Peretz, I. (2006). The Nature of Music from a Biological Perspective. *Cognition*, 100, 1–32. 27, 124

Peretz, I. & Zatorre, R. J. (2005). Brain Organization for Music Processing. *Annual Review of Psychology*, 56, 89–114. 10

Pickens, J. (2000). A Comparison of Language Modeling and Probabilistic Text Information Retrieval Approaches to Monophonic Music Retrieval. In *ISMIR*. 18

Pudil, P., Novovičivà, J., & Kittler, J. (1994). Floating Search Methods in Feature Selection. *Pattern Recognition Letters*, 15(11), 1119–25. 61, 66, 79

Rosch, E. (1973). Natural Categories. *Cognitive Psychology*, 4, 328–350. 39

Rubin, D. C. (1995). *Memory in Oral Traditions*. New York: Oxford University Press. 12

Sapp, C. (2005). Online Database of Scores in the Humdrum File Format. In *Proceedings of the 6th International Conference on Music Information Retrieval (ISMIR 2005)* London. 18

Schaffrath, H. (1992). The Retrieval of Monophonic Melodies and Their Variants—Concepts and Strategies for Computer-Aided Analysis. In A. Marsden & A. Pople (Eds.), *Computer Representations and Models in Music* (pp. 95–109). London: Academic Press. 16

Schaffrath, H., Ed. (1995). *The Essen Folksong Collection*. Stanford, CA: Center for Computer Assisted Research in the Humanities. 18, 20

Schaffrath, H. (1997). The Essen Associative Code: A Code for Folksong Analysis. In E. Selfridge-Field (Ed.), *Beyond Midi—The Handbook of Musical Codes* (pp. 343–361). Cambridge: The MIT Press. 16

Schmieder, W. (1990). *Thematisches-systematisches Verzeichnis der musikalischen Werke von Johann Sebastian Bach. Bach-Werke-Verzeichnis—2. überarbeitete und erweiterte Ausgabe*. Wiesbaden: Breitkopf & Härtel. 73, 78

Schweitzer, A. (1955). *J.S. Bach*. Wiesbaden: Breitkopf & Härtel, [reprint of the 1908 version] edition. 85

Shannon, C. E. (1948). A mathematical theory of communication. *Bell System Technical Journal*, 27, 379–423, 623–656. 77

Singer, J. (2004). Creating a nested melodic representation: competition and cooperation among bottom-up and top-down Gestalt principles. In *Proceedings of the 5th International Conference on Music Information Retrieval (ISMIR 2004)*. 18

Skalak, M., Han, J., & Pardo, B. (2008). Speeding Melody Search with Vantage Point Trees. In *Proceedings of the 9th International Conference on Music Information Retrieval (ISMIR 2008)* Philadelphia, PA, USA. 18

Smith, T. F. & Waterman, M. S. (1981). Identification of Common Molecular Subsequences. *Journal of Molecular Biology*, 147(1), 195–197. 126

Spitta, P. (1916). *Johann Sebastian Bach*. Leipzig: Breitkopf & Härtel, second edition. 85

Stein, E. (1979). The Use of Computers in Folklore and Folk Music: a Preliminary Bibliography. Accessed 1 June 2010 from http://www.loc.gov/folklife/guides/BibComputers.html. 16

Steinbeck, W. (1982). *Struktur und Ähnlichkeit—Methoden automatischer Melodieanalyse*, volume XXV of *Kieler Schriften zur Musikwissenschaft*. Kassel: Bärenreiter. 16, 29, 55, 56, 156, 157

Suchoff, B. (1981). Preface. In *The Hungarian Folk Song* (pp. ix–lv). Albany: State University of New York Press. 14

Suppan, W. & Stief, W., Eds. (1976). *Melodietypen des Deutschen Volksgesanges*. Tutzing: Hans Schneider. 15, 25

Sutcliffe, J. (1993). *Concept, Class, and Category in the Sense of Aristotle*, (pp. 33–65). Academic Press: London. 38

Tappert, W. (1890). *Wandernde Melodien: Eine musikalische Studie*. Leipzig: List und Francke. 124

Temperley, D. (2006). A Probabilistic Model of Melody Perception. In *Proceedings of the 7th International Conference on Music Information Retrieval (ISMIR 2006)* Victoria, BC. 18

Tittel, K. (1966). Welche unter J.S. Bachs Namen geführten Orgelwerke sind Johann Tobias bzw. Johan Ludwig Krebs zuzuschreiben? *Bach-Jahrbuch*, 52, 102–137. 88

Toiviainen, P. & Eerola, T. (2005). Classification of Musical Metre with Autocorrelation and Discriminant Functions. In *Proceedings of the 6th International Conference on Music Information Retrieval (ISMIR 2005)* London. 18

Trowbridge, L. (1982). *The Fifteenth-Century French Chanson: A Computer-Aided Study of Styles and Style Change*. PhD thesis, University of Illinois. 72

Trowbridge, L. (1985). Style Change in the Fifteenth-Century Chanson. *Journal of Musicology*, 4(2), 146–170. 72

Typke, R. (2007). *Music Retrieval based on Melodic Similarity*. PhD thesis, Utrecht University, Utrecht. 117

Typke, R., Giannopoulos, P., Veltkamp, R. C., Wiering, F., & van Oostrum, R. (2003). Using transportation distances for measuring melodic similarity. In *Proceedings of the 4th International Conference on Music Information Retrieval (ISMIR 2003)* Baltimore. 20, 23

Tzanetakis, G., Kapur, A., Schloss, W., & Wright, M. (2007). Computational Ethnomusicology. *Journal of interdisciplinary music studies*, 1(2), 1–24. 31

Uitdenbogerd, A. L. & Zobel, J. (1999). Melodic matching techniques for large music databases. In *ACM Multimedia '99* (pp. 57–66).: ACM, ACM Press. 93

Volk, A. (2008). Persistence and Change: Local and Global Components of Metre Induction using Inner Metric Analysis. *Journal of Mathematics and Computation in Music*, 2(2), 99–115. 101, 106

Weinberger, G., Ed. (1985). *Johann Ludwig Krebs, Sämtliche Orgelwerke*. Wiesbaden: Breitkopf & Härtel. 78

Wiering, F., Veltkamp, R. C., Garbers, J., Volk, A., & Van Kranenburg, P. (2009). Modelling Folksong Melodies. *Interdisciplinary Science Reviews*, 34(2–3), 154–171. xiii

Wiora, W. (1941). Systematik der musikalischen Erscheinungen des Umsingens. *Jahrbuch für Volksliedforschung*, 7, 128–195. 10, 25

Wittgenstein, L. (1953). *Philosophical Investigations*. London: Basil Blackwell. 39

Wollny, P. (1993). *Studies in the Music of Wilhelm Friedemann Bach: Sources and Style*. PhD thesis, Harvard University. 78

Ziv, N. & Eitan, Z. (2007). Themes as prototypes: Similarity judgments and categorization tasks in musical contexts. *Musicae Scientiae*, Discussion Forum 4A, 99–133. 39

Appendix A

The Contents of the Annotated Corpus

The following table provides an overview of the contents of the Annotated Corpus that is used in various chapters in this thesis. It shows the names of the tune families and the number of songs in each tune family.

Tune Family (short)	**Tune Family (long)**	**Size**
Heer	Daar ging een heer 1	16
Jonkheer	Daar reed een jonkheer 1	12
Ruiter 2	Daar was laatstmaal een ruiter 2	17
Maagdje	Daar zou er een maagdje vroeg opstaan 2	10
Dochtertje	Een Soudaan had een dochtertje 1	13
Lindeboom	Een lindeboom stond in het dal 1	9
Zoeteliefjes	En er waren eens twee zoeteliefjes	16
Ruiter 1	Er reed er eens een ruiter 1	27
Herderinnetje	Er was een herderinnetje 1	11
Koopman	Er was een koopman rijk en machtig	17
Meisje	Er was een meisje van zestien jaren 1	15
Vrouwtje	Er woonde een vrouwtje al over het bos	12
Femmes	Femmes voulez vous éprouver	13
Halewijn 2	Heer Halewijn 2	11
Halewijn 4	Heer Halewijn 4	11
Stavoren	Het vrouwtje van Stavoren 1	8

Tune Family (short)	Tune Family (long)	Size
Zomerdag	Het was laatst op een zomerdag	17
Driekoningenavond	Het was op een driekoningenavond 1	12
Stad	Ik kwam laatst eens in de stad	18
Stil	Kom laat ons nu zo stil niet zijn 1	11
Schipper	Lieve schipper vaar me over 1	15
Nood	O God ik leef in nood	8
Soldaat	Soldaat kwam uit de oorlog	17
Bruidje	Vaarwel bruidje schoon	11
Verre	Wat zag ik daar van verre 1	15
Boom	Zolang de boom zal bloeien 1	18

The following overview shows which songs belong to the tune families in the Annotated Corpus. The songs are identified by their recordnumber in the Dutch Song Database and by the strophe number. Thus, 70078_02 denotes the second strophe of song 70078. A song can be consulted by entering the record number in the search field on the website of the Dutch Song Database (http://www.liederenbank.nl).

Heer 72587_01 72587_02 72774_02 73046_01 73588_01 73672_01 73681_01 73743_01 73822_01 74004_01 74048_02 74227_01 75551_01 76625_01 76632_01 144072_01

Jonkheer 70801_01 72154_01 72912_01 72920_01 72946_01 73426_01 73929_01 74028_01 111656_01 112210_01 141251_01 144042_01

Ruiter 2 70493_01 72003_01 72015_01 72627_01 72690_01 72708_01 73287_01 73287_02 73337_02 73483_01 73628_01 73639_01 73990_01 74328_01 74427_01 74552_01 76211_01

Maagdje 15569_01 71441_01 71666_01 72299_01 72306_01 72311_01 72886_01 72886_02 73150_01 138219_01

Dochtertje 73120_01 73269_02 73324_01 73709_01 74308_01 74378_01 74452_01 74547_02 75273_01 75635_01 76426_01 76740_01 124573_01

Lindeboom 70089_01 70141_01 70144_01 70740_01 71958_01 73804_02 73866_01 74754_01 141648_01

Zoeteliefjes 70134_01 72585_01 72638_01 72823_01 73296_01 73331_01 74437_01 74533_01 74583_01 74649_01 75018_02 75040_01 75059_01 75174_01 75249_01 75612_01

Ruiter 1 70996_01 71957_03 72553_01 72559_01 72813_01 72837_01 72851_01 72851_02 72862_01 72883_01 72895_01 72898_01 72898_02 73076_01 73333_01 74246_01 74246_02 74349_01 74433_01 74575_01 75176_01 75184_01 75325_01 75325_02 145525_01 146608_01 162684_01

Herderinnetje 70238_01 70521_01 71016_01 72497_01 73339_01 75309_02 75318_01 112115_01 141407_01 141649_01 146728_01

Koopman 70079_01 70122_01 70475_01 70606_01 72441_01 72967_01 73031_01 73146_01 73788_01 73998_01 74948_01 75013_01 75313_01 75431_01 146699_01 151180_01 152778_01

Meisje 72355_01 72355_02 72355_12 72356_02 72357_01 72358_01 72359_01 72360_01 72457_01 73775_02 74260_01 74260_02 74334_01 74336_01 147463_01

Vrouwtje 73862_01 74309_01 74443_01 74593_01 75739_03 75742_01 75771_02 75848_01 75881_02 76076_01 76258_01 76271_01

Femmes 70526_01 72450_01 72871_01 74286_01 74390_01 74470_01 75034_01 75906_01 76118_01 76495_01 111478_01 125421_01 146731_01

Halewijn 2 72254_01 72255_01 72257_01 72378_01 73750_01 73754_01 73994_01 74156_03 74261_01 74426_02 74613_02

Halewijn 4 72248_01 72250_01 72253_01 72256_01 73326_01 74003_01 74216_01 74333_01 74603_01 143240_01 147912_01

Stavoren 70078_01 70125_01 70360_01 70535_01 70693_01 71227_01 72237_01 72500_01

Zomerdag 72482_01 72505_01 72567_01 72664_01 72754_01 73516_01 73777_01 73946_01 74104_01 74166_01 74342_01 74938_01 74956_01 75065_01 75074_01 75532_01 75616_01

Driekoningenavond 70033_01 71669_01 71974_02 72382_02 72614_01 72647_01 72881_01 73486_01 74276_01 74277_01 75063_01 75367_01

Stad 70137_01 71478_01 72721_01 73685_01 73803_01 73879_02 73895_01 73958_01 74077_02 74672_01 74769_01 74769_02 74840_01 74840_02 74860_01 75191_01 75831_01 76303_01

Stil 72085_01 73404_01 73562_01 73939_01 75021_01 75035_01 75057_01 75156_01 75167_01 75379_01 75525_01

Schipper 70411_01 70463_01 70532_01 70839_01 71237_01 72897_01 73374_01 76128_01 111484_01 112233_01 125427_01 134480_01 135273_01 145856_01 167193_01

Nood 72624_01 72968_01 74007_01 74038_01 75079_01 75307_03 146741_01 152784_01

Soldaat 72103_01 72283_01 72284_01 72285_01 72286_01 72287_01 72288_01 72289_01 73311_01 73393_01 73888_01 73992_01 74157_01 74161_01 74234_01 74468_01 74954_01

Bruidje 70732_01 71369_01 72499_01 73210_01 75158_01 76130_07 112123_01 134389_01 144100_01 162519_01 162526_01

Verre 70492_01 71944_01 72565_01 72665_01 72688_01 72691_01 72714_01 73286_01 73304_01 73771_02 73991_02 73997_01 74962_01 75064_01 75068_01

Boom 70053_01 70096_01 70748_01 71014_01 71064_01 71082_01 73225_01 73225_02 73277_01 73298_01 73626_01 73897_01 74100_01 74182_01 75073_01 111760_01 134474_01 141314_01

Appendix B

The Set of Global Features

The following features from the feature set of McKay are included in the set of global features that is used in this thesis.

Index	**Feature**	**Description as given by McKay (2004)**
1	Amount of Arpeggiation	Fraction of horizontal intervals that are repeated notes, minor thirds, major thirds, perfect fifths, minor sevenths, major sevenths, octaves, minor tenths or major tenths.
2	Average Melodic Interval	Average melodic interval (in semi-tones).
3	Chromatic Motion	Fraction of melodic intervals corresponding to a semi-tone.
4	Combined Strength of Two Strongest Rhythmic Pulses	The sum of the frequencies of the two beat bins of the peaks with the highest frequencies.
5	Direction of Motion	Fraction of melodic intervals that are rising rather than falling.
6	Distance Between Most Common Melodic Intervals	Absolute value of the difference between the most common melodic interval and the second most common melodic interval.
7	Dominant Spread	Largest number of consecutive pitch classes separated by perfect 5ths that accounted for at least 9% each of the notes.

Index	Feature	Description as given by McKay (2004)
8	Duration of Melodic Arcs	Average number of notes that separate melodic peaks and troughs in any channel.
9	Harmonicity of Two Strongest Rhythmic Pulses	The bin label of the higher (in terms of bin label) of the two beat bins of the peaks with the highest frequency divided by the bin label of the lower.
10	Interval Between Strongest Pitch Classes	Absolute value of the difference between the pitch classes of the two most common MIDI pitch classes.
11	Interval Between Strongest Pitches	Absolute value of the difference between the pitches of the two most common MIDI pitches.
12	Melodic Fifths	Fraction of melodic intervals that are perfect fifths.
13	Melodic Octaves	Fraction of melodic intervals that are octaves.
14	Melodic Thirds	Fraction of melodic intervals that are major or minor thirds.
15	Melodic Tritones	Fraction of melodic intervals that are tritones.
16	Most Common Melodic Interval	Melodic interval with the highest frequency.
17	Most Common Melodic Interval Prevalence	Fraction of melodic intervals that belong to the most common interval.
18	Most Common Pitch Class Prevalence	Fraction of Note Ons corresponding to the most common pitch class.
19	Number of Common Melodic Intervals	Number of melodic intervals that represent at least 9% of all melodic intervals.
20	Number of Common Pitches	Number of pitches that account individually for at least 9% of all notes.
21	Number of Moderate Pulses	Number of beat peaks with normalized frequencies over 0.01.
22	Number of Relatively Strong Pulses	Number of beat peaks with frequencies at least 30% as high as the frequency of the bin with the highest frequency.
23	Number of Strong Pulses	Number of beat peaks with normalized frequencies over 0.1.
24	Pitch Class Variety	Number of pitch classes used at least once.

Index	Feature	Description as given by McKay (2004)
25	Pitch Variety	Number of pitches used at least once.
26	Polyrhythms	Number of beat peaks with frequencies at least 30% of the highest frequency whose bin labels are not integer multiples or factors (using only multipliers of 1, 2, 3, 4, 6 and 8) (with an accepted error of +/- 3 bins) of the bin label of the peak with the highest frequency. This number is then divided by the total number of beat bins with frequencies over 30% of the highest frequency.
27	Range	Difference between highest and lowest pitches.
28	Relative Strength of Most Common Intervals	Fraction of melodic intervals that belong to the second most common interval divided by the fraction of melodic intervals belonging to the most common interval.
29	Relative Strength of Top Pitch Classes	The frequency of the 2nd most common pitch class divided by the frequency of the most common pitch class.
30	Relative Strength of Top Pitches	The frequency of the 2nd most common pitch divided by the frequency of the most common pitch.
31	Repeated Notes	Fraction of notes that are repeated melodically.
32	Size of Melodic Arcs	Average melodic interval separating the top note of melodic peaks and the bottom note of melodic troughs.
33	Stepwise Motion	Fraction of melodic intervals that corresponded to a minor or major second.
34	Strength of Second Strongest Rhythmic Pulse	Frequency of the beat bin of the peak with the second highest frequency.
35	Strength of Strongest Rhythmic Pulse	Frequency of the beat bin with the highest frequency.

Index	Feature	Description as given by McKay (2004)
36	Strength Ratio of Two Strongest Rhythmic Pulses	The frequency of the higher (in terms of frequency) of the two beat bins corresponding to the peaks with the highest frequency divided by the frequency of the lower.
37	Strong Tonal Centers	Number of peaks in the fifths pitch histogram that each account for at least 9% of all Note Ons.

The following features from the feature set of Steinbeck are included in the set of global features that is used in this thesis.

Index	Feature	Description (page numbers refer to Steinbeck, 1982.)
38	StdPitch	Standard deviation of the pitch (p.156ff).
39	Ambitus	Difference between the highest and lowest pitch in the melody (p.155).
40	MeanInterval	Mean of the size of the intervals. The intervals between the phrases are not taken into account (p.165ff).
41	StdInterval	Standard Deviation of the size of the intervals (p.165ff).
42	ChangingDirection	The fraction of the intervals that cause a change of direction (p.149f).
43	MeanSteepness	The steepness is the deviation in pitch between two turning points divided by the duration. This feature is the mean of these steepnesses (p.173ff).
44	FractionStressed	The sum of durations that start on a stressed beat as fraction of the total duration (p.178ff).
45	FractionDottedDuration	The fraction of transitions between pitches that has duration quotient 3:1 (p.152ff).
46	FractionHalfDuration	The fraction of transitions between pitches that has duration quotient 2:1 or 1:2 (p.152ff).
47	FractionEqualDurations	The fraction of transitions between pitches that has duration quotient 1:1 (p.152ff).

Index	Feature	Description (page numbers refer to Steinbeck, 1982.)
48	PitchLineCorrelation	The correlation of the pitch contours of the individual lines. For each line the maximum of the correlations with the other lines is taken. Of these values the mean is computed (p.299ff, p.93).
49	DurationLineCorrespondence	Similarity of the sequence of durations. This is computed in the same way as the previous feature, but instead of correlation the fraction of durations that corresponds is taken (p.299ff).

The following features from the feature set of Jesser (1991) are included in the set of global features that is used in this thesis.

Index	Feature	Description
50	prime	fraction of the melodic intervals that is a prime.
51	aminsecond	fraction of the melodic intervals that is an ascending minor second.
52	amajsecond	fraction of the melodic intervals that is an ascending major second.
53	aminthird	fraction of the melodic intervals that is an ascending minor third.
54	amajthird	fraction of the melodic intervals that is an ascending major third.
55	afourth	fraction of the melodic intervals that is an ascending perfect fourth.
56	aaugfourth	fraction of the melodic intervals that is an ascending augmented fourth.
57	afifth	fraction of the melodic intervals that is an ascending perfect fifth.
58	aminsixth	fraction of the melodic intervals that is an ascending minor sixth.
59	amajsixth	fraction of the melodic intervals that is an ascending major sixth.
60	aminseventh	fraction of the melodic intervals that is an ascending minor seventh.
61	amajseventh	fraction of the melodic intervals that is an ascending major seventh.

Index	Feature	Description
62	aoctave	fraction of the melodic intervals that is an ascending perfect octave.
63	ahuge	fraction of the melodic intervals that is larger than an ascending octave.
64	dminsecond	fraction of the melodic intervals that is a descending minor second.
65	dmajsecond	fraction of the melodic intervals that is a descending major second.
66	dminthird	fraction of the melodic intervals that is a descending minor third.
67	dmajthird	fraction of the melodic intervals that is a descending major third.
68	dfourth	fraction of the melodic intervals that is a descending fourth.
69	daugfourth	fraction of the melodic intervals that is a descending augmented fourth.
70	dfifth	fraction of the melodic intervals that is a descending perfect fifth.
71	dminsixth	fraction of the melodic intervals that is a descending minor sixth.
72	dmajsixth	fraction of the melodic intervals that is a descending major sixth.
73	dminseventh	fraction of the melodic intervals that is a descending minor seventh.
74	dmajseventh	fraction of the melodic intervals that is a descending major seventh.
75	doctave	fraction of the melodic intervals that is a descending perfect octave.
76	dhuge	fraction of the melodic intervals that is larger than an descending octave.
77	astep	fraction of the melodic intervals that is an ascending step.
78	aleap	fraction of the melodic intervals that is a ascending leap.
79	dstep	fraction of the melodic intervals that is a descending step.
80	dleap	fraction of the melodic intervals that is a descending leap.

Index	Feature	Description
81	shortestlength	shortest duration such that all durations are a multiple of this shortest duration, except for triplets.
82	doublelength	fraction of the notes with duration of twice the shortest duration.
83	triplelength	fraction of the notes with duration of three times the shortest duration.
84	quadruplelenght	fraction of the notes with duration of four times the shortest duration.
85	dotted	fraction of the notes that is dotted.
86	triplets	fraction of the notes that belongs to a triplet.
87	numlines	number of lines.
88	numpitchclasses	number of distinct pitch classes.

Summary

In order to develop a Music Information Retrieval system for folksong melodies, one needs to design an adequate computational model of melodic similarity, which is the subject of this Ph.D. thesis. Since fundamental understanding of melodies in oral culture as well as fundamental understanding of computational methods to model similarity relations between folksong melodies both are necessary, this problem needs a multidisciplinary approach.

Chapter 2 reviews the relevant academic background of both Folk Song Research (as sub-discipline of Ethnomusicology) and Music Information Retrieval. It also presents an interdisciplinary collaboration model in which Computational Musicology serves a 'man-in-the-middle' role. The particular task of Computational Musicology is to design computational models of concepts from Musicology. In the context of this thesis, the concept of *tune family* is the most important.

An important step towards the understanding of the concept of tune family is the method to annotate similarity relations between melodies that is presented in Chapter 3. It has been developed in close collaboration with musicological domain experts to make aspects of their intuitive similarity assessments explicit. 360 melodies in 26 tune families were 'manually' annotated, resulting in an Annotated Corpus that is a valuable resource for the study of melodic similarity and for the evaluation of computational models of melodic similarity. From the annotations we conclude that the relative importance of the various dimensions (global and local rhythm, global and local contour, motifs, lyrics) varies to a large extent in individual comparisons. Furthermore, the dimension of motifs seems the most important for recognizing melodies. This means that in many cases melodies are judged to be related based on shared characteristic melodic motifs.

In Chapter 4, 88 low-level, global, quantitative features of melody are used to

discriminate between tune families. It appears that such features can be used to recognize melodies within a relatively small corpus (such as the 360 melodies in the Annotated Corpus), but that these features lose their discriminative power in a larger dataset of thousands of melodies.

Chapter 5 uses the same kind of features in another musical domain: baroque organ fugues. Global features are used to assess authorship problems of fugues that are in the catalog of J.S. Bach. Several hypotheses from musicological literature about the authorship of several fugues are supported by findings using this method. The various degrees of success of the same computational method in the previous and the current chapters show that computational methods cannot 'blindly' be applied to musicological questions.

In Chapter 6, the potential of alignment algorithms for folk song melody retrieval is studied by incorporating musical knowledge in the algorithm in the form of appropriate, musically motivated, substitution scoring functions. This approach leads to good retrieval results both for a small (360 melodies) and a large (4830 melodies) dataset. Furthermore, domain experts were able to classify 'problematic' melodies using the results of the alignment algorithms.

The final chapter introduces a local, motif-based approach to the retrieval of folk song melodies using audio recordings rather than the symbolic transcriptions. The melodies are segmented by detecting breath and pauses. Using the resulting audio segments, it proves possible for most of the tune families in the Annotated Corpus to find a characteristic segment that has approximate local matches in other melodies from the same tune family. Retrieval results using these musically meaningful segments are better than using fixed-length segments. The overall retrieval results are not good enough to use in an end user retrieval system yet. However, there is much room for improvement.

This thesis contributes both to Folk Song Research and Music Information Retrieval by incorporating musical knowledge in computational models. The process of developing such models of similarity relations between folk song melodies, leads to better understanding of melodic similarity and, thus, of the concept of tune family, which is relevant for Folk Song Research. The models that have been developed, have successfully been used in the retrieval experiments. From the research that is presented in this thesis, it is clear that computational methods have a rich potential for the study of music; not as a replacement of 'traditional' methods, but as an extension of the research methods that are available to the musicologist.

Samenvatting

Om een zoeksysteem te maken voor volksliedmelodieën is het nodig een computationele methode te ontwikkelen om te berekenen in hoeverre melodieën op elkaar lijken. Dit is het onderwerp van deze dissertatie. Omdat zowel kennis vanuit volksliedonderzoek als vanuit de computerwetenschap nodig is, is een interdisciplinaire benadering noodzakelijk.

Hoofdstuk 2 geeft een overzicht van de relevante academische achtergrond van zowel het volksliedonderzoek (als sub-discipline van etnomusicologie) als onderzoek in 'Music Information Retrieval'. Ook wordt een rollenmodel gepresenteerd voor de samenwerking tussen deze disciplines, waarin de computationele muziekwetenschap een intermediaire rol vervult. Het is de taak van de computationele muziekwetenschap om concepten uit de muziekwetenschap te modelleren met behulp van datastructuren en algoritmes uit de computerwetenschap. In dit proefschrift is het concept 'tune family' (melodiefamilie) het belangrijkste.

Een belangrijke stap voor het begrijpen van dit concept is de nieuw ontwikkelde annotatiemethode die wordt gepresenteerd in hoofdstuk 3. Met behulp van deze methode worden impliciete aspecten van de beoordeling van melodische gelijkenis door domeindeskundigen, expliciet gemaakt. De resulterende annotaties kunnen worden gebruikt om meer inzicht in melodische gelijkenis als zodanig te krijgen en om de resultaten van algoritmes te evalueren. Er zijn 360 melodieën, verdeeld over 26 melodiefamilies geannoteerd. Vanuit deze annotaties wordt duidelijk dat de aspecten die een rol spelen bij het herkennen van melodieën van geval tot geval sterk kunnen verschillen. In het algemeen blijken terugkerende karakteristieke motieven het belangrijkst te zijn.

In hoofdstuk 4 worden 88 kwantitatieve, globale eigenschappen van melodieën gebruikt om automatisch onderscheid tussen melodiefamilies te maken. Het blijkt dat met deze eigenschappen onderscheid gemaakt kan worden tussen melodiefamilies in een kleine verzameling, maar dat het onderscheidend vermogen zeer sterk afneemt in grotere collecties van duizenden melodieën.

Gelijkaardige kwantitatieve eigenschappen worden gebruikt in hoofdstuk 5 om onderscheid te maken tussen orgelfuga's van J.S. Bach en andere componisten. Het blijkt dat deze methode een aantal auteurschapskwesties waarover discussie bestaat in muziekwetenschappelijke literatuur van nieuw bewijsmateriaal voorziet. Het wisselende succes van dezelfde methode in dit en het vorige hoofdstuk, laat zien dat het niet mogelijk is om een computationele methode zonder meer toe te passen op een muziekwetenschappelijk probleem.

In hoofdstuk 6 wordt een uitlijningsalgoritme gebruikt om de mate van gelijkenis tussen melodieën te berekenen. Dit algoritme laat toe dat muzikale kennis wordt geïntegreerd in de vorm van substitutiefuncties. Deze benadering leidt tot zeer goede zoekresultaten, zowel in een kleine als in een grote collectie melodieën. Bovendien kon een aantal 'problematische' melodieën met behulp van de resultaten van deze methode herkend worden en worden ingedeeld in de betreffende melodiefamilies.

Het laatste hoofdstuk introduceert een benadering die gebruik maakt van kleine melodische segmentjes om melodieën te herkennen. Hiervoor worden de geluidsopnamen gebruikt in plaats van het notenschrift. De melodieën worden opgesplitst op plaatsen waar ademhaling of stilte gedetecteerd wordt. Het blijkt voor de meeste melodiefamilies mogelijk om onder de resulterende melodiesegmenten die segmenten te vinden die karakteristiek zijn voor melodieën in de betreffende familie. Voorts blijkt dat met behulp van segmenten van vaste lengte minder melodieën herkend kunnen worden, dan met de muzikaal zinvolle segmentering. Hoewel de behaalde zoekresultaten nog niet goed genoeg zijn, biedt deze methode interessante aanknopingspunten voor verder onderzoek.

Dit proefschrift levert een bijdrage aan het volksliedonderzoek en aan Music Information Retrieval door muzikale kennis te integreren in computationele modellen. Het proces van modelleren leidt tot nieuwe kennis over waargenomen gelijkenis tussen melodieën, en dus over het concept melodiefamilie, wat relevant is voor het volksliedonderzoek. De in dit proefschrift ontwikkelde uitlijningsmodellen worden met succes gebruikt om melodieën te vinden in collectie volkliederen van het Meertens Instituut.

Het is duidelijk dat de inzet van dit soort computationele methoden een waardevolle toevoeging is aan de onderzoeksmethoden van de muziekwetenschap. De muziekwetenschap zou zichzelf tekort doen door deze methoden links te laten liggen.

Curriculum Vitæ

Peter van Kranenburg was born in Schiedam, Netherlands, on June 9th 1977. He received his preparatory education at Wartburg College, Rotterdam. He obtained masters degrees in Electrical Engineering (2003, Delft University of Technology) and Musicology (2004, Utrecht University). In his master studies he developed machine learning methods for studying authorship of musical compositions, with a particular application to the authorship of the organ fugue BWV 534/2, ascribed to J.S. Bach. From 2006 till 2010, he was Ph.D. researcher at Utrecht University under supervision of dr. Frans Wiering and prof.dr. Remco Veltkamp (Utrecht University) and prof.dr. Louis Grijp (Meertens Institute). The research, which is presented in this thesis, focuses on the computational modeling of similarity of folk-song melodies for retrieval purposes.

Acknowledgements

I would like to thank several people who played an important role in the realization of this thesis.

First of all my supervisors Frans Wiering, who also was the project leader, Remco Veltkamp and Louis Grijp. I learned a lot from all of them during the last four years. For reasons explained in the preface of this thesis, I was very happy to have a computational musicologist as project leader. I really enjoyed the many stimulating conversations with Frans about all kinds of musical subjects. His advices, directions and proof readings were very important. From Remco I learned to act and think strategically and purposefully when writing papers, doing reviews, preparing presentations, etc. I thank Louis for his interest and open-mindedness towards computational approaches. His willingness to commit resources and man hours of the song division of the Meertens Institute was all-important for the project and for the results that are achieved in this thesis.

Special thanks goes to my direct colleagues and fellow project members. The broad overview and the conscious scientific approach of Anja Volk have been inspiring. I enjoyed writing papers together with her. From collaboration with Jörg Garbers I learned a lot about software development, employment of algorithms, programming and the like. Furthermore, I could not have found a better travel companion. The many hours we spend together during and after conferences at various places on the globe are memorable. Both Anja and Jörg were perfect roommates at the Meertens Institute. I appreciated the conversations with and suggestions of current and former members of the music research group at Utrecht, Rainer Typke, Martijn Bosma, and Bas de Haas. The evaluation software of Bas has been important for this thesis to get a proper evaluation of retrieval results. I also like to thank the colleagues at the Meertens Institute, Martine de Bruin, Ellen van der Grijn Santen, Marieke Klein and

Mariet Kaptein for their cooperation and engagement. They all contributed much to the project. Without their efforts, we would not have such a valuable collection of digitized songs. Under their supervision and with help of musicology students, in the course of the project, thousands of songs and instrumental melodies have been digitized.

During a summer school at Maastricht University, Jeroen Donkers suggested me to have a look at alignment algorithms, which turned out to be a valuable suggestion. Donncha O'Maidin very kindly made his software library CPNView available to me, which I used to parse humdrum **kern files. I also thank George Tzanetakis, who was willing to have me as a visiting Ph.D. student at the University of Victoria (BC) for three months in 2009. I very much enjoyed our cooperation. A discussion with George is like being in a continuous flow of research ideas, which is very stimulating. I very much appreciated the company and involvement of Steven Ness, with whom I shared the lab at UVic. I thank the University of Victoria and Utrecht University for financial support of this visit.

Finally, I would like to thank my friends and relatives, who continuously showed their interest and support for what I was doing. Without people, research, ideas and music are meaningless.

SIKS Dissertatiereeks

1998-1 Johan van den Akker (CWI): DEGAS — An Active, Temporal Database of Autonomous Objects
1998-2 Floris Wiesman (UM): Information Retrieval by Graphically Browsing Meta-Information
1998-3 Ans Steuten (TUD): A Contribution to the Linguistic Analysis of Business Conversations within the Language/Action Perspective
1998-4 Dennis Breuker (UM): Memory versus Search in Games
1998-5 E.W. Oskamp (RUL): Computerondersteuning bij Straftoemeting
1999-1 Mark Sloof (VU): Physiology of Quality Change Modelling; Automated modelling of Quality Change of Agricultural Products
1999-2 Rob Potharst (EUR): Classification using decision trees and neural nets
1999-3 Don Beal (UM): The Nature of Minimax Search
1999-4 Jacques Penders (UM): The practical Art of Moving Physical Objects
1999-5 Aldo de Moor (KUB): Empowering Communities: A Method for the Legitimate User-Driven Specification of Network Information Systems
1999-6 Niek J.E. Wijngaards (VU): Re-design of compositional systems
1999-7 David Spelt (UT): Verification support for object database design
1999-8 Jacques H.J. Lenting (UM): Informed Gambling: Conception and Analysis of a Multi-Agent Mechanism for Discrete Reallocation.
2000-1 Frank Niessink (VU): Perspectives on Improving Software Maintenance
2000-2 Koen Holtman (TUE): Prototyping of CMS Storage Management
2000-3 Carolien M.T. Metselaar (UVA): Sociaal-organisatorische gevolgen van kennistechnologie; een procesbenadering en actorperspectief.
2000-4 Geert de Haan (VU): ETAG, A Formal Model of Competence Knowledge for User Interface Design
2000-5 Ruud van der Pol (UM): Knowledge-based Query Formulation in Information Retrieval.
2000-6 Rogier van Eijk (UU): Programming Languages for Agent Communication
2000-7 Niels Peek (UU): Decision-theoretic Planning of Clinical Patient Management
2000-8 Veerle Coup (EUR): Sensitivity Analyis of Decision-Theoretic Networks
2000-9 Florian Waas (CWI): Principles of Probabilistic Query Optimization
2000-10 Niels Nes (CWI): Image Database Management System Design Considerations, Algorithms and Architecture
2000-11 Jonas Karlsson (CWI): Scalable Distributed Data Structures for Database Management
2001-1 Silja Renooij (UU): Qualitative Approaches to Quantifying Probabilistic Networks
2001-2 Koen Hindriks (UU): Agent Programming Languages: Programming with Mental Models
2001-3 Maarten van Someren (UvA): Learning as problem solving
2001-4 Evgueni Smirnov (UM): Conjunctive and Disjunctive Version Spaces with Instance-Based Boundary Sets
2001-5 Jacco van Ossenbruggen (VU): Processing Structured Hypermedia: A Matter of Style
2001-6 Martijn van Welie (VU): Task-based User Interface Design
2001-7 Bastiaan Schonhage (VU): Diva: Architectural Perspectives on Information Visualization
2001-8 Pascal van Eck (VU): A Compositional Semantic Structure for Multi-Agent Systems Dynamics.
2001-9 Pieter Jan 't Hoen (RUL): Towards Distributed Development of Large Object-Oriented Models, Views of Packages as Classes
2001-10 Maarten Sierhuis (UvA): Modeling and Simulating Work Practice BRAHMS: a multiagent modeling and simulation language for work practice analysis and design
2001-11 Tom M. van Engers (VUA): Knowledge Management: The Role of Mental Models in Business Systems Design
2002-01 Nico Lassing (VU): Architecture-Level Modifiability Analysis
2002-02 Roelof van Zwol (UT): Modelling and searching web-based document collections
2002-03 Henk Ernst Blok (UT): Database Optimization Aspects for Information Retrieval
2002-04 Juan Roberto Castelo Valdueza (UU): The Discrete Acyclic Digraph Markov Model in Data Mining
2002-05 Radu Serban (VU): The Private Cyberspace Modeling Electronic Environments inhabited by Privacy-concerned Agents
2002-06 Laurens Mommers (UL): Applied legal epistemology; Building a knowledge-based ontology of the legal domain
2002-07 Peter Boncz (CWI): Monet: A Next-Generation DBMS Kernel For Query-Intensive Applications
2002-08 Jaap Gordijn (VU): Value Based Requirements Engineering: Exploring Innovative E-Commerce Ideas
2002-09 Willem-Jan van den Heuvel(KUB): Integrating Modern Business Applications with Objectified Legacy Systems
2002-10 Brian Sheppard (UM): Towards Perfect Play of Scrabble
2002-11 Wouter C.A. Wijngaards (VU): Agent Based Modelling of Dynamics: Biological and Organisational Applications
2002-12 Albrecht Schmidt (Uva): Processing XML in Database Systems
2002-13 Hongjing Wu (TUE): A Reference Architecture for Adaptive Hypermedia Applications
2002-14 Wieke de Vries (UU): Agent Interaction: Abstract Approaches to Modelling, Programming and Verifying Multi-Agent Systems
2002-15 Rik Eshuis (UT): Semantics and Verification of UML Activity Diagrams for Workflow Modelling
2002-16 Pieter van Langen (VU): The Anatomy of Design: Foundations, Models and Applications
2002-17 Stefan Manegold (UVA): Understanding, Modeling, and Improving Main-Memory Database Performance
2003-01 Heiner Stuckenschmidt (VU): Ontology-Based Information Sharing in Weakly Structured Environments
2003-02 Jan Broersen (VU): Modal Action Logics for Reasoning About Reactive Systems
2003-03 Martijn Schuemie (TUD): Human-Computer Interaction and Presence in Virtual Reality Exposure Therapy
2003-04 Milan Petkovic (UT): Content-Based Video Retrieval Supported by Database Technology
2003-05 Jos Lehmann (UVA): Causation in Artificial Intelligence and Law — A modelling approach
2003-06 Boris van Schooten (UT): Development and specification of virtual environments
2003-07 Machiel Jansen (UvA): Formal Explorations of Knowledge Intensive Tasks
2003-08 Yongping Ran (UM): Repair Based Scheduling
2003-09 Rens Kortmann (UM): The resolution of visually guided behaviour
2003-10 Andreas Lincke (UvT): Electronic Business Negotiation: Some experimental studies on the interaction between medium, innovation context and culture
2003-11 Simon Keizer (UT): Reasoning under Uncertainty in Natural Language Dialogue using Bayesian Networks
2003-12 Roeland Ordelman (UT): Dutch speech recognition in multimedia information retrieval
2003-13 Jeroen Donkers (UM): Nosce Hostem - Searching with Opponent Models
2003-14 Stijn Hoppenbrouwers (KUN): Freezing Language: Conceptualisation Processes across ICT-Supported Organisations
2003-15 Mathijs de Weerdt (TUD): Plan Merging in Multi-Agent Systems
2003-16 Menzo Windhouwer (CWI): Feature Grammar Systems — Incremental Maintenance of Indexes to Digital Media Warehouses
2003-17 David Jansen (UT): Extensions of Statecharts with Probability, Time, and Stochastic Timing
2003-18 Levente Kocsis (UM): Learning Search Decisions
2004-01 Virginia Dignum (UU): A Model for Organizational Interaction: Based on Agents, Founded in Logic
2004-02 Lai Xu (UvT): Monitoring Multi-party Contracts for E-business
2004-03 Perry Groot (VU): A Theoretical and Empirical Analysis of Approximation in Symbolic Problem Solving
2004-04 Chris van Aart (UVA): Organizational Principles for Multi-Agent Architectures
2004-05 Viara Popova (EUR): Knowledge discovery and monotonicity
2004-06 Bart-Jan Hommes (TUD): The Evaluation of Business Process Modeling Techniques
2004-07 Elise Boltjes (UM): Voorbeeldig onderwijs; voorbeeldgestuurd onderwijs, een opstap naar abstract denken, vooral voor meisjes
2004-08 Joop Verbeek(UM): Politie en de Nieuwe Internationale Informatiemarkt, Grensregionale politiële gegevensuitwisseling en digitale expertise
2004-09 Martin Caminada (VU): For the Sake of the Argument; explorations into argument-based reasoning
2004-10 Suzanne Kabel (UVA): Knowledge-rich indexing of learning-objects
2004-11 Michel Klein (VU): Change Management for Distributed Ontologies
2004-12 The Duy Bui (UT): Creating emotions and facial expressions for embodied agents
2004-13 Wojciech Jamroga (UT): Using Multiple Models of Reality: On Agents who Know how to Play
2004-14 Paul Harrenstein (UU): Logic in Conflict. Logical Explorations in Strategic Equilibrium

2004-15 Arno Knobbe (UU): Multi-Relational Data Mining
2004-16 Federico Divina (VU): Hybrid Genetic Relational Search for Inductive Learning
2004-17 Mark Winands (UM): Informed Search in Complex Games
2004-18 Vania Bessa Machado (UvA): Supporting the Construction of Qualitative Knowledge Models
2004-19 Thijs Westerveld (UT): Using generative probabilistic models for multimedia retrieval
2004-20 Madelon Evers (Nyenrode): Learning from Design: facilitating multidisciplinary design teams
2005-01 Floor Verdenius (UVA): Methodological Aspects of Designing Induction-Based Applications
2005-02 Erik van der Werf (UM)): AI techniques for the game of Go
2005-03 Franc Grootjen (RUN): A Pragmatic Approach to the Conceptualisation of Language
2005-04 Nirvana Meratnia (UT): Towards Database Support for Moving Object data
2005-05 Gabriel Infante-Lopez (UVA): Two-Level Probabilistic Grammars for Natural Language Parsing
2005-06 Pieter Spronck (UM): Adaptive Game AI
2005-07 Flavius Frasincar (TUE): Hypermedia Presentation Generation for Semantic Web Information Systems
2005-08 Richard Vdovjak (TUE): A Model-driven Approach for Building Distributed Ontology-based Web Applications
2005-09 Jeen Broekstra (VU): Storage, Querying and Inferencing for Semantic Web Languages
2005-10 Anders Bouwer (UVA): Explaining Behaviour: Using Qualitative Simulation in Interactive Learning Environments
2005-11 Elth Ogston (VU): Agent Based Matchmaking and Clustering — A Decentralized Approach to Search
2005-12 Csaba Boer (EUR): Distributed Simulation in Industry
2005-13 Fred Hamburg (UL): Een Computermodel voor het Ondersteunen van Euthanasiebeslissingen
2005-14 Borys Omelayenko (VU): Web-Service configuration on the Semantic Web; Exploring how semantics meets pragmatics
2005-15 Tibor Bosse (VU): Analysis of the Dynamics of Cognitive Processes
2005-16 Joris Graaumans (UU): Usability of XML Query Languages
2005-17 Boris Shishkov (TUD): Software Specification Based on Re-usable Business Components
2005-18 Danielle Sent (UU): Test-selection strategies for probabilistic networks
2005-19 Michel van Dartel (UM): Situated Representation
2005-20 Cristina Coteanu (UL): Cyber Consumer Law, State of the Art and Perspectives
2005-21 Wijnand Derks (UT): Improving Concurrency and Recovery in Database Systems by Exploiting Application Semantics
2006-01 Samuil Angelov (TUE): Foundations of B2B Electronic Contracting
2006-02 Cristina Chisalita (VU): Contextual issues in the design and use of information technology in organizations
2006-03 Noor Christoph (UVA): The role of metacognitive skills in learning to solve problems
2006-04 Marta Sabou (VU): Building Web Service Ontologies
2006-05 Cees Pierik (UU): Validation Techniques for Object-Oriented Proof Outlines
2006-06 Ziv Baida (VU): Software-aided Service Bundling - Intelligent Methods & Tools for Graphical Service Modeling
2006-07 Marko Smiljanic (UT): XML schema matching — balancing efficiency and effectiveness by means of clustering
2006-08 Eelco Herder (UT): Forward, Back and Home Again — Analyzing User Behavior on the Web
2006-09 Mohamed Wahdan (UM): Automatic Formulation of the Auditor's Opinion
2006-10 Ronny Siebes (VU): Semantic Routing in Peer-to-Peer Systems
2006-11 Joeri van Ruth (UT): Flattening Queries over Nested Data Types
2006-12 Bert Bongers (VU): Interactivation — Towards an e-cology of people, our technological environment, and the arts
2006-13 Henk-Jan Lebbink (UU): Dialogue and Decision Games for Information Exchanging Agents
2006-14 Johan Hoorn (VU): Software Requirements: Update, Upgrade, Redesign — towards a Theory of Requirements Change
2006-15 Rainer Malik (UU): CONAN: Text Mining in the Biomedical Domain
2006-16 Carsten Riggelsen (UU): Approximation Methods for Efficient Learning of Bayesian Networks
2006-17 Stacey Nagata (UU): User Assistance for Multitasking with Interruptions on a Mobile Device
2006-18 Valentin Zhizhkun (UVA): Graph transformation for Natural Language Processing
2006-19 Birna van Riemsdijk (UU): Cognitive Agent Programming: A Semantic Approach
2006-20 Marina Velikova (UvT): Monotone models for prediction in data mining
2006-21 Bas van Gils (RUN): Aptness on the Web
2006-22 Paul de Vrieze (RUN): Fundaments of Adaptive Personalisation
2006-23 Ion Juvina (UU): Development of Cognitive Model for Navigating on the Web
2006-24 Laura Hollink (VU): Semantic Annotation for Retrieval of Visual Resources
2006-25 Madalina Drugan (UU): Conditional log-likelihood MDL and Evolutionary MCMC
2006-26 Vojkan Mihajlovic (UT): Score Region Algebra: A Flexible Framework for Structured Information Retrieval
2006-27 Stefano Bocconi (CWI): Vox Populi: generating video documentaries from semantically annotated media repositories
2006-28 Borkur Sigurbjornsson (UVA): Focused Information Access using XML Element Retrieval
2007-01 Kees Leune (UvT): Access Control and Service-Oriented Architectures
2007-02 Wouter Teepe (RUG): Reconciling Information Exchange and Confidentiality: A Formal Approach
2007-03 Peter Mika (VU): Social Networks and the Semantic Web
2007-04 Jurriaan van Diggelen (UU): Achieving Semantic Interoperability in Multi-agent Systems: a dialogue-based approach
2007-05 Bart Schermer (UL): Software Agents, Surveillance, and the Right to Privacy: a Legislative Framework for Agent-enabled Surveillance
2007-06 Gilad Mishne (UVA): Applied Text Analytics for Blogs
2007-07 Natasa Jovanovic' (UT): To Whom It May Concern — Addressee Identification in Face-to-Face Meetings
2007-08 Mark Hoogendoorn (VU): Modeling of Change in Multi-Agent Organizations
2007-09 David Mobach (VU): Agent-Based Mediated Service Negotiation
2007-10 Huib Aldewereld (UU): Autonomy vs. Conformity: an Institutional Perspective on Norms and Protocols
2007-11 Natalia Stash (TUE): Incorporating Cognitive/Learning Styles in a General-Purpose Adaptive Hypermedia System
2007-12 Marcel van Gerven (RUN): Bayesian Networks for Clinical Decision Support: A Rational Approach to Dynamic Decision-Making under Uncertainty
2007-13 Rutger Rienks (UT): Meetings in Smart Environments; Implications of Progressing Technology
2007-14 Niek Bergboer (UM): Context-Based Image Analysis
2007-15 Joyca Lacroix (UM): NIM: a Situated Computational Memory Model
2007-16 Davide Grossi (UU): Designing Invisible Handcuffs. Formal investigations in Institutions and Organizations for Multi-agent Systems
2007-17 Theodore Charitos (UU): Reasoning with Dynamic Networks in Practice
2007-18 Bart Orriens (UvT): On the development an management of adaptive business collaborations
2007-19 David Levy (UM): Intimate relationships with artificial partners
2007-20 Slinger Jansen (UU): Customer Configuration Updating in a Software Supply Network
2007-21 Karianne Vermaas (UU): Fast diffusion and broadening use: A research on residential adoption and usage of broadband internet in the Netherlands between 2001 and 2005
2007-22 Zlatko Zlatev (UT): Goal-oriented design of value and process models from patterns
2007-23 Peter Barna (TUE): Specification of Application Logic in Web Information Systems
2007-24 Georgina Ramírez Camps (CWI): Structural Features in XML Retrieval
2007-25 Joost Schalken (VU): Empirical Investigations in Software Process Improvement
2008-01 Katalin Boer-SorbÃin (EUR): Agent-Based Simulation of Financial Markets: A modular,continuous-time approach
2008-02 Alexei Sharpanskykh (VU): On Computer-Aided Methods for Modeling and Analysis of Organizations
2008-03 Vera Hollink (UVA): Optimizing hierarchical menus: a usage-based approach
2008-04 Ander de Keijzer (UT): Management of Uncertain Data — towards unattended integration
2008-05 Bela Mutschler (UT): Modeling and simulating causal dependencies on process-aware information systems from a cost perspective
2008-06 Arjen Hommersom (RUN): On the Application of Formal Methods to Clinical Guidelines, an Artificial Intelligence Perspective
2008-07 Peter van Rosmalen (OU): Supporting the tutor in the design and support of adaptive e-learning
2008-08 Janneke Bolt (UU): Bayesian Networks: Aspects of Approximate

Inference

2008-09 Christof van Nimwegen (UU): The paradox of the guided user: assistance can be counter-effective

2008-10 Wauter Bosma (UT): Discourse oriented summarization

2008-11 Vera Kartseva (VU): Designing Controls for Network Organizations: A Value-Based Approach

2008-12 Jozsef Farkas (RUN): A Semiotically Oriented Cognitive Model of Knowledge Representation

2008-13 Caterina Carraciolo (UVA): Topic Driven Access to Scientific Handbooks

2008-14 Arthur van Bunningen (UT): Context-Aware Querying; Better Answers with Less Effort

2008-15 Martijn van Otterlo (UT): The Logic of Adaptive Behavior: Knowledge Representation and Algorithms for the Markov Decision Process Framework in First-Order Domains.

2008-16 Henriette van Vugt (VU): Embodied agents from a user's perspective

2008-17 Martin Op 't Land (TUD): Applying Architecture and Ontology to the Splitting and Allying of Enterprises

2008-18 Guido de Croon (UM): Adaptive Active Vision

2008-19 Henning Rode (UT): From Document to Entity Retrieval: Improving Precision and Performance of Focused Text Search

2008-20 Rex Arendsen (UVA): Geen bericht, goed bericht. Een onderzoek naar de effecten van de introductie van elektronisch berichtenverkeer met de overheid op de administratieve lasten van bedrijven

2008-21 Krisztian Balog (UVA): People Search in the Enterprise

2008-22 Henk Koning (UU): Communication of IT-Architecture

2008-23 Stefan Visscher (UU): Bayesian network models for the management of ventilator-associated pneumonia

2008-24 Zharko Aleksovski (VU): Using background knowledge in ontology matching

2008-25 Geert Jonker (UU): Efficient and Equitable Exchange in Air Traffic Management Plan Repair using Spender-signed Currency

2008-26 Marijn Huijbregts (UT): Segmentation, Diarization and Speech Transcription: Surprise Data Unraveled

2008-27 Hubert Vogten (OU): Design and Implementation Strategies for IMS Learning Design

2008-28 Ildiko Flesch (RUN): On the Use of Independence Relations in Bayesian Networks

2008-29 Dennis Reidsma (UT): Annotations and Subjective Machines - Of Annotators, Embodied Agents, Users, and Other Humans

2008-30 Wouter van Atteveldt (VU): Semantic Network Analysis: Techniques for Extracting, Representing and Querying Media Content

2008-31 Loes Braun (UM): Pro-Active Medical Information Retrieval

2008-32 Trung H. Bui (UT): Toward Affective Dialogue Management using Partially Observable Markov Decision Processes

2008-33 Frank Terpstra (UVA): Scientific Workflow Design; theoretical and practical issues

2008-34 Jeroen de Knijf (UU): Studies in Frequent Tree Mining

2008-35 Ben Torben Nielsen (UvT): Dendritic morphologies: function shapes structure

2009-01 Rasa Jurgelenaite (RUN): Symmetric Causal Independence Models

2009-02 Willem Robert van Hage (VU): Evaluating Ontology-Alignment Techniques

2009-03 Hans Stol (UvT): A Framework for Evidence-based Policy Making Using IT

2009-04 Josephine Nabukenya (RUN): Improving the Quality of Organisational Policy Making using Collaboration Engineering

2009-05 Sietse Overbeek (RUN): Bridging Supply and Demand for Knowledge Intensive Tasks — Based on Knowledge, Cognition, and Quality

2009-06 Muhammad Subianto (UU): Understanding Classification

2009-07 Ronald Poppe (UT): Discriminative Vision-Based Recovery and Recognition of Human Motion

2009-08 Volker Nannen (VU): Evolutionary Agent-Based Policy Analysis in Dynamic Environments

2009-09 Benjamin Kanagwa (RUN): Design, Discovery and Construction of Service-oriented Systems

2009-10 Jan Wielemaker (UVA): Logic programming for knowledge-intensive interactive applications

2009-11 Alexander Boer (UVA): Legal Theory, Sources of Law & the Semantic Web

2009-12 Peter Massuthe (TUE, Humboldt-Universitaet zu Berlin): perating Guidelines for Services

2009-13 Steven de Jong (UM): Fairness in Multi-Agent Systems

2009-14 Maksym Korotkiy (VU): From ontology-enabled services to service-enabled ontologies (making ontologies work in e-science with ONTO-SOA)

2009-15 Rinke Hoekstra (UVA): Ontology Representation — Design Patterns and Ontologies that Make Sense

2009-16 Fritz Reul (UvT): New Architectures in Computer Chess

2009-17 Laurens van der Maaten (UvT): Feature Extraction from Visual Data

2009-18 Fabian Groffen (CWI): Armada, An Evolving Database System

2009-19 Valentin Robu (CWI): Modeling Preferences, Strategic Reasoning and Collaboration in Agent-Mediated Electronic Markets

2009-20 Bob van der Vecht (UU): Adjustable Autonomy: Controling Influences on Decision Making

2009-21 Stijn Vanderlooy (UM): Ranking and Reliable Classification

2009-22 Pavel Serdyukov (UT): Search For Expertise: Going beyond direct evidence

2009-23 Peter Hofgesang (VU): Modelling Web Usage in a Changing Environment

2009-24 Annerieke Heuvelink (VUA): Cognitive Models for Training Simulations

2009-25 Alex van Ballegooij (CWI): "RAM: Array Database Management through Relational Mapping"

2009-26 Fernando Koch (UU): An Agent-Based Model for the Development of Intelligent Mobile Services

2009-27 Christian Glahn (OU): Contextual Support of social Engagement and Reflection on the Web

2009-28 Sander Evers (UT): Sensor Data Management with Probabilistic Models

2009-29 Stanislav Pokraev (UT): Model-Driven Semantic Integration of Service-Oriented Applications

2009-30 Marcin Zukowski (CWI): Balancing vectorized query execution with bandwidth-optimized storage

2009-31 Sofiya Katrenko (UVA): A Closer Look at Learning Relations from Text

2009-32 Rik Farenhorst (VU) and Remco de Boer (VU): Architectural Knowledge Management: Supporting Architects and Auditors

2009-33 Khiet Truong (UT): How Does Real Affect Affect Affect Recognition In Speech?

2009-34 Inge van de Weerd (UU): Advancing in Software Product Management: An Incremental Method Engineering Approach

2009-35 Wouter Koelewijn (UL): Privacy en Politiegegevens; Over geautomatiseerde normatieve informatie-uitwisseling

2009-36 Marco Kalz (OUN): Placement Support for Learners in Learning Networks

2009-37 Hendrik Drachsler (OUN): Navigation Support for Learners in Informal Learning Networks

2009-38 Riina Vuorikari (OU): Tags and self-organisation: a metadata ecology for learning resources in a multilingual context

2009-39 Christian Stahl (TUE, Humboldt-Universitaet zu Berlin): Service Substitution — A Behavioral Approach Based on Petri Nets

2009-40 Stephan Raaijmakers (UvT): Multinomial Language Learning: Investigations into the Geometry of Language

2009-41 Igor Berezhnyy (UvT): Digital Analysis of Paintings

2009-42 Toine Bogers: Recommender Systems for Social Bookmarking

2009-43 Virginia Nunes Leal Franqueira (UT): Finding Multi-step Attacks in Computer Networks using Heuristic Search and Mobile Ambients

2009-44 Roberto Santana Tapia (UT): Assessing Business-IT Alignment in Networked Organizations

2009-45 Jilles Vreeken (UU): Making Pattern Mining Useful

2009-46 Loredana Afanasiev (UvA): Querying XML: Benchmarks and Recursion

2010-01 Matthijs van Leeuwen (UU): Patterns that Matter

2010-02 Ingo Wassink (UT): Work flows in Life Science

2010-03 Joost Geurts (CWI): A Document Engineering Model and Processing Framework for Multimedia documents

2010-04 Olga Kulyk (UT): Do You Know What I Know? Situational Awareness of Co-located Teams in Multidisplay Environments

2010-05 Claudia Hauff (UT): Predicting the Effectiveness of Queries and Retrieval Systems

2010-06 Sander Bakkes (UvT): Rapid Adaptation of Video Game AI

2010-07 Wim Fikkert (UT): Gesture interaction at a Distance

2010-08 Krzysztof Siewicz (UL): Towards an Improved Regulatory Framework of Free Software. Protecting user freedoms in a world of software communities and eGovernments

2010-09 Hugo Kielman (UL): A Politiele gegevensverwerking en Privacy, Naar een effectieve waarborging

2010-10 Rebecca Ong (UL): Mobile Communication and Protection of Children

2010-11 Adriaan Ter Mors (TUD): The world according to MARP: Multi-Agent Route Planning

2010-12 Susan van den Braak (UU): Sensemaking software for crime analysis

2010-13 Gianluigi Folino (RUN): High Performance Data Mining using Bioinspired techniques

2010-14 Sander van Splunter (VU): Automated Web Service Reconfiguration

2010-15 Lianne Bodenstaff (UT): Managing Dependency Relations in Inter-Organizational Models

2010-16 Sicco Verwer (TUD): Efficient Identification of Timed Automata, theory and practice

2010-17 Spyros Kotoulas (VU): Scalable Discovery of Networked Resources: Algorithms, Infrastructure, Applications

2010-18 Charlotte Gerritsen (VU): Caught in the Act: Investigating Crime by Agent-Based Simulation

2010-19 Henriette Cramer (UvA): People's Responses to Autonomous and Adaptive Systems

2010-20 Ivo Swartjes (UT): Whose Story Is It Anyway? How Improv Informs Agency and Authorship of Emergent Narrative

2010-21 Harold van Heerde (UT): Privacy-aware data management by means of data degradation

2010-22 Michiel Hildebrand (CWI): End-user Support for Access to Heterogeneous Linked Data

2010-23 Bas Steunebrink (UU): The Logical Structure of Emotions

2010-24 Dmytro Tykhonov: Designing Generic and Efficient Negotiation Strategies

2010-25 Zulfiqar Ali Memon (VU): Modelling Human-Awareness for Ambient Agents: A Human Mindreading Perspective

2010-26 Ying Zhang (CWI): XRPC: Efficient Distributed Query Processing on Heterogeneous XQuery Engines

2010-27 Marten Voulon (UL): Automatisch contracteren

2010-28 Arne Koopman (UU): Characteristic Relational Patterns

2010-29 Stratos Idreos(CWI): Database Cracking: Towards Auto-tuning Database Kernels

2010-30 Marieke van Erp (UvT): Accessing Natural History — Discoveries in data cleaning, structuring, and retrieval

2010-31 Victor de Boer (UVA): Ontology Enrichment from Heterogeneous Sources on the Web

2010-32 Marcel Hiel (UvT): An Adaptive Service Oriented Architecture: Automatically solving Interoperability Problems

2010-33 Robin Aly (UT): Modeling Representation Uncertainty in Concept-Based Multimedia Retrieval

2010-34 Teduh Dirgahayu (UT): Interaction Design in Service Compositions

2010-35 Dolf Trieschnigg (UT): Proof of Concept: Concept-based Biomedical Information Retrieval

2010-36 Jose Janssen (OU): Paving the Way for Lifelong Learning; Facilitating competence development through a learning path specification

2010-37 Niels Lohmann (TUE): Correctness of services and their composition

2010-38 Dirk Fahland (TUE): From Scenarios to components

2010-39 Ghazanfar Farooq Siddiqui (VU): Integrative modeling of emotions in virtual agents

2010-40 Mark van Assem (VU): Converting and Integrating Vocabularies for the Semantic Web

2010-41 Guillaume Chaslot (UM): Monte-Carlo Tree Search

2010-42 Sybren de Kinderen (VU): Needs-driven service bundling in a multi-supplier setting — the computational e3-service approach

2010-43 Peter van Kranenburg (UU): A Computational Approach to Content-Based Retrieval of Folk Song Melodies