A Geometrical Distance Measure for Determining the Similarity of Musical Harmony

W. Bas De Haas

Frans Wiering

and Remco C. Veltkamp

Technical Report UU-CS-2011-015
May 2011

Department of Information and Computing Sciences
Utrecht University, Utrecht, The Netherlands
www.cs.uu.nl

ISSN: 0924-3275

Department of Information and Computing Sciences
Utrecht University
P.O. Box 80.089
3508 TB Utrecht
The Netherlands

Abstract

In the last decade, digital repositories of music have undergone an enormous growth. Therefore the availability for scalable and effective methods that provide content-based access to these repositories has become critically important. This study presents and tests a new geometric distance function that quantifies the harmonic distance between two pieces of music. Harmony is one of the most important aspects of music and we will show in this paper that harmonic similarity can significantly contribute to the retrieval of digital music. Yet, within the Music Information Retrieval field, harmonic similarity measures have received far less attention compared to other similarity aspects.The distance function we present, the Tonal Pitch Step Distance, is based on a cognitive model of tonality and captures the change of harmonic distance to the tonal center over time. This distance is compared to two other harmonic distance measures and, although it is not the best performing distance measure, the proposed measure is shown to be efficient for retrieving similar jazz standards and significantly outperforms a baseline string matching approach. Furthermore, we demonstrate in a case study how our harmonic similarity measure can contribute to the musicological discussion about melody and harmony in large-scale corpora.

A Geometrical Distance Measure for Determining the Similarity of Musical Harmony

W. BAS DE HAAS, FRANS WIERING and REMCO C. VELTKAMP
Utrecht University

1 Introduction

Content-based Music Information Retrieval (MIR[1]) is a rapidly expanding area within multimedia research. On-line music portals like last.fm, iTunes, Pandora, Spotify and Amazon disclose millions of songs to millions of users around the world. Propelled by these ever-growing digital repositories of music, the demand for scalable and effective methods for providing music consumers with the music they wish to have access to, still increases at a steady rate. Generally, such methods aim to estimate the subset of pieces that is relevant to a specific music consumer. Within MIR the notion of *similarity* is therefore crucial: songs that are similar in one or more features to a given relevant song are likely to be relevant as well. In contrast to the majority of approaches to notation-based music retrieval that focus on the similarity of the *melody* of a song, this paper presents a new method for retrieving music on the basis of its *harmony* structure.

Within MIR two main directions can be discerned: symbolic music retrieval and the retrieval of musical audio. The first direction of research stems from musicology and the library sciences and aims to develop methods that provide access to digitized musical scores. Here music similarity is determined by analyzing the combination of symbolic entities, such as notes, rests, meter signs, etc., that are typically found in musical scores. Musical audio retrieval arose when the digitization of audio recordings started to flourish and the need for different methods to maintain and unlock digital music collections emerged. Audio based MIR methods extract features from the audio signal and use these features for estimating whether two pieces of music are musically related. Often these features, e.g. chroma features Wakefield [1999] or Mel-Frequency Cepstral Coefficients [MFCCs, Logan 2000], do not directly translate to the notes, beats, voices and instruments that are used in the symbolic domain. Ideally, one would translate audio features into notes, beats and voices and use such a high level representation for similarity estimation. However, current automatic polyphonic music transcription systems have not matured enough for their output to be usable for determining music similarity. In this paper we focus on a symbolic musical representation that can be transcribed reasonably well from the audio signal using current technology: chord sequences. In other words, for applying our method to audio we assume a preprocessing step is made with one of the available chord labeling methods (See Section 2.2).

In this paper we present a novel similarity measure for chord sequences. We will show that such a method can be used to retrieve harmonically related pieces and can aid in musicological discussions. We will discuss related work on harmonic similarity and the research from music theory and music cognition that is relevant for our similarity measure in Section 2. Next, we will present the Tonal Pitch Step distance in Section 3. In Section 4 we show how our distance measure performs in practice and we show that it can also contribute to musicological discussions in Section 5. But first, we will give a brief introduction on what actually constitutes tonal harmony and harmonic similarity.

[1]Within this paper MIR will refer to *Music* (and not Multimedia) Information Retrieval

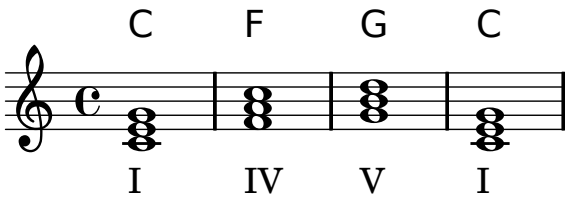

Figure 1: A very typical and frequently used chord progression in the key of C-major, often referred to as I-IV-V-I. Above the score the chord labels, representing the notes of the chords in the section of the score underneath the label, are printed. The roman numbers below the score denote the interval between the chord root and the tonic of the key.

1.1 What is Harmony?

The most basic element in music is a *tone*. A tone is a sound with a fixed frequency that can be described in a musical score with a *note*. All notes have a name, e.g. C, D, E, etc., and represent tones of specific frequencies. The distance between two notes is called an *interval* and is measured in semitones, which is the smallest interval in Western tonal music. Also intervals have names: minor second (1 semitone), second (2 semitones), minor third (3 semitones), etc., up to an octave (13 semitones). When two tones are an octave apart the highest tone will have exactly twice its frequency. These two tones are also perceived by the listeners as very similar, so similar even that all tones one or more octave apart have the same name. Hence, these tones are said to be in the same *pitch class*.

Harmony arises in music when two or more tones sound at the same time. These simultaneously sounding notes form chords, which can in turn be used to form chord sequences. A *chord* can be viewed as a group of tones that are often separated by intervals of roughly the same size. The most basic chord is the *triad* which consists of three pitch classes that are separated by two thirds. The two most important factors that characterize a chord are its structure, determined by the intervals between these notes, and the chord *root*. The root note is the note on which the chord is built. The root is often, but it does not necessarily have to be, the lowest sounding note. Figure 1 displays a frequently occurring chord sequence. The first chord is created by taking a C as root and subsequently a major third interval (E) and a minor third interval (G) are added, yielding a C-major chord. Above the score the names of the chords, which are based on the root of the chord, are printed. If the interval between the root of the chord and the third is major third, the chord is called a *major chord*, if it is a minor third, the chord is called a *minor chord*.

The internal structure of the chord has a large influence on the *consonance* or *dissonance* of a chord: some combinations of simultaneous sounding notes are perceived to have a more tense sound than others. Another important factor that contributes to perceived tension of a chord is the relation between the chord and the *key* of the piece. The key of a piece of music is the tonal center of the piece. It specifies the *tonic*, which is the most stable, and often the last, tone in that piece. Moreover, the key specifies the *scale*, which is set of pitches that will occur most frequently and that sound reasonably well together. A chord can be built up from pitches that are part of the scale or they can borrow notes from outside the scale, the latter being more dissonant. Especially the root note of a chord has a distinctive role, because the interval of the chord root and the key largely determines the *harmonic function* of the chord. The three most important harmonic functions are the dominant (V), that builds up tension, a sub-dominant (IV), that prepares a dominant and the tonic (I) that releases tension. In Figure 1 a Roman number that represents the interval between the root of the chord and the key, often called *scale degree*, is printed underneath the score.

Obviously, this is a rather basic view on tonal harmony. For a thorough introduction to tonal harmony we refer to Piston [1941]. Harmony is considered a fundamental aspect of Western tonal music by musicians and music researchers. For centuries, the analysis of harmony has aided composers and performers in understanding the tonal structure of music. The harmonic structure of a piece alone can reveal song structure through repetitions, tension and release patterns, tonal ambiguities, modulations (i.e. local key changes), and musical style. Therefore Western tonal harmony has become one of the most prominently investigated

topics in music theory and can be considered a feature of music that is equally distinctive as rhythm or melody. Nevertheless, harmonic structure as a feature for music retrieval has received far less attention than melody and rhythm within the MIR field.

1.2 Harmonic Similarity and Its Application in MIR

Harmonic similarity depends not only on the musical information, but also largely on the interpretation of this information by the human listener. Human listeners, musician and non-musician alike, have extensive culture-dependent knowledge about music that needs to be taken into account when modeling music similarity. It is important to realize that music only becomes music in the mind of the listener, and that not all information needed for making a good similarity judgment can be found in the musical data alone.

In this light we consider the harmonic similarity of two chord sequences to be the degree of agreement between structures of simultaneously sounding notes and the agreement between global as well as local relations between these structures in the two sequences as perceived by the human listener. By the agreement between structures of simultaneously sounding notes we denote the similarity that a listener perceives when comparing two chords in isolation and without surrounding musical context. However, chords are rarely compared in isolation and the relations to the global context–the key of a piece–and the relations to the local context play a very important role in the perception of tonal harmony. The local relations can be considered the relations between functions of chords within a limited time frame, for instance the preparation of a chord with a dominant function by means of a sub-dominant. All these factors play a role in the perception of tonal harmony and should be shared by two compared pieces up to certain extent to be considered similar.

In the context of this view on harmonic similarity, music retrieval based on harmony sequences clearly offers various benefits. It allows for finding different versions of the same song even when melodies vary. This is often the case in cover songs or live performances, especially when these performances contain improvisations. Moreover, playing the same harmony with different melodies is an essential part of musical styles like jazz and blues. Also, variations over standard basses in baroque instrumental music can be harmonically closely related, e.g. chaconnes.

1.3 contribution

We introduce a distance function that quantifies the dissimilarity between two sequences of musical chords. The distance function is based on a cognitive model of tonality and models the change of chordal distance to the tonic over time. The proposed measure can be computed efficiently and matches human intuitions about harmonic similarity. The retrieval performance is examined in an experiment on 5028 human-generated chord sequences, in which we compare it to two other harmonic distance functions. We furthermore show in a case study how the proposed measure can contribute to the musicological discussion about the relation between melody and harmony in melodically similar Bach chorales. The work presented here extends and integrates earlier harmony similarity work in [de Haas et al. 2008; 2010a].

2 Related Work

MIR methods that focus on the harmonic information in the musical data are quite numerous. Relevant for the current study is the work on polyphonic music transcription, e.g. Klapuri and Davy [2006], and automatic chord labeling, e.g. Mauch [2010], in the audio domain. Currently, the state-of-the-art in polyphonic music transcription does not produce transcriptions that are usable for the here presented distance measures. Nevertheless, it gave rise to new methods and ideas that are widely used in automatic chord labeling. These methods do not produce a complete score given a piece of musical audio but return a list of chord labels that can be directly matched with our distance measure. Within the symbolic domain the research seems to focus on complete polyphonic MIR systems, e.g. Bello and Pickens [2005]. By complete systems we mean systems that do chord labeling, segmentation, matching and retrieval all at once. The number of papers that purely focus on the development and testing of harmonic similarity measures is much smaller.

In the next Section we will review other approaches to harmonic similarity, in Section 2.2 we will discuss the current state of automatic chord labeling, in Section 2.3, and in 2.4 we elaborate on the cognition of tonality and the cognitive model relevant to the similarity measure that will be exposed in Section 3.

2.1 Harmonic Similarity Measures

All polyphonic similarity measures slice up a piece of music in segments that represent a single chord. Typical segment lengths range from the duration of a sixteenth note up to the duration of a couple of beats depending on the kind of musical data and the segmentation procedure.

An interesting symbolic MIR system based on the development of harmony over time is the one developed by Pickens and Crawford [2002]. Instead of describing a musical segment as a single chord, they represent a musical segment as a 24 dimensional vector describing the 'fit' between the segment and every major and minor triad, using the euclidean distance in the for dimensional pitch space as found by Krumhansl [1990] in her controlled listening experiments (see section 2.3). Pickens and Crawford then use a Markov model to model the transition distributions between these vectors for every piece. Subsequently, these Markov models are ranked using the Kullback-Leibler divergence to obtain a retrieval result.

Other interesting work has been done by Paiement et al. [2005]. They define a similarity measure for chords rather than for chord sequences. Their similarity measure is based on the sum of the perceived strengths of the harmonics of the pitch classes in a chord, resulting in a vector of twelve pitch classes for each musical segment. Paiement et al. subsequently define the distance between two chords as the euclidean distance between two of these vectors that correspond to the chords. Next, they use a graphical model to model the hierarchical dependencies within a chord progression. In this model they use their chord similarity measure for the calculation of the substitution probabilities between chords and not for estimating the similarity between sequences of chords.

Besides the similarity measure that we will elaborate on in this paper and which was earlier introduced in [de Haas et al. 2008; 2010a] there are two other methods that solely focus on the similarity of chord sequences: an alignment based approach to harmonic similarity Hanna et al. [2009] and a grammatical parse tree matching method de Haas et al. [2009]. The first two are quantitatively compared in an experiment in Section 4. The harmony grammar approach could, at the time of writing, not compete in this experiment because in its current state it is yet unable to parse all the songs in the used dataset.

The Chord Sequence Alignment System (CSAS) Hanna et al. [2009] is based on local alignment and computes similarity between two sequences of symbolic chord labels. By performing elementary operations the one chord sequence is transformed into the other chord sequence. The operations used to transform the sequences are deletion or insertion of a symbol, and substitution of a symbol by another. The most important part in adapting the alignment is how to incorporate musical knowledge and give these operations valid musical meaning. Hanna et al. experimented with various musical data representations and substitution functions and found a key relative representation to work well. For this representation they rendered the chord root as the difference in semitones between the chord root and the key; and substituting a major chord for a minor chord and vice versa yields a penalty. The total transformation from the one string into the other can be solved by dynamic programming in quadratic time. For a more elaborate description of the CSAS we refer to Hanna et al. [2009].

The third harmonic similarity measure using chord descriptions is a generative grammar approach de Haas et al. [2009]. The authors use a generative grammar of tonal harmony to parse the chord sequences, which results in parse trees that represent harmonic analyses of these sequences. Subsequently, a tree that contains all the information shared by the two parse trees of two compared songs is constructed and several properties of this tree can be analyzed yielding several similarity measures. Currently a parser can reject a sequence of chords as being ungrammatical. We expect this issue to be resolved in the near future by applying a error-correcting parser Swierstra [2009].

2.2 Finding Chord Labels

The application of harmony matching methods is extended by the extensive work on chord label extraction from musical audio and symbolic score data within the MIR community. Chord labeling algorithms extract chord labels from raw musical data and these labels can be matched using the distance measures presented in this paper.

Currently there are several methods available that derive these descriptions from raw musical data. Recognizing a chord in a musical segment is a difficult task: in case of audio data, the stream of musical must be segmented, aligned to a grid of beats, the different voices of the different instruments have to be recognized, etc. Even if such information about the notes, beats, voices, bar lines, key signatures, etc. is available, as it is in the case of symbolic musical data, finding the right description of the musical chord is not trivial. The algorithm must determine which notes are unimportant passing notes and sometimes the right chord can only be determined by taking the surrounding harmonies into account. Nowadays, several algorithms can correctly segment and label approximately 84 percent of a symbolic dataset (see for a review Temperley [2001]). Within the audio domain hidden Markov Models are frequently used for chord label assignment, e.g. Mauch [2010]; Bello and Pickens [2005]. Within the audio domain, the currently best performing methods have an accuracy of around 80 percent. Of course, these numbers depend on musical style and on the quality of the data.

2.3 Cognitive Models of Tonality

Only part of the information needed for sound similarity judgment can be found in the musical information. Musically schooled as well as unschooled listeners have extensive knowledge about music [Deliège et al. 1996; Bigand 2003] and without this knowledge it might not be possible to grasp the deeper musical meaning that underlies the surface structure. We strongly believe that music should always be analyzed within a broader music cognitive and music theoretical framework, and that systems without such additional musical knowledge are incapable of capturing a large number of important musical features de Haas et al. [2010b].

Of particular interest for the current research are the experiments of Carol Krumhansl 1990. Krumhansl is probably best known for her probe-tone experiments in which subjects rated the stability of a tone, after hearing a preceding short musical passage. Non surprisingly, the tonic was rated most stable, followed by the fifth, third, the remaining tones of the scale, and finally the non-scale tones. Also, Krumhansl did a similar experiment with chords: instead of judging the stability of a tone listeners had to judge the stability of all twelve major, minor and diminished triads[2]. The results show a hierarchical ordering of harmonic functions that is generally consistent with music-theoretical predictions: the tonic (I) was the most stable chord, followed by the subdominant (IV) and dominant (V) etc. For a more detailed overview we refer to [Krumhansl 1990; 2004].

These findings can very well be exploited in tonal similarity estimation. Therefore, we base our distance function on a model that not only captures the result found by Krumhansl quite nicely, but is also solidly rooted in music theory: the Tonal Pitch Space model.

2.4 Tonal Pitch Space

The Tonal Pitch Space (TPS) model Lerdahl [2001] builds on the seminal ideas in the *Generative Theory of Tonal Music* Lerdahl and Jackendoff [1996] and is designed to make music theoretical and music cognitive intuitions about tonal organization explicit. Hence, it allows to predict the proximities between musical chords that correspond very well to the findings of Krumhansl [1990]. Although the TPS can can be used to calculate distances between chords in different keys, it is more suitable for calculating distances within local harmonic contexts Bigand and Parncutt [1999]. Therefore the distance measure presented in the next Section only utilizes the parts of TPS needed for calculating the chordal distances within a given key. The TPS is an elaborate model of which we present an overview here, but additional information can be found in [Lerdahl 2001, pages 47 to 59].

[2] A diminished triad is a minor chord with a diminished fifth interval.

(a)	octave (root) level:	0							7					(0)
(b)	fifths level:	0							7					(0)
(c)	triadic (chord) level:	0				4			7					(0)
(d)	diatonic level:	0		2		4	5		7		9		11	(0)
(e)	chromatic level:	0	1	2	3	4	5	6	7	8	9	10	11	(0)
		C	C#	D	E♭	E	F	F#	G	G#	A	B♭	B	(C)

Table 1: The basic space of the tonic chord in the key of C major. The basic space of the tonic chord in the key of C major (C = 0, C# = 1, . . . , B = 11), from Lerdahl Lerdahl [2001].

		2									
		2							9		
		2			5				9		
0		2		4	5		7		9		11
0	1	2	3	4	5	6	7	8	9	10	11
C	C#	D	E♭	E	F	F#	G	G#	A	B♭	B

Table 2: A Dm chord represented in the basic space of C major. Level d is set to the diatonic scale of C major and the levels a-c represent the Dm chord, where the fifth is more stable than the third and the root more stable than the fifth.

The TPS model is a scoring mechanism that takes into account the perceptual importance of the different notes in a chord. The basis of the model is the *basic space* (see Table 1), which resembles a tonal hierarchy of an arbitrary key. In Table 1 the basic space is set to C major. Displayed horizontally are all twelve pitch classes, starting with 0 as C. The basic space comprises five hierarchical levels (a-e) consisting of pitch class subsets ordered from stable to unstable. The first and most stable level (a) is the root level, containing only the root of a chord. The next level (b) adds the fifth of a chord. The third level (c) is the triadic level containing all other pitch classes that are present in a chord. The fourth level (d) is the diatonic level consisting of all pitch classes of the diatonic scale of the current key. The last and least stable level (e) is the chromatic level containing all pitch classes. Chords are represented at level a-c and because the basic space is hierarchical, pitch classes present at a certain level will also present at subsequent levels. The more levels a pitch class is contained in, the more stable the pitch class is and the more consonant this pitch class is perceived by the human listener within the current key. For the C chord in Table 1 the root note (C) is the most stable note, followed by the fifth (G) and the third (E). It is no coincidence that the basic space strongly resembles Krumhansl's 1990 probe-tone data.

Table 2 shows how a Dm chord can be represented in the basic space of (the key of) C major. Now, we can use the basic space to calculate distances between chords by transforming the basic space of a certain chord into the basic space of another chord. In order to calculate the distance between chords, the basic space must first be set to the tonal context–the key–in which the two chords are compared. This is done by shifting pitch classes in the diatonic level (d) in such manner that they match the pitch classes of the scale of the desired key. The distance between two chords depends on two factors: the number of diatonic fifth intervals between the roots of the two compared chords and the number of shared pitch classes between the two chords. These two factors are captured in two rules: the Chord distance rule and the Circle-of-fifths rule (from Lerdahl [2001]):

> Chord distance rule: $d(x, y) = j + k$, where $d(x, y)$ is the distance between chord x and chord y. j is the minimal number of applications of the Circle-of-fifths rule in one direction needed to shift x into y. k is the number of distinctive pitch classes in the levels (a-d) within the basic space of y compared to those in the basic space of x. A pitch class is distinctive if it is present in the basic space of y but not in the basic space of x.

> Circle-of-fifths rule: move the levels (a-c) four steps to the right or four steps to the left (modulo 7) on level d. If the chord root is non-diatonic j receives the maximum penalty of 3.

		2																																	
		2										9																							
		2				5						9																							
0		2		4		5			7			9			11		0		2		4		5			7			9			11		
0	1	2	3	4		5	6		7	8		9	10		11		0	1	2	3	4		5	6		7	8		9	10		11		
C	C#	D	E♭	E		F	F#		G	G#		A	B♭		B		C	C#	D	E♭	E		F	F#		G	G#		A	B♭		B		
(a)																	(b)																		

Table 3: Two examples of the calculation of the TPS chord distance. Example 3a shows the distance between a C chord and a Dm chord. Example 3b shows the distance between C chord and G^7 chord both in the context of a C major key. The distinct pitch classes are underlined.

				4															2																	
				4										11					2										9							
				4				7						11					2				5						9						
0		2		4		5			7			9			11			1	2		4		5	6		7			9			11		
0	1	2	3	4		5	6		7	8		9	10		11		0	1	2	3	4		5	6		7	8		9	10		11		
C	C#	D	E♭	E		F	F#		G	G#		A	B♭		B		C	C#	D	E♭	E		F	F#		G	G#		A	B♭		B		
(a)																	(b)																		

Table 4: Two examples of the calculation of the TPS chord distance. Example 4a shows the chord distance calculation for a G and and a Em chord in the context of C major. Example 4b shows the calculation of the distance between D and a Dm chord in the context of a D major key. The distinct pitch classes are underlined.

The Circle-of-fifths rule makes sense music theoretically because the motion of fifths can be found in cadences throughout the whole of Western tonal music and is pervasive at all levels of tonal organization Piston [1941]. The TPS distance accounts for differences in weight between the root, fifth and third pitch classes by counting the distinctive pitch classes of the new chord at all levels. Two examples of calculation are given in Table 3. Example 3a displays the calculation of the distance between a C chord and a Dm chord in the key of C major. It shows the Dm basic space that has no pitch classes in common with the C major basic space (see Table 1) Therefore, all six underlined pitch classes at the levels c-e are distinct pitch classes. Furthermore, a shift from C to D requires two applications of the circle-of-fifth rule, which yields a total distance of 8. In example 3b one pitch class (G) is shared between the C major basic space and the G^7 basic space. With one application of the circle-of-fifth rule the total chord distance becomes 6.

Two additional examples are given in Table 4. Example 4a shows the calculation of the distance between a G and an Em chord in the key of C major. The G chord and the Em chord have two pitch classes in common and three applications of the circle-of-fifths rule are necessary to transform the G basic space into the Em basic space. Hence, the total distance is 7. Example 4b displays the distance between a D and a Dm in the context of a D major key. There is only one distinct, but non-diatonic, pitch class and no shift in root position yielding a distance of 2.

3 Tonal Pitch Step Distance

On the basis of the TPS chord distance rule, we define a distance function for chord sequences, named the Tonal Pitch Step Distance (TPSD). The TPSD compares two chord sequences and outputs a number between 0 and the maximum chordal distance 20. A low score indicates two very similar chord sequences and a high score indicates large harmonic differences between two sequences.

The central idea behind the TPSD is to compare the change of chordal distance to the tonic over time. This is done by calculating the TPS chord distance between each chord of the song and the tonic triad of the key of the song. The reason for doing so, is that if the distance function is based on comparing subsequent chords, the chord distance depends on the exact progression by which that chord was reached. This is undesirable because very similar but not identical chord sequences can then produce radically different scores.

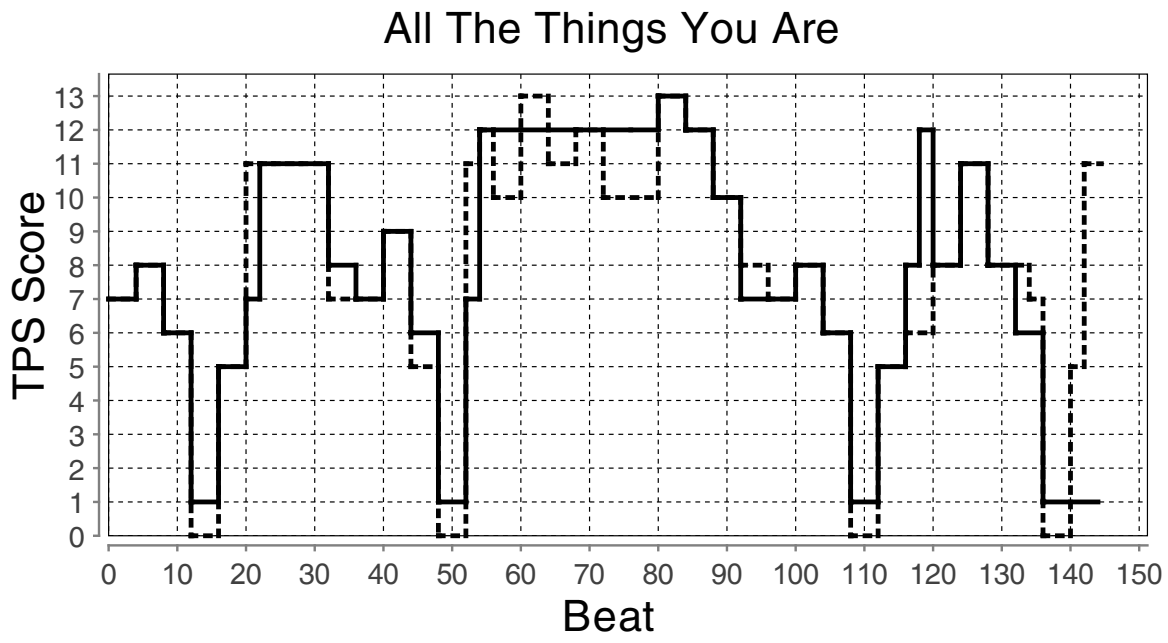

Figure 2: A plot demonstrating the comparison of two similar versions of *All the Things You Are* using the TPSD. The total area between the two step functions, normalized by the duration of the shortest song, represents the distance between both songs. A minimal area is obtained by shifting one of the step functions cyclically.

Plotting the chordal distance against the time results in a step function. The difference between two chord sequences can then be defined as the minimal area between the two step functions f and g over all possible horizontal shifts t of f over g (see Figure 2). These shifts are cyclic. To prevent longer sequences from yielding higher scores, the score is normalized by dividing it by the length of the shortest step function. Trivially, the TSPD can handle step functions of different length since the area between non-overlapping parts is always zero.

The calculation of the area between f and g is straightforward. It can be calculated by summing all rectangular strips between f and g, and trivially takes $O(n + m)$ time where n and m are the number of chords in f and g, respectively. An important observation is that if f is shifted along g, a minimum is always obtained when two vertical edges coincide. Consequently, only the shifts of t where two edges coincide have to be considered, yielding $O(nm)$ shifts and a total running time of $O(nm(n + m))$.

This upper bound can be improved. Arkin et al. [1991] developed an algorithm that minimized the area between two step functions by shifting it horizontally as well as vertically in $O(nm \log nm)$ time. The upper bound of their algorithm is dominated by a sorting routine. We adapted the algorithm of Arkin et al. in two ways for our own method: we shift only in the horizontal direction and since we deal with discrete time steps we can sort in linear time using counting sort Cormen et al. [2001]. Hence, we achieve an upper bound of $O(nm)$.

3.1 Metrical Properties of the TPSD

For retrieval and especially indexing purposes it has several benefits for a distance measure to be a metric. The TPSD would be a metric if the following four properties held, where $d(x, y)$ denotes the TPSD distance measure for all possible chord sequences x and y:

1. *non-negativity:* $d(x, y) \geq 0$ for all x and y.

										9							1																			
				4						9							1							7	8										
0				4						9				0	1	2	3	4	5	6		7	8		9	10	11								
0		2		4	5		7			9		11		0	1	2	3	4	5	6		7	8		9	10	11								
0	1	2	3	4	5	6	7	8		9	10	11		0	1	2	3	4	5	6		7	8		9	10	11								
C	C#	D	E♭	E	F	F#	G	G#		A	B♭	B		C	C#	D	E♭	E	F	F#		G	G#		A	B♭	B								
(a)														(b)																					

Table 5: An example of the minimal TPS chord distance and the maximal TPS chord distance. In example (a) two Am chords are compared yielding a distance of 0. In example (b) a C chord is compared to a C# chord with all possible additions resulting in a distance of 20. The distinct pitch classes are underlined.

2. *identity of indiscernibles:* $d(x, y) = 0$ if and only if $x = y$.
3. *symmetry:* $d(x, y) = d(y, x)$ for all x and y.
4. *triangle inequality:* $d(x, z) \leq d(x, y) + d(y, z)$ for all x, y and z.

Before we look at the TPSD it is good to observe that the TPS model has a minimum and a maximum (see Table 5). The minimal TPS distance can obviously be obtained by calculating the distance between two identical chords. In that case there is no need to shift the root and there are no uncommon pitch classes yielding a distance of 0. This maximum TPS distance can be obtained, for instance, by calculating the distance between a C major chord and C# chord containing all twelve pitch classes. The Circle-of-fifths rule is applied three times and the number of distinct pitch classes in the C# basic space is 17. Hence, the total score is 20.

The TPSD is clearly non-negative since the length of the compared pieces, a and b, will always be $a \geq 0$ and $b \geq 0$, the area between the two step functions and hence the TPSD will always be $d(x, y) \geq 0$. The TPSD is symmetrical: when we calculate $d(x, y)$ and $d(y, x)$ for two pieces x and y, the shortest of the two step functions is fixed and the other step function is shifted to minimize the area between the two, hence the calculation of $d(x, y)$ and $d(y, x)$ is identical. However, the TPSD does not satisfy the identity of indiscernibles property because more then one chord sequence can lead to the same step function, e.g. C G C and C F C in the key of C major, all with equal durations. The TPS distance between C and G and C and F is both 5, yielding two identical step functions and a distance of 0 between these two chord sequences. The TPSD also does not satisfy the triangle inequality. Consider three chord sequences, x, y and z, where x and z are two different chord sequences that the share one particular subsequence y. In this particular case the distances $d(x, y)$ and $d(y, z)$ are both zero, but the distance $d(x, z) > 0$ because x and y are different sequences. Hence, for these three chord sequences $d(x, z) \leq d(x, y) + d(y, z)$ does not hold.

4 Experiment

The retrieval capabilities of the TPSD were analyzed and compared to the CSAS in an experiment in which we tried to retrieve similar but not identical songs on the basis of similarity values of the tested distance functions. The distance TPSD between every sequence was calculated and for every song a ranking was constructed by sorting the other songs on the basis of their TPSD. Next, these rankings were analyzed. To place the performance of these distance functions and the difficulty of the task in perspective, the performance of the TPSD was compared with a baseline algorithm.

This baseline methods is the edit distance Levenshtein [1966] between the two chord sequences represented as a string. However, one might consider this an unfair comparison because the TPSD has more information it can exploit than the edit distance, namely temporal information, i.e. the length of the steps in the step functions. To make the comparison fair we incorporated temporal information in the strings that were matched with the edit distance: for every beat that did not contain a chord a dot was inserted into the string. Still for when matching a piece with a version of itself that is transposed to another key the edit distance would yield a disappointing performance. To overcome this problem we transpose all songs to C.

|Fm7 . . .	|Bbm7 . . .	|Eb7 . . .	|AbMaj7 . . . |
|DbMaj7 . . .	|Dm7b5 . G7b9 .	|CMaj7 . . .	|CMaj7 . . . |
|Cm7 . . .	|Fm7 . . .	|Bb7 . . .	|Eb7 . . . |
|AbMaj7 . . .	|Am7b5 . D7b9 .	|GMaj7 . . .	|GMaj7 . . . |
|A7 . . .	|D7 . . .	|GMaj7 . . .	|GMaj7 . . . |
|Gbm7 . . .	|B7 . . .	|EMaj7 . . .	|C+ . . . |
|Fm7 . . .	|Bbm7 . . .	|Eb7 . . .	|AbMaj7 . . . |
|DbMaj7 . . .	|Dbm7 . Gb7 .	|Cm7 . . .	|Bdim . . . |
|Bbm7 . . .	|Eb7 . . .	|AbMaj7 . . .	|. . . . |

Table 6: A leadsheet of the song *All The Things You Are*. A dot represents a beat, a bar represents a bar line, and the chord labels are presented as written in the Band-in-a-Box file.

Class Size	Frequency	Percent
1	3,253	82.50
2	452	11.46
3	137	3.47
4	67	1.70
5	25	.63
6	7	.18
7	1	.03
8	1	.03
10	1	.03
Total	5028	100

Table 7: The distribution of the song class sizes in the Chord Sequence Corpus

4.1 A Chord Sequence Corpus

For this experiment a large corpus of musical chord sequences was assembled. The Chord Sequence Corpus consists of 5,028 unique human-generated Band-in-a-Box files that are collected from the Internet. Band-in-a-Box is a commercial software package Gannon [1990] that is used to generate musical accompaniment based on a lead sheet. A Band-in-a-Box file stores a sequence of chords and a certain style, whereupon the program synthesizes and plays a MIDI-based accompaniment. A Band-in-a-Box file therefore contains a sequence of chords, a melody, a style description, a key description, and some information about the form of the piece, i.e. the number of repetitions, intro, outro etc. For extracting the chord label information from the Band-in-a-Box files we have extended software developed by Simon Dixon and Matthias Mauch 2007. An example of a chord sequence as found in a Band-in-a-Box file describing the chord sequence of *All the Things You Are* is given in Table 6.

All songs of the chord sequence corpus were collected from various Internet sources. These songs were labeled and automatically checked for having a unique chord sequence. All chord sequences describe complete songs and songs with fewer than 3 chords or shorter than 16 beats were removed from the corpus. The titles of the songs, which function as a ground-truth, as well as the correctness of the key assignments, were checked and corrected manually. The style of the songs is mainly jazz, latin and pop.

Within the collection, 1775 songs contain two or more similar versions, forming 691 classes of songs. Within a song class, songs have the same title and share a similar melody, but may differ in a number of ways. They may, for instance, differ in key and form, they may differ in the number of repetitions, or have a special introduction or ending. The richness of the chords descriptions may also diverge, i.e. a $C^{7\flat9\flat13}$ may be written instead of a C^7, and common substitutions frequently occur. Examples of the latter are relative substitution, i.e. Am instead of C, or tritone substitution, e.g. $F\#^7$ instead of C^7. Having multiple chord sequences describing the same song allows for setting up a *cover-song* finding experiment. The the title of the song is used as ground-truth and the retrieval challenge is to find the other chord sequences representing

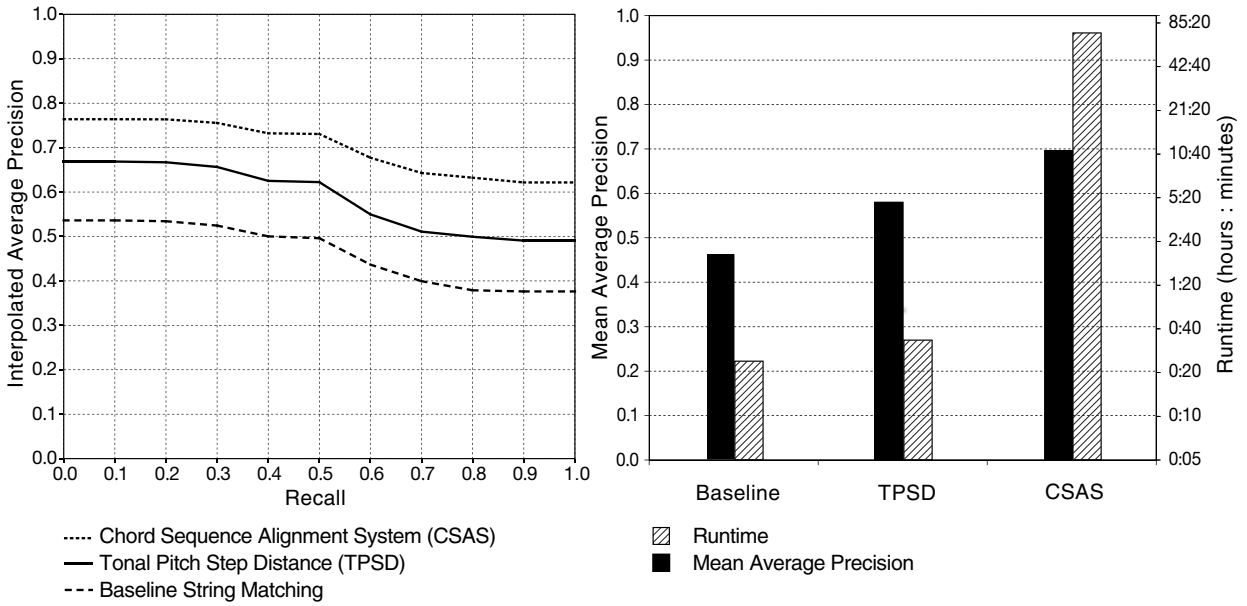

Figure 3: The graph on the left shows the average interpolated precision and recall graph of the baseline string matching approach (red), the TPSD (green) and the CSAS (blue). The plot on the right shows the MAP and Runtimes of the three algorithms. The MAP is displayed on the left axis and the runtimes are displayed on an logarithmic scale on the right axis.

the same song.

In de Haas et al. [2010a] we experimented with the amount of information in the chord that we used as input for the algorithms. The data contained a wealth of different rich chord descriptions, but using only the triad as input for our algorithms gave a significantly better retrieval performance. Discarding chord additions might be seen as a form of syntactical noise-reduction, since these additions, if they do not have a voice leading function, have a rather arbitrary character and can only add some harmonic spice. Hence, in the current experiment we also only used triadic chord information

The distribution of the song class sizes is displayed in Table 7 and gives an impression of the difficulty of the retrieval task. Generally, Table 7 shows that the song classes are relatively small and that for the majority of the queries there is only one relevant document to be found. It furthermore shows that 82.5% of the songs is in the corpus for distraction only. The chord sequence corpus is available to the research community on request.

4.2 Results

We analyzed the rankings of all 1775 queries with 11-point precision recall curves and Mean Average Precision (MAP, see Figure 3). We calculated the interpolated average precision as in Manning et al. [2008] and probed it at 11 different recall levels. In all evaluations the queries were excluded from the analyzed rankings. The graph shows clearly that the overall retrieval performance of all algorithms can be considered good, but that the CSAS outperforms the TPSD and both the TPSD and the CSAS outperform the baseline edit distance.

In Figure 3 we also present the MAP and the runtimes of the three algorithms on two different axes. The MAP is displayed on the left axis and the runtimes are shown on right axis that has an exponential scale doubling the amount of time at every tick. The MAP is a single-figure measure, which measures the precision at all recall levels and approximates the area under the (uninterpolated) precision recall graph Manning et al. [2008]. Having a single measure of retrieval quality makes it easier to evaluate the significance

of the differences between results. We tested whether the differences in MAP were significant by performing a non-parametric Friedman test, with a significance level of $\alpha = .05$. We chose the Friedman test because the underlying distribution of the data is unknown and in contrast to an ANOVA the Friedman does not assume a specific distribution of variance. There were significant differences between the runs, $\chi^2(2, N = 1775) = 274$, $p < .0001$. To determine which of the pairs of measurements differed significantly we conducted a post hoc Tukey HSD test[3]. Opposed to a T-test the Tukey HSD test can be safely used for comparing multiple means Downie [2008].

Also the MAP chart confirms the differences between algorithms, both in performance and in runtime. With a MAP of .70 the CSAS significantly outperforms the TPSD with a MAP of .58. Both the CSAS and the TPSD in significantly outperform the baseline string matching approach. The retrieval performance of the CSAS is good, but comes at a price. The CSAS run took about 73 hours which is considerably more than the 33 minutes of the TPSD or the 24 minutes of the edit distance. Hence, the TPSD offers the best quality-runtime ratio.

5 Case Study: Relating Harmony and Melody in Bach's Chorales

In this Section we show how a chord labeling algorithm can be combined with the TPSD and demonstrate how the TPSD can aid in answering musicological questions. More specifically, we will investigate whether melodically related chorale settings by J.S. Bach (1685-1750) are also harmonically related. Doing analyses of this kind by hand is very time consuming, especially when the corpus involved has a substantial size. Moreover, the question whether two pieces are harmonically related can hardly be answered with a simple yes or no. Pieces are harmonically similar up to a certain degree; forcing a binary judgment requires placing a threshold that is not trivial to choose and maybe not even meaningful from a musical point of view. However, for a well-trained musicologist determining whether two melodies stem from the same tune family is a relatively simple task.

Chorales are congregational hymns of the German Protestant church service Marshall and Leaver [2010]. Bach is particularly famous for the imaginative ways in which he integrated these melodies into his compositions. Within these chorale-based compositions, the so-called *Bach chorales* form a subset consisting of relatively simple four-voice settings of chorale melodies in a harmony-oriented style often described as 'Cantionalsatz' or 'stylus simplex'. Bach wrote most of these chorales as movements of large-scale works (cantatas, passions) when he was employed as a church musician in Weimar (1708-1717) and Leipzig (1723-1750) Wolff et al. [2010]. A corpus of Bach chorales consisting of 371 items was posthumously published by C.P.E. Bach and J.P. Kirnberger in 1784-87, but some more have been identified since. This publication had a didactic purpose: the settings were printed as keyboard scores and texts were omitted. Consequently, over the last two centuries, the chorales have been widely studied as textbook examples of tonal harmony. Nevertheless, they generally provide very sensitive settings of specific texts rather than stereotyped models and, despite their apparent homogeneity, there is quite some stylistic variation and evidence of development over time. Yet one can claim that Bach's chorale hamonizations were constrained by the general rules of tonal harmony in force in the first half of the 18^{th} century and that the range of acceptable harmonizations of a given melody was limited.

We hypothesize that if two melodies are related, the harmonizations are also related and melodically similar pieces can are also harmonically similar. To determine whether the melodies of two chorales are indeed related, we asked an expert musicologist to inspect the melodies that have the same title and to decide if these melodies belong to the same tune family. If they do, it should be possible to retrieve these settings by ranking them on the basis of their TPSD distance.

5.1 Experiment

To test whether the melodically related Bach Chorales were also harmonically related, we performed a retrieval experiment similar to the one in Section 4. We took 357 Bach Chorales and used the TPSD to

[3]All statistical tests were performed in Matlab 2009a.

Tune Family Size	Frequency	Percent
1	136	68.34
2	24	12.06
3	17	8.54
4	10	5.03
5	5	2.51
6	3	1.51
7	4	2.01
11	1	0.50
Total	357	100

Table 8: Tune family distribution in th Bach Chorales Corpus

determine how harmonically related these chorales were. Next, we used every chorale that belonged to a tune family, as specified by our musicological expert, as a query, yielding 219 queries, and created a ranking based on the TPSD. Subsequently we analyzed the rankings with standard retrieval performance evaluation methods to determine whether the melodically related chorales could be found on the basis of their TPSD.

The chorales scores are freely available[4] in MIDI format Loy [1985]. But as explained in the previous sections, the TPSD takes chords as input, not MIDI notes. We therefore use David Temperly's Chord root tracker Temperley [2001], which is part of the Melisma music analyzer[5]. The chord root tracker does not produce a label for a segment of score data like we have seen in the rest of this paper. It divides the piece into chord spans and it assigns a root label to each chord span. Thus, it does not produce a complete chord label, e.g. A♭m^9 but, this is not a problem, because the TPS model needs only to know which pitch class is the root and which one is the fifth. Once it is know which pitch class is the root, it is trivial to calculate which pitch class is the fifth. The remainder of the pitch classes in the chord is placed at level c of the basic space. The Melisma chord root tracker is a rule-based algorithm. It utilizes a metrical analysis of the piece performed by the meter analyzer, which is also part of the Melisma Music analyzer, and uses a small number of music theoretically inspired preference rules to determine the chord root. We segmented the score such that each segment contained at least two simultaneously sounding notes. Manually annotating a small random sample yields a correctness of the root tracker of approximately 80%, which is in line with the 84% as claimed in Temperley [2001].

The TPSD also requires to know the key of all chorales. The key information was generously offered by Martin Rohrmeier, who investigated the the distributions of the different chord transitions within the Chorales Corpus Rohrmeier and Cross [2008]. We selected the chorales of which both the MIDI data, a pdf score (for our musicological expert) and the key description was available. After preparation, which included checking for chorale doublets, the corpus contained 357 pieces.

5.2 results

We analyze the TPSD based rankings of Bach's chorales with a average interpolated precision versus recall plot, which is displayed in the graph in Figure 4. To place the results into context and have an idea of the structure of the corpus, we also printed the distribution of the sizes of the tune families in Table 8. The graph in Figure 4 shows clearly that a large section of the chorales that are based on the same melody can be found by analyzing only their harmony patterns. In general we can conclude that in some melodically similar pieces can be found by looking at their harmony alone. This is supported by a recognition rate, i.e. the percentage of queries that have a melodically related chorale at rank one (excluding the query), of .71. However, a considerable amount of pieces cannot be retrieved on the basis of their TPSD: in 24 percent of the queries had the first related chorale is not within the first ten retrieved chorales.

[4]See http://www.jsbchorales.net/ (accessed May 24, 2011) for more information.

[5]The source code of the Melisma Music Analyzer is freely available at: http://www.link.cs.cmu.edu/music-analysis/ (accessed May 24, 2011).

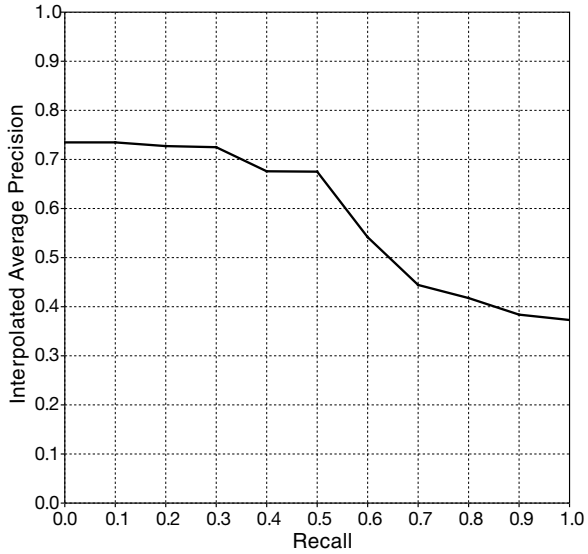

Figure 4: The average interpolated precision for ten different recall levels of the melodically related bach chorales retrieved on the basis of their TPSD scores.

This can have three reasons: the chorales are not harmonically related, the TPSD did not succeed in capturing the harmonic similarity well enough, or errors in the automatic chord labeling disturb the similarity measurement. We made a non-exhaustive analysis of the retrieval output in order to get a better idea of the issues at stake, focusing on the larger tune families. First, it appears that for some chorales the retrieval performance is very high. Perfect retrieval was attained for *Wo Gott, der Herr, nicht bei uns hält* (5 items), *Was Gott tut das ist wolgetan and Wenn mein Stundlein* (both 4 items). Tune families with near-perfect retrieval include *Jesu meine Freude; Werde munter, mein Gemüte* (both 6 items, 2 false positives in total) and *Auf meinen lieben Gott* (5 items, 1 false positive in total). Retrieval is also very good for the largest group, *O Welt, ich muß dich lassen* (11 items). For each member all of the top-5 hits are from the same tune family, and for most members all other items are ranked within the top-20. Only one item has more than one relevant item ranked below 20 (BWV[6] 394).

Herzlich tut mich verlangen (7 items) presents a musically interesting situation: there seem to be two clusters, one of four and one of three items. Chorales from each of the two clusters match very well to one another, but chorales from the other cluster are consistently ranked low. From a melodic point of view, there are only a few unimportant differences. The harmony is very different for the two groups, though. The four-item cluster consists of settings in the major mode, with the chorale melody ending on the third of the final chord. The three-item cluster contains settings that are in the minor mode: the chorale melody ends on the root of the final chord, but this chord itself acts as a V in the rest of the piece. Generally, the larger tune families seem to consist of a cluster of very similar items and one or two items that fall outside the clusters. These 'outliers' generally rank the clustered items relatively low. There are some factors that may explain outliers:

Different meter

The default meter for chorale melodies is $\frac{4}{4}$. However, variants in $\frac{3}{4}$ exist for several chorales. In these the basic rhythmic pattern of two quarter notes is changed into a half note followed by a quarter note. This

[6]The Bach-Werke-Verzeichnis (BWV) is a numbering system designed to order an identify the compositions by Johann Sebastian Bach. The works are numbered thematically, not chronologically and the prefix BWV, followed by the work's number has become a standard identifier for Bach's compositions.

has three effects: the total length of the melody changes, some chords are extended because they follow the durations of the melody notes, and extra chords may be inserted on the second part of the half notes. All three factors lead to a high TPSD score when comparing chorale settings from the same tune family with different meters. Examples include *Wie nach einer Wasserquelle* (two outliers in $\frac{3}{4}$ meter) and *Nun lob, mein Seel, den Herren* (three versions in $\frac{3}{4}$, one, the outlier, in $\frac{4}{4}$ meter).

Phrase length

Individual phrases in a chorale melody typically end with a note with a fermata, which may or may not have indicated a prolongation of this note in performance. Sometimes however, fermatas are written out, replacing a quarter note by a dotted half note. Also, notes within the phrase are sometimes extended. Both of these situations create an asynchrony in the step function that contributes to a higher TPSD score. Both situations occur in the two versions of the melody *O Ewigkeit, du Donnerwort*, so that the two settings match each other particularly badly.

Additional instrumental parts

Some of the chorales have additional instrumental parts. If they are written in the same style as the vocal parts, this seems to present no particular problems. However, when they are different, this may lead to a higher TPSD score. An example of this is *Die Wollust dieser Welt* (4 settings, 1 outlier). The outlier has an instrumental bass moving in eighth notes, which leads many additional chord labels on weak beats. Since these labels are often dissonant chords, the TPSD score with 'normal' settings–which would have the second half of a more consonant chord at the corresponding place–increases.

Differences in polyphony

There are a number of settings that are much more polyphonic than most of the others. Some of these may actually be instrumental organ works written out in four voices. The rhythmic and melodic behavior of the voices is very different. An example is *Christ lag in Todesbanden* (5 items, 2 outliers). Of the outliers, BWV 278 is particularly noticeable for its inner voices moving often in sixteenth notes and chromaticism. Here too a likely explanation is that extra, often dissonant chord labels are generated.

The last two points are related to a limitation of the TPSD, namely that all chords are considered equally important to the overall perception of harmonic similarity. In fact, chords have hierarchical relationships to each other, and in addition their contribution depends on the metric position of their onsets.

False negatives, items that get a high rank but belong to a different tune family, are informative as well. Sometimes these indeed appear to have an interesting relationship, as in the case of *Was mein Gott will*. Two settings of this melody also retrieve items with the melody *Wo Gott, der Herr, nicht bei uns hält*. It appears that the harmony of the first eight bars is very similar and that the melodies themselves also could be considered related. However, most of the false negatives are difficult to interpret. One reason is the cyclic shifting, which causes an arbitrary alignment between items that disrupts the phrase structure or may even lead to a match that includes a jump from the end of the piece to its beginning. Another reason is that different chords may have the same TPSD score, and that similar step functions may be generated by chord sequences that are musically quite different.

A different way of analyzing false negatives is by looking into the average rank of each item over all queries. Ideally, the average rank should be normally distributed over all items in the collection, with a mean of half the collection size and a small standard deviation. Deviations from this ideal indicate that the similarity measure is sensitive to certain properties in the collection. In particular, items with a high average rank are likely have certain properties that make them match to a large number of unrelated items. We studied the 15 pieces with the highest average rank and the 15 pieces with the lowest average rank and found clear patterns. The 15 pieces with the highest rank were all pieces in a minor key and the pieces with the lowest average rank were mainly major. Also, the pieces with a low average rank tend to be relatively

long and the high-ranked ones tend to be relatively short. This indicates that the TPSD is susceptible to differences in length and key.

Nevertheless, we can conclude that a considerable number of pieces of the Bach chorales corpus that share the same melody could be shown to be also harmonically related.

6 Concluding Remarks

We presented a new geometric distance measure that captures the harmonic distance between two sequences of musical harmony descriptions, named the Tonal Pitch Step Distance (TPSD). This distance is based on the changes of the distance between chord and key as given by using Lerdahl's Tonal Pitch Space model. This cognitive model correlates with empirical data from psychology and matches music-theoretical intuitions. A step function is used to represent the change of chordal distance to its tonic over time and the distance between two chord progressions is defined as the minimal area between two step functions. The TPSD is a distance measure that is simple to use, does not require a lot of parameter tuning, is key invariant, can be computed efficiently, and matches human intuitions about harmonic similarity.

The performance of the TPSD can still be considered good, especially if one considers the size of the test corpus used in the experiment and the relatively small classsizes (see Table 7). We compared the performance of the TPSD to the performance of baseline string matching approach and a Chord Sequence Alignment System (CSAS). Both the TPSD and the CSAS significantly outperform the baseline string matching approach. In turn, the CSAS outperforms the TPSD significantly, but is prohibitively costly to use. Hence, the TPSD has a better performance-runtime ratio than the CSAS. We furthermore demonstrated how the TPSD can contribute to the musicological discussions on melody and harmony in Bach's Chorales in a case study. In this case study we showed that a for a considerable number of Bach Chorales that share a melody also are harmonically related.

Nevertheless, there is still room for improvement. The TPSD cannot deal with large structural changes, e.g. adding repetitions, a bridge, etc. A prior analysis of the structure of the piece combined with partial matching could improve the retrieval performance. Another important issue is that the TPSD treats all chords as equally important. This is musicologically not plausible. Considering the musical function in the local as well as global structure of the chord progression, like is done in de Haas et al. [2009] might improve retrieval results.

The performance of the TPSD was only tested on symbolic data in this paper. Nevertheless, the application TPSD is not limited to symbolic music and audio applications are currently investigated. Especially the recent developments in chord label extraction are very promising because the output of these methods could be matched directly with the systems here presented. The good performance of the TPSD lead us to believe that also in other musical domains, such as audio, retrieval systems will benefit from harmonic similarity based matching in the near future.

7 Acknowledgments

We would like to thank Peter van Kranenburg for comparing and annotating the similarity of the melodies of the Bach chorales used in Section 5 and Martin Rohrmeier for providing the information about musical key of the same corpus.

References

Arkin, E., Chew, L., Huttenlocher, D., Kedem, K., and Mitchell, J. (1991). An efficiently computable metric for comparing polygonal shapes. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 13(3):209–216.

Bello, J. and Pickens, J. (2005). A robust mid-level representation for harmonic content in music signals. In *Proceedings of the 6th International Symposium on Music Information Retrieval (ISMIR)*, pages 304–311.

Bigand, E. (2003). More about the musical expertise of musically untrained listeners. *Annals of the New York Academy of Sciences*, 999:304–312.

Bigand, E. and Parncutt, R. (1999). Perceiving musical tension in long chord sequences. *Psychological Research*, 62(4):237–254.

Cormen, T., Leiserson, C., Rivest, R., and Stein, C. (2001). *Introduction to Algorithms*. MIT press.

Deliège, I., Mélen, M., Stammers, D., and Cross, I. (1996). Musical schemata in real time listening to a piece of music. *Music Perception*, 14(2):117–160.

Downie, J. (2008). The music information retrieval evaluation exchange (2005–2007): A window into music information retrieval research. *Acoustical Science and Technology*, 29(4):247–255.

Gannon, P. (1990). Band-in-a-Box. PG Music.

de Haas, W. B., Robine, M., Hanna, P., Veltkamp, R. C., and Wiering, F. (2010a). Comparing Approaches to the Similarity of Musical Chord Sequences. In *Proceedings of the 7th International Symposium on Computer Music Modeling and Retrieval (CMMR)*, pages 299–315.

de Haas, W. B., Rohrmeier, M., Veltkamp, R. C., and Wiering, F. (2009). Modeling harmonic similarity using a generative grammar of tonal harmony. In *Proceedings of the 10th International Conference on Music Information Retrieval (ISMIR)*.

de Haas, W. B., Veltkamp, and Wiering, F. (2010b). Hooked on Music Information Retrieval. *Empirical Musicology Review*, page in press.

de Haas, W. B., Veltkamp, R. C., and Wiering, F. (2008). Tonal pitch step distance: A similarity measure for chord progressions. In *Proceedings of the 9th International Conference on Music Information Retrieval (ISMIR)*, pages 51–56.

Hanna, P., Robine, M., and Rocher, T. (2009). An alignment based system for chord sequence retrieval. In *Proceedings of the 2009 Joint International Conference on Digital Libraries*, pages 101–104. ACM New York, NY, USA.

Klapuri, A. and Davy, M. (2006). *Signal processing methods for music transcription*. Springer, New York.

Krumhansl, C. (1990). *Cognitive Foundations of Musical Pitch*. Oxford University Press, USA.

Krumhansl, C. (2004). The cognition of tonality - as we know it today. *Journal of New Music Research*, 33(3):253–268.

Lerdahl, F. (2001). *Tonal Pitch Space*. Oxford University Press.

Lerdahl, F. and Jackendoff, R. (1996). *A Generative Theory of Tonal Music*. MIT Press.

Levenshtein, V. I. (1966). Binary codes capable of correcting deletions, insertions, and reversals. *Cybernetics and Control Theory*, 10(8):707–710.

Logan, B. (2000). Mel frequency cepstral coefficients for music modeling. In *Proceedings of the 1th Society for Music Information Retrieval Conference (ISMIR)*.

Loy, G. (1985). Musicians make a standard: the MIDI phenomenon. *Computer Music Journal*, 9(4):8–26.

Manning, C. D., Raghavan, P., and Schütze, H. (2008). *Introduction to Information Retrieval*. Cambridge University Press, New York, NY, USA.

Marshall, R. L. and Leaver, R. A. (accessed December 24, 2010). Chorale. In *Grove Music Online*. Oxford Music Online, http://www.oxfordmusiconline.com/subscriber/article/grove/music/05652.

Mauch, M. (2010). *Automatic chord transcription from audio using computational models of musical context*. PhD thesis, Queen Mary University of London.

Mauch, M., Dixon, S., Harte, C., Casey, M., and Fields, B. (2007). Discovering chord idioms through beatles and real book songs. In *Proceedings of the 8th International Conference on Music Information Retrieval (ISMIR)*, pages 255–258.

Paiement, J.-F., Eck, D., and Bengio, S. (2005). A probabilistic model for chord progressions. In *Proceedings of the 6th International Conference on Music Information Retrieval (ISMIR)*, pages 312–319, London, UK.

Pickens, J. and Crawford, T. (2002). Harmonic models for polyphonic music retrieval. In *Proceedings of the 11th International Conference on Information and Knowledge Management*, pages 430–437. ACM New York, NY, USA.

Piston, W. (1941). *Harmony*. Norton, W. W. & Company, New York.

Rohrmeier, M. and Cross, I. (2008). Statistical properties of tonal harmony in Bachs chorales. In *Proceedings of the 10th International Conference on Music Perception and Cognition (ICMPC)*, pages 619–627.

Swierstra, S. D. (2009). *Combinator Parsing: A Short Tutorial*, pages 252–300. Springer-Verlag.

Temperley, D. (2001). *The Cognition of Basic Musical Structures*. Cambridge, MA, MIT Press.

Wakefield, G. H. (1999). Mathematical representation of joint time-chroma distributions. In *Part of the Society of Photographic Instrumentation Engineers Conference on Advanced Signal Processing Algorithms, Architectures, and Implementations (SPIE)*, pages 637–645.

Wolff, C., Emery, W., Wollny, P., Leisinger, U., and Roe, S. (accessed December 24, 2010). Bach. In *Grove Music Online*. Oxford Music Online, http://www.oxfordmusiconline.com/subscriber/article/grove/music/40023pg10.