Int J Multimed Info Retr (2013) 2:189–202
DOI 10.1007/s13735-013-0036-6

REGULAR PAPER

A geometrical distance measure for determining the similarity of musical harmony

W. Bas de Haas · Frans Wiering · Remco C. Veltkamp

Received: 17 January 2012 / Revised: 25 July 2012 / Accepted: 6 February 2013 / Published online: 18 June 2013

Abstract In the last decade, digital repositories of music have undergone an enormous growth. Therefore, the availability of 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. We show that it can be efficiently used for retrieving similar chord sequences, and that it significantly outperforms a baseline string matching approach. Although the proposed method is not the best performing distance measure, it offers the best quality–runtime ratio. Furthermore, we demonstrate in a case study how our harmonic similarity measure can contribute to the musicological discussion of the melody and harmony in large-scale corpora.

Keywords Harmony · Music information retrieval · Similarity · Step function · Tonality

W. B. de Haas is supported by the Netherlands Organization for Scientific Research, NWO-VIDI grant 276-35-001. This article has been made open access with funding of the NWO Stimuleringsfonds Open Access (036.002.298).

W. B. de Haas (✉) · F. Wiering · R. C. Veltkamp
Utrecht University, PO Box 80.089, 3508 TB Utrecht, The Netherlands
e-mail: W.B.deHaas@uu.nl

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, provide access to 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 access to these repositories 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*.

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. These features, e.g., chroma features [29] or

[1] Within this paper, MIR refers to *music* (and not multimedia) information retrieval.

Mel-Frequency Cepstral coefficients MFCCs [19], do not directly translate to the notes, beats, voices and instruments that are used in the symbolic domain. Of course, much depends on the application or task at hand, but we believe that for judging the musical content of an audio source, translating the audio features into a high-level representation, which contains descriptors that can be musically interpreted, should be preferred. Although much progress has been made, automatic polyphonic music transcription is a difficult problem, and is currently too unreliable to use as a preprocessing step for similarity estimation. Hence, 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. As a consequence, for applying our method to audio data, we rely on one of the available chord labeling methods (See Sect. 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 Sect. 2. Next, we will present the Tonal pitch step distance in Sect. 3. In Sect. 4, we show how our distance measure performs in practice and we show that it can also contribute to musicological discussions in Sect. 5. But first, we will give a brief introduction on what actually constitutes tonal harmony and harmonic similarity.

1.1 What is harmony?

Within Western tonal music, it is common to represent a sound with a fixed frequency by a *note*. All notes have a name, e.g., C, D, E, etc. 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 (12 semitones). When two notes are an octave apart, the highest note will have exactly twice the frequency of the lower. These two notes are perceived by listeners as very similar, so similar even that all notes one or more octave apart have the same name. Hence, these notes are said to be in the same *pitch class*.

Harmony arises in music when two or more notes sound at the same time.[2] These simultaneously sounding notes form *chords*, which can in turn be used to form chord sequences. The two most important factors that characterize a chord are its structure, determined by the intervals between the notes, and the chord's *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. The most basic chord is the *triad*, which consists of a root and two pitch classes a third and a fifth interval above the root. If the third interval in a triad is a major third, the triad is called a *major triad*, if it is a minor third, the triad is called a *minor triad*. 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 (C–E) and a fifth interval (C–G) are added, yielding a C-major chord. Above the score the names of the chords, which are based on the root notes, are printed.

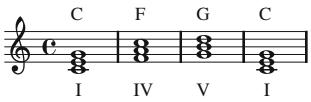

Fig. 1 A very typical and frequently used chord progression in the key of C-major, often referred to as I-IV-V-I. Above the score the chord labels, representing the notes of the chords in the section of the score underneath the label, are printed. The roman numbers below the score denote the interval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

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, pitch class in that piece. Moreover, the key specifies the *scale*, which is the set of pitches that occur most frequently, and that sound reasonably well together. Chords can be created from pitches that belong to the scale, or they can borrow notes from outside the scale, the latter being more dissonant. The root note of a chord has an especially distinctive role, because the interval of the chord root and the key largely determine the *harmonic function* of the chord. The most important harmonic functions are the dominant (V) that builds up tension, a sub-dominant (IV) that can prepare a dominant, and the tonic (I) that releases tension. In Fig. 1, roman numbers denote the interval between the root of the chord and the key, often called *scale degrees*, are printed underneath the score.

Obviously, this is a rather basic view of tonal harmony. For a thorough introduction to tonal harmony, we refer the reader to [26]. 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. For this reason, Western tonal harmony has become one of the most prominently investigated topics in music

[2] One can even argue that notes played successively within a short time frame also induce harmony.

theory and can be considered a feature of music that is quite as distinctive as rhythm or melody. Nevertheless, harmonic structure as a feature for music retrieval has received far less attention than melody or rhythm.

1.2 Harmonic similarity and its application in MIR

Harmonic similarity depends not only on musical information, but also largely on the interpretation of this information by the human listener. Musicians as well as non-musicians have extensive culture-dependent knowledge about music that needs to be taken into account while modeling music similarity [4,6]. Hence, we believe that music only becomes music in the mind of the listener, and that not all information needed for making good similarity judgments can be found in the musical data alone [10].

In this light, we consider the harmonic similarity of two chord sequences to be the degree of agreement between structures of simultaneously sounding notes including the agreement between global as well as local relations between these structures 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 thus contribute to the harmonic similarity of musical works.

Harmonic similarity also has practical value and 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 can be used to retrieve harmonically related chord sequences. The retrieval performance is examined in an experiment on 5,028 human-generated chord sequences, in which we compare it to two other harmonic distance functions and measure the effect of the chord representation. Although the proposed distance measure is not the best performing measure, it is much faster and offers the best quality–runtime ratio. We furthermore show in a case study how the proposed measure can contribute to the musicological discussion of the relation between melody and harmony in melodically similar Bach chorales. The work presented here extends and integrates the earlier harmonic similarity work [11,13].

2 Related work

MIR methods that focus on the harmonic information in the musical data are quite numerous. After all, a lot of music is polyphonic, and limiting a retrieval system to melodic data considerably restricts its application domain. Most research seems to focus on complete polyphonic MIR systems e.g., [3]. By complete systems, we mean systems that do chord transcription, 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 Sect. 2.2, we will discuss the current state of automatic chord transcription; in Sects. 2.3 and 2.4, we elaborate on the cognition of tonality and the cognitive model relevant to the similarity measure that will be presented in Sect. 3.

2.1 Harmonic similarity measures

An interesting symbolic MIR system based on the development of harmony over time is the one developed by Pickens and Crawford [25]. Instead of describing a musical segment as a single chord, the authors 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 4-dimensional pitch space as found by Krumhansl [15] in her controlled listening experiments (see Sect. 2.3). The authors 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. [24]. 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 12 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 representing 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 calculating the substitution probabilities between chords and not for estimating the similarity between sequences of chords.

Besides the distance measure that we will elaborate on in this paper, which was earlier introduced in [11,13], there exist two other methods that solely focus on the similarity of chord sequences: an alignment-based approach to harmonic similarity [14] and a grammatical parse tree matching method [12]. The first two are quantitatively compared in in Sect. 4.

The chord sequence alignment system (CSAS) [14] 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; 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. Note that the key relative representation of Hanna et al. requires the global key to be known.

The third harmonic similarity measure using chord descriptions is a generative grammar approach [12]. The authors use a generative grammar of tonal harmony to parse the chord sequences, which result 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. However, the rejection of ungrammatical harmonies by the parser is problematic, but can be resolved by applying an error-correcting parser [9].

2.2 Automatic chord transcription

The application of harmony matching methods is extended by the extensive work on chord label extraction from raw musical data within the MIR community. Chord transcription algorithms extract chord labels from either musical scores or musical audio. Given a symbolic score, automatically deriving the right chord labels is not trivial. Even if information about the notes, beats, voices, bar lines, key signatures, etc., is available, the algorithm must determine which notes are unimportant passing notes. Moreover, sometimes the right chord can only be determined by taking the surrounding harmonies into account. Several algorithms can correctly segment and label approximately 84 % of a symbolic dataset (for review, see [28]).

Although the extraction of chord labels from score data is interesting, most digital music repositories store music as (compressed) digitized waveforms. Therefore, to be able to apply the methods presented in this paper to the audio domain, automatic audio transcription methods are necessary. Ideally, a piece of audio would be automatically transcribed into a representation similar to a musical score. However, although much progress has been made, multiple fundamental frequency (F0) estimation, the holy grail in polyphonic music transcription, is still considered too unreliable and imprecise for many MIR tasks. Hence, automatic chord transcription has offered a welcome alternative, which transforms polyphonic audio into musically feasible symbolic annotations, and can be used for serious MIR tasks.

In general, most chord transcription systems have a similar outline. First, the audio signal is split into a series of overlapping *frames*. A frame is a finite observation interval specified by a windowing function. Next, *chroma vectors* [29], representing the intensities of the 12 different pitch classes, are calculated for every frame. Finally, the chroma vectors are matched with chord profiles, which is often done using the Euclidean distance. The chord structure that best matches the chroma vector is selected to represent the frame. Although the digital signal processing-specific parameters may vary, most approaches toward automatic chord transcription use a chroma vector-based representation and differ in other aspects like chroma tuning, noise reduction, chord transition smoothing and harmonics removal. For an elaborate review of the related work on automatic chord transcription, we refer the reader to [22].

2.3 Cognitive models of tonality

Only part of the information needed for reliable similarity judgment can be found in the musical information. Untrained as well as musically trained listeners have extensive knowledge about music [4,6]; 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 [10].

Of particular interest for the current research are the experiments of Krumhansl [15]. 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. Not surprisingly, the tonic was rated most stable, followed by the fifth, third, the remaining tones of the scale, and finally the non-scale tones. Krumhansl also

Table 1 The basic space of a C-major triad in the key of C-major (C = 0, C♯ = 1,…, B = 11), from [16]

(a)	Root level:	0											
(b)	Fifths level:	0						7					
(c)	Triadic level:	0				4		7					
(d)	Diatonic level:	0		2		4	5		7		9		11
(e)	Chromatic level:	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

		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

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

did a similar experiment with chords: instead of judging the stability of a tone listeners had to judge the stability of all 12 major, minor and diminished triads.[3] The results show a hierarchical ordering of harmonic functions that are generally consistent with music-theoretical predictions: the tonic (I) was the most stable chord, followed by the subdominant (IV) and dominant (V), etc.

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 [16] is built on the seminal ideas in the *Generative Theory of Tonal Music* [17] and is designed to make music theoretical and music cognitive intuitions about tonal organization explicit. Hence, it allows prediction of the proximities between musical chords that correspond very well to the findings of Krumhansl [15]. The TPS model is an elaborate one: it supports the estimation of chord proximities within a single key, but also across different key regions. In the distance measure we present in the next section, we only use the *within region* TPS model, i.e., the part that predicts chord proximities within a single key. We present an overview of the within region model here, but also briefly address the full model.

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 allows for representing any possible chord within any arbitrary key. In Table 1, the basic space is set to a C major chord in the context of the C major key. Displayed horizontally are all 12 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 the chord. The next level (b) adds the fifth of the chord. The third level (c) is the triadic level (Lerdahl's term; actually this level concerns all chord types) containing all other pitch classes that are present in the chord. The fourth level (d) is the diatonic level consisting of all pitch classes of the diatonic scale of the key. The basic space can be set to represent any key by cyclically shifting level (d) to match the diatonic scale of the preferred 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 be 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 major space in Table 1, the root note (C) is the most stable, followed by the fifth (G) and the third (E). It is no coincidence that the basic space strongly resembles Krumhansl's [15] probe-tone data.

Table 2 shows how a Dm chord can be represented in the context of the C major key. We can use the basic space to calculate distances between chords within one tonal context by transforming the basic space. First, the basic space must be set to the tonal context 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 [16]):

> *Chord distance rule*: $d_c(x, y) = j + k$, where $d_c(x, y)$ is the distance between chord x and chord y in the context of key c. 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 distinct 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 distinct if it is present in the basic space of y but not in the basic

[3] A diminished triad consists of a root, a minor third, and a diminished fifth (six semitones).

Table 3 The basic space transformation from a C chord to a Dm chord (a), in the context of the C major key and to a G^7 chord (b), also in a C major context

(a)											
		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
(b)											
							7				
		2					7				
		2			5		7				11
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

The distinct pitch classes are underlined

Table 4 The basic space transformation from a G to a Em chord, in the context of C major (a) and the basic space transformation from a D chord to a Dm chord, in the context of the D major key (b)

(a)											
				4							
				4							11
				4			7				11
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
(b)											
		2									
		2							9		
		2			5				9		
	1	2		4	5	6	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

The distinct pitch classes are underlined

space of x. If the chord root is non-diatonic, j receives the maximum penalty of 3.

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.

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 [26]. The TPS distance accounts for differences in weight between the root, fifth and third pitch classes by counting the distinct pitch classes of the transformed basic space at all levels. Two examples of calculation are given in Table 3. Table 3(a) displays the calculation of the distance between a C chord and a Dm chord in the key of C major. The Dm basic space that has no pitch classes in common with the C major basic space at level (a–c, see Table 1). Therefore, all six underlined pitch classes at the levels a–c 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 Table 3(b), 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. Table 4(a) shows the calculation of the distance between a G and an Em chord in the key of C major. The basic space of the G chord and the basic space of the Em chord in the context of C major have four distinct pitch classes (the chords have two pitch classes in common) and three applications of the circle-of-fifths rule are necessary to transform G into an E. Hence, the total distance is 7. Table 4(b) 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.

The original TPS model also supports changes of key by augmenting the chord distance rule that quantifies the number of fifth leaps[4] of the diatonic level (d) to match a new region, i.e., key [16, p. 60, termed *full version*]. By shifting the diatonic level, the tonal context is changed and a modulation is established. Next, the model as described above is applied in the same manner, but with the diatonic level shifted to match the new key. A difficulty of the regional extension is that it features a rather liberal modulation policy, which allows for the derivation of multiple different modulation sequences. We do not use the regional chord distance rule for the distance measures here presented and we will explain why in the next section. Hence, explaining the regional distance rules is beyond the scope of this article and we refer the reader to [16] for the details of the full model.

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). 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. Hereby, we deviate from the TPS model in two ways: first, we only use the within region chord distance rule and discard the regional shifts; second, we apply the

[4] Shifts of seven steps on the chromatic level (e).

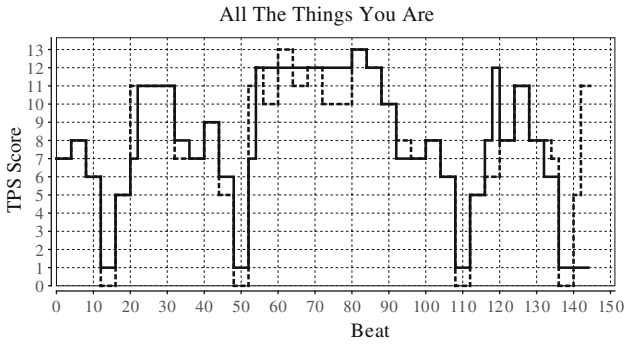

Fig. 2 A plot demonstrating the comparison of two similar versions of *All the Things You Are* using the TPSD. The *total area* between the two step functions, normalized by the duration of the shortest song, represents the distance between both songs. A minimal area is obtained by shifting one of the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

chord distance rule not to subsequent chords, but calculate the distance between each chord and the tonic triad of the global key of the song.

The choice for calculating the TPS chord distance between each chord of the song and the tonic triad of the key of the song, was a representational one: 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 similar but not identical chord sequences can then produce radically different scores.

When we plot the chordal distance to the tonic over time, a step function appears. In this, we assume that time is represented discretely, and the duration of the chords is available in the score. Next, the difference between two chord sequences can then be defined as the minimal area between the two step functions, f and g, of the sequences over all possible horizontal shifts of f over g (see Fig. 2). These shifts are cyclic. If the step functions have different lengths, the difference in length, i.e., the non-overlapping part of the longer step function, is defined to be zero. To prevent longer sequences from yielding higher scores, the score is normalized by dividing it by the length of the shortest step function. Because step functions represent the tonal distance to the tonic, their representation is key-relative and requires information about the global key. Moreover, if a musical piece has many or large key changes, information about these key changes is required as well.

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 occurs when two vertical edges coincide. Consequently, all shifts where two edges coincide have to be considered, yielding $O(nm)$ shifts because in the worst case every vertical edge in f has to be aligned to every vertical edge in g. Hence, the total running time is $O(nm(n+m))$.

Table 5 An example of the minimal TPS chord distance and the maximal TPS chord distance

(a)											
									9		
				4					9		
0				4					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
(b)											
	1										
	1						7	8			
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

In (a), two Am chords are compared yielding a distance of 0. In (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. Note that pitch classes present a certain level are also present at subsequent levels

This upper bound can be improved. Arkin et al. [2] 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 [5]. Hence, we achieve an upper bound of $O(nm)$.

3.1 Metrical properties of the TPSD

For retrieval and indexing purposes, there are several benefits if a distance measure is 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.
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.

We 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 12 pitch classes. The circle-of-fifths rule yields the maximum score of 3, 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 than 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 5 in each case, 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 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 chord sequences, $d(x, z) \leq d(x, y) + d(y, z)$ does not hold.

4 Experiment 1

The retrieval capabilities of the TPSD were analyzed and compared to the CSAS in an experiment in which we aimed to retrieve similar but not identical songs. For every query, a ranking was created on the basis of the values obtained by the evaluated similarity measures. 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 an algorithm we call BASELINE. To measure the impact of the chord representation, we compared three different flavors of both the TPSD as well as the CSAS: in the first task, only the root note of the chord was available to the algorithms; in the second task, we presented the root note and the triad of the chord (major, minor, augmented and diminished); and in the third task, we presented the full chord with all extensions as they are found in the data. Note that all evaluated similarity measures use a key relative representation.

For the triad and full chord tasks, we used the TPSD as described in the previous section. We will denote these variants of the TPSD by TPSDTRIAD and TPSDFULL, respectively. For the tasks where only the chord root was available, we used a different step function representation. In these tasks, the interval between the chord root and the root note of the key defined the step height and the duration of the chord again defined the step length. This matching method is very similar to the melody matching approach by [1]. Note that the latter was never evaluated in practice.

We also evaluated different variants of the CSAS. The first variant, CSASROOT, was evaluated in the root only task. In this variant, +2 was added to the total alignment score if the root note matched and −2 otherwise. In the chord triad task, the same procedure was followed: if the triad matched, +2 was added and −2 if the triads did not match; this CSAS variant is named CSASTRIAD. In the full chord task, the within region TPS model was used as a substitution function, this variant is denoted by CSASFULL.

The BASELINE similarity measure used the edit distance [18] between the two chord sequences represented as a string, with a chord label at every beat, to quantify the similarity between the two chord sequences. However, one might consider this an unfair comparison because the TPSD and CSAS have more information they can exploit than the edit distance, namely information about the key. To make the comparison more fair, we transposed all songs to C major and C minor before matching the strings.

4.1 A chord sequence corpus

For the experiment, a large corpus of musical chord sequences was assembled, consisting of 5,028 unique Band-in-a-Box files that were created by music enthusiasts and collected from the Internet. Band-in-a-Box is a commercial software package [8] that is used to generate musical accompaniment based on a lead sheet.[5] A Band-in-a-Box file stores a sequence of chords and a certain style, whereupon the program synthesizes 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 [23].

These songs were labeled and duplicate sequences were removed. All chord sequences describe complete songs; those 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, were checked and corrected manually. However, the size of the corpus is too large to check all sequences manually, and because the data is mainly created by non-professional users, the corpus might still contain

[5] A lead sheet is a score that shows only the melody, the chord sequence, and the lyrics (if any) of a composition.

Table 6 The distribution of songs and song class sizes in the chord sequence corpus

Class size	No. of classes	Songs in class	100 × Classes/ total classes	100 × Songs/ total songs
1	3,253	3,253	64.7	82.5
2	452	904	18.0	11.5
3	137	411	8.2	3.5
4	67	268	5.3	1.7
5	25	125	2.5	0.6
6	7	42	0.8	0.2
7	1	7	0.1	0.0
8	1	8	0.2	0.0
10	1	10	0.2	0.0
Total	3,944	5,028	100	100

some harmonically atypical sequences or wrong key assignments. The style of the songs is mainly jazz, latin and pop.

Within the collection, 1,775 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., $F\sharp^7$ instead of C^7. Having multiple chord sequences describing the same song allows for setting up a retrieval experiment in which we aim to retrieve the other versions of a song. The title of the song is used as ground-truth and the retrieval challenge is to find the other chord sequences representing the same song.

The distribution of the song class sizes is displayed in Table 6 and gives an impression of the difficulty of the retrieval task. Generally, Table 6 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 in corpus are non-relevant background items. The chord sequence corpus is available to the research community and can be obtained from the first author on request.

4.2 Results of experiment 1

We analyzed the rankings of all 1,775 queries. Figure 3 shows the interpolated average precision calculated at 11 standard recall levels, calculated as described in [20]. 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, and that

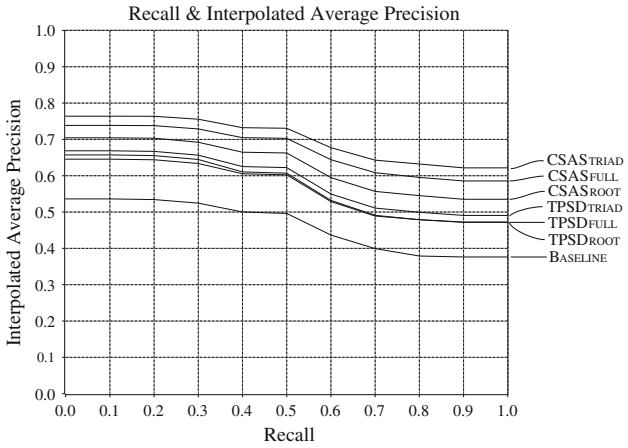

Fig. 3 The interpolated average precision measured at 11 recall levels of the BASELINE, CSAS and TPSD. The latter two are evaluated in three tasks in which the amount of chord information is varied

the CSAS outperforms the TPSD, and both the TPSD and the CSAS outperform the BASELINE.

We also calculated the mean average precision (MAP). 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 [20] (Table 7). 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 $\alpha = 0.01$. We chose the Friedman test because the underlying distribution of the data is unknown and, in contrast to an ANOVA, the Friedman test does not assume a specific distribution of variance. There were significant differences between the runs, $\chi^2(6, N = 1{,}775) = 896$, $p < 0.0001$. To determine which of the pairs of measurements differed significantly, we conducted a post hoc Tukey HSD test.[6] As opposed to a T-test, the Tukey HSD test can be safely used for comparing multiple means [7]. Table 8 displays the pairwise differences.

Most differences can be considered statistically significant. Only the differences between CSASROOT and TPSDTRIAD, between CSASROOT and TPSDFULL, and between TPSDFULL and TPSDROOT were not statistically significant.[7] Hence, we can conclude that both the CSAS and

[6] All statistical tests were performed in Matlab 2011b.

[7] The non-significant differences between CSASROOT and TPSDTRIAD or TPSDFULL may seem counterintuitive. This lack of statistical significance can be explained by the fact that the number of queries in which the CSASROOT outperforms the TPSD-based measures is only slightly higher than the number of queries in which the TPSD-based measures outperform the CSASROOT. However, when the CSASROOT outperforms the TPSD-based measures, the differences in average precision are larger, resulting in a higher MAP.

Table 7 The mean average precision (MAP) of the rankings based on the compared similarity measures and the running times (hours:minutes)

	CSASTRIAD	CSASFULL	CSASROOT	TPSDTRIAD	TPSDFULL	TPSDROOT	BASELINE
MAP	0.696	0.666	0.622	0.580	0.565	0.559	0.459
Runtime	72:57	95:54	74:45	0:33	0:37	0:28	0:24

Table 8 The pairwise statistical significance between all similarity measures

A + denotes a statistically significant difference and a − denotes a non-significant difference. The + and − signs were derived by pairwise comparison of the confidence intervals

	CSASFULL	CSASROOT	TPSDTRIAD	TPSDFULL	TPSDROOT	BASELINE
CSASTRIAD	+	+	+	+	+	+
CSASFULL		+	+	+	+	+
CSASROOT			−	−	+	+
TPSDTRIAD				+	+	+
TPSDFULL					−	+
TPSDROOT						+

the TPSD outperform the BASELINE method and that, irrespective of the kind of chord representation, the CSAS outperforms the TPSD. This does not mean that the chord representation does not have an effect. It is surprising to observe that the triad representation significantly outperforms the other representations for both the CSAS and the TPSD. It is furthermore interesting to see that using only the root of the chord already yields very good results, which in some cases is not even statistically different from using the full chord specification. Apparently, discarding chord additions acts as a form of syntactical noise reduction, since these additions, if they do not have a voice-leading function, they tend to differ between versions and mainly add harmonic spice. The retrieval performance of the CSAS is good, but comes at a price. The CSAS run took on average about 81 h which is considerably more than the average of 33 min of the TPSD or the 24 min of the BASELINE. 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 investigate whether melodically related chorale settings by Bach (1685–1750) are also harmonically related. Doing analyses of this kind by hand is very time-consuming, especially when corpora have a substantial size.

Chorales are congregational hymns of the German Protestant church service [21]. 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) [30]. A corpus of Bach chorales consisting of 371 items was posthumously published by Bach and Kirnberger (1784–1787), 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 a fair amount of stylistic variation and evidence of development over time. Yet, one can claim that Bach's chorale harmonizations were constrained by the general rules of tonal harmony in force in the first half of the 18th century and that the range of acceptable harmonizations of a given melody was limited.

We hypothesize that if two melodies belong to the same tune family, the harmonizations of these melodies are very similar as well. Hence, we expect that melodically similar pieces can also be retrieved on the basis of their harmonic similarity. To determine whether the melodies of two chorales are part of the same tune family, 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.[8] If they do, it should be possible to retrieve these settings by ranking them on the basis of their TPSD distance.

[8] Note that manually doing the 357^2 harmonic or melodic similarity assessments is infeasible.

5.1 Experiment 2

To test whether the melodically related Bach chorales were also harmonically related, we performed a retrieval experiment similar to the one in Sect. 4. We took 357 Bach chorales and used the TPSD to 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 221 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[9] in MIDI format. But as explained in the previous sections, the TPSD takes chords as input, not MIDI notes. We therefore use David Temperley's Chord root tracker [28], which is part of the Melisma music analyzer.[10] 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 known 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. The score was segmented such that each segment contained at least two simultaneously sounding notes. Manually annotating a small random sample yielded a correctness of the root tracker of approximately 80 %, which is in line with the 84 % claimed in [28].

The TPSD also needs knowledge of the keys of all chorales. The key information was generously offered by Martin Rohrmeier, who investigated the distributions of the different chord transitions within the chorales corpus [27]. We selected the chorales for which the MIDI data, a pdf score (for our musicological expert) and the key description were available. After preparation, which included checking for duplicate chorale, the corpus contained 357 pieces.

[9] See http://www.jsbchorales.net/ (accessed 11 Feb 2013) for more information.

[10] The source code of the Melisma Music Analyzer is freely available at: http://www.link.cs.cmu.edu/music-analysis/ (accessed 11 Feb 2013).

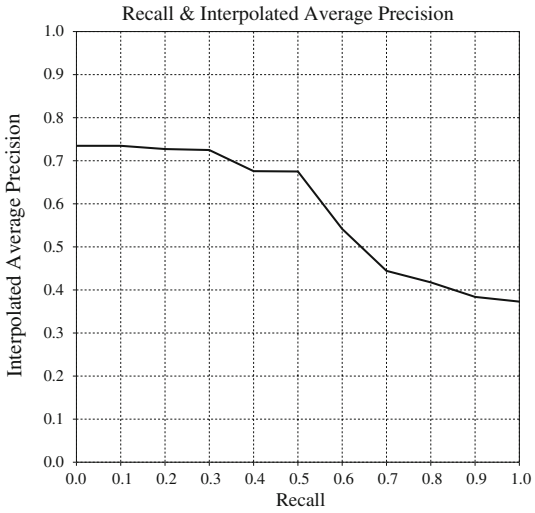

Fig. 4 The average interpolated precision for 11 different recall levels of the melodically related chorales retrieved on the basis of their TPSD scores

Table 9 Tune and tune family distribution in the Bach chorales corpus

Family size	No. of families	Tunes in family	100 × Families/ total families	100 × Tunes/ total tunes
1	136	136	38	68
2	24	48	13	12
3	17	51	14	8.5
4	10	40	11	5
5	5	25	7	2.5
6	3	18	5	1.5
7	4	28	8	2
11	1	11	3	0.5
Total	200	357	100	100

5.2 Results of experiment 2

We analyze the TPSD-based rankings of Bach's chorales with an average interpolated precision versus recall plot, displayed in Fig. 4. To give an idea of the structure of the corpus, we also printed the distribution of the sizes of the tune families in Table 9. The graph in Fig. 4 shows clearly that a large proportion 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 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 0.71. However, a considerable number of pieces cannot be retrieved on the basis of their TPSD: in 24 % of the queries, the first related chorale is not within the first 10 retrieved chorales.

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 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 four 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[11] 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.

5.2.1 Different meter

The default meter for chorale melodies is 4/4. However, variants in 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 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 chorales from the same tune family with different meters. Examples include *Wie nach einer Wasserquelle* (two outliers in 3/4 meter) and *Nun lob, mein Seel, den Herren* (three versions in 3/4, one, the outlier, in 4/4 meter).

5.2.2 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.

5.2.3 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 lead to 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.

5.2.4 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 their contribution to perceived similarity depends on metric position and duration as well.

False positives, 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

[11] The Bach-Werke-Verzeichnis (BWV) is a numbering system designed to order and 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.

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 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 to 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 major and minor keys were distributed fairly equally over the dataset (54 % major keys and 46 % minor keys), and the lengths of the chorales are not correlated with the key of the piece, $r = 0.022$.[12] The 15 pieces with the highest rank were all pieces in a minor key, and those 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. The differences in length make sense because the length-based normalization penalizes long chorales. The effect of key is more difficult to explain. Possibly Bach's pieces in minor keys yield a more pronounced step function that boosts the retrieval of these pieces.

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. This distance is based on the changes of the distance between chord and key as estimated by Lerdahl's Tonal pitch space model. The TPS 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 the tonic over time and the distance between two chord progressions is defined as the minimal area between two step functions. The TPS model that lies at the basis of the TPSD is very well grounded in both music theory and music cognition. It correlates very well with the listening experiments of Krumhansl [15], and it is therefore likely that the TPSD captures at least some of the perceptual relations between chords and their tonal contexts. The TPSD is a distance measure that is simple to use, requires little parameter tuning, is key invariant, and can be computed efficiently.

The performance of the TPSD can be considered good, especially if one takes the size of the test corpus into account and the relatively small class sizes (see Table 6). We compared the performance of the TPSD to the performance of the 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 significantly outperforms the TPSD statistically. However, the TPSD has a better performance–runtime ratio than the CSAS. Surprisingly, only using the root note of a chord gives good retrieval results. In the case of the TPSD, the difference between using only the root is not even statistically different from using full chord specifications. Removing all chord additions and using only the triad significantly improve these results for both similarity measures. We furthermore demonstrated how the TPSD can contribute to the musicological discussions on melody and harmony in Bach's chorales in a case study. We showed that a considerable number of Bach chorales that share a melody are also harmonically more similar.

Nevertheless, there is still room for improvement. The TPSD does not handle large structural differences between pieces very well, e.g., having extra repetitions or a bridge, etc. A prior analysis of the structure of the piece combined with partial matching could improve the retrieval performance. Also smaller asynchronies between step functions can be harmful; in future versions of the TPSD, these might be removed by allowing a small number of horizontal edges to be lengthened or shortened. Still, within the TPSD, all chords are treated equally important, which is musicologically not plausible. Hence, we expect that exploiting the musical function in the local as well as global contexts, as done in [12], will improve harmonic similarity estimation.

Although we showed that the cover versions of chord sequences can be retrieved quickly with the TPSD, potential users might also be interested in songs that share musical properties other than harmony. Hence, a straightforward, but interesting, extension to a content-based retrieval system would be the inclusion of other musical similarity measures. First and foremost, melody should be added, but also timbre or rhythmical similarity could be musically satisfying additions. This directly raises questions about how one should combine these different similarity measures. How should they be weighted, and how can user feedback be taken into account? Also, it might not always be similarity that a

[12] We used a Pearson correlation with key as a binary variable.

user is looking for; perhaps a user wants to retrieve songs that share the same melody, but are harmonically very different. This requires notions of harmonic dissimilarity, which might not simply be the inverse of the distance measure presented in this paper. Maybe a user is searching for surprising and not necessarily similar music. These issues present some challenging directions for future MIR research, illustrating that content-based retrieval of music is not yet a solved problem.

The retrieval performance of the TPSD was evaluated on symbolic data in this paper. Nevertheless, recent developments in audio chord and key transcription extend its application to audio data because the output of these methods can be matched directly with the similarity measures here presented. The good performance of harmonic similarity measures leads us to believe that also in other musical domains, such as the audio domain, retrieval systems will benefit from chord sequence-based similarity measures.

Acknowledgments We would like to thank Peter van Kranenburg for comparing and annotating the similarity of the melodies of the Bach chorales used in Sect. 5, Martin Rohrmeier for providing information about the musical key of the same corpus, and Marcelo E. Rodríguez López and three anonymous reviewers for valuable comments on earlier versions of this article.

References

1. Aloupis G, Fevens T, Langerman S, Matsui T, Mesa A, Nuñez Y, Rappaport D, Toussaint G (2004) Algorithms for computing geometric measures of melodic similarity. Comput Music J 30(3): 67–76
2. Arkin E, Chew L, Huttenlocher D, Kedem K, Mitchell J (1991) An efficiently computable metric for comparing polygonal shapes. IEEE Trans Pattern Anal Mach Intell 13(3):209–216
3. Bello J, 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), pp 304–311
4. Bigand E (2003) More about the musical expertise of musically untrained listeners. Ann N Y Acad Sci 999:304–312
5. Cormen T, Leiserson C, Rivest R, Stein C (2001) Introduction to algorithms. MIT Press, Cambridge
6. Deliège I, Mélen M, Stammers D, Cross I (1996) Musical schemata in real time listening to a piece of music. Music Percept 14(2): 117–160
7. Downie J (2008) The music information retrieval evaluation exchange (2005–2007): a window into music information retrieval research. Acoust Sci Technol 29(4):247–255
8. Gannon P (1990) Band-in-a-Box. PG Music. http://www.pgmusic.com/
9. de Haas WB (2012) Music information retrieval based on tonal harmony. PhD thesis, Utrecht University, Utrecht
10. de Haas WB, Wiering F (2010) Hooked on music information retrieval. Empir Musicol Rev 5(4):176–185
11. de Haas WB, Veltkamp RC, 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), pp 51–56
12. de Haas WB, Rohrmeier M, Veltkamp RC, 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), Kobe
13. de Haas WB, Robine M, Hanna P, Veltkamp RC, Wiering F (2010) Comparing approaches to the similarity of musical chord sequences. In: Proceedings of the 7th International Symposium on Computer Music Modeling and Retrieval (CMMR), pp 299–315
14. Hanna P, Robine M, Rocher T (2009) Joint international conference on digital libraries. ACM, New York
15. Krumhansl C (1990) Cognitive foundations of musical pitch. Oxford University Press, New York
16. Lerdahl F (2001) Tonal pitch space. Oxford University Press, New York
17. Lerdahl F, Jackendoff R (1996) A generative theory of tonal music. MIT Press, Cambridge
18. Levenshtein VI (1966) Binary codes capable of correcting deletions, insertions, and reversals. Cybern Control Theory 10(8): 707–710
19. Logan B (2000) Mel frequency cepstral coefficients for music modeling. In: Proceedings of the 1th Society for Music, Information Retrieval Conference (ISMIR), Plymouth
20. Manning CD, Raghavan P, Schütze H (2008) Introduction to information retrieval. Cambridge University Press, New York
21. Marshall RL, Leaver RA (2012) Chorale. In: Grove Music Online, Oxford Music Online. http://www.oxfordmusiconline.com/subscriber/article/grove/music/05652. Accessed 16 Jan 2012
22. Mauch M (2010) Automatic chord transcription from audio using computational models of musical context. PhD thesis, Queen Mary University of London, London
23. Mauch M, Dixon S, Harte C, Casey M, Fields B (2007) Discovering chord idioms through beatles and real book songs. In: Proceedings of the 8th International Conference on Music Information Retrieval (ISMIR), pp 255–258
24. Paiement JF, Eck D, Bengio S (2005) A probabilistic model for chord progressions. In: Proceedings of the 6th International Conference on Music Information Retrieval (ISMIR), London, pp 312–319
25. Pickens J, Crawford T (2002) Harmonic models for polyphonic music retrieval. In: Proceedings of the 11th International Conference on Information and Knowledge Management, pp 430–437
26. Piston W (1941) Harmony. Norton, W. W. & Company, New York
27. Rohrmeier M, Cross I (2008) Statistical properties of tonal harmony in Bach's chorales. In: Proceedings of the 10th International Conference on Music Perception and Cognition (ICMPC), pp 619–627
28. Temperley D (2001) The cognition of basic musical structures. MIT Press, Cambridge
29. Wakefield GH (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), pp 637–645
30. Wolff C, Emery W, Wollny P, Leisinger U, Roe S (2012) Bach. In: Grove Music Online, Oxford Music Online. http://www.oxfordmusiconline.com/subscriber/article/grove/music/40023pg10. Accessed 16 Jan 2012