

Master Thesis

**Towards Automated Retrieval of Audio Recordings
based on
Musical Themes**

submitted by

Lukas Lamprecht

submitted

November 5, 2014

Supervisor / Advisor

Prof. Dr. Meinard Müller

Dipl.-Ing. Stefan Balke

Thomas Prätzlich, M.Sc.

Reviewers

Prof. Dr. Meinard Müller

Erklärung

Ich versichere, dass ich die Arbeit ohne fremde Hilfe und ohne Benutzung anderer als der angegebenen Quellen angefertigt habe und dass die Arbeit in gleicher oder ähnlicher Form noch keiner anderen Prüfungsbehörde vorgelegen hat und von dieser als Teil einer Prüfungsleistung angenommen wurde. Alle Ausführungen, die wörtlich oder sinngemäß übernommen wurden sind als solche gekennzeichnet.

Erlangen, 5. November 2014

Lukas Lamprecht

Acknowledgements

I would like to express my gratitude to a crowd of people who supported me in the time of writing my thesis.

First of all, I want to thank Meinard Müller for supervising my thesis. The discussions with him and his Ph.D. students gave me inspirations for my master thesis and guided me into the right direction.

I also want to thank Stefan Balke, for advising me and proof reading my thesis. I want to thank Thomas Prätzlich for providing and explaining source code to me for this thesis. He always gave me good advice when I had questions regarding technical issues. Furthermore, I want to thank Jonathan Driedger and Nanzhu Jiang for proof reading my thesis and giving me feedback and inspirations for my master thesis. I want to thank one of my fellow students, Sanu Pulimootil. He was always available for questions and discussions. Furthermore, he provided me image data for the database and my thesis.

I also want to express my sincere gratitude to Lena Krauß and Vlora Arifi-Müller for helping me with the creation of the database I used for this master thesis. I want to thank Stefan Balke for helping me to collect the audio and MIDI data of the database. It was very time-consuming to collect organize and check data for such a huge database.

Finally, I want to thank my family and my friends, who supported me during the writing of my thesis.

Abstract

The research area of music information retrieval deals with the development of automated methods for analyzing, organizing and searching musical content in a robust and intelligent manner. In particular, music retrieval based on the query-by-example paradigm has received a lot of research attention: Given a music representation or a fragment of it (used as query or example), the task is to automatically retrieve documents from a music database containing parts or aspects that are similar to it.

In this thesis, we consider a cross-modal retrieval scenario, where the queries are symbolic encodings of musical themes and the database documents are audio recordings of musical performances. The task is to identify all audio recordings that contain the musical theme specified by a query. Within this scenario, we consider several challenges. First, there may be significant variations in the global and local tempo between query and the audio recordings. Second, the musical theme is monophonic, whereas the corresponding section in an audio recording may be polyphonic. Third, the audio recording may be transposed or detuned compared to the query.

The main contributions of this thesis are as follows. First, using methods from signal processing and information retrieval, we adapt an existing retrieval pipeline to cope with the various challenges. Second, we develop a large database consisting of several thousand musical themes and audio recordings of Western classical music. Third, we conduct extensive experiments by systematically adjusting parameters for the feature computation and retrieval step. As a result, we do not only improve the retrieval results, but also gain a deeper understanding of the underlying musical data.

Contents

Erklärung	i
Acknowledgements	iii
Abstract	v
1 Introduction	3
1.1 Retrieval Scenario	4
1.2 Main Contributions	6
1.3 Thesis Organization	6
2 Audio Matching	9
2.1 Chroma Features	9
2.1.1 CENS Features	10
2.1.2 36-dimensional CENS Features	11
2.2 Subsequence-Dynamic Time Warping	13
2.3 Cyclic shift	15
2.4 Evaluation Measures	16
2.4.1 Top N Match	16
2.4.2 Mean Rank	17
3 Matching Symbolic Themes to Audio	19
3.1 Matching Procedure	20
3.2 Datasets	24
4 Experiments	31
4.1 Parameter selection using the subset BM-Small	31
4.1.1 Chroma resolution and feature resolution	32
4.1.2 Step size condition	33
4.1.3 Fixed query length	33
4.1.4 Weight factors	34
4.1.5 Key shifts	34
4.2 Evaluation of the subset BM-Medium	35
4.3 Analysis of specific queries of the subset BM-Mini	37
4.3.1 Specificity	37
4.3.2 Feature resolution	38
4.3.3 Tuning	39
4.3.4 Global and local tempo variations	39

CONTENTS

4.3.5	Degree of Polyphony	41
<hr/>		
5	Conclusions	45
A	Source Code	49
B	BM-Subsets	53
B.1	BM-Mini Matching functions	53
B.2	BM-Mini ranking results	63
B.3	Ensemble and Instrumentation IDs	66
B.4	Audio Recordings for BM-Mini	67
B.5	EDM-MIDIs of BM-Mini	67
B.6	Audio Recordings for BM-Small	68
B.7	EDM-MIDIs of BM-Small	72
B.8	BM-Medium	75
B.9	Errors	75
	Bibliography	81

Chapter 1

Introduction

Today, we live in a world in which large music collections are omnipresent. In the last years, the demand for online music collections and services strongly increased (see Figure 1.1). Services like Spotify¹, SHAZAM² and the International Music Score Library Project³ (IMSLP) became more and more popular. The services and the underlying types and formats of data which include text, symbolic data, images and audio provided by these three platforms are quite different. The IMSLP for example contains 81,547 musical works and 281,541 scores by 11,250 composers, and 30,440 audio recordings by 282 performers [15]⁴. But dealing with such large collections of data also bares problems. For that reason, research groups from all over the world have considered issues in the field of *music information retrieval* (MIR), a research area that opens up new possibilities regarding the organization of large databases of music in an automated or semi-automated fashion. The organization of large music databases can be simplified by *automated retrieval* techniques. By analyzing, correlating and comparing the multimodal data, the aim is to identify the data and establish semantically meaningful relationships between different types of data, such that not only the database can be used in an intuitive fashion, but also searches lead to good results, even for non-sophisticated users.

To find data within a database, the user has to specify his information needs by means of a so called *query*. The retrieval system should then retrieve all documents from a data collection that are somehow related to the query. The retrieved *documents* are then displayed on a ranked list in a sorted fashion. For many *Information Retrieval* tasks, the query and the documents are available in the same format. An example for this scenario is to specify a keyword as query in order to search through an e-book. Using such text based retrieval systems for music collections require the audio material to be enriched with suitable metadata. In general however, the query and the documents do not necessarily have to be available in the same format. For example, someone, who only remembers a short melody, wants to find the underlying musical work without the knowledge of its composer and title. The query is then specified by singing or humming this melody into a microphone (*query-by-humming*). To cope with this task, a *content based retrieval* system is needed, that only makes use of the raw music data, rather than relying on manually generated metadata. The term *content* in this context refers to any kind of information

¹www.spotify.com

²www.shazam.com

³www.imslp.org

⁴Retrieved on 23.07.2014

1. INTRODUCTION

that can be directly derived from the music material to be queried, compared and retrieved. In particular, content-based retrieval strategies that follow the *query-by-example* paradigm are of great importance: Given a music representation or a fragment of it (used as query or example), the task is to automatically retrieve documents from a music collection containing parts or aspects that are similar to it.

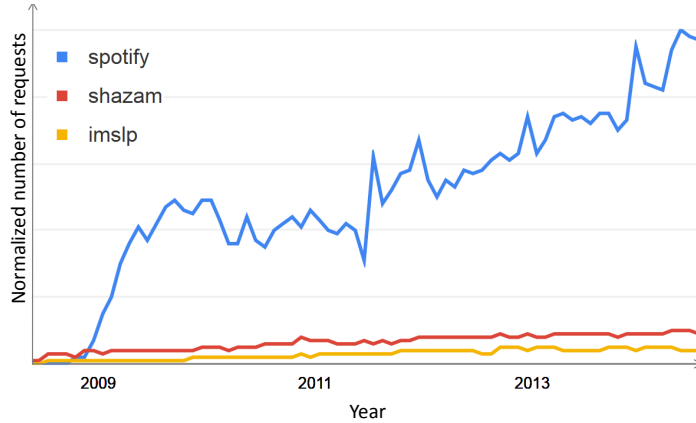


Figure 1.1: Demand on music platforms and services measured by the number of search requests on Google. Retrieved from www.google.de/trends.

1.1 Retrieval Scenario

The task of this thesis is the adaptation of a query-by-example task on datasets based on Western classical music. On the one hand, we consider musical themes given in some symbolic format, and on the other hand audio recordings of entire pieces of music. Then, given a musical theme used as query, the task is to identify the audio recording of the musical work containing this theme. The retrieved documents are then displayed on a ranked list.

In this thesis, we use three different music representations. *Sheet music* available as digital images visually represents a score or a piece of music. Furthermore, *symbolic* music data available in MIDI⁵ format is used. Finally, we use *audio recordings* in WAV format that encode acoustic sound waves [10, pp.17-27]. The symbolic music data that is used in this thesis is derived from the book “A DICTIONARY OF MUSICAL THEMES” by Harold Barlow and Sam Morgenstern [1] (which we will refer to as **BM** in this thesis). This dictionary contains roughly 10,000 musical themes of mainly instrumental Western classical music which we will refer to as “BM-themes”. The **BM** was designed as a reference book for trained musicians and professional performers. It was published in the year 1948 and is an early approach for query-by-example tasks. By playing a theme in the key of C Major or Minor (for themes in major or minor respectively), the note sequence can be looked up in the so called notation index in the **BM**. With the notation index, we can find the theme, the theme name and the corresponding composer in the **BM** [1, pp.viii-ix]. Most of the **BM**-themes are also available as MIDI-files provided by the Electronic

⁵MIDI is a standard that is used for electronic instruments and communication between electronic instruments and computers [9].

Dictionary of Musical Themes (EDM)⁶ [18], which are used as queries in this thesis. We refer to these themes as “EDM-themes”. Finally, a collection of audio recordings from musical works that are also listed in the BM is used as dataset for this thesis. In particular we create three datasets which we refer to as subsets. Every subset contains several audio recordings and the corresponding EDM-themes. The content of these subsets will be described in more detail in Chapter 3.2.

Our retrieval scenario offers several challenges, which are described hereafter, where the corresponding technical terms are specified in bold print.

The underlying data is **cross-modal**, as the database contains EDM-themes available in a symbolic music format as well as audio recordings in WAV file format. Furthermore, the EDM-themes are monophonic consisting of a melody without accompanying harmony. In contrast, the audio recordings are typically homophonic (music with one dominant melodic voice accompanied by chords) or polyphonic (music with two or more simultaneous lines of independent melody) (**degree of polyphony**). The EDM-themes are of very short duration and are looked up in a large database, so they may be of low **specificity** (having a low discriminative power). The audio recording may be **transposed** and thus be available in another key than the EDM-theme. The audio recordings may be **detuned**. The **tempo** of the audio recordings may differ widely from the tempo specified in the corresponding EDM-theme. There may be high local **tempo variations** in the audio recording which are not present in the corresponding EDM-theme. The **dynamics** of the audio recordings may be quite high. Finally, the **quality** of some audio recordings may be quite low, especially for old recordings which are often noisy. However the task of improving audio quality and dynamics reduction is not treated in this thesis.

In this thesis we create a so called pipeline, which performs the processing of the entire matching procedure. Furthermore we systematically analyze the underlying data, which makes it possible to improve the retrieval results by adjusting parameters of the pipeline.

In the following, we will have a look at the specific steps of the pipeline needed to realize the query-by-example task. An overview of the pipeline is given in Figure 1.2. The EDM-theme corresponding to Beethoven’s 5th Symphony, 1st Movement, 1st Theme is used as query. The database consists of several audio recordings. As these two different formats cannot be compared directly, a mid-level representation is needed. Therefore, the query as well as the database are converted into chroma-based features. These features are aligned by using a subsequence-dynamic time warping (SDTW) strategy to compensate tempo differences between the query and the database. The SDTW leads to a matching function. In case the matching process was successful, the minimum of this function indicates an excerpt of the database that is semantically related to the query. This matching function is then analyzed and a ranked list is created, which shows the retrieval results for this query. Finally, an evaluation is performed to validate the retrieval quality. To this end, evaluation measures like the Top N Match and the Mean Rank are calculated for many different queries by using ground truth data. Similar processing pipelines have already been used before, e. g. in [4] for the task of sheet music-audio identification. Related music matching tasks have been considered in [3] and [21].

⁶<http://www.multimedialibrary.com/barlow/>

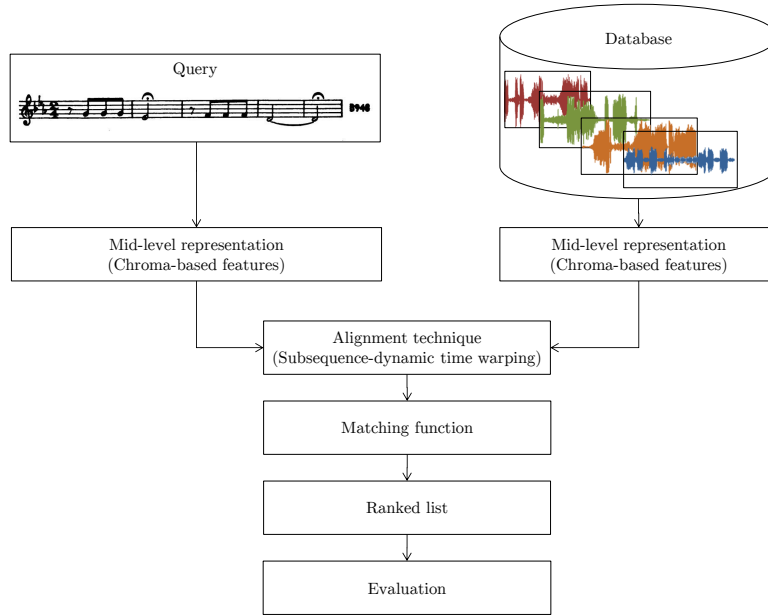


Figure 1.2: Overview of the pipeline, showing the different steps of the matching process.

1.2 Main Contributions

In this thesis we modify an existing pipeline to evaluate its performance depending on several parameter settings. Furthermore several intermediate steps of the pipeline are analyzed for selected data to improve the retrieval results. Therefore, we introduce an enhanced version of chroma features that takes tuning errors into account. Several strategies are implemented and evaluated to compensate great tempo differences between query and audio recording. Moreover, we adjust different parameter settings to further improve the retrieval.

In this thesis we create subsets containing EDM-themes from the website [18] and corresponding audio recordings. We generate ground truth data by excluding corrupted EDM-themes and apply a mapping between the EDM-themes and the corresponding audio recordings. Several exceptions in the data that occur are documented and problems are analyzed and discussed.

1.3 Thesis Organization

In this section, we give an overview of the structure of this thesis and briefly describe the content of each chapter.

Chapter 2 describes the intermediate steps that are essential for matching MIDI files to audio recordings like the calculation of chroma features and alignment techniques. In this Chapter we introduce an enhanced version of chroma features. Furthermore, evaluation measures are introduced.

In Chapter 3 we explain the pipeline for the matching process step by step, using the techniques we introduced in Chapter 2 as a foundation. Additionally, we introduce and describe the subsets.

In Chapter 4 we tune the parameters of our pipeline, whereas we use a database of 100 audio recordings. We then fix one parameter setting to perform and evaluate an experiment on a database of 1113 audio recordings. By analyzing several statistical measures we get insight in the underlying data. Finally, we analyze our pipeline for specific queries with respect to certain properties, like tuning, tempo variations and polyphony.

Chapter 5 concludes the thesis by recapitulating the main achievements. Besides, it provides motivations for future research in this area.

Chapter 2

Audio Matching

In this chapter, we introduce the intermediate steps we perform to match symbolic music data to audio recordings in a database. Symbolic music data cannot be compared to audio recordings directly, as the underlying data types are different. To resolve this issue, each of the files is converted into a mid-level representation. With this mid-level representation we can compare the content of the different data formats on an abstract level. We use chroma features as mid-level representation which are described in Section 2.1 closely following [10, pp.60-64]. Chroma features were published for the first time in [5], where they are called Pitch Class Profiles (PCP). In [6], [17] and [19] they are called Harmonic Pitch Class Profiles (HPCP). In Section 2.1.2 we introduce an enhanced version of chroma features.

To compensate tempo deviations and tempo changes between the mid-level representations of query and database we use subsequence-dynamic time warping (SDTW) as alignment technique for the audio matching which is described in Section 2.2, closely following [10, pp.69-82]. In Section 2.3 we introduce the cyclic shift operator which allows us to compensate errors in tuning and transposition. Finally, we describe evaluation measures in Section 2.4.

2.1 Chroma Features

Human auditory perception is similar for pitches that differ by one or more octaves [10, p. 60]. “A pitch can be separated into two components, which are referred to as *tone height* and *chroma*. The tone height refers to the octave number and the chroma to the respective pitch spelling attribute contained in the set $\{C, C\#, D, \dots, B\}$ ” [10, p.60] as used in Western music notation [11, p.2]. A *chroma feature* is a twelve dimensional vector consisting of one numeric value for each chroma. This value indicates the portion of the underlying music material which can be assigned to the respective chroma. We can compute chroma features both from audio signals, as well as from MIDI files. Given an audio signal, the signal is segmented into equidistant overlapping frames of fixed length. Each of the frames is converted into a chroma feature vector, resulting in a chromagram. A MIDI file can be processed in a similar way. Figure 2.1(e) and 2.1(f) show chromagrams derived from a MIDI file and an audio recording. Both kinds of music representations are derived from Beethoven’s 5th Symphony, 1st Movement, 1st Theme. For that reason both chromagrams are quite similar. We can see that the chromagram of the MIDI file

2. AUDIO MATCHING

contains almost only zeros and ones (ignoring some gray entries which occur due to smoothing effects). The reason is, that a MIDI file describes the onset and offset time, and the key velocity (which is correlated to the loudness) for each pitch. The chromagram of the audio recording shows the same dominant melody containing the pitches $\{G, D\#, F, D\}$. Furthermore, we see gray entries in the chromagram that indicate that other pitches are also active. This is because the audio recording is polyphonic, whereas the MIDI file is monophonic.

Figure 2.1(a) visualizes the content of the previously mentioned MIDI file in a piano roll representation. We now compute pitch features, as shown in Figure 2.1(c). The pitch features are directly derived from the MIDI file. We use the 88 pitches on a piano keyboard as a basis for the calculation of the chroma features, which correspond to the MIDI pitches $p = 21$ to $p = 108$. A chromagram is derived by summing up all values of the same chroma, for each pitch p and time instant $n \in \mathbb{Z}$, from the pitch features $\mathcal{Y}_{\text{LF}} : \mathbb{Z} \times [21 : 108] \rightarrow \mathbb{R}_{\geq 0}$:

$$\mathcal{C}(n, c) := \sum_{\{p \in [21:108] \mid p \bmod 12 = c\}} \mathcal{Y}_{\text{LF}}(n, p) \quad (2.1)$$

for $c \in [0 : 11]$.

We calculate the chromagram for the audio recording in the same way, whereas the pitch features have to be calculated first. As an audio recording encodes acoustic waves in time domain (see Figure 2.1(b)), we apply a multirate filterbank with 88 bandpass-filters and a -3dB bandwidth of one semitone per bandpass-filter. The center frequencies of the 88 frequency bands correspond to the MIDI pitches $p = 21$ to $p = 108$. This leads to a sequence of 88 dimensional pitch features where the entries correspond to the mentioned MIDI pitches [11, p.3], see Figure 2.1(d).

2.1.1 CENS Features

Chroma energy normalized statistics (CENS) are derived from the chroma features we introduced in Section 2.1. First of all, each chroma vector x is replaced by its l^1 -norm $x/\|x\|_1$ to express the relative distribution of the energy of the signal in the 12 chroma bands, where

$$\|x\|_1 := \sum_{i=0}^{11} |x(i)| \quad (2.2)$$

is the l^1 -norm of x . This normalization step makes the features more robust to differences in sound intensity or dynamics. During passages of very low energy, we replace the chroma vector x by the uniform distribution to avoid random energy distributions. Then a logarithmic quantization is applied on each of the twelve chroma values per chroma vector. This quantization step models the logarithmic sensation of the sound intensity of the human ear. Then the features are smoothed over a window of length $\omega \in \mathbb{N}$ and downsampled by a factor of d . Finally the resulting features are normalized with respect to the l^2 -norm

$$\|x\| := \left(\sum_{i=0}^{11} |x(i)|^2 \right)^{1/2} \quad (2.3)$$

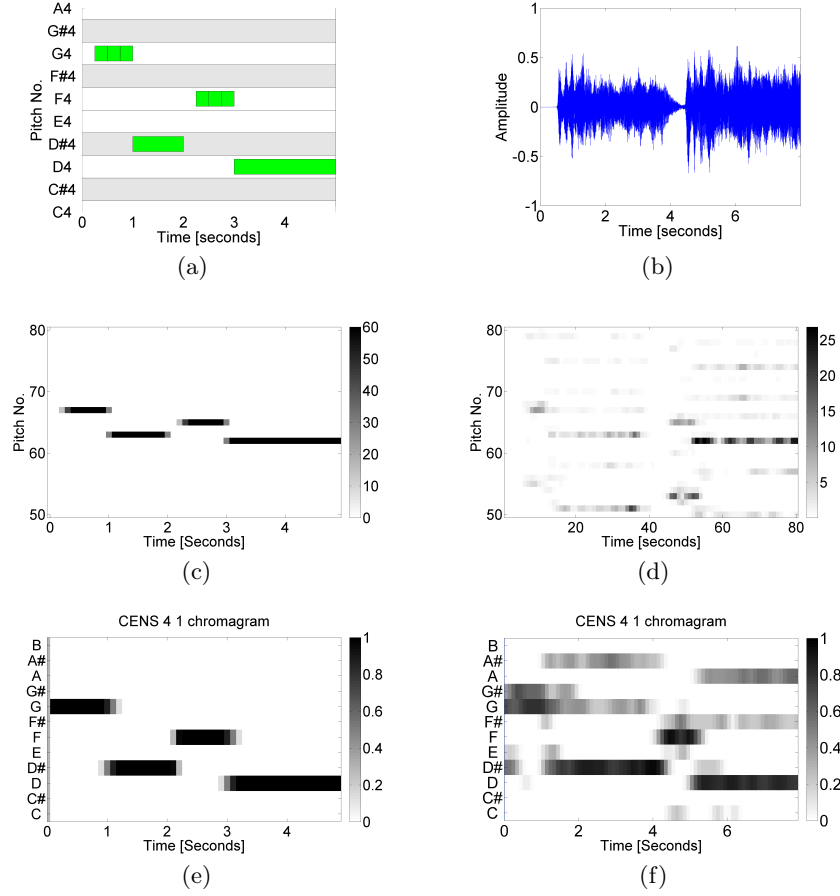


Figure 2.1: Derivation of chroma features from MIDI and audio for Beethoven’s 5th Symphony. (a) MIDI (b) Audio (c) Pitch features derived from MIDI. (d) Pitch features derived from audio. (e) CENS features derived from MIDI. (f) CENS features derived from audio.

and denoted by CENS_d^ω or $\text{CENS } \omega \ d$ where ω is the length of the smoothing window and d is the downsampling factor [10, p.61], [12, pp.2-3].

2.1.2 36-dimensional CENS Features

A recurring problem in our concerned retrieval scenario is that audio recordings may strongly vary in their tuning. When computing chroma features for an audio recording this often leads to chroma features that are smeared between chroma bands. This is the case since chroma features are designed for a specified tuning (A4 of 440 Hz) and cannot adapt to the exact tuning of the audio recording. When computed for an audio recording having a different tuning, the energy of a certain pitch class of this recording spreads across several chroma bands in the chroma features since the range of the bandpass-filters do not match the pitch classes of the audio recording. For an example, see Figure 2.2(a).

To resolve this issue we revert to 36-dimensional chroma features. Here, the idea is to have a finer chroma resolution than one semitone, as it is the case for 12-dimensional chroma features.

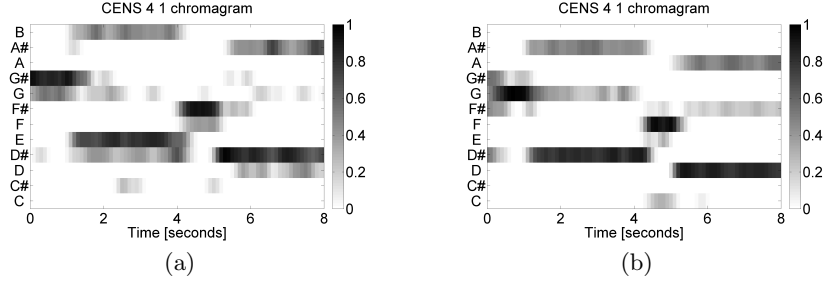


Figure 2.2: (a) Chromagram of an audio recording that is detuned by $+1/3$ semitone. (b) Chromagram of an audio recording that is detuned by $+1/3$ semitone using a filterbank that is also detuned by $+1/3$ semitone.

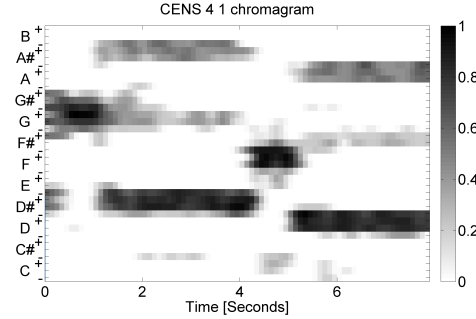


Figure 2.3: 36-dimensional CENS features derived from Beethoven's 5th Symphony, 1st Movement.

Instead of having one band for each of the 12 chroma, we extend this notion to three bands per chroma. These bands reflect the respective chroma for the standard tuning, a tuning of $-1/3$ semitone, and a tuning of $+1/3$ semitone. Still keeping the width of each chroma band to be equivalent to one semitone, this allows us to capture three different tunings simultaneously. To indicate the tuning we add a '-' or '+' sign to the respective chroma labels, see Figure 2.3.

In order to compute 36-dimensional chroma features we first compute the 12-dimensional chroma features. Therefore we first calculate pitch features, using the same filterbank as in Section 2.1. Then we compute two further sequences of 12-dimensional chroma features, whereas we use a filterbank for which every bandpass-filter is shifted by $-1/3$ of a semitone and $+1/3$ of a semitone, respectively. We then interleave these three sequences of 12-dimensional chroma features according to the scheme $\{C-, C, C+, D#-, D#, D#+, \dots, B-, B, B+\}$. In Figure 2.3 we see a 36-dimensional CENS feature sequence of an audio recording of Beethoven's 5th Symphony. In order to compute 36-dimensional chroma features for a MIDI file, we first compute the corresponding 12-dimensional chroma feature sequence. Then we create two sequences, each of the same size containing only zero values. Then we interleave these three feature sequences according to the scheme $\{0, C, 0, 0, D#, 0, \dots, 0, B, 0\}$. This definition yields the property that the scalar product of a 36-dimensional chroma vector derived from an audio recording and of a 36-dimensional chroma vector derived from a MIDI file results in the same value as the scalar product of the respective 12-dimensional chroma vectors.

2.2 Subsequence-Dynamic Time Warping

Subsequence-Dynamic Time Warping (SDTW) is an alignment technique used to find a subsequence X within a longer sequence Y that best fits sequence Y [10, p.79]. Using SDTW, we can match a short query (for example an EDM-theme) with a large database of audio recordings whereas tempo changes and tempo differences can be compensated up to some extent.

In the following, we describe SDTW step by step, whereas we need a couple of definitions:

Given two *feature sequences* $X = (x_1, x_2, \dots, x_N)$ and $Y = (y_1, y_2, \dots, y_M)$ with $M \gg N$, we calculate a *cost matrix* $\mathbf{C}(n, m) = c(x_n, y_m)$ for $n \in [1 : N]$ and $m \in [1 : M]$, see [10, p.79]. Its entries are calculated by computing a so called *cost measure* c . In our experiments, we use the *cosine distance* defined by

$$c(x_n, y_m) := 1 - \frac{\langle x|y \rangle}{\|x\| \cdot \|y\|} \quad (2.4)$$

for two non-zero vectors x, y . Furthermore we define two different step size conditions expressed by the sets

$$\Sigma_1 := (1, 0), (0, 1), (1, 1) \quad (2.5)$$

and

$$\Sigma_2 := (2, 1), (1, 2), (1, 1) \quad (2.6)$$

The step size condition constrains the possible warping paths by fixing the maximum tempo variations between X and Y . Σ_1 allows infinite tempo variations. Σ_2 however allows only tempo variations of factor two. This means that SDTW can only find the subsequence X in Y , if the length of the corresponding sequence in Y is in the bounds $N/2$ and $2N$. The accumulated cost matrix \mathbf{D} is derived from \mathbf{C} in three steps. The first row of \mathbf{D} is initialized by assigning it with the values of the first row of \mathbf{C} :

$$\mathbf{D}(1, m) := \mathbf{C}(1, m) \quad (2.7)$$

for $m \in [1 : M]$. With this initialization step, a subsequence can start anywhere in m , taking into account only the local cost of \mathbf{C} at $m \in [1 : M]$.

Then, the first column of \mathbf{D} is calculated by computing

$$\mathbf{D}(n, 1) = \sum_{k=1}^n \mathbf{C}(k, 1) \quad \text{for } n \in [1 : N]. \quad (2.8)$$

All remaining entries are then calculated by using the formula

$$\mathbf{D}(n, m) = \mathbf{C}(n, m) + \min \begin{cases} \mathbf{D}(n-1, m-1) \\ \mathbf{D}(n-1, m) \\ \mathbf{D}(n, m-1) \end{cases} \quad (2.9)$$

for $n \in [1 : N]$ and $m \in [1 : M]$, when using the set Σ_1 . For calculating the values for the first row and the first column of \mathbf{D} we extend \mathbf{D} by appending one row and one column each given the respective index 0. Then we initialize $\mathbf{D}(n, 0) := \infty$ for $n \in [1 : N]$ and $\mathbf{D}(0, m) := \infty$ for $m \in [1 : M]$.

When we use the set Σ_2 , we use the formula

$$\mathbf{D}(n, m) = \mathbf{C}(n, m) + \min \begin{cases} \mathbf{D}(n-1, m-1) \\ \mathbf{D}(n-2, m-1) \\ \mathbf{D}(n-1, m-2) \end{cases} \quad (2.10)$$

with $n \in [1 : N]$ and $m \in [1 : M]$. For this step size condition we extend \mathbf{D} by appending two rows and two columns each given the respective indices -1 and 0 . Then we initialize $\mathbf{D}(n, -1) := \mathbf{D}(n, 0) := \infty$ for $n \in [1 : N]$ and $\mathbf{D}(-1, m) := \mathbf{D}(0, m) := \infty$ for $m \in [1 : M]$.

We specify the weight factors $w_d, w_h, w_v \in \mathbb{R}$ to favour the diagonal, horizontal, or vertical direction in the alignment [10, p.76]. SDTW has the property that the subsequence X is much shorter than Y . When we calculate the matrix \mathbf{D} , the more steps we take in n -direction, the earlier we reach N and the less steps we need in total. So in general a path that describes a short subsequence is preferred over a longer and probably semantically more meaningful subsequence. We specify different weight factors w_d, w_h, w_v , to compensate this issue.

Then the accumulated cost matrix

$$\mathbf{D}(n, m) = \min \begin{cases} \mathbf{D}(n-1, m-1) + w_d \cdot \mathbf{C}(n, m) \\ \mathbf{D}(n-1, m) + w_v \cdot \mathbf{C}(n, m) \\ \mathbf{D}(n, m-1) + w_h \cdot \mathbf{C}(n, m) \end{cases} \quad (2.11)$$

with $n \in [1 : N]$ and $m \in [1 : M]$. is calculated for the step sizes Σ_1 , using Equation (2.7) and (2.8) as initialization steps. Furthermore we initialize \mathbf{D} in the same way as for Equation (2.9).

Using the step sizes Σ_2 , the accumulated cost matrix

$$\mathbf{D}(n, m) = \min \begin{cases} \mathbf{D}(n-1, m-1) + w_d \cdot \mathbf{C}(n, m) \\ \mathbf{D}(n-2, m-1) + w_v \cdot \mathbf{C}(n, m) \\ \mathbf{D}(n-1, m-2) + w_h \cdot \mathbf{C}(n, m) \end{cases} \quad (2.12)$$

is calculated using the initialization step in Equation (2.7). Here we initialize \mathbf{D} in the same way as for Equation (2.10). [10, pp.70-81].

Figure 2.4(a) shows an accumulated cost matrix, derived from Beethoven's 5th Symphony, whereas we use Beethoven's 5th Symphony, 1st Movement, 1st Theme as query. A match between the query and the database is indicated by a diagonal in the matrix with low cost-values. The query occurs twice in the database. These two passages are highlighted by red frames. The two diagonals in dark gray highlighted by red frames indicate subsequences with low cost. These sequences best match the query.

The uppermost row in \mathbf{D} defines a so called matching function

$$\Delta_{\text{DTW}}(m) := \frac{1}{N} \mathbf{D}(N, m) \quad (2.13)$$

with $\Delta_{\text{DTW}} : [1 : M] \rightarrow \mathbb{R}$ and $m \in [1 : M]$ to identify similar subsequences highlighted in Figure 2.4(a) by a red frame. The matching function corresponding to the accumulated cost matrix in Figure 2.4(a) is shown in Figure 2.4(b). Local minima with low cost values indicate the end position of a matching subsequence. In Figure 2.4(b) we can see two of these local minima. With a backtracking procedure the starting point of the subsequence is obtained. The

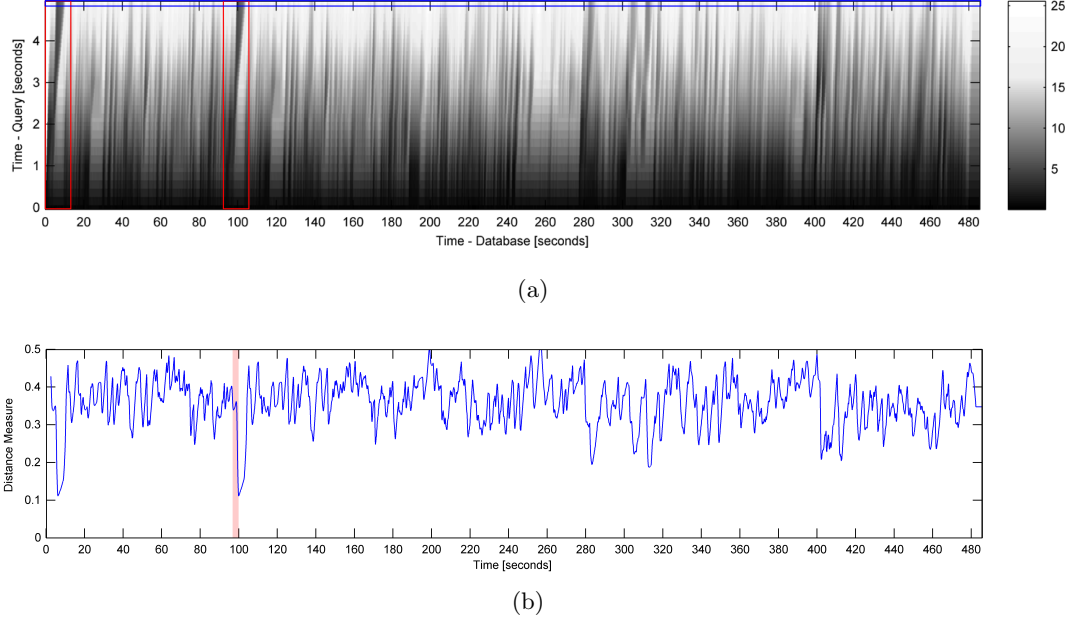


Figure 2.4: (a) Accumulated cost matrix \mathbf{D} . (b) Matching function Δ_{DTW} . The computed subsequence is highlighted in light red color.

subsequence that matches best is highlighted in light red color [10, pp.70-81].

2.3 Cyclic shift

In this section we introduce the cyclic shift operator. With this operator we can take into account transposed versions of the query in the database. Therefore we shift each of the chroma vectors of the query, for example by one semitone according to the formula

$$\sigma(x) := (x(11), x(0), x(1), \dots, x(10))^T. \quad (2.14)$$

with $\sigma : \mathbb{R}^{12} \rightarrow \mathbb{R}^{12}$ for the 12-dimensional chroma features and we musically transpose each of the chroma vectors of the query, for example by 1/3 semitone according to the formula

$$\sigma(x) := (x(35), x(0), x(1), \dots, x(34))^T. \quad (2.15)$$

with $\sigma : \mathbb{R}^{36} \rightarrow \mathbb{R}^{36}$ for the 36-dimensional chroma features before performing the SDTW. Depending on the dimensionality of the chroma features, we only take into account transpositions for the 12-dimensional chroma features, or we take into account both transpositions and global tuning variations using the 36-dimensional chroma features. Figure 2.5 shows matching functions for several transpositions of the query. We can see that the query occurs in the same key as in the database twice, at time instants 30 seconds and 85 seconds. Additionally, the query occurs in a transposed version. It is transposed by +2 semitones in the audio recording occurring at time instant 130 seconds.

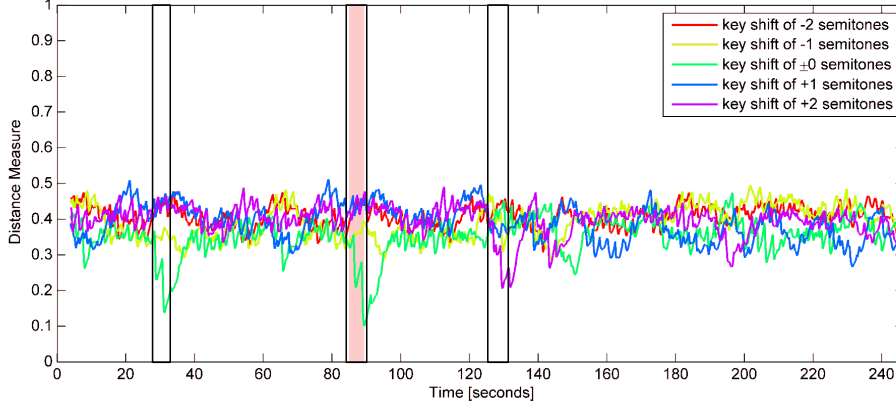


Figure 2.5: Matching functions for Beethoven Op.2 No.1 for 5 transpositions. Sequences that match the query are highlighted by a black frame. The best matching sequence is highlighted in light red color.

2.4 Evaluation Measures

In this section we introduce measures, that we will use for evaluation of the performance of the pipeline in Chapter 4. With these evaluation measures, we can compare the retrieval results for different parameter settings. So the parameters can be tuned to optimize results.

We use the Top N Match and the Mean Rank as evaluation measures, inspired by [17, p.55].

2.4.1 Top N Match

The goal of our retrieval system is, given a MIDI file (query) to identify the semantically corresponding audio recording (document) in a database. Similar to retrieval systems like Google¹, our system provides not only a single answer but a ranked list of documents. Formally, given S queries to a database with K documents, a retrieval system creates S ranked lists with N entries, $N \leq K$. We now want to quantify for how many of the S queries the correct document is contained in the respective ranked list. Let M be the number of queries fulfilling this condition. The Top N Match measure can be defined as:

$$\text{Top N Match} = \frac{M}{S} \cdot 100\%$$

We refer to the Top 1 Match also as Top Match. This evaluation measure is also used in [22, p.188] where it is referred to as “Top-N accuracy”.

¹www.google.com

2.4.2 Mean Rank

The Mean Rank specifies the rank on which the correct document occurs on a ranked list on average. This measure is related to the mean reciprocal rank used in [17]. We use the Mean Rank, because it is a semantically more meaningful measure to the user of a retrieval system. We want to know, how many entries in a ranked list we have to check on average until we find the document we are looking for. The mean reciprocal rank however only returns a number in the range of $[0, 1]$. The disadvantage of the Mean Rank is that the size of the database is not taken into account when we use this measure. For example a Mean Rank of 20 means that on average a user has to search through the top 20 entries of a ranked list to find the match. However, this measure is strongly dependent on the size of the database. For a database containing 100 documents, a Mean Rank of 20 is much better, than for a database containing only 50 documents.

The Mean Rank is calculated as follows:

$$\text{Mean Rank} = \frac{1}{S} \sum_{i=1}^S R_i$$

where R_i is the rank for the i^{th} query and S is the number of queries.

Note that the Mean Rank is a measure that takes into account all K entries of the retrieved ranked lists, whereas the Top N Match only takes into account the respective N entries.

Chapter 3

Matching Symbolic Themes to Audio

After describing the fundamental techniques, the task is to match monophonic themes to polyphonic audio recordings. In Section 3.1 we first describe the matching procedure, followed by introducing two enhancement strategies. The first strategy aims to cope with musical key shifts or different tunings and the second one with high tempo variations between the query and the database. As we will perform our following experiments based on the book by Barlow and Morgenstern (BM) and its digitized version from the EDM-website, we will further investigate in these datasets. Due to computation time and potential pitfalls when considering the entire dataset, we introduce three subsets in Section 3.2. Retrieval tasks that are similar to the matching task of this thesis can be found in [3, 14, 17].

Before we describe the structure of the pipeline for the matching process we make several assumptions about the underlying data: The used subsets contain Western classical music, so we assume an equal tempered scale containing 12 semitones per octave as a foundation for chroma features. Furthermore, we assume in general that the melody specified by the EDM-theme we use as query is dominant in the polyphonic audio recording. This assumption is met because of several reasons: The musical themes in the BM are famous themes which we would remember from listening to the audio recording and then may want to look them up. Therefore a theme must be well audible which implicates that it is the dominant melody. The theme can only be well audible if the degree of polyphony is low, or if several instruments play the theme at the same time. Furthermore, many of the themes are repeated several times in an audio recording. This also leads to a memorizing effect by the listener. It is highly possible, that the theme occurs as dominant melody at least once in the audio recording. We also assume that the queries do not contain playing errors. We assume that the audio recordings are of fair quality. Finally, we assume that every query can be mapped to one audio recording uniquely. For example we have a query a which corresponds to audio recording A and query b corresponds to audio recording B . So we assume that A does not contain the melody specified by b and B does not contain the melody specified by a . This also holds for transposed and detuned versions of b and a which may also differ in tempo.



Figure 3.1: Sheet music of the “Fate-Theme” from Beethoven’s 5th Symphony.

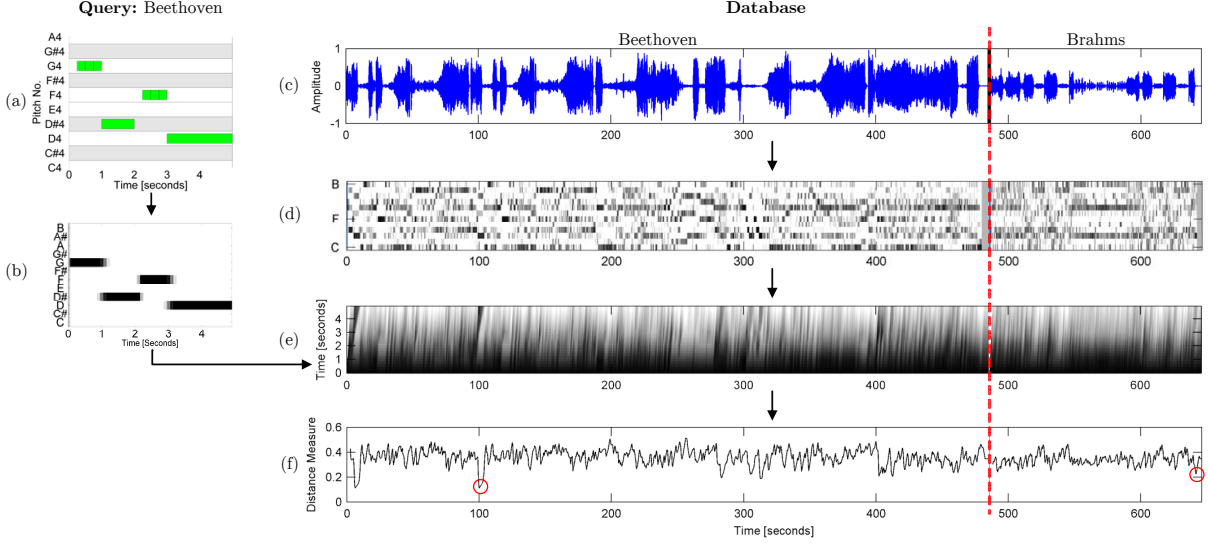


Figure 3.2: Outline of the overall pipeline of this thesis showing the matching procedure for one query with a database that contains two audio recordings. (a) Piano roll representation of the “Fate-Theme” by Beethoven. (b) Chromagram of the query. (c) Waveform of the database consisting of Beethoven’s 5th Symphony and Brahms’ Hungarian Dance No.5. (d) Chromagram of the database. (e) Accumulated cost-matrix. (f) Matching function for the given query in the database.

3.1 Matching Procedure

Following these assumptions, we present our matching procedure. Let our goal be to find the theme given in Figure 3.1 in a database of audio recordings. Figure 3.2 shows the pipeline for this particular example. The piano roll representation of the query can be seen in Figure 3.2(a). The corresponding database is shown in Figure 3.2(c), consisting of a recording from Beethoven’s 5th Symphony and Brahms’ Hungarian Dance No.5. The core idea of our matching procedure is to convert both modalities into a mid-level representation and then find an optimum alignment between the respective feature representations of the query and a subsequence of the database. As mid-level representation, we use chroma features, in particular CENS features, as introduced in Section 2.1.1. The implementation is given by the Chroma-Toolbox, provided by Müller et. al. [2]. The conversion result is shown in Figure 3.2(b) resp. Figure 3.2(d). We define $X \in \mathbb{R}^{\{12,36\}}$ as chroma sequence of the query and $Y \in \mathbb{R}^{\{12,36\}}$ as chroma sequence of the database. The database Y is built by simply concatenating the chroma representations of the audio recordings. To prevent the algorithm which performs SDTW from aligning paths across the end of one audio

Rank	ComposerID	WorkID	Minimum
1	Beethoven	Op067-01	0.11
2	Brahms	HungarianDances-05	0.23

Table 3.1: Results of the matching procedure shown as a ranked list. Beethoven was used as query and correctly identified by the algorithm.

recording and the start of another audio recording in the database, the chroma sequences are separated by two infinity columns $y_\infty \in \mathbb{R}^{12,36}$. In our example, the feature sequence of the database looks like this:

$$Y = [Y_1 \ y_\infty \ y_\infty \ Y_2]. \quad (3.1)$$

The chroma sequence Y of the database is visualized in Figure 3.2(d). With these two feature sequences we perform SDTW as explained in Section 2.2. The accumulated cost matrix is shown in Figure 3.2(e). To get the optimum alignment, we extract the matching function from the accumulated cost matrix, see Figure 3.2(f). To extract the best match for the given query in the database, we consider each database entry separately. In this example, we yield two minima from the matching function: One for the audio recording by Beethoven and a second one for the audio recording by Brahms, highlighted by red circles in Figure 3.2(f). The audio recordings are sorted in a ranked list, according to the respective minimum values in ascending order (see Table 3.1). For this matching procedure, the first entry of the list is the best matching candidate. In our example, the best match is at 100 seconds (see Figure 3.2(f)). This is where the theme is played for the second time in the Symphony.

In this example, the procedure is successful. This is not the case for every query, especially when there are significant time stretches or detuning of the audio recording. As we already mentioned, we therefore extend this pipeline by using two enhancement strategies.

The first strategy takes into account transposed and detuned versions of the query by using the cyclic shift operator defined in the Equations (2.14) and (2.15) to shift the feature sequence of the query. Then we derive matching functions as explained previously using all shifted versions of the query. Figure 3.3 shows matching functions for five transpositions. The black function in

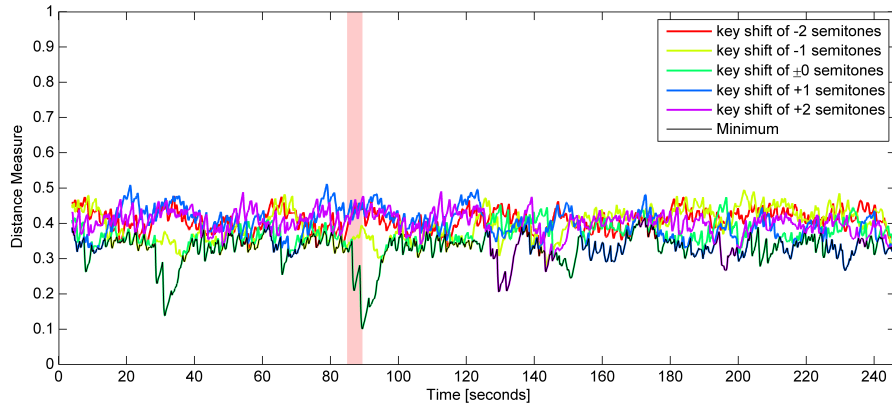


Figure 3.3: Matching functions for Beethoven's Op.2 No.1 for 5 transpositions. The sequence of lowest cost is highlighted in light red color.

3. MATCHING SYMBOLIC THEMES TO AUDIO

this figure is derived by taking the minimum value of these five matching functions for each time instant as kind of a lower envelope. In this thesis we refer to this function as minimum function. In order to evaluate musically transposed or detuned versions of a query, we apply the cyclic shift operator to the chroma representation of the query once or several times and derive the matching functions. Then we derive the minimum function and perform the evaluation as explained in the previous paragraph. In the following, we define a parameter that denotes the number of key shifts for a certain type of chroma features. We use the parameter κ_0^{12} or κ_0^{36} to denote that we only use the standard tuning without any key shifts as query, using 12-dimensional chroma features or 36-dimensional chroma features, respectively. Furthermore we use the parameter κ_2^{12} to denote that we use 12-dimensional chroma features including all musical key shifts from +2 semitones to -2 semitones. Accordingly, we use the parameter κ_2^{36} to denote that we use 36-dimensional chroma features including all musical key shifts from +2.3 semitones to -2.3 semitones. Note that we compare the results for the key shifts of ± 2 semitones for the 12-dimensional chroma features to the respective key shifts of ± 2.3 semitones for the 36-dimensional chroma features to also consider audio recordings that are transposed by 2 semitones and additionally detuned. We also define a parameter $W := (w_d, w_h, w_v)$ that denotes the used weight factors as follows: $W_1 := (1, 1, 1)$, $W_2 := (1, 0.7, 1.4)$, $W_3 := (1, 0.5, 2)$, $W_4 := (1, 0.3, 3)$.

The second strategy is used to compensate high tempo variations between query and database. Therefore we stretch or shrink the query to a fixed duration before performing the retrieval task. Figure 3.4(a) shows the piano roll representation of the EDM-theme “Prelude No.1” from the “Well-tempered Clavichord Book I” by Bach. In the Figures 3.4(b) to 3.4(f) we can see chromagrams of this EDM-theme with different durations. Besides the original duration (denoted as $t_\delta = \text{orig}$) specified by the EDM-theme shown in Figure 3.4(b) we use the durations $t_\delta = \{5 \text{ s}, 10 \text{ s}, 15 \text{ s}, 20 \text{ s}\}$ for the Figures 3.4(c) to 3.4(f), respectively. Smoothing and low duration cause blending effects between chroma vectors (see Figure 3.4(b) and 3.4(c)). This effect is undesirable, as it reduces the specificity of the query.

With the introduced strategies we make our pipeline very robust against problems in the dataset, such as detuned and transposed audio recordings, as well as high tempo variations. Using the cyclic shift operator, we take into account transposed and detuned versions of the query. Furthermore, SDTW compensates global and local tempo variations between query and database to some extent. High tempo variations are compensated by stretching or shrinking the query to a fixed duration. Finally, we can improve the specificity of a query by increasing the feature resolution for the CENS features. So we cover all the challenges described in Section 1.1 except for the degree of polyphony. Therefore we meet the assumption that the melody specified by the EDM-theme we use as query is dominant in the polyphonic audio recording.

As we will use many different parameter settings for our experiments in Chapter 4, we fix a notation for a parameter setting. We use the notation $\{\text{CENS}_d^\omega, \Sigma_s, t_\delta, W\}$ to specify one parameter setting including the window-length smoothing ω , down-sampling factor d , step size $s \in \{1, 2\}$ (see definition in Equation (2.5) and (2.6)), duration t_δ and the weight factors defined by W . In this thesis the default pitch feature resolution is 10 Hz. So CENS_{10}^{41} features have a feature resolution of 1 Hz and CENS_1^4 features have a feature resolution of 10 Hz. If not explicitly specified for the respective experiment, we use the setting $\{\text{CENS}_{10}^{41}, \Sigma_1, t_\delta = \text{orig}, W_1\}$.

Figure 3.5 gives an overview on the input and output behavior of the pipeline. It shows the dependencies between the input parameters and particular parts of the pipeline. Given a subset

which contains queries and a database of audio recordings, the documents of the database are concatenated. Then we derive the database sequence Y , using the specified windowlength smoothing factor ω and the downsampling factor d for the CENS features of the audio recordings. Furthermore we adjust the query taking into account the duration t_δ and the key shifts κ . We then calculate the CENS feature representation X of the query, also taking into account ω and d . The SDTW is performed on the sequences X and Y and the step size condition defined by Σ_s . As a result we get a matching function whereas we derive a minimum value for each document of the database. Then we create a ranked list containing the documents of the database, sorted by the respective minimum in ascending order. For each query we derive a ranked list. The output of the pipeline are the evaluation measures Mean Rank and Top Match, which are derived from the ranked lists, see Section 2.4.

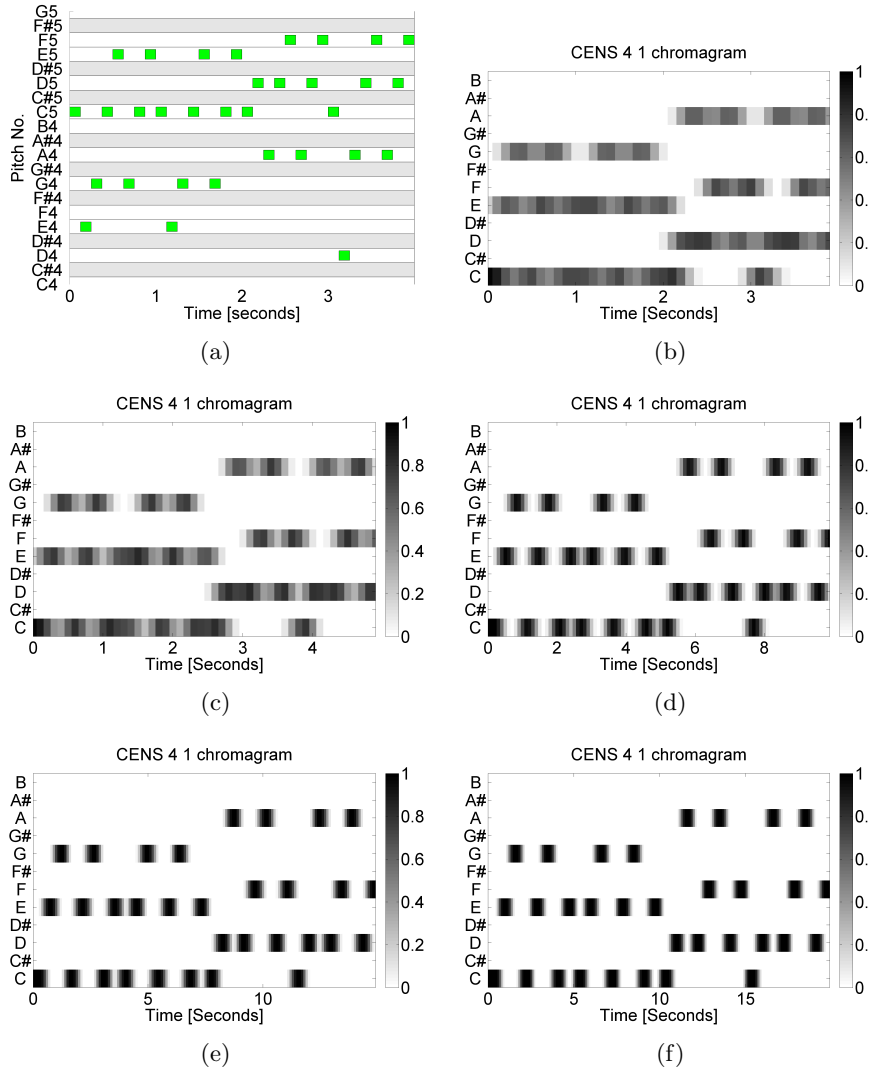


Figure 3.4: (a) Piano roll representation of the EDM-theme from the “Well-tempered Clavi-chord Book I”, “Prelude No.1” by Bach. (b) to (f): Chromagrams of the durations $t_\delta = \{\text{orig}, 5 \text{ s}, 10 \text{ s}, 15 \text{ s}, 20 \text{ s}\}$, respectively.

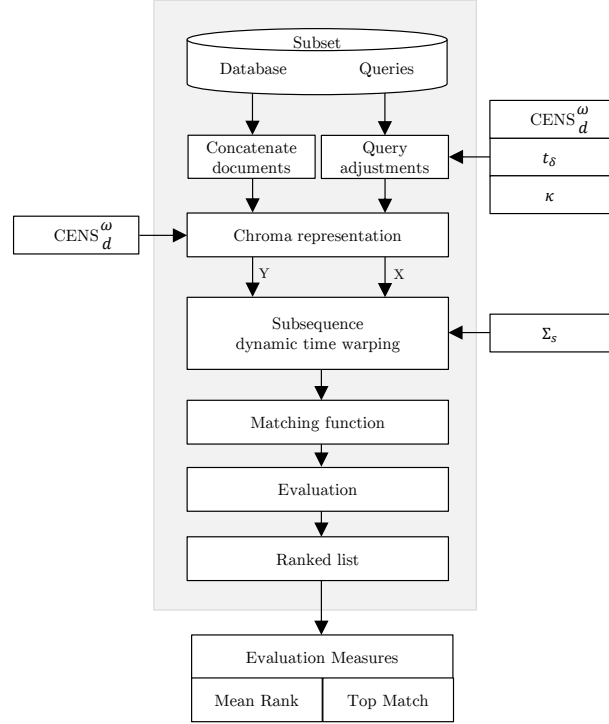


Figure 3.5: Overview of the input and output behavior of the pipeline, including the dependencies of the parameters on the appropriate parts of the pipeline.

3.2 Datasets

In order to do experiments on the pipeline we introduced, we need several datasets which we will refer to as subsets. We decided to create three subsets, each with a different number of audio recordings, to then compare the retrieval results for these subsets. The music data of the subsets is derived from the book “A DICTIONARY OF MUSICAL THEMES” by Harold Barlow and Sam Morgenstern. The BM was designed as a reference book for trained musicians and professional performers [1]. It contains roughly 10,000 musical themes from outstanding compositions in Western tradition including orchestral music, chamber music, and works for solo instruments. Each theme is specified by a visual score representation of the notes. Furthermore, the respective composer, the underlying musical work, and the movement are specified. Within the book, the themes are systematically organized and suitably indexed. Most of the BM-themes are also available as MIDI-files provided by the Electronic Dictionary of Musical Themes (EDM)¹. In Figure 3.6 we can see one exemplary page of the BM containing musical works by Beethoven. The musical work “Sonata No.1, in F Minor Op.2, No.1, Pft” is highlighted by a blue frame. It contains four movements. The first movement is represented by two themes, highlighted by a red frame.

In the following, we use the term “audio recording” to refer to what one typically finds as a track of a commercial CD of Western classical music. The tracks, in turn, usually correspond to individual movements or numbers of a multi-movement musical work (such as a sonata, symphony,

¹<http://www.multimedialibrary.com/barlow/>

BEETHOVEN 58 B686—B702

Sonata in F
Op. 17, Horn & Pft.

1st Movement B686

2nd Movement B686a

3rd Movement B686b

Sonata No. 1, in F Minor
Op. 2, No. 1, Pft.

1st Movement B687

2nd Movement B688

3rd Movement B688a

1st Movement 1st Theme B689

1st Movement 2nd Theme B690

2nd Movement B691

3rd Movement 1st Theme Minuet B692

3rd Movement 2nd Theme B693

4th Movement 1st Theme B694

4th Movement 2nd Theme B695

Sonata No. 2, in A
Op. 2, No. 2, Pft.

1st Movement 1st Theme B696

1st Movement 2nd Theme B697

2nd Movement B698

3rd Movement 1st Theme B699

3rd Movement 2nd Theme B700

4th Movement 1st Theme B701

4th Movement 2nd Theme B702

Figure 3.6: One bookpage of the BM [1], whereas one musical work and one movement are highlighted in different colors.

opera, suite), or to the entire piece of music if it consists of only a single movement. In general, a single movement may contain several BM-themes.

We create the subsets by matching data types as follows: First we choose movements from the BM. Second we add one version of a corresponding audio recording to the subset. Third we add all the corresponding EDM-themes from the EDM-website [18] to our subset. In this thesis we will refer to one movement, a corresponding audio recording and all the corresponding EDM-themes as *BM data type*. According to this procedure we create three subsets, whereas we only use data that is complete, meaning there is at least one audio recording available for the movement that contains

3. MATCHING SYMBOLIC THEMES TO AUDIO

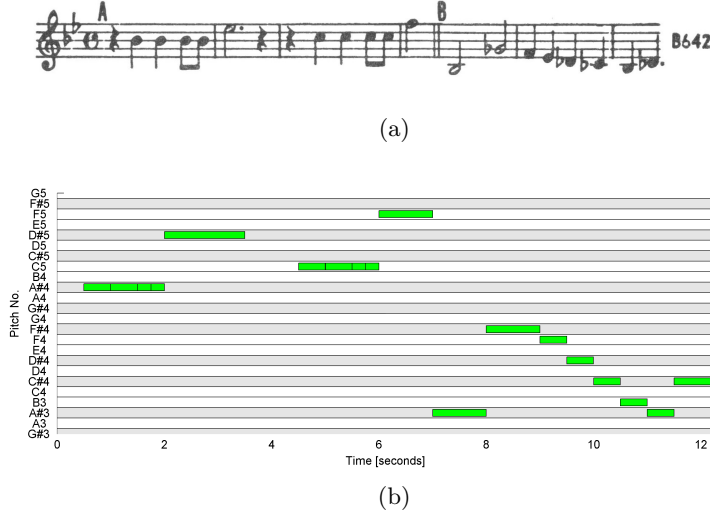


Figure 3.7: Example for Midi Joint Themes. (a) BM-theme of Beethoven’s Quartet in B Flat, Op130-01, 1st Movement, derived by concatenating the parts ‘A’ and ‘B’. (b) Piano roll representation of the EDM-theme.

all its BM-themes and all corresponding EDM-themes are available. Table 3.2 gives an overview on the three subsets including the number of audio recordings, the number of EDM-themes and the total duration of all audio recordings.

Some of the EDM-themes are problematic with regard to the usage for our pipeline. In this thesis we will refer to these two exceptions as *Midi Joint Theme* and *Midi Part*. In the following, we explain these two irregularities. Figure 3.7 shows a BM-theme and the piano roll representation of its corresponding EDM-theme. The capital letters ‘A’ and ‘B’ in Figure 3.7(a) indicate that these 6.5 measures which we can see in the sheet music are not played consecutively in the audio recording. The first four measures which are denoted by the letter ‘A’ are played in the audio recording at first, followed by a second part which is not notated in the sheet music. Then the 2.5 measures are played in the audio recording which are denoted by the letter ‘B’. In Figure 3.7(b) we can see that in the EDM-theme the parts ‘A’ and ‘B’ are played consecutively, whereas the second part is missing. As we use the EDM-theme as a query, we ignore the part enclosed by the parts ‘A’ and ‘B’ and thus our query is wrong, because the melody specified in the query does not occur consecutively in the audio recording.

Figure 3.8 shows another kind of exception. In this case one BM-theme corresponds to two EDM-themes, that means we have two queries for one BM-theme. Generally, these queries are not wrong compared to a Midi Joint Theme, but by splitting one query into two queries, the queries

Subset	No. Audio	No. EDM	Dur. DB [h:min:s]	Dur. Queries [h:min:s]	Mean Dur. per Audio [min:s]	Mean Dur. per Query [s]	Mean Notes per Query
BM-Mini	14	26	01:10:41	00:03:11	05:02	7.3	16.2
BM-Small	100	177	10:50:50	00:21:57	06:30	7.4	20.5
BM-Medium	1113	2039	119:10:44	04:00:45	06:24	7.1	19.1

Table 3.2: Basic information for the three created subsets including the number of audio recordings, the number of EDM-themes and the duration of the respective database.

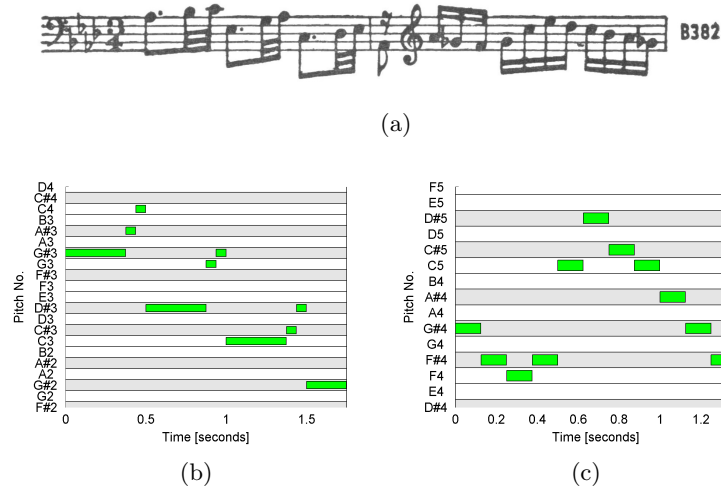


Figure 3.8: Example of Midi Parts. (a) BM-theme of Bach’s “Well Tempered Clavichord Book II”, Prelude No.17. (b) and (c): Visualization of the corresponding EDM-themes in a piano roll representation.

get shorter and thus loose specificity which makes the retrieval task harder. In this thesis we refer to this kind of exception as *Midi Part*.

In the following we describe each subset in more detail and explain the rules we applied to create the subsets.

The first subset BM-Mini contains 26 famous EDM-themes of Western classical music and 14 audio recordings, see Table 3.2. The mean duration of an EDM-theme in this subset is roughly 7 seconds. Figure 3.10(a) and Figure 3.10(b) show distributions of the query duration and of the number of notes, respectively. This database has a duration of roughly one hour. It is statistically not representative for doing evaluations and was created to analyze queries which represent edge cases for a specified parameter setting, and to test if our pipeline works. The advantage of this subset is that experiments require only a short time for such a small number of audio recordings. In this subset we do not use BM data types that contain a Midi Joint Theme or a Midi Part. In Appendix B.4 we provide a table containing all audio recordings of the subset BM-Mini, including Composer, WorkID², name of the album or CD-box, CD and track number, instrumentation and EnsembleID. Furthermore, in Appendix B.5 we provide a table containing all EDM-themes of the subset BM-Mini, including an index number for each EDM-theme, Composer, WorkID, BMID³, ThemeID and the MidiNo⁴. In this thesis we refer to specific queries of the subset BM-Mini using this index number. For example by the term **query #1** we refer to Bach’s “Well Tempered Clavichord Book I”, Prelude No.1, BWV 846. Finally, Appendix B.1 provides matching functions for every query of the subset BM-Mini with specified feature settings.

The second subset BM-Small contains 177 EDM-themes and 100 audio recordings, see Table 3.2. The total duration of the database is roughly 11 hours and thus much larger than BM-Mini. The

²An identifier that uniquely identifies one movement in combination with the Composer.

³The BMID is an identifier, that uniquely identifies one BM-theme. The BMID is the number provided on the right side of each BM-theme in the BM.

⁴A number that uniquely identifies each EDM-theme provided by the EDM-website.

3. MATCHING SYMBOLIC THEMES TO AUDIO

mean duration of a query however, is also roughly 7 seconds. One query consists of roughly 20 notes in mean. Distributions of the query duration and of the number of notes are shown in the respective Figures 3.10(c) and 3.10(d). BM-Small contains the subset BM-Mini. Furthermore, we select many more audio recordings considering certain constraints, to make it statistically as representative for the entire BM as possible - however it contains only roughly 2% of the BM-themes. The first constraint is to choose about 100 movements in total from many different composers and different periods. The second constraint is to use movements with many different instrumentations and ensembles. In this subset we do not use BM data types that contain a Midi Joint Theme or a Midi Part. In Appendix B.6 and B.7 we provide tables containing all audio recordings and all EDM-themes of BM-Small. For most of the experiments we use this subset, as it is (assumed to be) representative for the entire BM and the runtime for most of the experiments is still acceptable.

The third subset BM-Medium contains 2039 EDM-themes and 1113 audio recordings, see Table 3.2. The total duration of the database is roughly 119 hours and thus roughly ten times larger than BM-Small. The mean duration per query is roughly 7 seconds, just like for the previous subsets. In this subset one query also consists of roughly 20 notes in mean. So from a statistical point of view, the queries in the subset BM-Small are quite similar to the subset BM-Medium. Distributions of the query duration and of the number of notes are shown in the respective Figures 3.10(c) and Figure 3.10(d). Figure 3.10(g) shows a histogram of the distribution of the query length with respect to the number of notes of the query. There is no noticeable correlation between the number of notes and the query length. In general, we would expect a correlation between the number of notes and the query length. As this is not the case, this indicates that the query length of a noticeable number of EDM-themes is wrong. BM-Medium contains the subset BM-Small. It allows us to perform evaluations on a big dataset. We create this subset by selecting several complete cycles, including all BM data types derived by the 24 preludes and fugues from “The Well-Tempered Clavichord I” and the 24 preludes and fugues from “The Well-Tempered Clavichord II” by Bach, all BM data types by Beethoven, all BM data types derived by the Symphonies and the Concertos by Brahms, all BM data types derived by the Mazurkas and by the Nocturnes by Chopin and several famous pieces by Bartók, Debussy, Dvořák, Haydn, Khachaturian, Mahler, Mozart, Schubert, Schumann, Shostakovich and Tchaikovsky. In the subset BM-Medium we allow BM data types containing Midi Joint Theme and Midi Part if they are part of a complete cycle. In Appendix B.8 we provide a statistic that shows each composer and the respective number of audio recordings that are in the subset BM-Medium.

In general, we do not exclude BM data types that have tempo errors and playing errors in some of the respective EDM-themes which are recorded in Appendix B.9. Finally, we created three subsets that we will use for our experiments on the pipeline in Chapter 4.

To this end, we discuss reasonable parameter settings for the created subsets, to then also consider these parameter settings in our experiments. Therefore we have a look at two queries which we considered to be extreme examples. In Figure 3.9 we can see two examples of BM-themes, where the according EDM-themes have a duration of only roughly two seconds. The EDM-theme corresponding to the BM-theme Bach’s Toccata in D Minor shown in Figure 3.9(a) has a duration of 1.6 seconds. The corresponding section of the audio recording however has a duration of roughly 15 seconds. This means a tempo difference of more than factor 9. To overcome such high global tempo variations, we do the fixed duration sampling for the queries, whereas we use the durations $t_\delta = \{5 \text{ s}, 10 \text{ s}, 15 \text{ s}, 20 \text{ s}, \text{orig}\}$. On the one hand fixed duration sampling for short



(a) Bach's Toccata in D Minor, BWV0565-01

(b) Chopin's Fantaisie-Impromptu, 1st Theme, Op066

Figure 3.9: BM-themes where corresponding EDM-themes are of short duration

queries leads to a time stretched query. So this query which is often erroneously too short, is sampled to a duration that is closer to the duration of the corresponding BM-theme in the audio recording. This also prevents us from choosing an extremely high feature resolution. As the EDM-theme contains 20 notes⁵, we would have to choose a feature resolution of at least 13 Hz⁶, to ensure that every note is represented by at least one feature vector.

In the second example, the excerpt in the audio recording has a duration of only roughly 1.7 seconds, which means that it is even shorter than the corresponding EDM-theme with 2 seconds duration. This BM-theme contains 15 notes. This means, that we need a feature resolution of at least 7.5 Hz for both, the audio recordings and the EDM-themes. So a sufficient feature resolution is 10 Hz. To obtain this feature resolution, we use CENS₁⁴ features and CENS₁¹ features. Due to high computational load for such a high feature resolution, we also consider lower feature resolutions of 1 Hz, 2 Hz and 5 Hz, using the respective features CENS₁₀⁴¹, CENS₅²¹ and CENS₂⁹. Here we also see the downside of fixed duration sampling. By sampling this EDM-theme to a duration of 10 seconds, we only can overcome the global tempo variation of factor 6 with the step size condition Σ_1 which lead to a high loss in specificity. So one task is to find a suitable combination of the step size condition and the duration t_δ to compensate the high global tempo variations and thus get good retrieval results. We will tune these and other parameters in Chapter 4 by analyzing selected parameter settings and evaluating the respective performance of our pipeline.

⁵In this thesis we determine the number of notes of a BM-theme by counting the number of MIDI events of the corresponding EDM-theme. This number may deviate from the actual number of notes of the BM-theme.

⁶Therefore we assume that the EDM-theme contains no rests and every note has the same duration.

3. MATCHING SYMBOLIC THEMES TO AUDIO

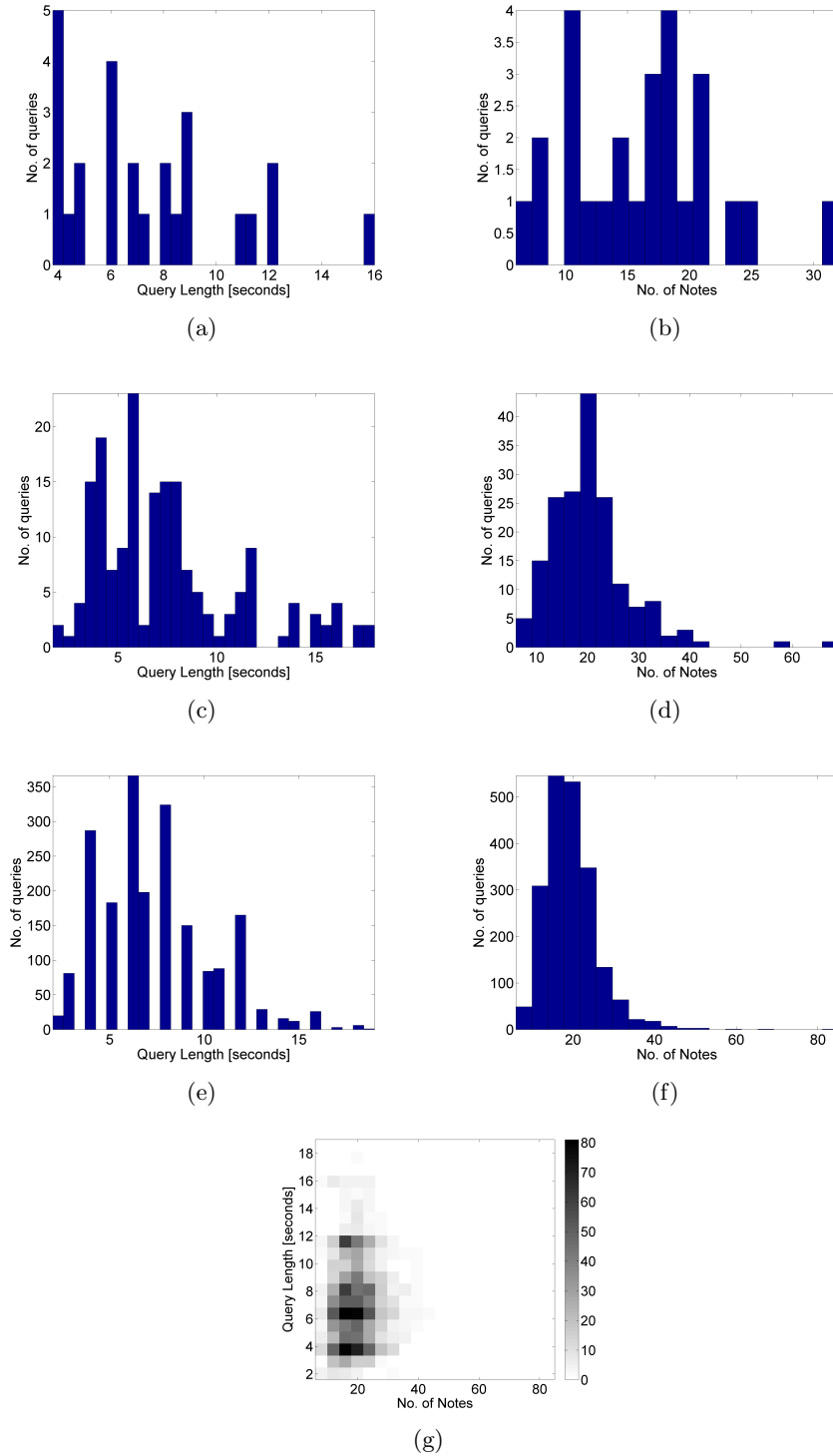


Figure 3.10: Histograms of the subsets. (a) Query lengths of BM-Mini. (b) Number of notes of BM-Mini. (c) Query lengths of BM-Small. (d) Number of notes of BM-Small. (e) Query lengths of BM-Medium. (f) Number of notes of BM-Medium. (g) Query length in dependence of the number of notes of the query for BM-Medium.

Chapter 4

Experiments

In this chapter, we present the results of our experiments using the pipeline and subsets which we introduced in Chapter 3. For each subset the task is to retrieve the audio recording specified by the corresponding query. As a subset has several queries, all these queries are processed. The retrieval results are then evaluated, taking into account all the queries. Therefore we use the evaluation measures Mean Rank and Top Match, as introduced in Section 2.4.

We start with the smallest subset **BM-Mini**. This subset is only used to test our algorithm and study certain behavior. We can see the matching results for each of the themes in Appendix B.1. Furthermore, retrieval results for many different parameter settings are shown in Appendix B.2. The results in this chapter are based on the subset **BM-Small** which contains 100 audio recordings and 177 queries. We will use these results to evaluate different parameters of the algorithm. We then fix a set of parameters which will be applied to the subset **BM-Medium** which contains 1113 audio recordings and 2039 queries.

Finally, we will pick some specific examples for cases where the matching procedure had problems in identifying the correct audio recording in the database to further understand the algorithm and the underlying data.

4.1 Parameter selection using the subset **BM-Small**

In this section we determine our best possible parameter setting for the subset **BM-Small** by doing experiments in the following order: First, we fix one parameter setting and evaluate the results by calculating the Mean Rank and the Top Match for this parameter setting. Second, we do this experiment again, whereas we vary one parameter according to the different settings defined in Section 3.1. Then we compare the results of these experiments, determine our best value for this parameter and fix this value for the next experiment. Then we repeat this procedure, whereas we vary the next parameter, until in the end we fixed all parameters.

4. EXPERIMENTS

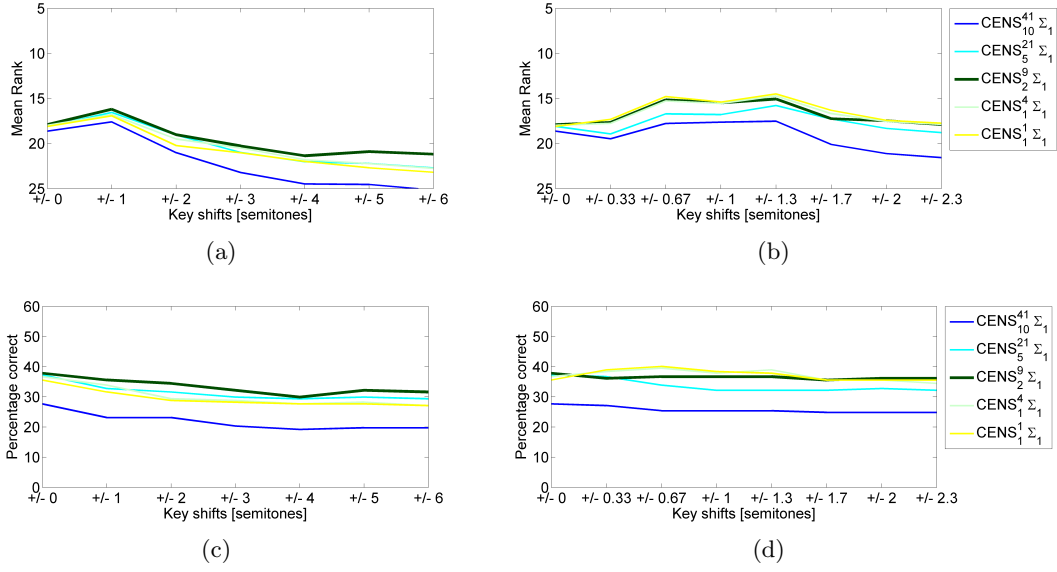


Figure 4.1: Dependency of evaluation results on the feature resolution. The evaluation was performed on the subset BM-Small, using the settings $\{\Sigma_1, W_1\}$. (a) Mean Rank with 12-dimensional chroma features. (b) Mean Rank with 36-dimensional chroma features. (c) Top Match with 12-dimensional chroma features. (d) Top Match with 36-dimensional chroma features.

4.1.1 Chroma resolution and feature resolution

In this section we determine the chroma resolution and the feature resolution that perform best for our experimental setup. First, we compare the 12-dimensional chroma features to the 36-dimensional chroma features. Figure 4.1(a) shows the Mean Rank for the 12-dimensional chromas with respect to the number of applied key shifts. The figure contains one function for several different feature resolutions. The lowest Mean Rank of 16 is achieved with the parameter setting $\text{CENS}_2^9 \Sigma_1$ with the key shifts of ± 1 semitone. This indicates, that several audio recordings are transposed by ± 1 semitones. For the key shifts of ± 2 semitones we get a Mean Rank of 19. The more number of key shifts we apply, the more unspecific our query gets and thus the Mean Rank gets worse. Figure 4.1(c) shows the Top Match for the 12-dimensional chromas. Our best Top Match is also achieved using the parameter setting $\text{CENS}_2^9 \Sigma_1$. Using the key shifts of ± 0 semitones we get a Top Match of 38%. For the key shifts of ± 1 semitones the Top Match is still 36%. In the following, we compare these results to the 36-dimensional chromas. Using the parameter setting $\text{CENS}_2^9 \Sigma_1$ we get slightly better results than for the 12-dimensional chromas (see Figure 4.1(b) and 4.1(d)). By contrast, for the 36-dimensional chromas the setting $\text{CENS}_1^1 \Sigma_1$ performs better with regard to the Mean Rank and the Top Match. So it implies that for a higher feature resolution, also the tuning gets more relevant. In the following, we restrict our experiments to the 36-dimensional chromas. We achieve the lowest Mean Rank of 15 and the highest Top Match of 40% using the parameter setting $\{\text{CENS}_1^1 \Sigma_1\}$ and key shifts of ± 1.3 semitones and ± 0.67 semitones respectively (see Figure 4.1). So the $\text{CENS}_1^1 \Sigma_1$ features perform better, even though these features are not suitable for larger datasets due to the high computational cost. So the $\text{CENS}_2^9 \Sigma_1$ features are a good trade-off between computational cost and retrieval performance.

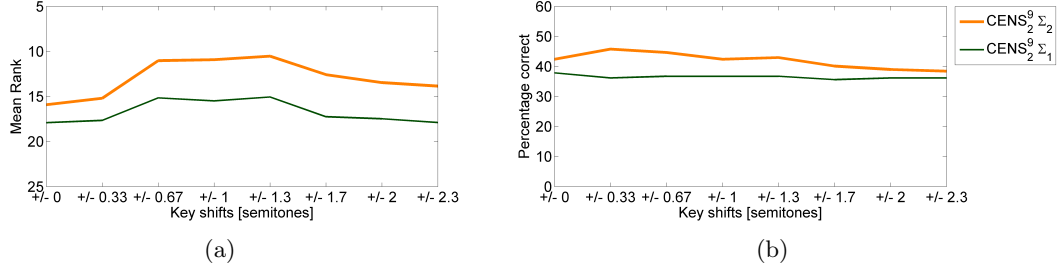


Figure 4.2: Dependency of evaluation results on the step size. The evaluation was performed on the subset BM-Small, using the settings $\{\text{CENS}_2^9, \text{orig}, W_1, \kappa^{36}\}$. (a) Mean Rank (b) Top Match

4.1.2 Step size condition

In the following, we compare the step sizes Σ_1 to the step sizes Σ_2 for the subset BM-Small. In Figure 4.2 we can see, that the lowest Mean Rank of 11 as well as the highest Top Match of 46% is achieved for the step sizes Σ_2 . The step sizes Σ_2 leads to much better results because this parameter strongly constrains the warping path of the SDTW (see Section 2.2). This constraint in turn increases the specificity which is a very important property for a database of 100 audio recordings. Furthermore, it indicates that there are many queries that have global tempo variations of less than a factor of two.

4.1.3 Fixed query length

In the next step we evaluate the results for different query durations. The Mean Rank is slightly better for a fixed duration of $t_\delta = 10$ s than for the so far used original duration (see Figure 4.3). This indicates that there are several queries having a high global tempo variation which is compensated by the usage of fixed duration sampling. The queries of the length 10 seconds lead to the better results for the Mean Rank of 9 as well as for the Top Match of 46% (see Figure 4.3). To this end, we infer that the matching excerpts of the audio recordings also have a mean duration of 10 seconds. Furthermore, we infer, that many of the queries have a wrong tempo, as the mean duration of the queries is only roughly 7 seconds (see Table 3.2). Using fixed

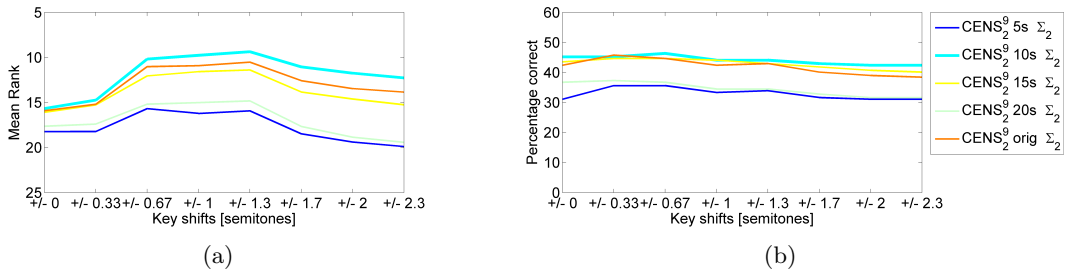


Figure 4.3: Dependency of evaluation results on the duration of the queries. The evaluation was performed on the subset BM-Small, using the settings $\{\text{CENS}_2^9, \Sigma_2, W_1, \kappa^{36}\}$. (a) Mean Rank (b) Top Match

4. EXPERIMENTS

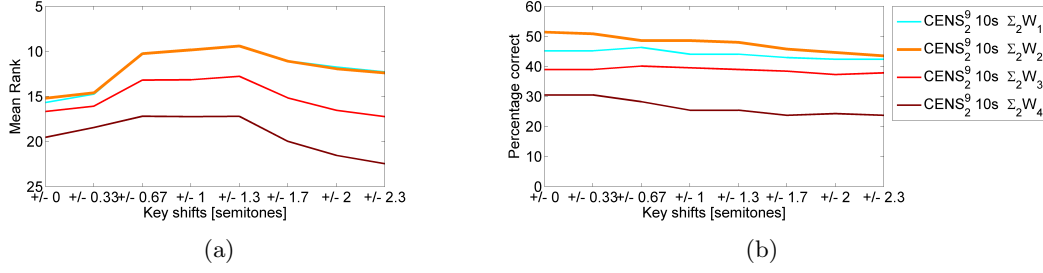


Figure 4.4: Dependency of evaluation results on the weight factor. The evaluation was performed on the subset BM-Small, using the settings $\{\text{CENS}_2^9, 10 \text{ s}, \Sigma_2, \kappa^{36}\}$. (a) Mean Rank (b) Top Match

duration sampling with $t_\delta = 10 \text{ s}$ and the step size condition Σ_2 seems to be a good trade-off between both parameters.

4.1.4 Weight factors

In the following, we evaluate different weight factors. Using the weight factors W_2 we also get a Mean Rank of 9. Furthermore we get a Top Match of 51% (see Figure 4.4), which is a high gain by comparison to the so far used weight factors W_1 . So we finally fix the parameter setting $\{\text{CENS}_2^9, 10 \text{ s}, \Sigma_2, W_2, \kappa^{36}\}$.

4.1.5 Key shifts

Finally considering several key shifts, we have a look not only on the Top Match, but also on the ranks below. Therefore, we use the previously mentioned parameter setting. Figure 4.5 shows the percentage of correct matches as a function of the Top N Match. The Top 10 Matches differ for the evaluated key shifts from 72% to 78% (see Figure 4.5). The Top 50 Matches however differ more widely from 87% to 95%. In the following, we consider the Top 50 Matches. Therefore we subdivide the different number of key shifts into three categories:

- Category I: Key shifts of ± 0 semitones and ± 0.33 semitones
- Category II: Key shifts of ± 0.67 semitones, ± 1 semitones, and ± 1.3 semitones
- Category III: Key shifts of ± 1.7 semitones, ± 2 semitone and ± 2.3 semitones

For Category I we get the lowest Top 50 Matches in the range of 87% and 89%, because several audio recordings are transposed or detuned. Thus they are not detected by the algorithm if we apply only key shifts of up to ± 0.33 semitones. We get our highest Top 50 Matches of roughly 95% for Category II because these key shifts compensate transposed and detuned audio recordings and thus lead to better retrieval results for several queries. Finally, the results for Category III are slightly worse than that for Category II. Transposed and detuned audio recordings are also detected by Category III, however the higher number of key shifts also leads to a lower specificity and thus degrades the results. Our best Top 5 Match is achieved for key shifts of

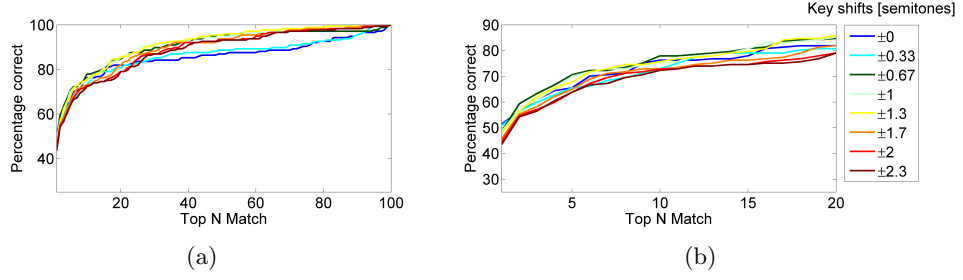


Figure 4.5: Dependency of evaluation results on the number of key shifts. The evaluation was performed on the subset BM-Small, using the settings $\{\text{CENS}_2^9, 10\text{ s}, \Sigma_2, W_2, \kappa^{36}\}$. (a) Overview of the entire database. (b) Extract showing the Top 20 Matches.

± 0.67 semitones at 70% (see Figure 4.5(b)). So the Top 5 Match is much higher than the Top Match with 51% at most.

4.2 Evaluation of the subset BM-Medium

In this section we discuss the results for the subset BM-Medium and compare them to BM-Small. Furthermore, we consider histograms for several measures to get deeper insight into the data of BM-Medium. For BM-Medium we apply the parameter setting $\{\text{CENS}_2^9, 10\text{ s}, \Sigma_2, W_2, \kappa^{36}\}$ which we considered to be a good trade-off for the subset BM-Small. Figure 4.6 shows the evaluation results for the subset BM-Medium. We get a Mean Rank of 110 and a Top Match of 29% when

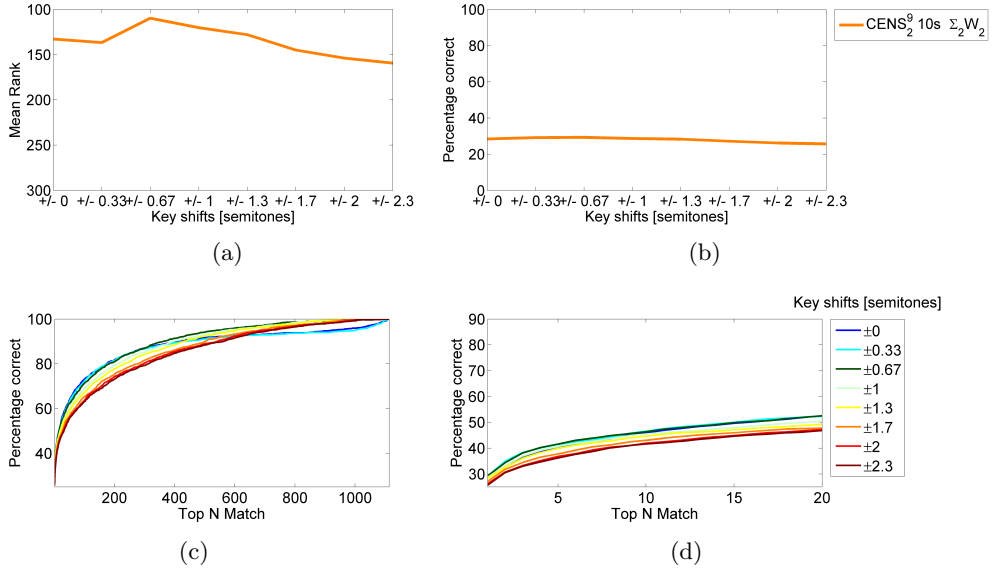


Figure 4.6: Dependency of evaluation results on the number of key shifts. The evaluation was performed on the subset BM-Medium, using the settings $\{\text{CENS}_2^9, 10\text{ s}, \Sigma_2, W_2, \kappa^{36}\}$. (a) Mean Rank. (b) Top Match. (c) Overview of the entire database. (d) Extract showing the Top 20 Matches.

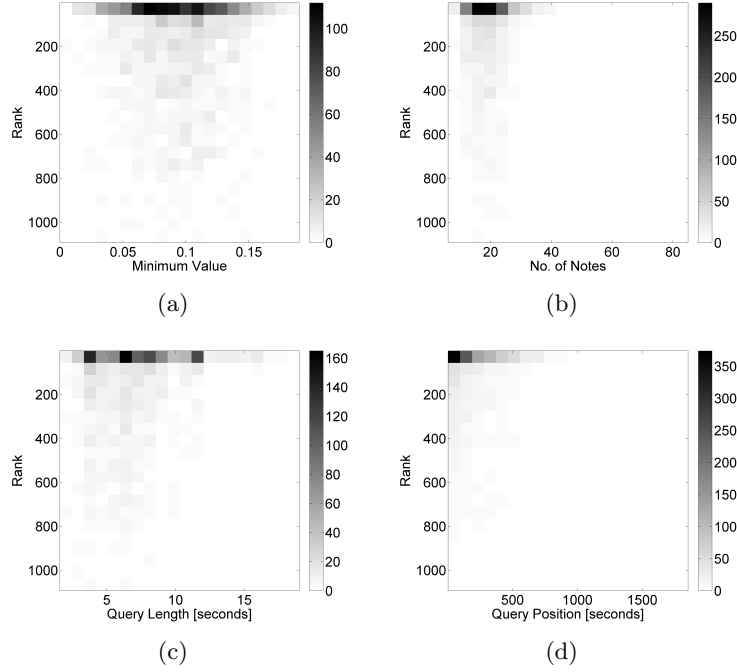


Figure 4.7: Histograms of the distribution of the Ranks in dependence of: (a) Minimum values (b) Number of notes (c) Query length (d) Position of the query within the audio. The evaluation was performed on the subset **BM-Medium** using the parameter setting $\{\text{CENS}_2^9, \text{orig}, W_1, \kappa_2^{36}\}$.

we use the key shifts of ± 0.67 semitones. These results differ widely from the corresponding measures for the subset **BM-Small**, however the subset **BM-Medium** is 11 times larger than the subset **BM-Small**. The Mean Rank of **BM-Medium** is roughly 12 times higher than the Mean Rank of **BM-Small**. So the algorithm scales approximately to the size of the underlying subset. In the following, we compare measures for the Top Matches between both subsets. We take into account the size of the respective subset, so the Top 20 Matches of **BM-Small** correspond to the Top 220 Matches of **BM-Medium**, for instance. **BM-Small** has a Top 20 Match of at most 85% for key shifts of ± 1.3 semitones, whereas **BM-Medium** has a Top 220 Match of at most 83% for the key shifts of ± 0.67 semitones (see Figure 4.6(c)). So both subsets behave similar, with respect to their size. Considering Figure 4.6(c) more closely, we can also subdivide three categories of key shifts, as for the subset **BM-Small**. Category I performs better than the other categories from the Top Match to the Top 300 Match. The reason is, that specificity is more important for large databases.

In the following, we look at the correlation of different measures. Figure 4.7 shows several two-dimensional histograms, whereas every histogram shows the correlation between two measures. In Figure 4.7(a), we see the minimum value of the corresponding matching function in dependence of the rank. We can see that for most of the queries the matching document is among the ranks 1 to 50. We infer that the rank is independent of the minimum value. So in general we cannot exclude documents from a ranked list by considering the minimum value. By excluding documents in a ranked list which have a high minimum value, we would also exclude matches. Figure 4.7(b) shows the correlation between the rank and the number of notes of the query. It also shows that the rank does not depend on the number of notes. This is an unexpected observation because we would expect that the number of notes influences the specificity and thus

a query that contains many notes has a better ranking result. Figure 4.7(c) shows the correlation between the rank and the length of the query. This Figure gives an unexpected result. The rank does not depend on the query length. In general we would expect, that the rank is lower for long queries, as these queries are supposed to be more specific. In Figure 4.7(d), we can see the correlation between the query position and the rank. By the term query position, we mean the time after which the query occurs in an audio recording. Here we see that for most of the audio recordings the query occurs in the beginning. So we can cut all audio recordings in the database to say 700 seconds at most and use the cut audio recordings for the retrieval system. This will save calculation time and also increases the specificity of our database. In practice however, this will either have not much effect on our database because less than 10% of the audio recordings have a duration of more than 700 seconds, or it will definitely result in a loss of Top Matches when we cut the audio recordings too much.

4.3 Analysis of specific queries of the subset BM-Mini

In this section, we consider several specific queries for specific parameter settings. Using the subset BM-Mini as foundation, we analyze several matching functions that do not result in a Top Match. This is done to get deeper insight in the limits of our algorithm and find out how the retrieval can be improved in future. The matching functions shown in this section and in Appendix B.1 stick to the following scheme: The matching function is plotted in black color. Gray vertical lines denote the ending of one audio recording and the beginning of another audio recording. The audio recording which we expect our algorithm to retrieve is highlighted in light red color (ground truth document). The global minimum of the matching function is highlighted by a horizontal and a vertical blue dotted line. The vertical dotted line shows whether the global minimum of the matching function is in the ground truth document or not. The horizontal dotted line is meant to show the distance between the global minimum and the minimum of the ground truth document. It gives a feeling of the gap between the ground truth document and the retrieved document.

4.3.1 Specificity

In this section we analyze the 2nd Theme of the Hungarian Dance No.5 by Brahms (#16). The corresponding matching function is shown in Figure 4.8(a). This matching function has no deep minimum peaks. The function has values of around 0.4 which is quite low compared to other matching functions. In the following, we compare the corresponding chromagrams. Figure 4.8(b) shows the 36-dimensional chromagram of the query, and Figure 4.8(c) shows the chromagram of the corresponding excerpt of the database. Here we see that there are mainly three active chromas G+, B+, D+ like in a chord of G Major. So the chromagram of the database is very unspecific. The reason therefore is probably that the ensemble of this audio recording is an entire orchestra. So the main melody is accompanied by deep string instruments which leads to this smoothened chromagram. By comparing it to the chromagram of the query in Figure 4.8(b), we see that they are very dissimilar. So this query will obviously not lead to good matching results. The audio recording seems to be detuned by ± 1.3 semitones with respect to the query. In Appendix B.27 and B.28 we see that this query never results in a Top Match, no matter which

4. EXPERIMENTS

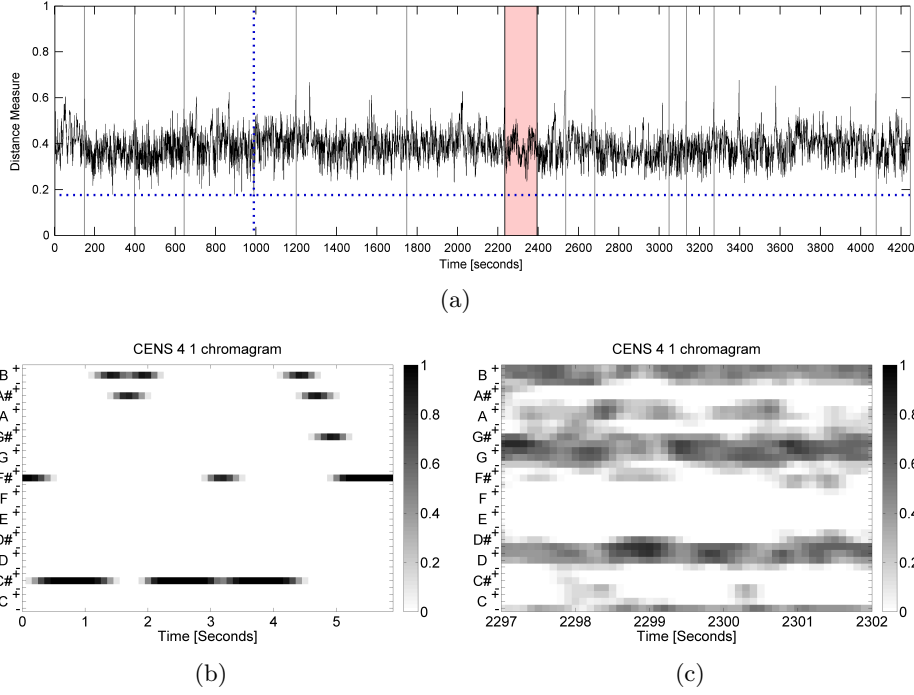


Figure 4.8: Query #16: Example for low specificity. Parameter setting: $\{\text{CENS}_1^4, \Sigma_1, \text{orig}, W_1, \kappa_2^{36}\}$. (a) Matching function. (b) Chromagram of the query. (c) Chromagram of the database corresponding to (b) (excerpt).

parameter setting is applied. This indicates that the features for the database are not sufficient and may be improved by filtering low frequency components.

4.3.2 Feature resolution

In this section we analyze different feature resolutions for Schubert's Symphony No.8, 1st Movement, 1st Theme (#25). In Figure 4.9 we can see similar matching functions corresponding to the query. Despite the chromagrams look quite different for different feature resolutions, this does not lead to great differences for the corresponding matching functions. So in this case, increasing the feature resolution has only a very small effect on the matching functions. Using a feature resolution of 1 Hz, the algorithm does not detect a match (see Figure 4.9(a)). However, when we use a feature resolution of 5 Hz, the algorithm detects the match (see Figure 4.9(b)). We conclude that increasing the feature resolution increases the specificity and thus improves the performance of the algorithm. On the downside, increasing the feature resolution also increases the computational cost. Furthermore, the curve shapes of both matching functions differ only slightly (see Figure 4.9(a) and Figure 4.9(b)). So increasing the feature resolution will only slightly improve the retrieval results.

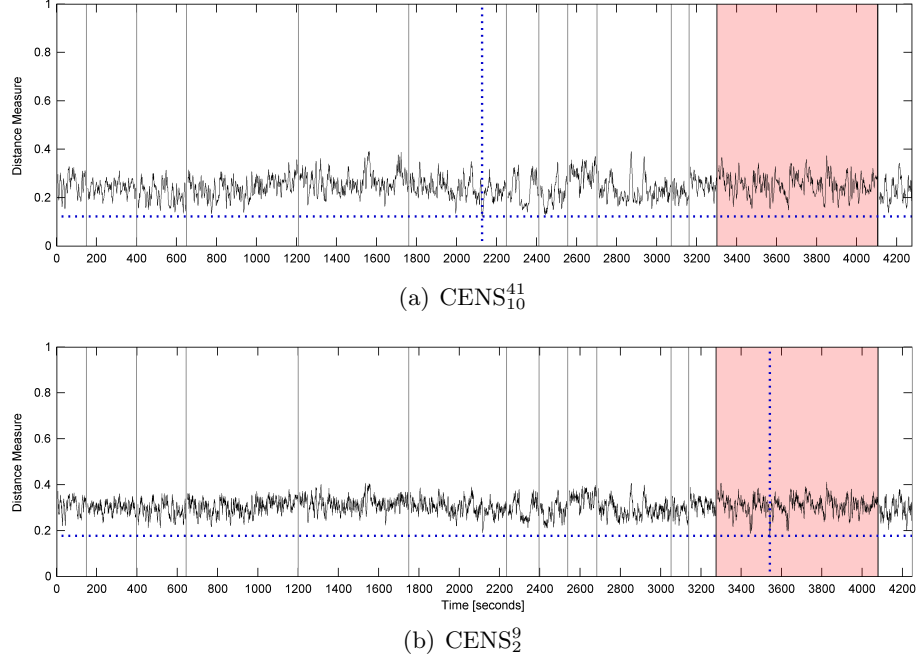


Figure 4.9: Query #25: Example for different feature resolutions. Parameter setting: $\{15 \text{ s}, \Sigma_2, W_3, \kappa_2^{12}\}$ (a) CENS₁₀⁴¹. (b) CENS₂⁹.

4.3.3 Tuning

In this section we analyze Brahms' Hungarian Dance No.5 (# 15). The audio recording is detuned by +1.3 semitones as we can see by comparing the chromagrams of query and database (see Figure 4.10(c) and 4.10(d)). So if we apply key shifts of ± 0 semitones, we do not retrieve a match (see Figure 4.10(a)). However if we apply key shifts of ± 1.3 semitones, we retrieve the matching sequence, see Figure 4.10(b) at time 2350 seconds in the database. Having a closer look at both matching functions, we see that the mean value of the matching function in Figure 4.10(a) is roughly 0.7. However, if we apply key shifts of ± 1.3 semitones, this value is roughly 0.4. Thus the gap between the minima for certain documents gets smaller. So this example shows that the specificity is greatly influenced by the number of key shifts.

4.3.4 Global and local tempo variations

In this section we analyze several parameters which define the tolerated local and global tempo variations between query and database. There are two parameters that have influence on tempo variations, these are the step size condition and the duration of the query. So we analyze two different EDM-themes, each with two different step sizes and two different query durations. First of all, we consider the matching function of Beethoven's Op.2 No.1, 1st Movement, 1st Theme (#3). Using original query duration, the matching function results in a Top Match for both, the step size condition Σ_2 and Σ_1 . The original duration of the matching excerpt in this audio recording is roughly 2.7 seconds, whereas the corresponding EDM-theme has a duration of 4 seconds, so both step sizes are convenient for this scenario. We now analyze the corresponding matching

4. EXPERIMENTS

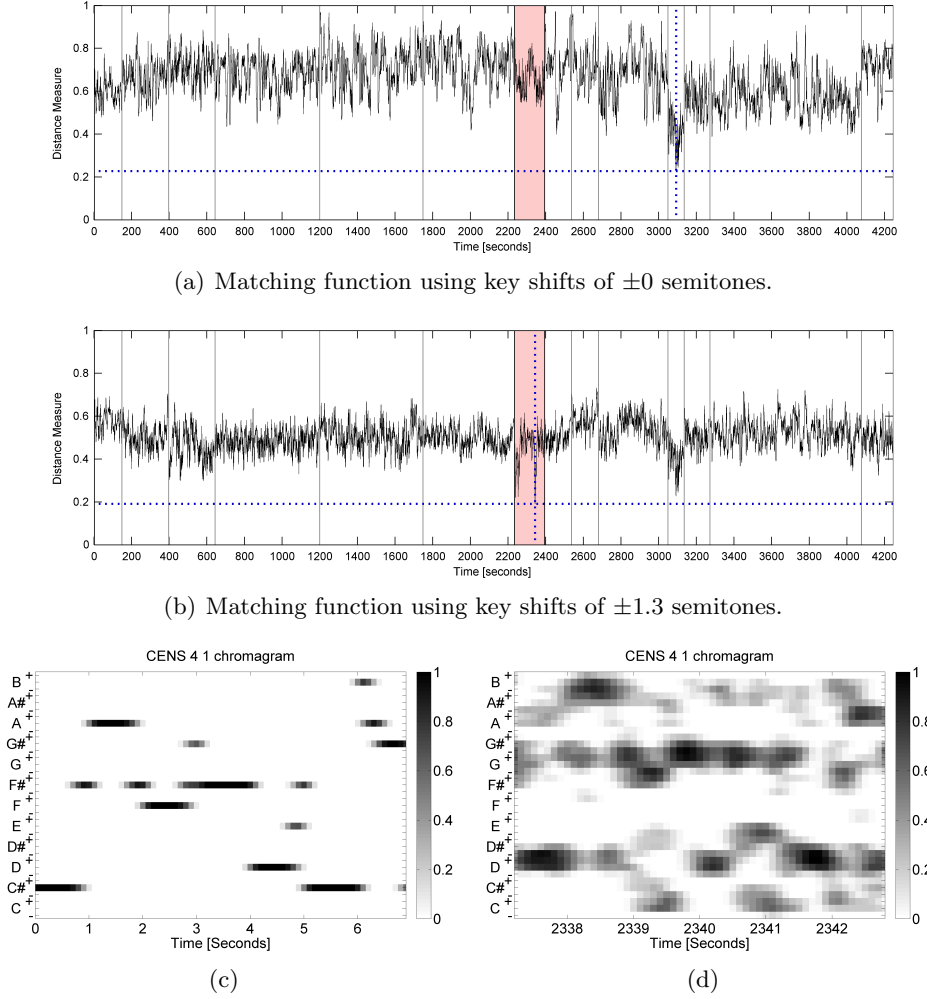


Figure 4.10: Query # 15: Example for a detuned audio recording. Parameter setting: $\{\text{CENS}_1^4, \Sigma_1, \text{orig}, W_1, \kappa^{36}\}$. (a) Matching function using key shifts of ± 0 semitones. (b) Matching function using key shifts of ± 1.3 semitones. (c) Chromagram of the query. (d) Chromagram of the database (excerpt).

function shown in Figure 4.11, where we use queries of a fixed duration of 10 seconds and the step size condition Σ_2 . In this scenario, we get a matching function containing no distinctive minima and no Top Match result, because of the high global tempo variations. These high tempo variations can be compensated with the step size condition Σ_1 . Then we get a Top Match. On the downside, the specificity decreases and thus this approach is not suitable for larger databases.

In the following, we analyze another EDM-theme which behaves quite different regarding the matching result for the specified parameter settings. Therefore we examine the matching function of Beethoven's Op.13, 1st Movement, 1st Theme (#7) for the step sizes Σ_1 and original query duration. The duration of the matching excerpt in the audio recording is roughly 4.4 seconds, the EDM-theme however has a duration of 12 seconds. So the matching procedure fails when we use the step size condition Σ_2 , as this parameter setting can compensate tempo variations of factor two at most (see Figure 4.12). However when we use the step size condition Σ_1 , we receive a Top

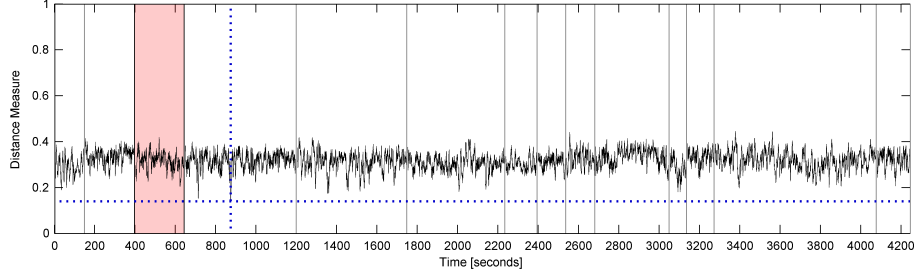


Figure 4.11: Query #3: Example for low tempo variations. Parameter setting : $\{\text{CENS}_1^4, 10 \text{ s}, \Sigma_2, W_1, \kappa_2^{12}\}$.

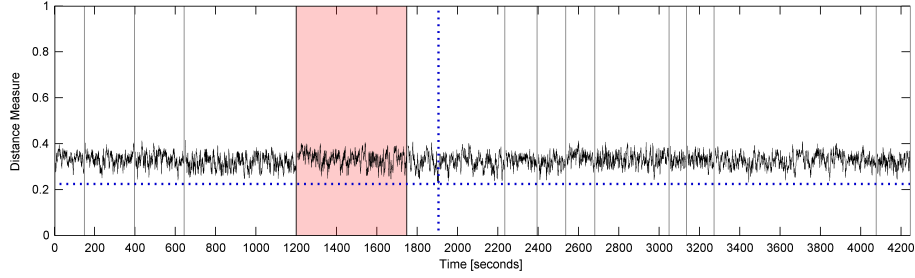


Figure 4.12: Query #7: Example for high tempo variations. Parameter setting : $\{\text{CENS}_1^4, \text{orig}, \Sigma_2, W_1, \kappa_2^{12}\}$.

Match. When we use a fixed query duration of 10 seconds, the matching procedure is successful for both step size conditions. So this is a border case for the step size condition Σ_2 as the Top Match is still identified despite the global tempo difference is slightly higher than a factor of two.

4.3.5 Degree of Polyphony

In the following, we show one example of a query, that corresponds to a solo piano audio recording, which is polyphonic. By comparing the chromagram of the query in Figure 4.13(c) to the chromagram of the database excerpt in Figure 4.13(d) we observe big deviations between both chromagrams. The chroma band G^\sharp of the database contains noticeably more energy. In order to understand this difference of both chromagrams, we examine the corresponding excerpt of the score and compare it with the BM-theme, see Figure 4.14(b), resp. Figure 4.14(a). The melody shown in the BM-theme is played by the right hand in the score. This melody is accompanied by several chords, especially by an A^b which is played by the left hand. This is the reason why the chroma band $G^{\sharp 1}$ has so much energy, see Figure 4.13(d). This example shows us that even a solo piano recording can have a high degree of polyphony and thus degrades the specificity and the retrieval result. The retrieval may be improved by filtering out low frequency coefficients. However we cannot assume that this improves the retrieval results in general, as there are also queries containing low pitches. The algorithm does not retrieve the right document in this scenario when we use the parameter setting $\{\text{CENS}_1^4, \Sigma_1, \text{orig}, W_1, \kappa_2^{36}\}$, see the matching function in Figure 4.13(a). However if we change the parameter which specifies the number of key

¹The notes A^b and G^\sharp are enharmonically equivalent notes.

4. EXPERIMENTS

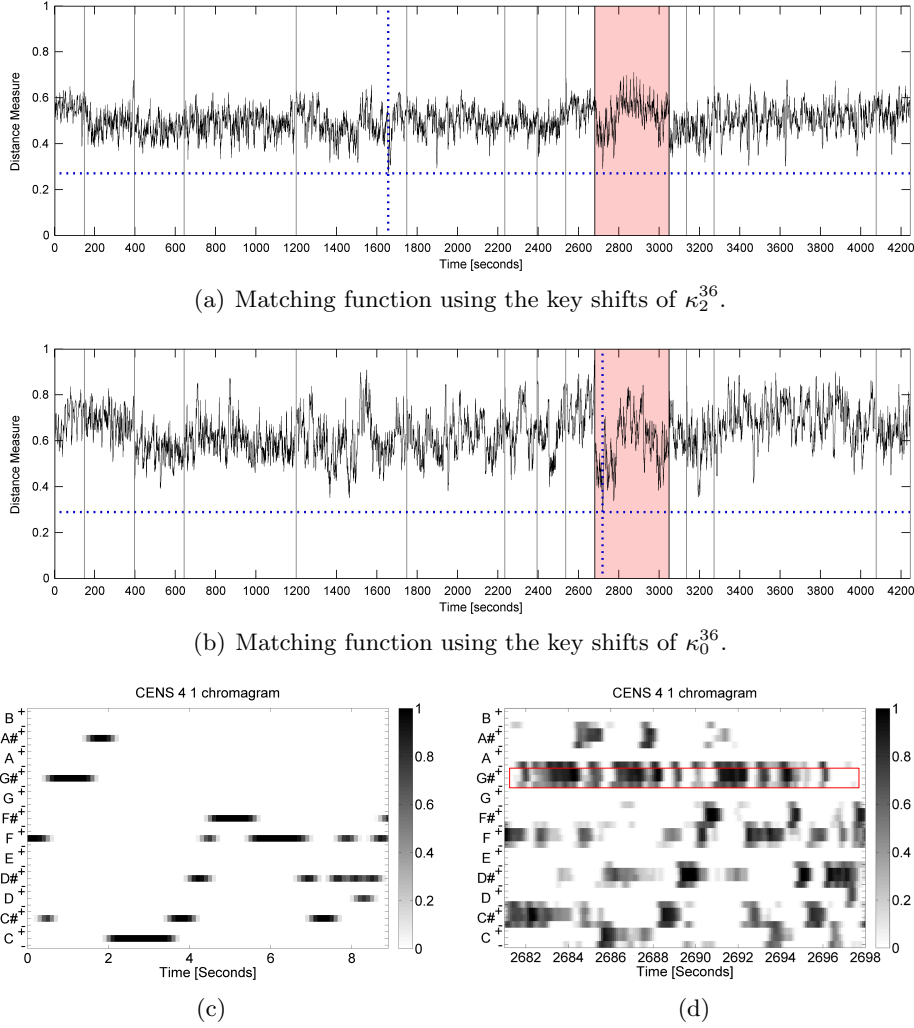


Figure 4.13: Query #19: Example for a polyphonic audio recording. Parameter setting: $\{\text{CENS}_1^4, \Sigma_1, \text{orig}, W_1\}$. (a) Matching function using the key shifts of κ_2^{36} . (b) Matching function using the key shifts of κ_0^{36} . (c) Chromagram of the query. (d) Chromagram of the database (excerpt).

shifts to κ_0^{36} , the algorithm finds the corresponding document in the database, see Figure 4.13(b). This however only works if the audio recording is neither transposed nor detuned. As a result, this scenario shows us that the performance of the algorithm is highly dependent on the specificity. The specificity strongly depends on the number of key shifts.

In the following, we shortly summarize the achievements of Chapter 4. Depending on the underlying datasets of Western classical music, we have to trade-off between several parameters. As the subset BM-Medium is a very large dataset, specificity is the most important aspect. Increasing the feature resolution leads to higher specificity. On the downside, this highly increases the computational cost. Using the step size condition Σ_2 , we also gain specificity, by allowing global tempo variations of a factor of two at most. On the downside, this leads to mismatches for several queries which have much higher global tempo variations. Therefore we stretch every query to a duration of 10 seconds to compensate these high global tempo variations. Increasing



(a)



(b)

Figure 4.14: Query #19: Example for a polyphonic piano score. (a) BM-theme. (b) Score excerpt [15].

the number of key shifts decreases specificity. So we apply key shifts of ± 0.67 semitones at most. Finally, polyphony also highly influences matching results of our algorithm.

Chapter 5

Conclusions

In this thesis, we modified an existing pipeline for a query-by-example task. We added several strategies to improve the retrieval results for the underlying data of Western classical music. Furthermore, we introduced an enhanced version of chroma features that takes tuning errors into account and thus performs better than standard chroma features. We tested our pipeline on a database of 100 audio recordings and tuned the parameters of our pipeline. Then we tested our pipeline on a large database of 1113 audio recordings using the previously tuned parameter setting. By comparing the results of both experiments, we observed that the results scale with respect to the size of the respective database. By analyzing the retrieval results and the correlation of several measures of the subset BM-Medium we gained further information about the underlying data. Furthermore we analyzed particular queries with respect to the properties specificity, tuning, feature resolution, global tempo variations and the degree of polyphony. To this end, we discussed the relation between these properties and the parameters of the pipeline.

The database and the pipeline which were created in this master thesis can be used in a content based search engine for Western classical music with the following properties: Given a query of 7 seconds in mean, the document we search is retrieved as Top Match in 29%. On average, it appears among the top 110 entries. The achievement of this search engine is as follows: Instead of listening through a database of 119 hours of music, we only need to listen to $6:24 \text{ min} \cdot 110 = 11:44$ hours of audio recordings in mean. If we further restrict to listening only to the retrieved sequences (we therefore assume that the algorithm always finds the right sequence within an audio recording) and not the entire recordings, this time further decreases to roughly 18 minutes of audio excerpts.

As future work one could perform experiments on the entire BM collection. Therefore, strategies will be needed to decrease the running time of our algorithm. This can be done by cutting audio recordings either to a fixed duration or to a specific percentage of the original length as a preprocessing step. Furthermore, the calculation time can be improved by adding a segmentation step to our pipeline, as introduced in [13, 20]. With this segmentation every audio recording can be analyzed with regard to repeating segments. Then all repetitions can be removed so that all the remainig segments are unique. So the idea is to reduce the size of the database, by removing “redundancy” of the audio recordings. Further future work can be done to improve the retrieval performance of our pipeline. Therefore, the degree of polyphony of the audio recordings can be reduced by a preprocessing step. This can be done by filtering out low frequency coefficients.

5. CONCLUSIONS

The degree of polyphony can also be reduced by using predominant F0 estimation for the audio recordings, as in [7, 8, 16, 17]. Finally, results of chroma based and fundamental frequency based matching procedures can be combined using a descriptor fusion strategy to further improve the performance, see [17].



Lukas Lamprecht obtained a B.Eng. in Electrical Engineering from Baden-Wuerttemberg Cooperative State University, Mosbach, Germany, in 2012. Currently, he is a final year student completing a Masters Degree in Electrical Engineering at Friedrich-Alexander Universität, Erlangen, Germany. He is doing his master thesis at the International Audio Laboratories Erlangen in the group for semantical audio processing headed by Prof. Dr. Meinard Müller. His research interests include multimedia signal processing and information retrieval tasks. His particular interests include music information retrieval and automated retrieval tasks for huge databases. He is also interested in the analysis of these tasks especially focusing on musical aspects.

Appendix A

Source Code

In this chapter, the headers of selected MATLAB functions created during the writing of this thesis are reproduced. The headers contain information about the name of the described function and its input/output behavior.

Feature Extraction

The `calc_pitch_features` function is used as a wrapper for several low-level functions that perform feature extraction and storage for audio and MIDI files.

Sample usage:

```
calc_pitch_features(db_wav,db_midi,parameter);
```

```
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% Name: calc_pitch_features
% Date: 2014
% Programmer: Lukas Lamprecht
%
% Description:
% Calculates and saves pitch features for all wav files and all midi files in
% specified folders, if they were not calculated so far.
%
% Input:
% db_wav: The first column of this cell array specifies the filenames for
%         the wav files to be processed.
% db_midi: The first column of this cell array specifies the filenames for
%          the midi files to be processed.
%
%          The path for the wav files is specified in
%          parameter.directory.abs.wav
%
%          The output path for the wav features is specified in
%          parameter.directory.abs.featurewav
%
%          The path for the midi files is specified in
%          parameter.directory.abs.midi
%
%          The output path for the midi features is specified in
%          parameter.directory.abs.featuremidi
%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
```

Evaluation Measures

The `eval_top_n` function calculates the Top N Match as described in Section 2.4.1.

Sample usage:

```
[ pr_match, ranks ] = eval_top_n( gt_vector, doc_matrix );

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% Name: eval_top_n
% Date: 2014
% Programmer: Lukas Lamprecht
%
% Description:
% Searches the values of a column vector gt_vector rowwise in all n
% columns of a matrix doc_matrix and returns the number of the column in the
% vector ranks, and the probability pr_match that the gt_vector entries
% are equal to one of the n columns in the matrix doc_matrix.
% The first dimension of both vectors must be equal.
%
% Input: gt_vector: Column vector. The nth entry contains the index of the
%                  groundtruth document, that corresponds to query n.
%          doc_matrix: First dimension is the number of documents.
%                     Second dimension is the number of queries.
%                     Each column contains the according documents that were
%                     found for this query in a sorted fashion.
%
% Output: pr_match: Top N Match evaluation measure.
%          ranks:   Column vector which contains the actual rank for every
%                  query.
%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
```

The `mean_rank` function calculates the mean rank as described in Section 2.4.2.

Sample usage:

```
[ m_rank ] = mean_rank(gt_vector, doc_matrix);

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% Name: mean_rank
% Date: 2014
% Programmer: Lukas Lamprecht
%
% Description:
% Calculates the mean rank for a given groundtruth vector gt_vector and a
% matrix doc_matrix. The n-th column of the matrix is a ranking list for
% the query n. It contains the indices of the most relevant documents in
% descending order.
%
% Input:  gt_vector: The gt_vector is a column vector of height m. It
%                   contains the index of the ground truth document for
%                   the m-th query.
%          doc_matrix: Matrix, that contains a ranking list for each
%                     query. Each column is a ranking list for one query.
%                     Every entry is the number of a document in the
%                     database.
%
% Output: m_rank: Mean rank evaluation measure.
%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
```

Creation of L^AT_EX-Code

The `cellArray2latex` function is a low-level function which generates L^AT_EX-code for tables provided as cell arrays. This function is called by the function `latex2file` which saves the L^AT_EX-code in a file.

Sample usage:

```
[ latex ] = cellArray2latex( cell_array, column_headers, entry, order );
```

```
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% Name: cellArray2latex
% Date: 2014
% Programmer: Lukas Lamprecht
%
% Description:
% Converts a cell array into a latex table, using the specified column headers.
%
% Input:  cell_array:    The cell array, that contains the data to be
%                        converted into a latex table.
%          column_headers: Headers of the columns, that will be shown in the
%                        latex table.
%          entry:        Specifies the name of the latex table. Has to be
%                        different for every table.
%          order:        Vector that specifies the order in which the columns
%                        are positioned.
%
% Output: latex:        Cell array, that contains the latex code.
%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
```

Sample usage:

```
latex2file( cell, headers, save_dir, filename, expr_rep );
```

```
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% Name: latex2file
% Date: 2014
% Programmer: Lukas Lamprecht
%
% Description:
% Converts and saves a cell array into a latex file, using the specified column
% headers.
%
% Input: cell:          Cell array with table entries
%          headers:      Table Headers provided in a cell array.
%          save_dir:      Latex file is saved in this directory.
%          filename:      Latex file is saved with the specified filename.
%          expr_rep:      Optional - Cell array of size (N,2). Here you can pass a
%                        cell array of regular expressions, that are replaced.
%                        The content of the first row is the expression that is
%                        searched, whereas the second row specifies the replaced
%                        expression.
%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
```


Appendix B

BM-Subsets

In the following we provide selected data, that was created during the writing of this thesis. The data includes matching functions for all queries of the subset **BM-Mini**, whereas we use the feature settings that lead to the best matching results. Furthermore we provide an overview of the retrieval results for the subset **BM-Mini** for several parameter settings. Moreover we provide one table containing all audio recordings and another table containing all musical themes for the subset **BM-Mini** and **BM-Small** respectively. For the subset **BM-Medium** we provide statistics on the composers. Finally we provide tables that document errors in the subsets, and also in EDM-themes.

B.1 BM-Mini Matching functions

In the following, we show the matching results by presenting the corresponding matching functions for the subset **BM-Mini**. Therefore we use the parameter setting: $\{\text{CENS}_1^4, \Sigma_1, \text{orig}, W_1, \kappa_2^{36}\}$. The matching functions shown in this section stick to the following scheme: The matching function is plotted in black color. Gray vertical lines denote the ending of one audio recording and the beginning of another audio recording. The audio recording which we expect our algorithm to retrieve is highlighted in light red color (ground truth document). The global minimum of the matching function is highlighted by a horizontal and a vertical blue dotted line. The vertical dotted line shows whether the global minimum of the matching function is in the ground truth document or not. The horizontal dotted line is meant to show the distance between the global minimum and the minimum of the ground truth document. It gives a feeling of the gap between the ground truth document and the retrieved document.

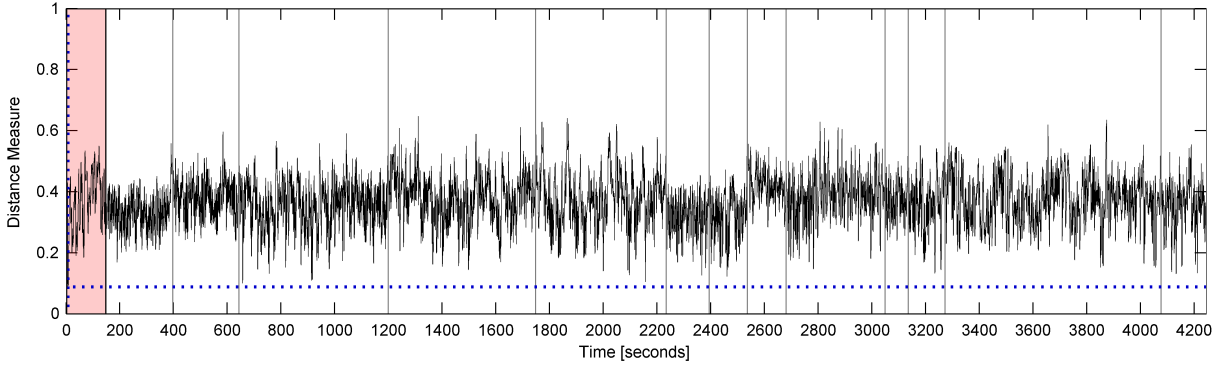


Figure B.1: Matching function for the query **Bach_BWV0846-01_EDM0394_BM0389** returns the document **Bach_BWV0846-01_Belder_BM0389-BM**.

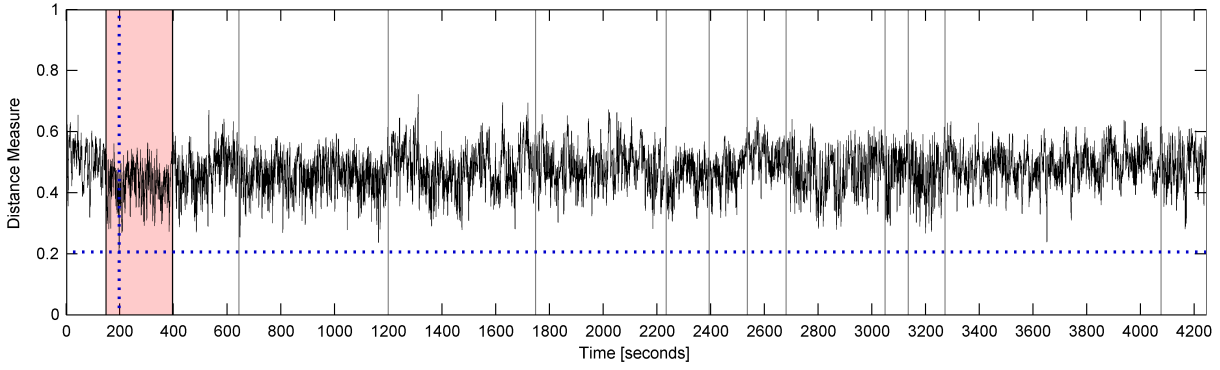


Figure B.2: Matching function for the query **Bach_BWV1041-01_EDM0174_BM0169** returns the document **Bach_BWV1041-01_Sitkovetsky_BM0169-BM**.

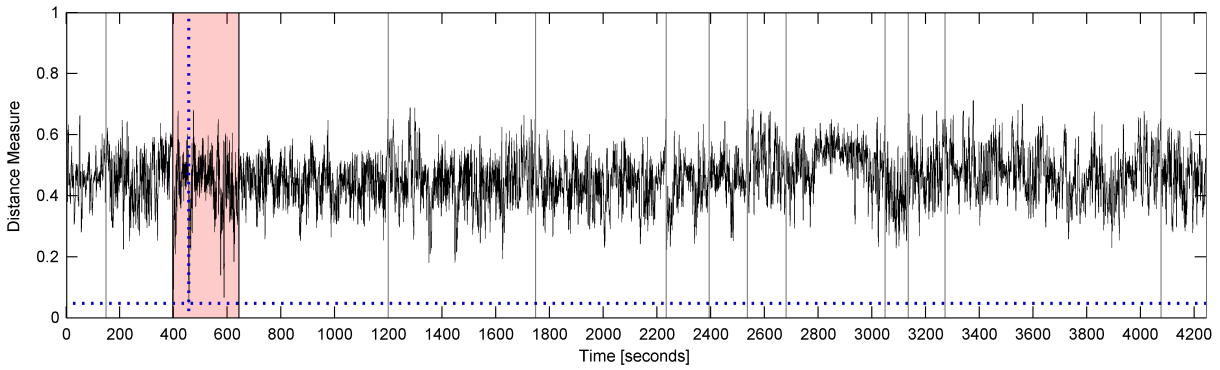


Figure B.3: Matching function for the query **Beethoven_Op002No1-01_EDM0813_BM0807** returns the document **Beethoven_Op002No1-01_Brendel_BM0807-BM**.

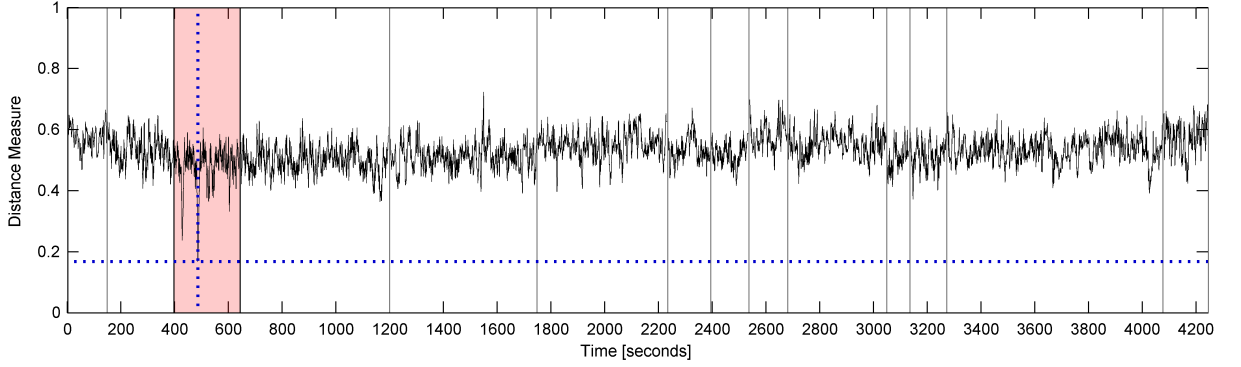


Figure B.4: Matching function for the query **Beethoven_Op002No1-01_EDM0814_BM0808** returns the document **Beethoven_Op002No1-01_Brendel_BM0807-BM**.

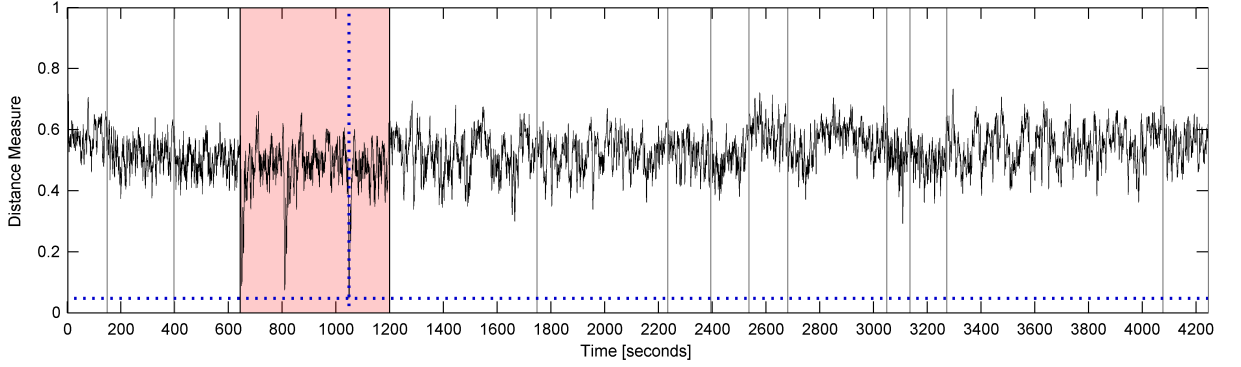


Figure B.5: Matching function for the query **Beethoven_Op011-01_EDM1156_BM1149** returns the document **Beethoven_Op011-01_Berkes_BM1149-BM**.

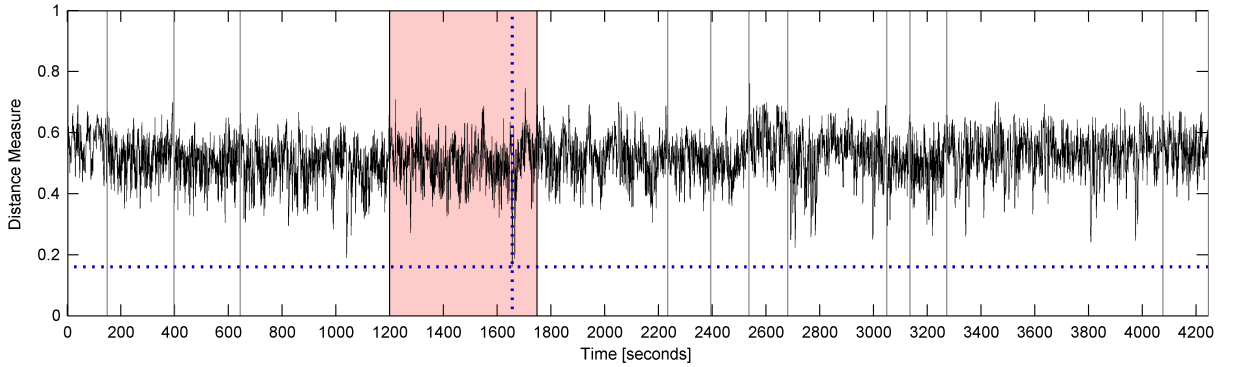


Figure B.6: Matching function for the query **Beethoven_Op013-01_EDM0852_BM0846** returns the document **Beethoven_Op013-01_Brendel_BM0846-BM**.

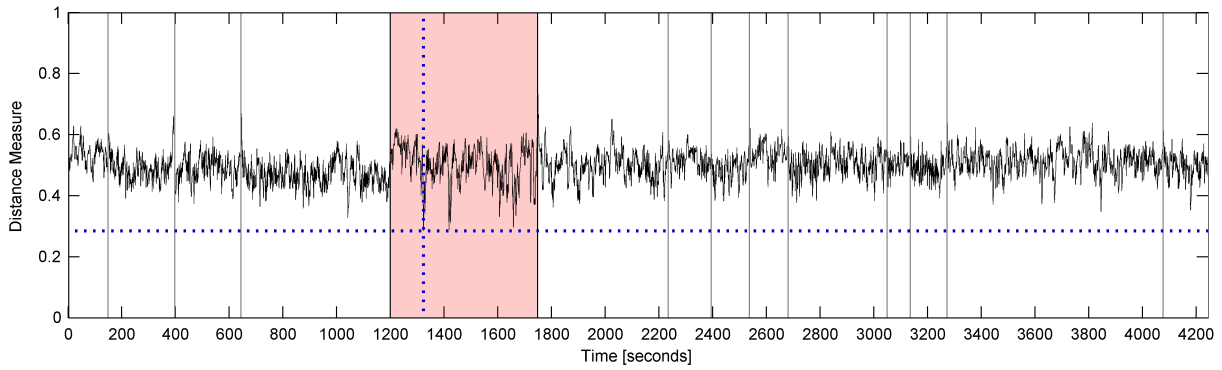


Figure B.7: Matching function for the query **Beethoven_Op013-01_EDM0853_BM0847** returns the document **Beethoven_Op013-01_Brendel_BM0846-BM**.

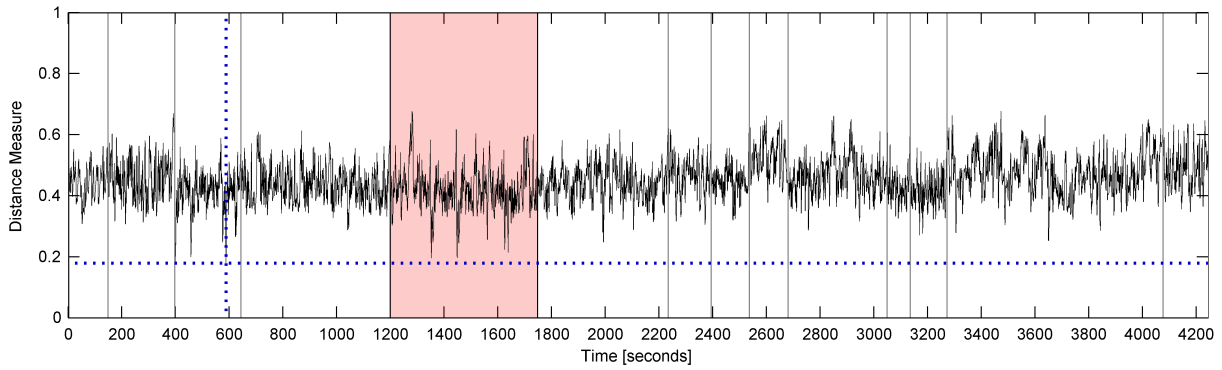


Figure B.8: Matching function for the query **Beethoven_Op013-01_EDM0854_BM0848** returns the document **Beethoven_Op002No1-01_Brendel_BM0807-BM**.

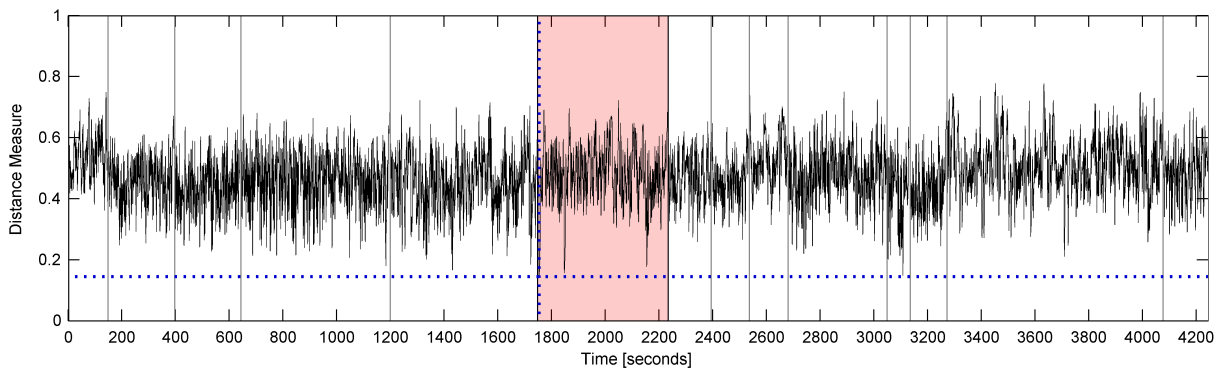


Figure B.9: Matching function for the query **Beethoven_Op067-01_EDM1072_BM1066** returns the document **Beethoven_Op067-01_Blomstedt_BM1066-BM**.

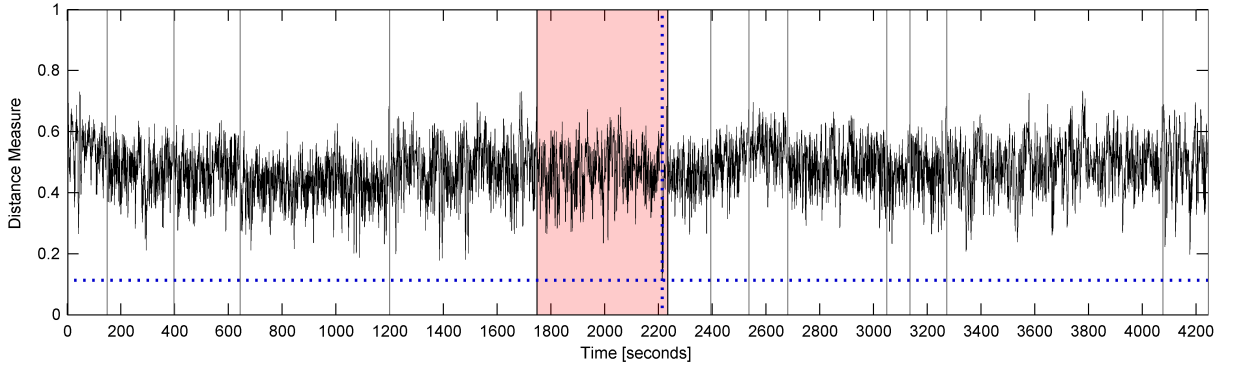


Figure B.10: Matching function for the query **Beethoven_Op067-01_EDM1073_BM1067** returns the document **Beethoven_Op067-01_Blomstedt_BM1066-BM**.

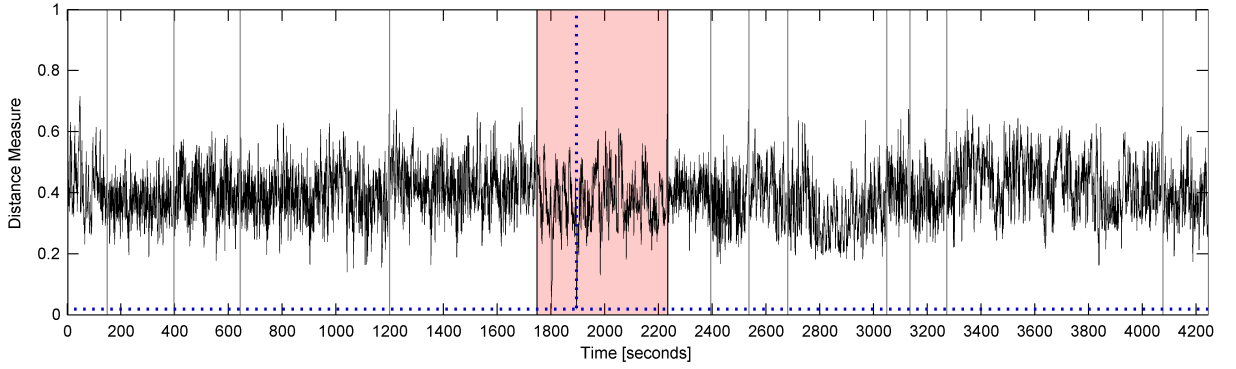


Figure B.11: Matching function for the query **Beethoven_Op067-01_EDM1074_BM1068** returns the document **Beethoven_Op067-01_Blomstedt_BM1066-BM**.

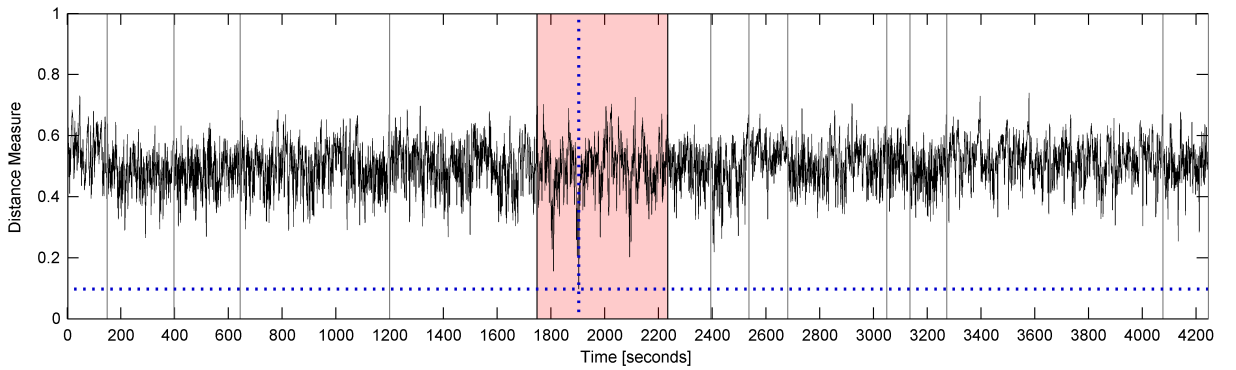


Figure B.12: Matching function for the query **Beethoven_Op067-01_EDM1075_BM1069** returns the document **Beethoven_Op067-01_Blomstedt_BM1066-BM**.

B. BM-SUBSETS

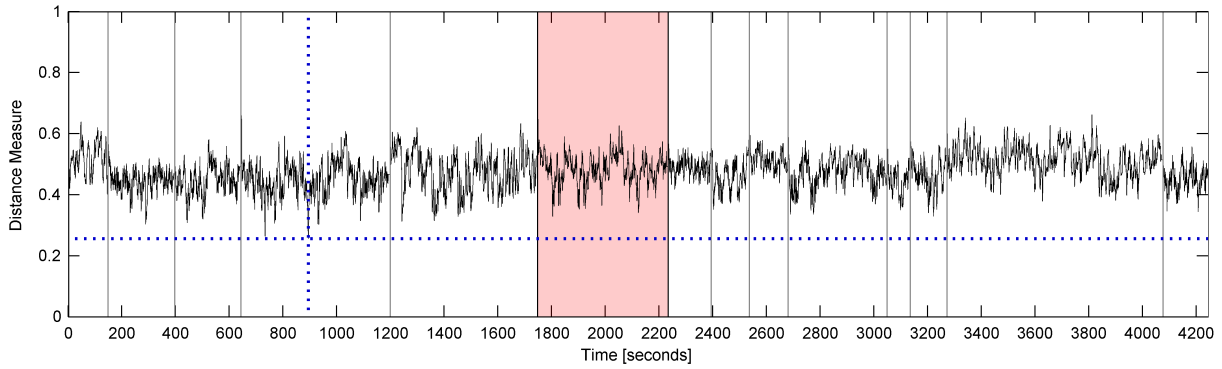


Figure B.13: Matching function for the query **Beethoven_Op067-01_EDM1076_BM1070** returns the document **Beethoven_Op011-01_Berkes_BM1149-BM**.

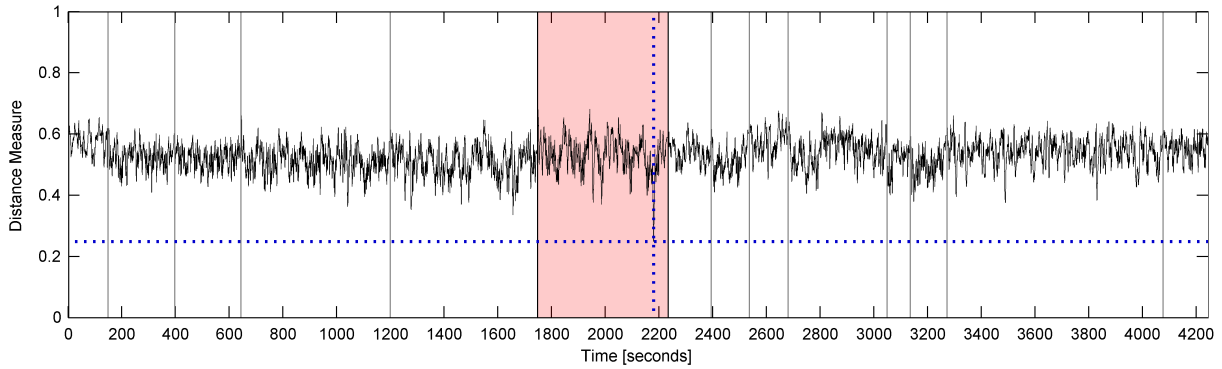


Figure B.14: Matching function for the query **Beethoven_Op067-01_EDM1077_BM1071** returns the document **Beethoven_Op067-01_Blomstedt_BM1066-BM**.

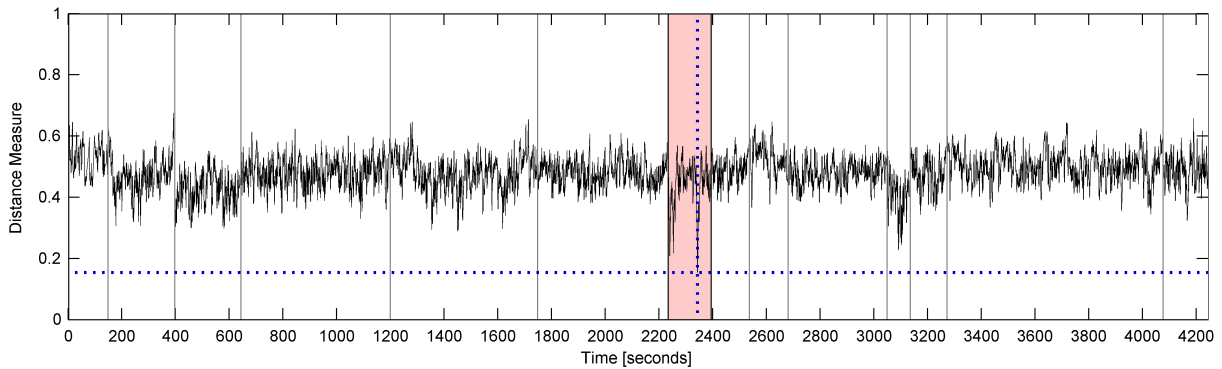


Figure B.15: Matching function for the query **Brahms_HungarianDances-05_EDM8508_BM1511** returns the document **Brahms_HungarianDances-05_SchmidtIsserstedt_BM1511-BM**.

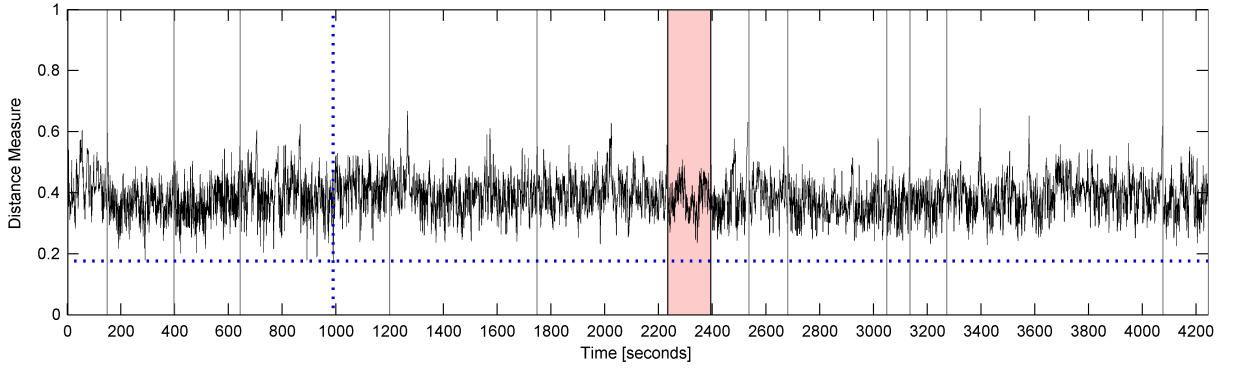


Figure B.16: Matching function for the query **Brahms_HungarianDances-05_EDM8509_BM1512** returns the document **Beethoven_Op011-01_Berkes_BM1149-BM**.

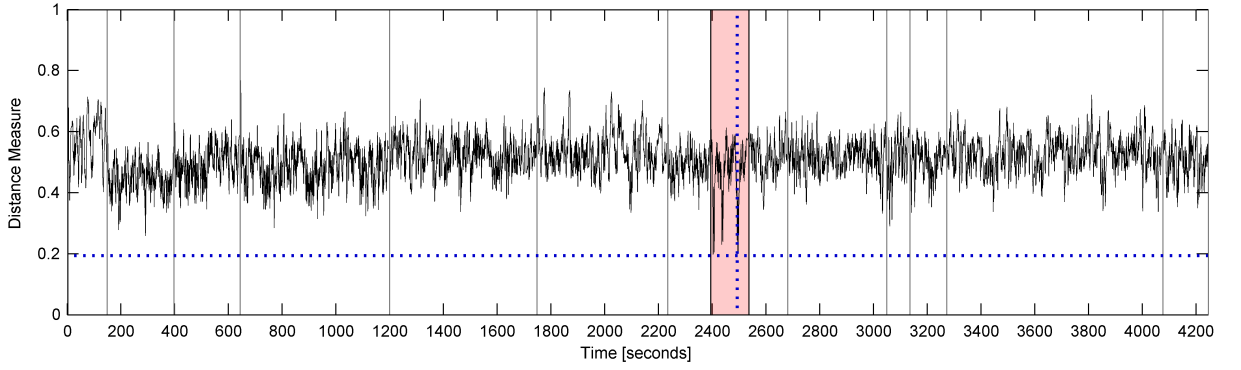


Figure B.17: Matching function for the query **Chopin_Op024-02_EDM9222_BM2219** returns the document **Chopin_Op024-02_Groot_BM2219-BM**.

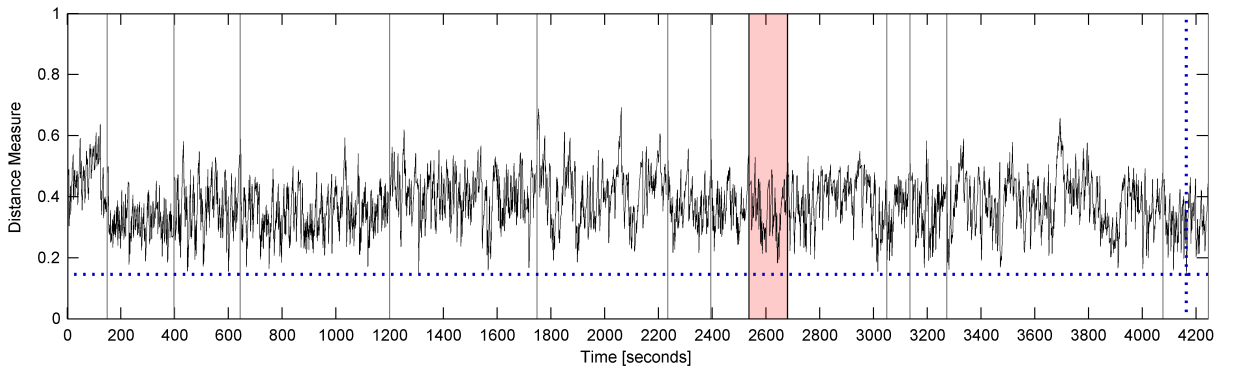


Figure B.18: Matching function for the query **Chopin_Op028-04_EDM9280_BM2276** returns the document **Schumann_Op015-07_Horowitz_BM7940-BM**.

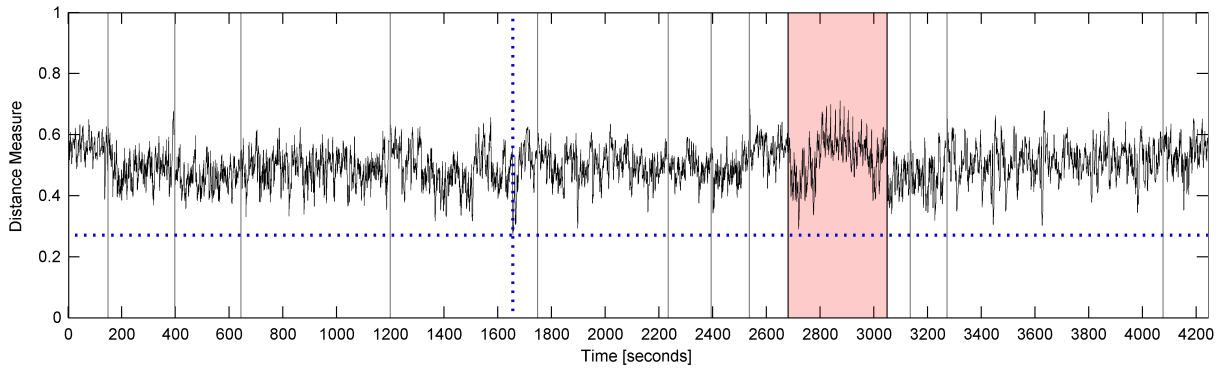


Figure B.19: Matching function for the query **Chopin_Op028-15_EDM9291_BM2287** returns the document **Beethoven_Op013-01_Brendel_BM0846-BM**.

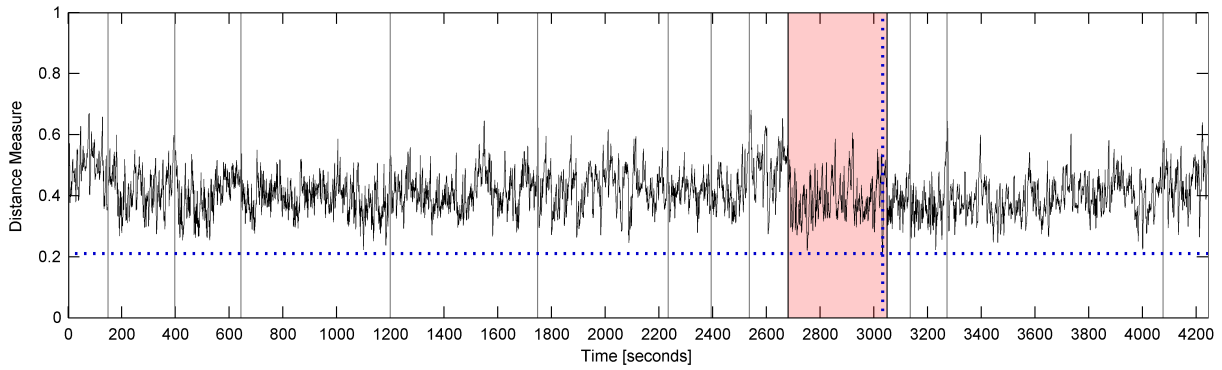


Figure B.20: Matching function for the query **Chopin_Op028-15_EDM9292_BM2288** returns the document **Chopin_Op028-15_Davidovich_BM2287-BM**.

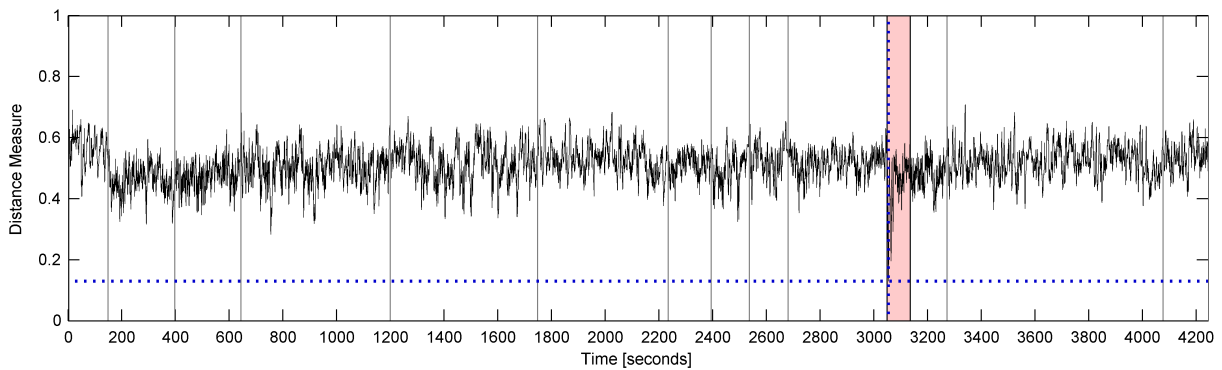


Figure B.21: Matching function for the query **Chopin_Op030-02_EDM9224_BM2221** returns the document **Chopin_Op030-02_Groot_BM2221-BM**.

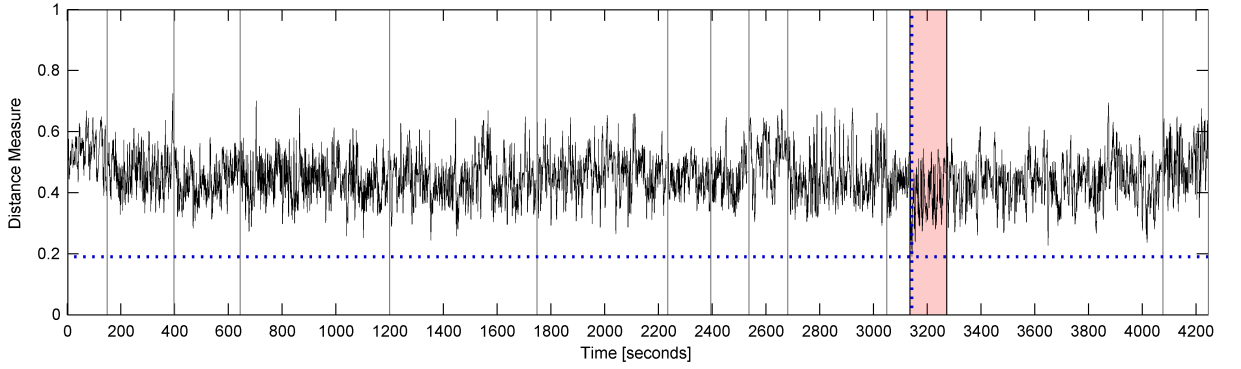


Figure B.22: Matching function for the query **Chopin_Op063-03_EDM9239_BM2236** returns the document **Chopin_Op063-03_Groot_BM2236-BM**.

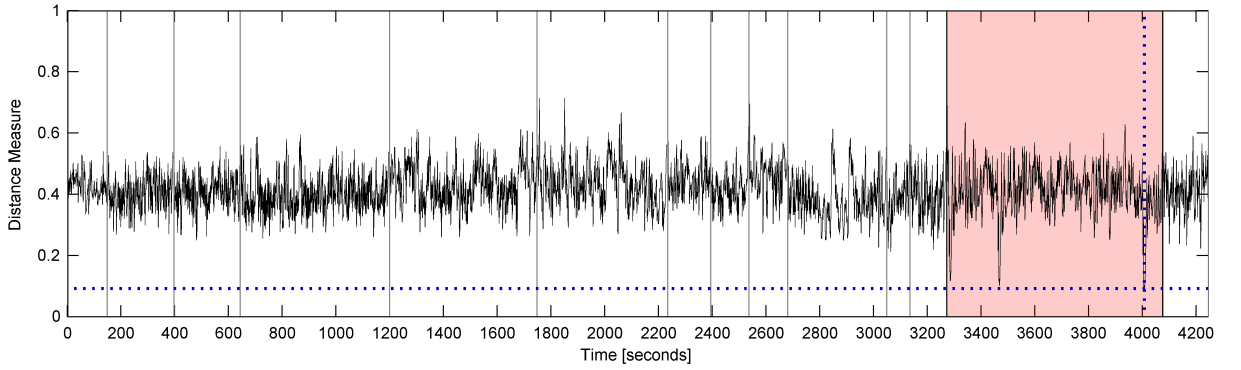


Figure B.23: Matching function for the query **Schubert_D0759-01_EDM6120_BM7752** returns the document **Schubert_D0759-01_Goodman_BM7752-BM**.

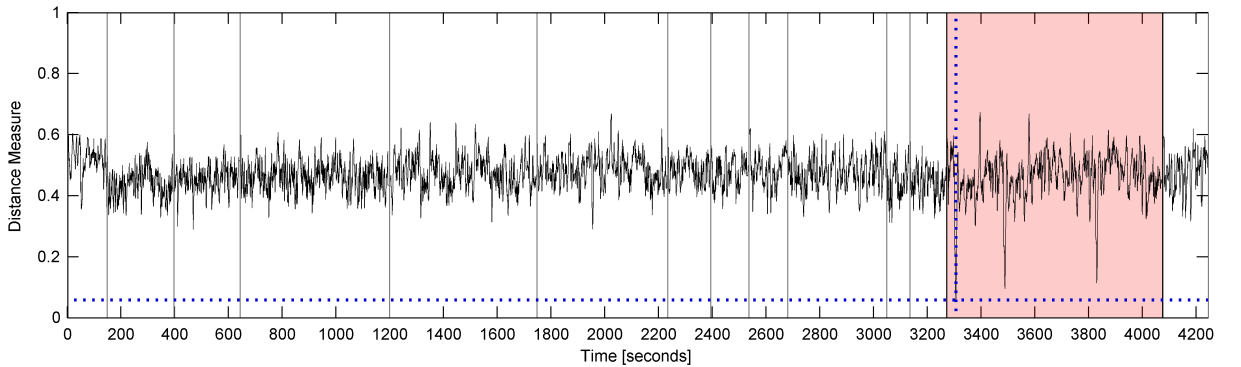


Figure B.24: Matching function for the query **Schubert_D0759-01_EDM6121_BM7753** returns the document **Schubert_D0759-01_Goodman_BM7752-BM**.

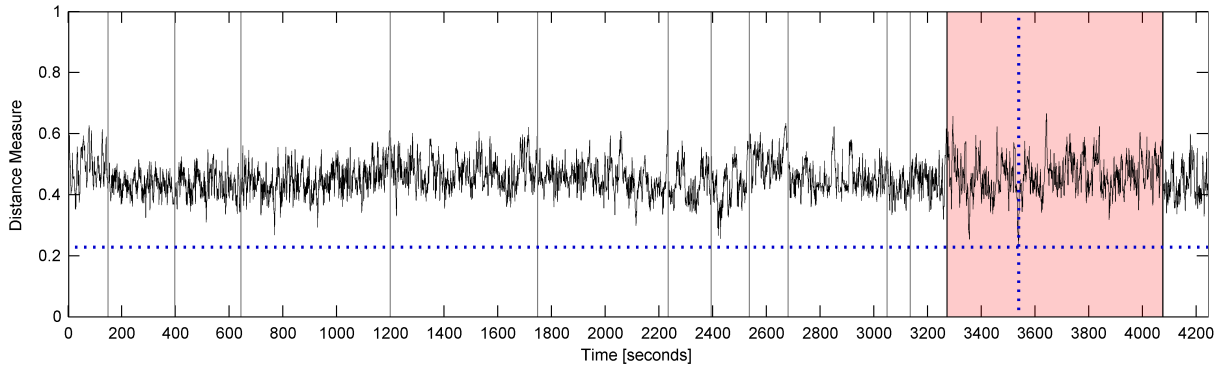


Figure B.25: Matching function for the query **Schubert_D0759-01_EDM6122_BM7754** returns the document **Schubert_D0759-01_Goodman_BM7752-BM**.

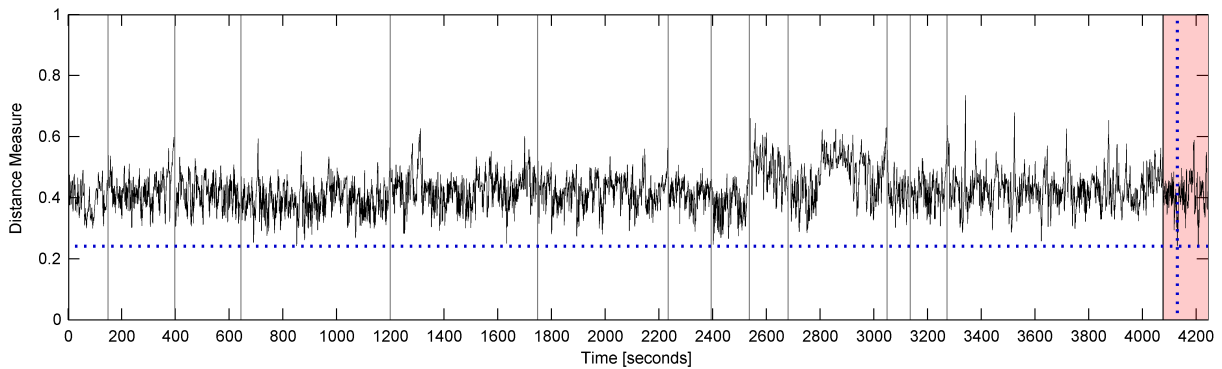


Figure B.26: Matching function for the query **Schumann_Op015-07_EDM6308_BM7940** returns the document **Schumann_Op015-07_Horowitz_BM7940-BM**.

B.2 BM-Mini ranking results

This Appendix provides figures that give an overview of the retrieval results of the subset **BM-Mini** for several parameter settings. For every query of the subset, the rank on which the ground truth document appears is color coded. White spots indicate a Top Match, whereas dark spots indicate a bad matching result. With these figures many different parameter settings can be compared directly.

B. BM-SUBSETS

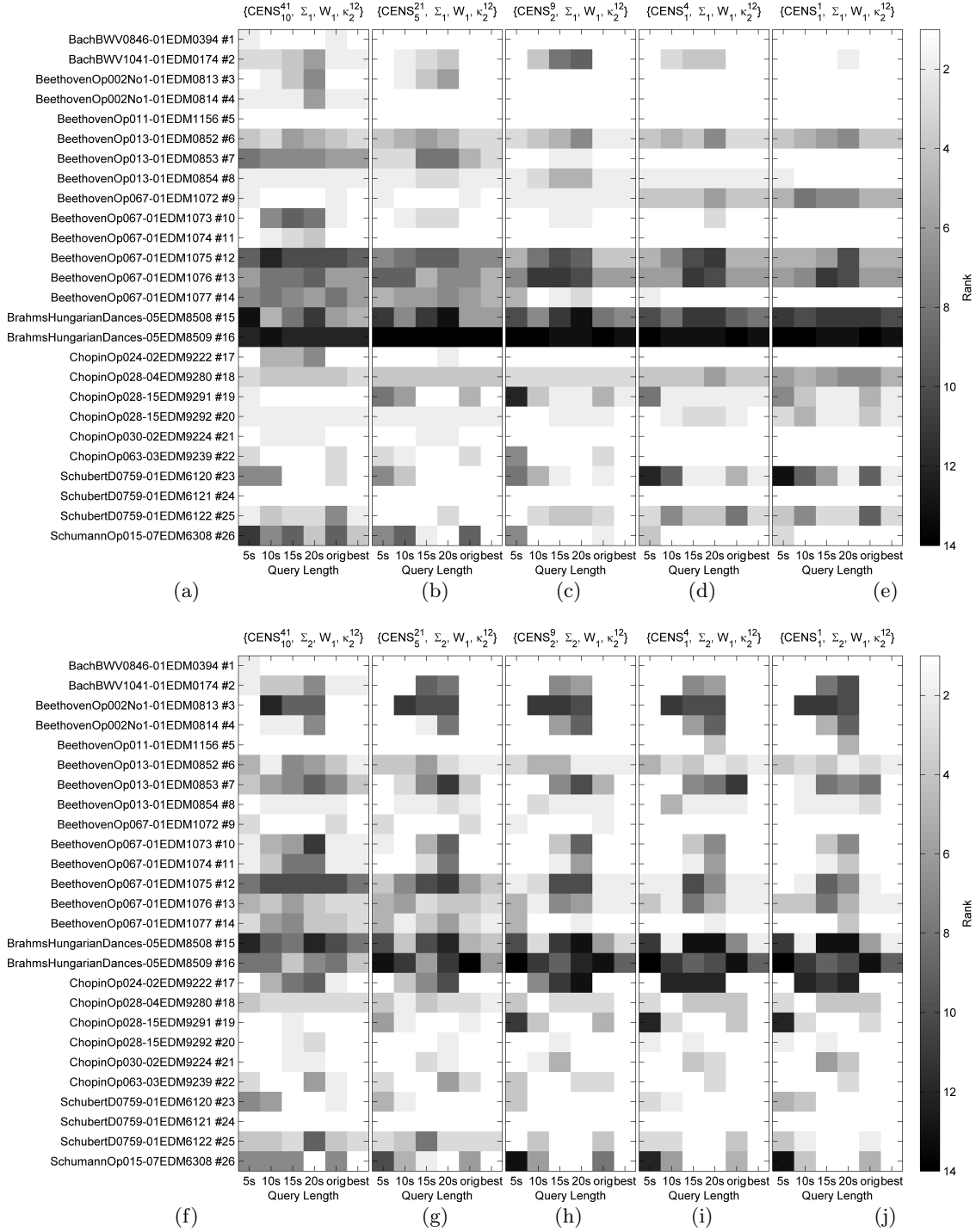
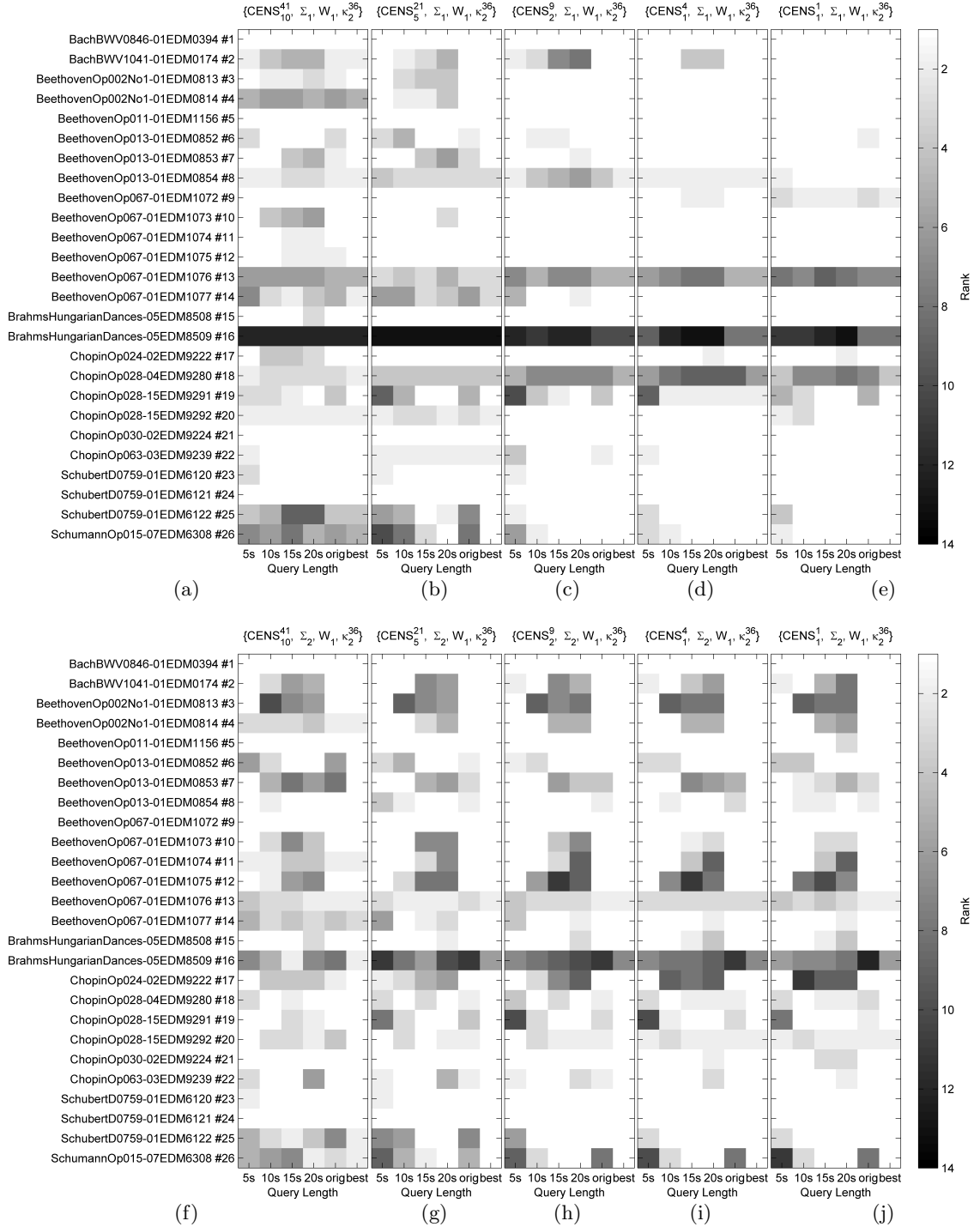


Figure B.27: Ranking results for subset BM-Small using the parameter setting $\{\kappa_2^{12}\}$.

Figure B.28: Ranking results for subset BM-Small using the parameter setting $\{\kappa_2^{36}\}$.

B.3 Ensemble and Instrumentation IDs

EnsembleID	Number of instruments
Solo	1
Duo	2
Trio	3
Quartet	4
Quintet	5
Sextet	6
Septet	7
Octet	8
Nonet	9
Decet	10
Concerto	solo instrument and orchestra
Orchestra	orchestra

InstrumentID	Definition
altorecorder	Altorecorder
bassoon	Bassoon
cello	Cello
clarinet	Clarinet
doublebass	Doublebass
englishhorn	Englishhorn
flute	Flute
guitar	Guitar
harp	Harp
harpsichord	Harpsichord
horn	Horn
oboe	Oboe
orchestra	Orchestra
organ	Organ
piano	Piano
piccolotrumpet	Piccolotrumpet
strings	Strings or string chamber orchestra
timpani	Timpani
trumpet	Trumpet
viola	Viola
violagamba	Viola da Gamba
violin	Violin
windsection	Windsection
xylophone	Xylophone

B.4 Audio Recordings for BM-Mini

Composer	WorkID	Folder	CD	Tr	Name	Ensemble	Instrumentation	Duration
Bach	BWV0846-01	Bach Complete Edition BC	24	1	Prelude	Solo	Harpsichord	02:28
Bach	BWV1041-01	Bach Complete Edition BC	5	1	Violin Concert in A minor	Concerto	Violin Harpsichord Strings	04:08
Beethoven	Op002No1-01	Beethoven Complete Edition BC	49	1	Sonata no. 1	Solo	Piano	04:06
Beethoven	Op011-01	Beethoven Complete Edition BC	20	4	Gassenhauer-Trio	Trio	Clarinet Cello Piano	09:16
Beethoven	Op013-01	Beethoven Complete Edition BC	52	1	Pathétique	Solo	Piano	09:08
Beethoven	Op067-01	Beethoven Complete Edition BC	3	1	Symphony No. 5	Orchestra	Orchestra	08:05
Brahms	Hungarian Dances-05	Various MasterpiecesClassicalMusic MEMBRAN	15	8		Orchestra	Orchestra	02:40
Chopin	Op024-02	Chopin Complete Works BC	7	15	Mazurka	Solo	Piano	02:22
Chopin	Op028-04	Chopin Complete Works BC	16	4	Prelude	Solo	Piano	02:23
Chopin	Op028-15	Chopin Complete Works BC	16	15	Prelude Raindrop	Solo	Piano	06:09
Chopin	Op030-02	Chopin Complete Works BC	7	19	Mazurka	Solo	Piano	01:26
Chopin	Op063-03	Chopin Complete Works BC	8	18	Mazurka	Solo	Piano	02:16
Schubert	D0759-01	Schubert TheMasterworks BC	2	1	Unfinished Symphony	Orchestra	Orchestra	13:24
Schumann	Op015-07	Schumann MasterworksEdition SONY	5	7	Träumerei	Solo	Piano	02:50

B.5 EDM-MIDIs of BM-Mini

IDX	Composer	WorkID	BMID	ThemeID	MidiNo	Notes	Duration [s]
1	Bach	BWV0846-01	B301		394	32	4
2	Bach	BWV1041-01	B83		174	15	3.7
3	Beethoven	Op002No1-01	B689	1st Theme	813	10	4
4	Beethoven	Op002No1-01	B690	2nd Theme	814	16	8
5	Beethoven	Op011-01	B1031		1157	12	5.5
6	Beethoven	Op013-01	B728	Intro	852	12	3.7
7	Beethoven	Op013-01	B729	1st Theme	853	17	12
8	Beethoven	Op013-01	B730	2nd Theme	854	14	8.5
9	Beethoven	Op067-01	B948	1st Theme A	1072	8	5
10	Beethoven	Op067-01	B949	1st Theme B	1073	10	4.2
11	Beethoven	Op067-01	B950	1st Theme C	1074	6	5
12	Beethoven	Op067-01	B951	2nd Theme	1075	8	4
13	Beethoven	Op067-01	B952	3rd Theme	1076	25	7.2
14	Beethoven	Op067-01	B953	4th Theme	1077	17	9
15	Brahms	Hungarian Dances-05	B1375	1st Theme	8508	18	7
16	Brahms	Hungarian Dances-05	B1376	2nd Theme	8509	20	6
17	Chopin	Op024-02	C232		9222	24	6
18	Chopin	Op028-04	C289		9280	18	16
19	Chopin	Op028-15	C300	1st Theme	9291	21	9
20	Chopin	Op028-15	C301	2nd Theme	9292	19	11.5
21	Chopin	Op030-02	C234		9224	18	6
22	Chopin	Op063-03	C249		9239	13	6
23	Schubert	D0759-01	S533	Intro	6120	10	12
24	Schubert	D0759-01	S534	1st Theme	6121	17	11
25	Schubert	D0759-01	S535	2nd Theme	6122	21	8.7
26	Schumann	Op015-07	S713	Träumerei	6308	21	8

B.6 Audio Recordings for BM-Small

Composer	WorkID	Folder	CD	Tr	Name	Ensemble	Instrumentation	Duration
Bach	BMW1048-01	Bach Complete Edition BC	1	8	Brandenburg Concerto no. 3	Concerto	Violin Viola Cello Harpsichord Strings	05:19
Bach	BWV0543-01	Bach Complete Edition BC	157	16	Prelude	Solo	Organ	03:29
Bach	BWV0565-01	Bach Complete Edition BC	152	1	Toccatà	Solo	Organ	02:49
Bach	BWV0565-02	Bach Complete Edition BC	152	2	Fugue	Solo	Organ	06:17
Bach	BWV0808-01	Bach Complete Edition BC	35	20	English Suite No 3	Solo	Harpsichord	03:22
Bach	BWV0846-01	Bach Complete Edition BC	24	1	Prelude	Solo	Harpsichord	02:28
Bach	BWV0846-02	Bach Complete Edition BC	24	2	fugue	Solo	Harpsichord	01:59
Bach	BWV0847-01	Bach Complete Edition BC	24	3	WTC: prelude	Solo	Harpsichord	01:28
Bach	BWV0847-02	Bach Complete Edition BC	24	4	fugue	Solo	Harpsichord	01:31
Bach	BWV1001-01	Bach Complete Edition BC	10	1	Sonata: Adagio	Solo	Violin	04:56
Bach	BWV1002-03	Bach Complete Edition BC	10	9	Partita: Sarabande	Solo	Violin	02:53
Bach	BWV1009-05	Bach Complete Edition BC	12	11	Cello Suite	Solo	Cello	03:44
Bach	BWV1030-01	Bach Complete Edition BC	14	16	Sonata	Duo	Flute Harpsichord	07:19
Bach	BWV1041-01	Bach Complete Edition BC	5	1	Violin Concert in A minor	Concerto	Violin Harpsichord Strings	04:08
Bach	BWV1046-01	Bach Complete Edition BC	1	1	Brandenburg Concerto no 1	Concerto	Oboe Violin Horn Harpsichord Strings	03:54
Bach	BWV1065-01	Bach Complete Edition BC	7	13	Concerto	Concerto	Harpsichord Strings	04:00
Bartok	Sz112-01	Bartok OrchestralWorks EMI	3	1	Concerto	Concerto	Violin Orchestra	16:09
Beethoven	Op002No1-01	Beethoven Complete Edition BC	49	1	Sonata no. 1	Solo	Piano	04:06
Beethoven	Op011-01	Beethoven Complete Edition BC	20	4	Gassenhauer-Trio	Trio	Clarinet Cello Piano	09:16
Beethoven	Op013-01	Beethoven Complete Edition BC	52	1	Pathétique	Solo	Piano	09:08
Beethoven	Op017-01	Beethoven Complete Edition BC	20	7		Duo	Horn Piano	08:06
Beethoven	Op018No4-01	Various MasterworksHeritage SONY	15	1		Quartet	Violin Viola Cello	06:32
Beethoven	Op027No2-01	Beethoven Complete Edition BC	50	8	Moonlight Sonata	Solo	Piano	06:02
Beethoven	Op067-01	Beethoven Complete Edition BC	3	1	Symphony No. 5	Orchestra	Orchestra	08:05
Beethoven	WoO059	Beethoven Complete Edition BC	56	8	Für Elise	Solo	Piano	02:48
Brahms	Hungarian Dances-05	Various MasterpiecesClassicalMusic MEMBRAN	15	8		Orchestra	Orchestra	02:40
Brahms	Op015-01	Various MasterworksHeritage SONY	23	1		Concerto	Piano Orchestra	21:20
Brahms	Op025-04	Brahms Complete Works BC	10	4	Rondo alla Zingarese	Quartet	Piano Violin Viola Cello	08:09
Britten	Op002	Britten TheCollectorsEdition EMI	10	1	Phantasy-Quartet	Quartet	Oboe Violin Viola Cello	13:18
Chopin	Op009-02	Chopin Complete Works BC	3	2		Solo	Piano	04:56
Chopin	Op010-12	Chopin Complete Works BC	2	12	Etude Revolutionary	Solo	Piano	02:49
Chopin	Op024-02	Chopin Complete Works BC	7	15	Mazurka	Solo	Piano	02:22

B.6 AUDIO RECORDINGS FOR BM-SMALL

Composer	WorkID	Folder	CD	Tr	Name	Ensemble	Instrumentation	Duration
Chopin	Op028-04	Chopin Complete Works BC	16	4	Prelude	Solo	Piano	02:23
Chopin	Op028-15	Chopin Complete Works BC	16	15	Prelude Raindrop	Solo	Piano	06:09
Chopin	Op030-02	Chopin Complete Works BC	7	19	Mazurka	Solo	Piano	01:26
Chopin	Op063-03	Chopin Complete Works BC	8	18	Mazurka	Solo	Piano	02:16
Chopin	Op066	Chopin Complete Works BC	1	8	Fantasie-Improptu	Solo	Piano	05:05
Debussy	L075-03	Debussy TheDebussyEdition DG	8	3	Suite Bergamasque: Clair de Lune	Solo	Piano	05:52
Debussy	L095-01	Debussy TheDebussyEdition DG	8	8	Pour le Piano	Solo	Piano	04:12
Debussy	L103-01	Debussy TheDebussyEdition DG	3	3		Concerto	Harp Strings	04:38
Dukas	ApprentiSorcier	Various MasterpiecesClassicalMusic MEMBRAN	51	2	Zauberlehrling	Orchestra	Orchestra	09:26
Dvorak	B078-08	Dvorak Complete Symphonies BC	9	8	Slav. Dances	Orchestra	Orchestra	04:08
Dvorak	B178-01	Various MasterworksHeritage SONY	22	5	Aus der neuen Welt	Orchestra	Orchestra	08:41
Dvorak	B178-04	Various MasterworksHeritage SONY	22	8		Orchestra	Orchestra	10:55
Gershwin	AmericanParis	Gershwin GeorgeGershwin TIM	2	6		Orchestra	Orchestra	15:48
Granados	Op037-02	Bream Complete RCACollection SONY	29	21	Spanish Dance	Solo	Guitar	05:02
Grieg	Op036-01	Grieg Edition BC	7	1		Duo	Cello Piano	09:46
Handel	HWV-368a-01	Handel Portrait BC	15	1	Sonata	Quartet	Flute Violin Cello Harpsichord	10:54
Handel	HWV287-01	Handel Portrait BC	2	6	Concerto	Concerto	Oboe Strings Harpsichord	02:37
Handel	HWV289-01	Handel Portrait BC	8	1	Concerto	Concerto	Organ Strings Windsection Harpsichord	05:14
Handel	HWV361-01	Handel Portrait BC	14	1		Trio	Violin Cello Harpsichord	07:20
Handel	HWV365-01	Handel Portrait BC	18	1		Duo	Altorecorder Harpsichord	02:35
Haydn	Hob03No006-01	Haydn Edition BC	93	6	Quartet	Quartet	Violin Viola Cello	01:54
Haydn	Hob03No031-01	Haydn Edition BC	91	9	Streichquartett	Quartet	Violin Viola Cello	06:25
Haydn	Hob07bNo002-01	Haydn Edition BC	39	4	Concerto	Concerto	Cello Strings Windsection Harpsichord	13:30
Haydn	Hob15No027-01	Haydn Edition BC	111	1	Trio	Trio	Piano Violin Cello	07:51
Khachaturian	ConcertoViolin DMinor-01	Khachaturian ComposerConductorPianist SUP	1	1	Concerto	Concerto	Violin Orchestra	13:08
Kodaly	GalantaDances	Kodaly OrchestralWorks BC	2	3		Orchestra	Orchestra	15:34
Liszt	S541-03	Liszt TheComplete PianoMusic HYP	24	3	Liebstraum	Solo	Piano	04:14
Mendelssohn	Op030-03	Mendelssohn Portrait BC	22	9	Lied ohne Worte	Solo	Piano	02:39
Mozart	KV145	Mozart Complete Edition BC	68	5		Concerto	Violin Strings	02:55
Mozart	KV186-02	Mozart Complete Edition BC	43	7		Decet	Clarinet Oboe Englishhorn Bassoon Horn	02:05
Mozart	KV188-01	Mozart Complete Edition BC	45	21		Septet	Trumpet Piccolotrumpet Timpani	01:46
Mozart	KV219-01	Mozart Complete Edition BC	28	4	turkish	Concerto	Violin Strings Windsection	09:37

B. BM-SUBSETS

Composer	WorkID	Folder	CD	Tr	Name	Ensemble	Instrumentation	Duration
Mozart	KV298-01	Mozart Complete Edition BC	58	8		Quartet	Flute Violin Viola Cello	05:56
Mozart	KV315	Mozart Complete Edition BC	24	4	Andante	Concerto	Flute Strings Windsection	06:26
Mozart	KV334-01	Mozart Complete Edition BC	30	10		Concerto	Violin Strings Windsection	06:38
Mozart	KV361-01	Mozart Complete Edition BC	46	1		Orchestra	Windsection	09:58
Mozart	KV370-01	Mozart Complete Edition BC	53	4		Quintet	Oboe Violin Viola Cello	07:17
Mozart	KV378-01	Mozart Complete Edition BC	66	1		Duo	Violin Piano	12:34
Mozart	KV447-01	Mozart Complete Edition BC	26	4		Concerto	Horn Strings Windsection Harpsichord	06:40
Mozart	KV452-01	Mozart Complete Edition BC	54	1		Quintet	Piano Clarinet Oboe Horn Bassoon	09:29
Mozart	KV550-01	Mozart Complete Edition BC	11	1	Symphony No 40	Orchestra	Orchestra	07:29
Mozart	KV581-01	Mozart Complete Edition BC	53	7		Quintet	Clarinet Violin Viola Cello	09:03
Mozart	KV622-01	Mozart Complete Edition BC	23	1	Klarinettenkonzert	Concerto	Clarinet Orchestra	12:00
Paganini	MS025-24	Paganini Accardo-PlaysPaganini DG	5	24	Caprice: Vn.	Solo	Violin	04:23
Piston	IncredibleFlutist-01	Various CollectorsEdition2 MERCURY	35	4		Orchestra	Flute Orchestra	01:09
Rachmaninoff	Op019-01	Rachmaninoff Edition BC	19	7		Duo	Cello Piano	13:02
Ravel	MR081	Various MasterpiecesClassicalMusic MEMBRAN	26	3	Bolero	Orchestra	Orchestra	15:30
Saint	CarnavalAnimaux-12	Saint KlarheitUndRaffinesse MEMBRAN	4	12		Orchestra	Clarinet Xylophone Piano Violin Viola Cello Doublebass	01:28
Saint	Op028	Saint KlarheitUndRaffinesse MEMBRAN	7	5	introduction rondo Capriccioso	Concerto	Violin Orchestra	08:42
Saint	Op033-01	Various MasterpiecesClassicalMusic MEMBRAN	40	4	Cello concert	Concerto	Cello Orchestra	05:29
Saint	Op040	Saint KlarheitUndRaffinesse MEMBRAN	3	4	Danse Macabre	Orchestra	Orchestra	07:28
Saint	Op065-01	Various Classical-Masterworks SONY	20	17		Septet	Trumpet Strings Piano	04:49
Scarlatti	K009	Scarlatti Complete KeyboardSonatas BC	1	9		Solo	Harpsichord	03:23
Schubert	D0667-04	Schubert TheMasterworks BC	13	4		Quintet	Piano Violin Viola Cello doublebass	07:59
Schubert	D0759-01	Schubert TheMasterworks BC	2	1	Unfinished Symphony	Orchestra	Orchestra	13:24
Schubert	D0780-03	Schubert TheMasterworks BC	19	7	Moments musicaux	Solo	Piano	02:08
Schubert	D0810-01	Schubert TheMasterworks BC	10	5	der Tod und das Mädchen	Quartet	Violin Viola Cello	16:12
Schubert	D0899-03	Schubert TheMasterworks BC	26	3	Impromptu	Solo	Piano	05:38
Schumann	Op015-07	Schumann MasterworksEdition SONY	5	7	Trauemerei	Solo	Piano	02:50
Schumann	Op094-01	Schumann MasterworksEdition SONY	20	14	Three Romances	Duo	Oboe Piano	03:29
Shostakovich	Op040-01	Shostakovich Edition BC	24	1		Duo	Cello Piano	11:23
Smetana	MyCountry-02	Smetana BedrichSmetana DOC	1	2	Moldau	Orchestra	Orchestra	11:30
Stravinsky	DumbartonOaks-01	Stravinsky WorksOffgorStravinsky SONY	11	10		Orchestra	Strings Windsection	04:22

B.6 AUDIO RECORDINGS FOR BM-SMALL

Composer	WorkID	Folder	CD	Tr	Name	Ensemble	Instrumentation	Duration
Stravinsky	OctetWinds-01	Stravinsky Works Of Igor Stravinsky SONY	12	3		Orchestra	Windsection	03:57
Telemann	TWV043-e4-01	Telemann Edition BC	23	13		Quintet	Flute Violin Viola Cello Harp	05:08
Tschaikovsky	Op071-13	Tschaikovsky Edition BC	16	19	Blumenwaltzer (Nussknacker- suite)	Orchestra	Orchestra	06:26
Turina	Op036	Bream Complete RCACollection SONY	5	11	Fandanguillo	Solo	Guitar	05:13
Visee	SuiteDMinor-01	Bream Complete RCACollection SONY	9	9	Petites Suite	Solo	Guitar	00:51

B.7 EDM-MIDIs of BM-Small

IDX	Composer	WorkID	BMID	ThemeID	MidiNo	Notes	Duration [s]
1	Bach	BMW1048-01	B40		131	25	4
2	Bach	BWV0543-01	B93		184	20	2.6
3	Bach	BWV0565-01	B298		391	20	1.6
4	Bach	BWV0565-02	B299		392	28	3.5
5	Bach	BWV0808-01	B245		338	24	3.9
6	Bach	BWV0846-01	B301		394	32	4
7	Bach	BWV0846-02	B302		395	18	3.6
8	Bach	BWV0847-01	B303		396	24	3
9	Bach	BWV0847-02	B304		397	20	4.1
10	Bach	BWV1001-01	B175		266	20	3.5
11	Bach	BWV1002-03	B156		247	15	5.2
12	Bach	BWV1009-05	B209		302	23	8
13	Bach	BWV1030-01	B171		262	19	4.2
14	Bach	BWV1041-01	B83		174	15	3.7
15	Bach	BWV1046-01	B30		121	25	4.1
16	Bach	BWV1065-01	B81		172	31	7.5
17	Bartok	Sz112-01	B461	1st Theme	563	20	7.2
18	Bartok	Sz112-01	B462	2nd Theme	564	19	4
19	Beethoven	Op002No1-01	B689	1st Theme	813	10	4
20	Beethoven	Op002No1-01	B690	2nd Theme	814	16	8
21	Beethoven	Op011-01	B1031		1157	12	5.5
22	Beethoven	Op013-01	B728	Intro	852	12	3.7
23	Beethoven	Op013-01	B729	1st Theme	853	17	12
24	Beethoven	Op013-01	B730	2nd Theme	854	14	8.5
25	Beethoven	Op017-01	B687		810	20	7.5
26	Beethoven	Op018No4-01	B592	1st Theme	694	12	5.5
27	Beethoven	Op018No4-01	B593	2nd Theme	695	17	8
28	Beethoven	Op027No2-01	B762	Intro.	886	12	3
29	Beethoven	Op027No2-01	B763		887	11	7
30	Beethoven	Op067-01	B948	1st Theme A	1072	8	5
31	Beethoven	Op067-01	B949	1st Theme B	1073	10	4.2
32	Beethoven	Op067-01	B950	1st Theme C	1074	6	5
33	Beethoven	Op067-01	B951	2nd Theme	1075	8	4
34	Beethoven	Op067-01	B952	3rd Theme	1076	25	7.2
35	Beethoven	Op067-01	B953	4th Theme	1077	17	9
36	Beethoven	WoO059	B554		656	17	2.7
37	Brahms	Hungarian Dances-05	B1375	1st Theme	8508	18	7
38	Brahms	Hungarian Dances-05	B1376	2nd Theme	8509	20	6
39	Brahms	Op015-01	B1337	1st Theme	8470	69	15.5
40	Brahms	Op015-01	B1338	2nd Theme	8471	14	17.5
41	Brahms	Op015-01	B1339	3rd Theme	8472	21	7.5
42	Brahms	Op015-01	B1340	4th Theme	8473	21	12
43	Brahms	Op025-04	B1415	1st Theme	8548	22	6
44	Brahms	Op025-04	B1416	2nd Theme	8549	32	4
45	Brahms	Op025-04	B1417	3rd Theme	8550	32	6
46	Britten	Op002	B1711j	1st Theme	8854	24	8
47	Britten	Op002	B1711k	2nd Theme	8855	18	17
48	Britten	Op002	B1711l	3rd Theme	8856	25	17.2
49	Britten	Op002	B1711m	4th Theme	8857	29	8.5
50	Chopin	Op009-02	C258		9248	17	6
51	Chopin	Op010-12	C198		9186	17	8.6
52	Chopin	Op024-02	C232		9222	24	6
53	Chopin	Op028-04	C289		9280	18	16
54	Chopin	Op028-15	C300	1st Theme	9291	21	9
55	Chopin	Op028-15	C301	2nd Theme	9292	19	11.5
56	Chopin	Op030-02	C234		9224	18	6
57	Chopin	Op063-03	C249		9239	13	6
58	Chopin	Op066	C223		9213	15	2
59	Chopin	Op066	C224		9214	16	7
60	Debussy	L075-03	D144	Clair de Lune 1st Theme	9721	22	11.2
61	Debussy	L075-03	D145	Clair de Lune 2nd Theme	9722	15	8.2
62	Debussy	L095-01	D97	Prelude 1st Theme	9674	23	6.5
63	Debussy	L095-01	D98	2nd Theme	9675	19	9.5
64	Debussy	L103-01	D30		9607	9	15
65	Dukas	ApprentiSorcier	D262	Intro	9850	10	2.9
66	Dukas	ApprentiSorcier	D263	1st Theme	9851	19	6.7
67	Dukas	ApprentiSorcier	D264	2nd Theme	9852	25	6
68	Dvorak	B078-08	D372	1st Theme	1296	21	11

IDX	Composer	WorkID	BMID	ThemeID	MidiNo	Notes	Duration [s]
69	Dvorak	B078-08	D373	2nd Theme	1297	13	18
70	Dvorak	B178-01	D434	1st Theme	1358	23	7.2
71	Dvorak	B178-01	D435	2nd Theme	1359	20	4
72	Dvorak	B178-01	D436	3rd Theme	1360	22	8
73	Dvorak	B178-04	D443	1st Theme	1367	11	7
74	Dvorak	B178-04	D444	2nd Theme	1368	21	4
75	Dvorak	B178-04	D445	3rd Theme	1369	19	14
76	Gershwin	AmericanParis	G39	1st Theme	1777	33	8
77	Gershwin	AmericanParis	G40	2nd Theme	1778	16	7
78	Gershwin	AmericanParis	G41	3rd Theme Blues Theme	1779	22	14
79	Gershwin	AmericanParis	G42	4th Theme	1780	21	8
80	Granados	Op037-02	G206	1st Theme	1955	34	10.5
81	Granados	Op037-02	G207	2nd Theme	1956	21	6
82	Grieg	Op036-01	G310	1st Theme	2060	20	13.5
83	Grieg	Op036-01	G311	2nd Theme	2061	12	16
84	Handel	HWV-368a-01	H163		2282	18	3.5
85	Handel	HWV287-01	H9		2128	21	4.2
86	Handel	HWV289-01	H13	1st Theme	2132	37	6
87	Handel	HWV289-01	H14	2nd Theme	2133	19	5
88	Handel	HWV361-01	H183		2302	20	4.7
89	Handel	HWV365-01	H147		2266	14	3.5
90	Haydn	Hob03No006-01	H320		2439	17	5.2
91	Haydn	Hob03No031-01	H336		2455	34	7.5
92	Haydn	Hob07bNo002-01	H300	1st Theme	2420	30	5.7
93	Haydn	Hob07bNo002-01	H301	2nd Theme	2421	17	3.7
94	Haydn	Hob15No027-01	H703	1st Theme	2824	21	3.5
95	Haydn	Hob15No027-01	H704	2nd Theme	2825	23	3.7
96	Khachaturian	ConcertoViolin DMinor-01	K30	1st Theme A	3201	58	11.5
97	Khachaturian	ConcertoViolin DMinor-01	K31	1st Theme B	3202	32	6
98	Khachaturian	ConcertoViolin DMinor-01	K32	2nd Theme	3203	25	7.7
99	Kodaly	GalantaDances	K63	Intro	3234	15	6
100	Kodaly	GalantaDances	K64		3235	27	6.5
101	Kodaly	GalantaDances	K65		3236	21	5.1
102	Kodaly	GalantaDances	K66		3237	23	4
103	Kodaly	GalantaDances	K67	1st Theme	3238	16	6
104	Kodaly	GalantaDances	K68	2nd Theme	3239	21	4
105	Kodaly	GalantaDances	K69	1st Theme	3240	29	8
106	Kodaly	GalantaDances	K70	2nd Theme	3241	20	3.7
107	Liszt	S541-03	L189		3483	11	11.5
108	Mendelssohn	Op030-03	M289		3899	22	7.5
109	Mozart	KV145	M822		4448	14	8.5
110	Mozart	KV186-02	M557		4181	16	11.5
111	Mozart	KV188-01	M561		4185	22	6
112	Mozart	KV219-01	M522a		4145	15	3.5
113	Mozart	KV219-01	M523		4146	21	6
114	Mozart	KV219-01	M524		4147	22	8
115	Mozart	KV298-01	M615		4240	23	7
116	Mozart	KV315	M429		4051	24	4.5
117	Mozart	KV334-01	M579		4203	19	4.5
118	Mozart	KV361-01	M736		4362	16	11.5
119	Mozart	KV370-01	M619		4244	20	5.7
120	Mozart	KV378-01	M840		4466	21	7
121	Mozart	KV447-01	M445		4067	18	7.5
122	Mozart	KV447-01	M446		4068	23	7
123	Mozart	KV452-01	M687		4312	17	4.5
124	Mozart	KV452-01	M688		4313	14	3.7
125	Mozart	KV550-01	M972		4598	20	9
126	Mozart	KV550-01	M973		4599	14	7.5
127	Mozart	KV581-01	M682		4307	15	11
128	Mozart	KV622-01	M434		4056	14	7.2
129	Paganini	MS025-24	P37		4772	20	3.7
130	Piston	IncredibleFlutist-01	P114	Intro	4854	21	14.7
131	Rachmaninoff	Op019-01	R53		5181	22	14
132	Rachmaninoff	Op019-01	R54		5182	15	8
133	Ravel	MR081	R125	Theme A	5254	36	8
134	Ravel	MR081	R126	Theme B	5255	28	7.5
135	Saint	CarnavalAnimaux-12	S14	Fossiles	5598	20	8.5
136	Saint	Op028	S63	Intro	5648	26	5.5
137	Saint	Op028	S64	1st Theme	5649	22	7
138	Saint	Op028	S65	2nd Theme	5650	41	5.7
139	Saint	Op028	S66	3rd Theme	5651	30	6
140	Saint	Op033-01	S17	1st Theme	5601	19	7
141	Saint	Op033-01	S18	2nd Theme	5602	20	14

B. BM-SUBSETS

IDX	Composer	WorkID	BMID	ThemeID	MidiNo	Notes	Duration [s]
142	Saint	Op040	S47	1st Theme	5632	40	9.5
143	Saint	Op040	S48	2nd Theme	5633	18	11.5
144	Saint	Op065-01	S89	Preamble	5674	15	8.5
145	Scarlatti	K009	S214	Pastorale	5800	33	6
146	Schubert	D0667-04	S387	Theme and Variation on Die Forelle	5974	14	4
147	Schubert	D0759-01	S533	Intro	6120	10	12
148	Schubert	D0759-01	S534	1st Theme	6121	17	11
149	Schubert	D0759-01	S535	2nd Theme	6122	21	8.7
150	Schubert	D0780-03	S319		5906	20	4
151	Schubert	D0810-01	S370	1st Theme	5957	12	8
152	Schubert	D0810-01	S371	2nd Theme	5958	22	8.1
153	Schubert	D0899-03	S301		5888	13	15.5
154	Schumann	Op015-07	S713	Träumerei	6308	21	8
155	Schumann	Op094-01	S802		6398	22	7.5
156	Shostakovich	Op040-01	S870a	1st Theme	6467	21	10
157	Shostakovich	Op040-01	S870b	2nd Theme	6468	15	16
158	Smetana	MyCountry-02	S1114	1st Theme	6723	23	11.2
159	Smetana	MyCountry-02	S1115	2nd Theme	6724	21	3.7
160	Smetana	MyCountry-02	S1116	3rd Theme	6725	12	16
161	Smetana	MyCountry-02	S1117	4th Theme	6726	10	15
162	Stravinsky	DumbartonOaks-01	S1534a	1st Theme	7154	29	4
163	Stravinsky	DumbartonOaks-01	S1534b	2nd Theme	7155	16	5.5
164	Stravinsky	DumbartonOaks-01	S1534c	3rd Theme	7156	18	4.5
165	Stravinsky	OctetWinds-01	S1541	Intro.	7167	15	5.6
166	Stravinsky	OctetWinds-01	S1542	1st Theme	7168	23	7.2
167	Stravinsky	OctetWinds-01	S1543	2nd Theme	7169	9	7.5
168	Telemann	TWV043-e4-01	T58	Prelude	7362	28	9
169	Telemann	TWV043-e4-01	T59	1st Theme	7363	38	7
170	Telemann	TWV043-e4-01	T60	2nd Theme	7364	40	4.7
171	Tschaikovsky	Op071-13	T198	1st Theme	7541	14	10.5
172	Tschaikovsky	Op071-13	T199	2nd Theme	7542	13	5.7
173	Tschaikovsky	Op071-13	T200	3rd Theme	7543	13	9.5
174	Tschaikovsky	Op071-13	T201	4th Theme	7544	14	10.5
175	Turina	Op036	T301	1st Theme	7645	30	7
176	Turina	Op036	T302	2nd Theme	7646	22	6
177	Visee	SuiteDMinor-01	V123	Prelude	7787	16	4.7

B.8 BM-Medium

In the following we provide a statistic that shows each composer and the respective number of audio recordings that are in the subset BM-Medium.

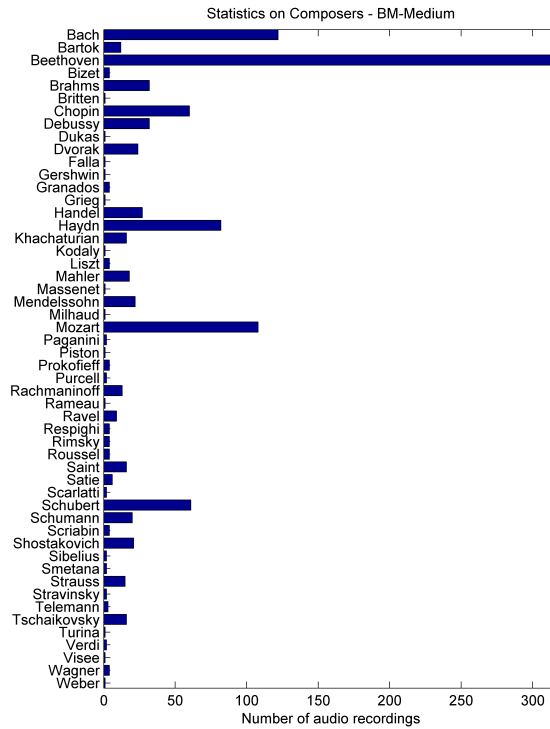


Figure B.29: Number of audio recordings by composer for the subset BM-Medium

B.9 Errors

In the following we provide a table that contains errors which affect the subset BM-Medium. For each of the entries, there is one audio recording per theme and two themes per movement. Due to the naming conventions for the audio recordings this leads to an error which causes that one of the two audio recordings is overwritten by the other one.

Composer	WorkID	Folder	CD	Tr	ErrorType
Beethoven	Op130-04	Beethoven CompleteEdition BC	42	4	Movement
Beethoven	Op130-04	Beethoven CompleteEdition BC	42	5	Movement
Beethoven	Op005No2-01	Beethoven CompleteEdition BC	28	3	Movement
Beethoven	Op005No2-01	Beethoven CompleteEdition BC	28	4	Movement
Beethoven	Op110-03	Beethoven CompleteEdition BC	46	8	Movement
Beethoven	Op110-03	Beethoven CompleteEdition BC	46	9	Movement

B. BM-SUBSETS

In the following we provide a table that contains EDM-themes that are not within the subset BM-Medium, despite the corresponding audio recordings are in the subset. These EDM-themes were either found after the creation of the subset BM-Medium, or they are not available on the EDM-website.

Composer	WorkID	BMID	ThemeID	MidiNo	ErrorType
Bach	BWV0886-01	B382		476	EDM-Theme not in subset
Beethoven	Op002No1-04	B694	1st Theme	818	EDM-Theme not in subset
Beethoven	Op084-00	B561	2nd Theme	663	EDM-Theme not in subset
Chopin	Op034-02	C344		9338	EDM-Theme not in subset
Chopin	Op037-02	C268		9259	EDM-Theme not in subset
Wagner	WWV086C-02-02-02	W47	5th Theme	7893	EDM-Theme not in subset
Beethoven	Op092-01	B988	1st Theme		No EDM-Theme on EDM Website

In the following we provide a table that contains errors regarding EDM-themes. The specific error is specified by a keyword in the column ‘ErrorType’. For some of the errors, we provide a more detailed description in the column ‘ErrorDescription’. For large tempo differences between the EDM-theme and the audio recording we use the keyword ‘MidiErrTempo’. Playing errors in the EDM-theme are stated by the keyword ‘MidiErr’. The keywords ‘MidiJointThemes’ and ‘MidiPart’ are explained in Section 3.2.

Composer	WorkID	BMID	ThemeID	MidiNo	ErrorType	ErrorDescription
Bach	BWV0846-01	B301		394	MidiErrTempo	
Bach	BWV0543-01	B93		184	MidiErrTempo	
Bach	BWV1001-01	B175		266	MidiErrTempo	
Bach	BWV0565-01	B298		391	MidiErrTempo	
Beethoven	WoO059	B554		656	MidiErrTempo	
Beethoven	Op018No4-01	B593	2nd Theme	695	MidiErr	
Brahms	Op025-04	B1415	1st Theme	8548	MidiErr	
Chopin	Op010-12	C198		9186	MidiErr	
Chopin	Op009-02	C258		9248	MidiErr	
Gershwin	AmericanParis	G39	1st Theme	1777	MidiErr	
Handel	HWV361-01	H183		2302	MidiErrTempo	
Mozart	KV315	M429		4051	MidiErrTempo	
Scarlatti	K009	S214	Pastorale	5800	MidiErr	
Schubert	D0780-03	S319		5906	MidiErr	
Stravinsky	OctetWinds-01	S1541	Intro	7167	MidiErr	
Bach	BWV0851-01	B311		404	MidiErrTempo	
Bach	BWV0855-01	B319		412	MidiErr MidiErrTempo	missing trill
Bach	BWV0857-01	B323		416	MidiErrTempo	
Bach	BWV0861-01	B331		424	MidiErrTempo	
Bach	BWV0862-02	B334		427	MidiErrTempo	
Bach	BWV0863-02	B336		429	MidiErr	double sharp
Bach	BWV0865-02	B340		433	MidiErrTempo	
Bach	BWV0866-01	B341		434	MidiErrTempo	
Bach	BWV0872-01	B354		447	MidiErr	
Bach	BWV0877-01	B364		457	MidiErr	double sharp
Bach	BWV0877-02	B365		458	MidiErr	double sharp
Bach	BWV0882-01	B374		467	MidiErrTempo	
Bach	BWV0883-01	B376		469	MidiErrTempo	
Bach	BWV0885-01	B380		473	MidiErrTempo	
Bach	BWV0886-01	B382		475476	MidiPart	
Bach	BWV0887-01	B384		478	MidiErr	double sharp
Bach	BWV0889-02	B389		483	MidiErr	wrong note
Bartok	Sz040-02	B484	1st Theme	586	MidiErr	wrong note
Bartok	Sz067-01	B489		591	MidiErr	wrong notes
Bartok	Sz087-01	B495	1st Theme	597	MidiErrTempo	

Composer	WorkID	BMID	ThemeID	MidiNo	ErrorType	ErrorDescription
Bartok	Sz087-02	B498	2nd Theme	600	MidiErr	missing ornaments
Beethoven	Op019-03	B532		634	MidiErr	
Beethoven	Op037-02	B535		637	MidiErrTempo	
Beethoven	Op058-02	B539		641	MidiErrTempo	
Beethoven	Op073-03	B546	1st Theme	648	MidiErr	wrong note
Beethoven	Op073-03	B547	2nd Theme	649	MidiErr	missing ornament
Beethoven	Op138	B566	2nd Theme	668	MidiErr	missing grace note
Beethoven	Op018No2-01	B579	1st Theme	681	MidiErrTempo MidiErr	missing grace note
Beethoven	Op018No2-02	B581	1st Theme	683	MidiErr	missing notes
Beethoven	Op018No4-04	B597	1st Theme	699	MidiErr	missing grace note
Beethoven	Op018No6-02	B606		709	MidiErrTempo MidiErr	missing notes
Beethoven	Op059No1-03	B612		715	MidiErrTempo MidiErr	missing notes
Beethoven	Op095-04	B633	Intro	736	MidiErrTempo	
Beethoven	Op130-01	B642		745	MidiJointThemes	A B
Beethoven	Op131-06	B654		758	MidiErr	wrong note
Beethoven	Op132-01	B657	Intro and 1st Theme	761	MidiJointThemes	Intro1st Theme
Beethoven	Op135-01	B663a	1st Theme	770	MidiErrTempo MidiErr	wrong notes
Beethoven	Op135-04	B666	Intro	774	MidiJointThemes	
Beethoven	Op016-01	B667a	Intro	776	MidiErrTempo	
Beethoven	Op016-02	B667c		778	MidiErrTempo	
Beethoven	Op020-06	B678		790	MidiErrTempo	
Beethoven	Op008-02	B679b		793	MidiErrTempo	
Beethoven	Op005No2-02	B680a		801	MidiErrTempo	
Beethoven	Op102No1-02	B686a		808	MidiErrTempo MidiErr	wrong note
Beethoven	Op017-03	B688a		812	MidiErrTempo	
Beethoven	Op002No1-02	B691		815	MidiErr MidiErrTempo	missing grace note
Beethoven	Op002No2-04	B701	1st Theme	825	MidiErr	missing grace note
Beethoven	Op002No3-02	B706		830	MidiErrTempo	
Beethoven	Op007-04	B713		837	MidiErrTempo	
Beethoven	Op010No1-02	B716		840	MidiErr	missing ornament
Beethoven	Op014No1-01	B737	2nd Theme	861	MidiErr	double sharp
Beethoven	Op027No1-02	B759	2nd Theme	883	MidiErr	wrong note
Beethoven	Op028-02	B772	2nd Theme	896	MidiErrTempo	
Beethoven	Op081a-02	B823	l'absence	947	MidiErrTempo	
Beethoven	Op101-03	B829		953	MidiErrTempo	
Beethoven	Op110-01	B842	1st Theme A	966	MidiErr	missing note
Beethoven	Op111-01	B847	Intro	971	MidiErrTempo	
Beethoven	Op012No3-02	B859		983	MidiErr	missing grace note missing ornament
Beethoven	Op023-01	B861		985	MidiErr	missing grace notes
Beethoven	Op030No1-01	B869	2nd Theme	993	MidiErr	missing grace notes
Beethoven	Op030No2-03	B875	1st Theme	999	MidiErr	missing grace notes
Beethoven	Op036-02	B912	4th Theme	1036	MidiErrTempo	
Beethoven	Op055-02	B924	1st Theme	1048	MidiErrTempo	
Beethoven	Op067-02	B954	1st Theme	1078	MidiErrTempo	
Beethoven	Op068-03	B975	1st Theme A	1099	MidiErr	missing grace note
Beethoven	Op093-02	B1007	1st Theme	1132	MidiErrTempo	
Beethoven	Op093-02	B1008	2nd Theme	1133	MidiErrTempo MidiErr	missing trills
Beethoven	Op125-01	B1016	1st Theme	1141	MidiErrTempo	
Beethoven	Op001No3-01	B1034	1st Theme A	1159	MidiErr	Missing ornament
Beethoven	Op070No1-02	B1044		1169	MidiErrTempo	
Brahms	HungarianDances-07	B1378		8511	MidiErrTempo	
Brahms	Op073-01	B1604	5th Theme	8737	MidiErr	double sharp
Brahms	Op073-03	B1612	2nd Theme	8745	MidiErr	missing bar
Chopin	Op050-02	C243		9233	MidiErr	missing grace note
Chopin	Op009-03	C259		9249	MidiErr	wrong notes
Chopin	Op037-02	C268		92589259	MidiPart	
Chopin	Op053	C284		9275	MidiErr	missing grace notes
Chopin	Op034-02	C344		93379338	MidiPart MidiErr	B C double sharp
Chopin	Op064-02	C350		93449345	MidiPart	
Debussy	L066-02	D18	1st Theme	9595	MidiErr	missing repetition

B. BM-SUBSETS

Composer	WorkID	BMID	ThemeID	MidiNo	ErrorType	ErrorDescription
Debussy	L113-03	D24	Serenade of the Doll	9601	MidiErr	missing grace notes
Debussy	L113-06	D28	2nd Theme	9605	MidiErr	missing grace note
Debussy	L100-03	D38	Jardins sous la Pluie 1st Theme	9615	MidiErr	wrong note
Debussy	L122No2-01	D51	5th Theme	9628	MidiErrTempo	
Debussy	L110-01	D63	1st Theme	9640	MidiJointThemes	A B
Debussy	L109-01	D69	1st Theme	9646	MidiErr	missing tie
Debussy	L109-02	D75	1st Theme	9652	MidiErrTempo	
Debussy	L109-02	D76	2nd Theme	9653	MidiErr	missing grace note
Dvorak	B191-01	D278	1st Theme	1193	MidiErr	wrong note
Dvorak	B147-04	D384	1st Theme	1308	MidiErrTempo	
Dvorak	B147-04	D385	2nd Theme	1309	MidiErrTempo	
Dvorak	B147-06	D389	2nd Theme	1313	MidiErr	wrong note
Granados	Op037-04	G208	1st Theme	1957	MidiErr	missing grace notes
Haydn	Hob17No006	H294	1st Theme	2414	MidiErrTempo	
Haydn	Hob03No032-02	H342	2nd Theme	2461	MidiErr MidiErrTempo	missing grace note
Haydn	Hob03No034-02	H346		2465	MidiErr MidiErrTempo	missing grace note
Haydn	Hob03No039-02	H362	2nd Theme	2481	MidiErr	wrong note
Haydn	Hob16No023-01	H508		2628	MidiErrTempo	
Haydn	Hob16No034-02	H513		2633	MidiErrTempo	
Haydn	Hob16No036-01	H520		2640	MidiErrTempo	
Haydn	Hob16No036-03	H523	1st Theme	2643	MidiErr	wrong note
Haydn	Hob16No037-02	H526		2646	MidiErrTempo	
Haydn	Hob15No025-01	H689		2809	MidiErr	wrong grace note
Haydn	Hob15No025-02	H690	1st Theme	2810	MidiErr	missing grace note
Haydn	Hob15No026-02	H700		2820	MidiErrTempo MidiErr	wrong grace note
Haydn	Hob15No029-02	H708		2829	MidiErr	double sharp
Khachaturian	ConcertoPianoDFlat-01	K23	1st Theme	3194	MidiErr	double b
Khachaturian	ConcertoPianoDFlat-01	K24	2nd Theme	3195	MidiErr	missing grace notes
Khachaturian	ConcertoPianoDFlat-02	K25	Intro	3196	MidiErr	missing grace notes
Khachaturian	GayaneBalett-09	K41	Lullaby	3212	MidiErr	missing grace notes
Liszt	S244-02	L149	1st Theme	3443	MidiErrTempo	
Liszt	S244-02	L155	7th Theme	3449	MidiErr	wrong note
Mahler	SymphonyNo02-01	M59	2nd Theme	3667	MidiErr	wrong note
Mahler	SymphonyNo04-02	M80	2nd Theme	3688	MidiErrTempo	midi much faster
Mahler	SymphonyNo05-04	M90	Adagietto	3698	MidiErrTempo	midi much faster
Mahler	SymphonyNo09-01	M93	3rd Theme	3701	MidiErr	wrong note
Mahler	SymphonyNo09-03	M102	2nd Theme	3710	MidiErrTempo	audio much faster
Mahler	SymphonyNo09-04	M106	Intro	3714	MidiErrTempo	midi much faster
Mozart	KV320-03	M731		4357	MidiErr	wrong note
Mozart	KV525-02	M755		4381	MidiErr	wrong note
Mozart	KV543-02	M968		4594	MidiErr	missing note
Mozart	KV550-04	M980		4606	MidiErr	one note too much
Mozart	KV551-01	M981		4607	MidiErr	wrong note
Paganini	MS025-05	P23		4758	MidiErr	wrong note
Ravel	MR072	R208	7th Theme	53385339	MidiPart	
Schubert	D0667-02	S384		5971	MidiErrTempo	audio much slower
Schubert	D0667-03	S385	1st Theme	5972	MidiErrTempo	audio much faster
Schubert	D0894-02	S414	1st Theme	6001	MidiErrTempo	
Schubert	D0894-02	S415	2nd Theme	6002	MidiErrTempo	
Schubert	D0960-03	S438		6025	MidiErrTempo	audio much faster
Schubert	D0574-02	S455		6042	MidiErrTempo	audio much faster
Schubert	D0574-03	S456		6043	MidiErrTempo	missing grace note midi much faster
Schubert	D0485-01	S499		6086	MidiErrTempo	audio much faster
Shostakovich	Op010-01	S873	2nd Theme	6476	MidiErr	
Shostakovich	Op047-03	S891	1st Theme	6494	MidiErrTempo	audio much slower
Shostakovich	Op047-03	S893	3rd Theme	6496	MidiErr	
Shostakovich	Op054-01	S897	2nd Theme	6500	MidiErrTempo	
Shostakovich	Op054-03	S900	1st Theme	6503	MidiErrTempo	audio much faster
Shostakovich	Op070-01	S914	1st Theme	6517	MidiErrTempo	audio much faster
Shostakovich	Op070-01	S915	2nd Theme	6518	MidiErrTempo	audio much faster
Shostakovich	Op070-03	S918	1st Theme	6521	MidiErrTempo	audio much faster
Shostakovich	Op070-04	S920		6523	MidiErrTempo	audio much slower

Composer	WorkID	BMID	ThemeID	MidiNo	ErrorType	ErrorDescription
Strauss	Op418	S1308	No. 4 2nd Theme	6917	MidiErr	
Tschaikovsky	Op071-12b	T195		7538	MidiJointThemes	
Wagner	WWV086C-02-02-02	W47	5th Theme	78927893	MidiPart	

Bibliography

- [1] H. BARLOW AND S. MORGENSTERN, *A Dictionary of Musical Themes*, Crown Publisher, New York, 1962.
- [2] CHROMA TOOLBOX. <http://www.mpi-inf.mpg.de/~mmueller/chromatoolbox/>, Retrieved 20.08.2014.
- [3] S. EWERT, M. MÜLLER, AND M. CLAUSEN, *Musicmatching bei Variabilitäten in der Harmonik und Polyphonie*, in Proceedings of the 36th Deutsche Jahrestagung für Akustik (DAGA), Berlin, Germany, 2010.
- [4] C. FREMERIEY, M. CLAUSEN, S. EWERT, AND M. MÜLLER, *Sheet music-audio identification*, in Proceedings of the International Society for Music Information Retrieval Conference (ISMIR), Kobe, Japan, 2009, pp. 645–650.
- [5] T. FUJISHIMA, *Realtime chord recognition of musical sound: A system using common lisp music*, in Proceedings of the International Computer Music Conference (ICMC), Beijing, 1999, pp. 464–467.
- [6] E. GÓMEZ AND P. HERRERA, *Estimating the tonality of polyphonic audio files cognitive versus machine learning modelling strategies*, 2004.
- [7] M. GOTO, *A real-time music-scene-description system: Predominant-F0 estimation for detecting melody and bass lines in real-world audio signals*, Speech Communication (ISCA Journal), 43 (2004), pp. 311–329.
- [8] A. P. Klapuri, *Multiple fundamental frequency estimation by summing harmonic amplitudes*, in International Society for Music Information Retrieval Conference (ISMIR), 2006, pp. 216–221.
- [9] MIDI MANUFACTURERS ASSOCIATION, *Midi manufacturers association - the official source of information about midi*, 09 2014.
- [10] M. MÜLLER, *Information Retrieval for Music and Motion*, Springer Verlag, 2007.
- [11] M. MÜLLER AND S. EWERT, *Towards timbre-invariant audio features for harmony-based music*, IEEE Transactions on Audio, Speech, and Language Processing, 18 (2010), pp. 649–662.
- [12] ———, *Chroma Toolbox: MATLAB implementations for extracting variants of chroma-based audio features*, in Proceedings of the 12th International Conference on Music Information Retrieval (ISMIR), Miami, USA, 2011, pp. 215–220.
- [13] M. MÜLLER, N. JIANG, AND P. GROSCHE, *A robust fitness measure for capturing repetitions in music recordings with applications to audio thumbnailing*, IEEE Transactions on Audio, Speech & Language Processing, 21 (2013), pp. 531–543.
- [14] M. MÜLLER, F. KURTH, AND M. CLAUSEN, *Audio matching via chroma-based statistical features*, in Proceedings of the 6th International Conference on Music Information Retrieval (ISMIR), 2005, pp. 288–295.

BIBLIOGRAPHY

- [15] PROJECT PETRUCCI LLC, *International Music Score Library Project (IMSLP) Petrucci Music Library*, 07 2014.
- [16] J. SALAMON AND E. GÓMEZ, *Melody extraction from polyphonic music signals using pitch contour characteristics*, IEEE Transactions on Audio, Speech, and Language Processing, 20 (2012), pp. 1759–1770.
- [17] J. SALAMON, J. SERRÀ, AND E. GÓMEZ, *Tonal representations for music retrieval: from version identification to query-by-humming*, International Journal of Multimedia Information Retrieval, (2012), pp. 45–58.
- [18] J. T. SCHWARTZ AND D. SCHWARTZ, *The electronic dictionary of musical themes*, 07 2014.
- [19] J. SERRÀ, E. GÓMEZ, P. HERRERA, AND X. SERRA, *Chroma binary similarity and local alignment applied to cover song identification*, IEEE Transactions on Audio, Speech and Language Processing, 16 (2008), pp. 1138–1151.
- [20] J. SERRÀ, M. MÜLLER, P. GROSCHE, AND J. L. ARCOS, *Unsupervised detection of music boundaries by time series structure features*, in Proceedings of the AAAI International Conference on Artificial Intelligence, Toronto, Ontario, Canada, 2012.
- [21] I. S. SUYOTO, A. L. UITDENBOGERD, AND F. SCHOLER, *Searching musical audio using symbolic queries*, IEEE Transactions on Audio, Speech & Language Processing, 16 (2008), pp. 372–381.
- [22] W. H. TSAI, H. M. YU, AND H. M. WANG, *A query-by-example technique for retrieving popular songs with similar melodies*, in Proceedings of ISMIR, 2005.