Analysis of Chord Progression Data

Brandt Absolu, Tao Li, and Mitsunori Ogihara

Abstract. Harmony is an important component in music. Chord progressions, which represent harmonic changes of music with understandable notations, have been used in popular music and Jazz. This article explores the question of whether a chord progression can be summarized for music retrieval. Various possibilities for chord progression simplification schemes, N-gram construction schemes, and distance functions are explored. Experiments demonstrate that such profiles can be used for artist grouping and for composition retrieval via top-k queries.

1 Introduction

The chord progression is an important component in music. Musicians and listeners speak of novel and influential chord progressions. A well-known example of famous chord progressions is the Tristan Chord of Richard Wagner, the very first two chords in the First Act Prelude of "Tristan und Isolde" and a motif that reappears over and over again in the ensuing four hours of drama. Another example is "Because" by The Beatles, whose main theme runs on an eight-bar chord sequence that is sometimes rumored to have been produced by reversing the chord progression for the main theme of the "Moonlight" Piano Sonata, one of the most famous piano compositions by Ludwig van Beethoven (Sonata Op.27 No.2). Yet another example is "Giant Steps" by John Coltrane, which uses a combination of dominant-to-tonic `[V7 - I]` cadence and repeatedly raises the key by major third.

Brandt Absolu
Maritime and Science Technology High School, Key Biscayne, FL, USA
e-mail: bla8291@aol.com

Tao Li
School of Computer Science
Florida International University, Miami, FL, USA
e-mail: taoli@cs.fiu.edu

Mitsunori Ogihara
Department of Computer Science
University of Miami, Coral Gables, FL, USA
e-mail: ogihara@cs.miami.edu

Z.W. Raś and A.A. Wieczorkowska (Eds.): Adv. in Music Inform. Retrieval, SCI 274, pp. 165–184.
springerlink.com

Among many genres of music the role of chord progressions appears to be the most significant in Jazz. The performance in Jazz takes the form of Theme - Improvisation - Theme, where the middle part is improvisation in which the melody is spontaneously created while the chord progression of the main theme is being played repeatedly. Keeping in mind that the melody has to be created spontaneously, Jazz performers select tunes with chord progressions having certain characteristics.

Many Jazz compositions are based on one of two well-known chord progression forms. One is the 12-bar blues progressions and the other is “I Got Rhythm” by George Gershwin, where the chord progression of a tune is constructed out of the base progression and a new melody is played over the new progression. The abundance of such tunes witnesses the fact that Jazz music is highly improvisational (the tunes themselves might have been composed spontaneously by way of improvisation) and the fact that chord progressions play an extremely important role in that genre.

Quite often in studio recordings and live performances of Jazz, their performance programs consist of many tunes. Sometimes the tunes are compositions by the performers themselves, but more frequently they are compositions by someone else. If one surveys a large collection of Jazz recordings he/she will notice that many of them contain compositions from a small set of famous Jazz composers, such as Duke Ellington, Wayne Shorter, and Thelonius Monk. The high popularity of these composers, along with the fact that Jazz performers select tunes based on chord progression, suggests that there are Jazz composers with a unique chord progression style. Thus we here hypothesize:

there is a group of popular Jazz composers whose composition style can be well represented by chord progressions.

This article explores this hypothesis from the perspectives of clustering (the problem of grouping data according to their similarity) and similarity search (the problem of finding data objects similar to an input data object).

Fundamental to this exploration is a method for assigning a distance value given two chord progressions. An approach for designing a distance measure is sequence alignment, which is often used in melody-based music retrieval systems (see, e.g., [2, 6, 12, 13]). The basis for the sequence-alignment approach is a theory that models transformation of a chord progression to another (see, e.g., [9, 11]). Such a generative theory can offer a highly understandable explanation as to why two progressions are similar or why they aren’t, but has a substantial limitation that computing a transformational path might be very difficult for tunes that have musically little to do with each other. It may be possible to deal with this issue by the use of partial alignments, as has been done in other scientific disciplines, but such a solution for chord progression analysis is yet to be established.

Also, the sequence alignment approach has a limitation that the pairwise similarity does not enable calculation of the mean—the chord progression that represents a collection of progressions as a whole—that is an essential component in clustering. Furthermore, computation of pairwise distance via sequence alignment is very expensive, which might limit its practical usage if alignment must be computed on the spot for a large number of chord progression pairs.

This consideration suggests the use of statistics to summarize chord progressions and then to compare chord progressions. The simplest of such statistics will be frequency counts of chords, i.e., how often the chords or chord components are used in a composition. The use of statistics has two major advantages. One, a single scan will be sufficient to compute frequencies of chords in a progression, so such statistics are easy to compute. Two, the instrument a composer uses for composing may result in a certain bias in the statistics. However, the simple statistics are insufficient for our purpose, since the frequencies chords do not provide information about the order in which chords appear.

We address the above issue by using of N-grams—the patterns consisting of N-consecutive chords that appear in a chord progression. The N-gram is a standard tool in natural language understanding (see, e.g., [5]), and has been used in the area of music information retrieval, in particular, in the melodic contour analysis [3, 4, 10]. Notably, the recent work of Mauch et al. [7] use 4-grams of triads to compare the compositions by The Beatles and Jazz tunes. The work used triads because a large portion of the chords in the Beatles compositions are simply triads.

The present article, extending an earlier work by a subset of the authors [8], considers the use of other chord tones in the analysis. Highly prominent in the Jazz harmony are the tone group of the 6th, 7th and major 7th notes and the group of tension notes (the 9th, the 11th, and the 13th notes). The former signifies the functions that chords possess while the latter adds color to triads. Chord progression analysis in terms of triads is likely to enable fundamental understanding of the chord structure. However, deeper understanding perhaps cannot be obtained without examining these non-triad notes, in particular, for comparing Jazz compositions.

While the chord progressions are an important subject in musicology, one might ask how chord progressions can be successfully incorporated into a music information retrieval system. One possible scenario is where tunes are retrieved by fragments of chord progression and accompanying metadata. In such a system, the user provides a chord sequence (either typed or copy-pasted from a sequence on screen) as input and the system retrieves tunes that contain a part with either exactly the same as (with the possibility of allowing transposition of the key) or similar to the input sequence, where the input chord progression is specified using an unambiguous notation system (such as the one in [1]). Also, the accompanying metadata (artist, genre, etc.) is used to narrow the scope of the search.

Another possible scenario is the retrieval of tunes with a certain set of chords as a constraint. In such a system, the user specifies the set of chords he/she can play well on his/her instrument (for example, a guitar) and metadata (again, artist, genre, etc.), and the system retrieves the tunes that meet the criteria. For example, the user may say "I need a Beatles song that uses chords only from { `G`, `EMI`, `A`, `C`, `B7` }" and then the system retrieves "Run For Your Life" (which actually is based on the five chords with `BMI` in place of `B7`).

1.1 Contributions of This Article

This article presents a novel concept of N-gram profiles for the purpose of computing numerical representation of a single chord progression as well as a collection of chord progressions. First, the efficacy of the proposed profiles is tested using hierarchical clustering of famous Jazz composers according to their profiles. There are too many possible N-grams because the native chord space is gigantic. Thus, the chord name space has to be reduced using some chord simplification. With respect to two selected N-gram formats the hierarchical representation of Jazz composers reflects very well the historic development of the Jazz compositional idioms. Next, the use of N-gram profiles for similarity search is tested using a larger set of compositions and composers. Here different formats of profiles can be mixed together to represent a composition. The best combination of formats is searched for using a greedy algorithm with the effectiveness of top-K queries as the guide. Finally, the combinatorial search for the top-K queries is applied to the problem of identifying the composer given a composition as input.

1.2 Organization of the Article

This article is organized as follows. The next section discusses in detail how a chord is defined and how N-grams of a chord progression can be computed. Section 3 presents a proof-of-concept analysis of chord progression profiles via hierarchical clustering of Jazz composers. Section 4 presents exploration of best chord progression profiles using top-K query analysis.

2 Chord and N-grams Profiles

2.1 Chord Name Space

The Oxford University Press defines a chord as: “Any simultaneous combination of notes, but usually of not fewer than 3. The use of chords is the basic foundation of harmony.” This definition readily accepts as a chord any multiple number of simultaneously played notes and thus a smash of keys on the piano is considered to be a chord. However, since the focus of this article is chord progression analysis that is useful for retrieving tunes, the definition of chords must be narrowed so that all the chords can be presented using a compact and clear notational scheme without specifying a chord as the collection of pitch names that are present in it. Chords presented with such a scheme are instrument-independent in the following manner. Given any instrument or any ensemble of instruments, as long as all the chord notes are presented in the harmony and no others are, we must think that the chord is correctly presented.

Such notational schemes indeed exist. Books of popular music often present chord names in addition to the the melody, lyrics, piano accompaniment chart, and somewhat less frequently present guitar tabs. There even exist “fakebooks” that present only the melody, lyrics, and chords, which are often used by Jazz musicians for free-style interpretation of compositions. While the chord names that appear in these books have their basis in the Western classical music theory, as pointed by Brandt and Roemer in [1], there exists conspicuous, and curious, ambiguity in the notation. The 7th chord of `G` with the augmented fifth and with the 9th note (that is, the chord consisting of notes `G`, `B`, `D`$^\sharp$, `F` and `A` with the `G` at the bottom), can be written in six different ways: `G AUG 9`, `G+9`, `G9(`$\sharp$`5)`, `G AUG 7(9)`, `G+7(9)` and `G7(`$\sharp$`5 9)`. Since the first three are also used for the same chord without the 7th, the coexistence of various chord name scheming is very confusing. To resolve this issue Brandt and Romer [1] proposed a unified chord naming scheme that is both succinct and compact. The chord names considered in this article are all representable using this unified scheme.

In the chord notation scheme by Brandt and Roemer, a chord consists of four major parts: (1) the triad (the root, the 3rd, the 5th), (2) the 6th/7th, (3) the tension notes (the 9th, the 11th, the 13th), and (4) the added bass note. In the proposed chord notation the names start with the root and the triad together (the triad being denoted as empty for the major triad, `MI` for the minor triad, and `SUS` for the suspended 4th triad) followed by the 6th/7th note specification (`6` for the sixth, `7` for the seventh, and `MA7` for the major 7th). This is followed by additional information presented within a pair parentheses, which consists of alterations to the 3rd and the 5th notes and of the tension notes, and then a special keyword “on” and the bass note name if there is an added bass note. Also, the combinations (the 7th and the 9th), (the 7th, the 9th and the 11th), and (the 7th, the 9th, the 11th, and the 13th) are respectively represented by the numbers 9, 11, and 13 attached immediately after the root name for short-hand, with the exception that (a) if the triad is major then the second combination will be used and the third combination does not include the 11th; and (b) the 7th note may be the major 7th note, in which case, the numbers 9, 11, and 13 will be preceded by letters “MA”. For example, `DMIMA11` is equivalent to `DMIMA7 (9 11)`. (The interested reader is encouraged to consult with the book for more detail.) An implicit restriction here is that not more than one note can be present from each of the four note groups: the 6th/7th notes, the 9th notes, the 11th notes, and the 13th notes.

2.2 Chord N-grams

2.2.1 Formal definition of an N-gram

For a set of symbols, U, for an integer $N \geq 1$, an N-gram over U is an ordered N-tuple $(u_1, \ldots, u_N)$ such that $u_1, \ldots, u_N \in U$. An N-gram $(u_1, \ldots, u_N)$ is said to be *proper* if for all i, $1 \leq i \leq N-1$, it holds that $u_i \neq u_{i+1}$.

2.3 Chord Simplification

The chord name space defined in Section 2.1 is enormous. There are twelve possible choices for the root (without distinguishing between two notes that refer to the same note in the equal temperament); four for the 3rd (Minor, Major, Suspended 4th, and Omitted 3rd); four for the 5th ($\flat$5, $\natural$5, $\sharp$5, and Omitted 5th); four for the 6th/7th (6th, Minor 7th, Major 7th, and none of the three being used); four for the 9th ($\flat$9, $\natural$9, $\sharp$9, and none of the three being used); four for the 11th ($\natural$11, $\sharp$11, and neither of the two being used); four for the 13th ($\flat$13, $\natural$13, and neither of the two being used); and finally 12 for the added bass note. These make the total number of choices more than 320,000. This means that the total number of possible N-grams is more than 100 billions for $N = 2$ and 27 trillions for $N = 3$.

One must, however, be cautioned that although the space of N-grams is enormous, the N-grams that actually appear in a chord progression are very small in quantity. In fact, for a sequence of M chords, there are only $M-N+1$ positions from which an N-gram can be started, the number of unique N-grams appearing in the sequence is at most $M-N+1$. Even though the distributions of chords are often very skewed (towards certain keys and towards chords without tension notes), the vastness may make it unlikely for the N-gram profile of a chord progression with highly enriched chords to intersect with the N-gram profile of another chord progression. This problem can be overcome by simplifying chords.

The concept of chord simplification corresponds well with the concept of stemming in document processing, which is the process of removing modifiers of words thereby making words generated from the same root with difference modifiers treated as identical words. The process of simplifying a chord can be divided into two parts: (1) turning a chord with an attached bass note (such as `AMI7 on B`) into a non-fractional chord and (2) simplifying the tensions and the use of 6th and 7th notes.

There are three options for the first part:

- (B_0) simply removing the bass note (for example, `AMI7 on B` is changed to `AMI7`),
- (B_1) reorganizing the chord notes so that the bass note becomes the root (for example, `AMI7 on B` is changed to `B7SUS4 (`$\sharp 5 \flat 9$`))`, and
- (B_2) incorporating the bass note as a tension (for example, `AMI7 on B` is changed to `AMI9`).

There are three options for the second part:

- (T_0) removing entirely the tensions and the 6th/7th note,
- (T_1) removing entirely the tensions but keeping the 6th/7th note, and
- (T_2) replacing the whole tension notes with a single bit of information as to whether the chord has any tension and keeping the 6th/7th note.

Also included in the list of possibilities are the possibility to keep the bass note, which will be denoted by B_3, and the possibility to keep all the tensions intact and keeping the 6th/7th note, which will be denoted by T_3.

The simplification options that are considered here then can be denoted by a pair (B_i, T_j) such that $0 \leq i \leq 3$ and $0 \leq j \leq 3$. The most aggressive simplifications are $(B_i, T_0), 0 \leq i \leq 3$. Each of these simplifications has the effect of reducing any chord to a triad or a chord that is a proper subset of a triad and therefore reduces the number of possibilities for a chord name to 192. For a progression Π and a simplification method τ, we will use $\tau(\Pi)$ to denote the progression Π after applying τ. Table 1 shows an example of how these simplifications work. The listed are the chords generated by simplifying `AMI7(11) on F`. For T_2 simplification, we will use the symbol of `(9)` to show that there is a tension. Note that the incorporation

Table 1 Various simplifications of the chord `AMI7(11) on F`

	T_0	T_1	T_2	T_3
B_0	AMI	AMI7	AMI7(9)	AMI7(11)
B_1	AMI	AMI7	AMI7(9)	AMI7(11 ♭13)
B_2	F	FMA7	FMA7(9)	FMA7(9 11 13)
B_3	AMI on F	AMI7 on F	AMI7(9) on F	AMI7(11) on F

of the bass note as a chord note required in the B_1 simplification may make the accompanying melody inconsistent with the chord. This characteristic is highly more prominent in the T_2 simplification, where all the tension notes are represented by the `(9)` tension symbol.

2.3.1 Measuring the Length of an N-gram

There are two important issues to consider when defining chord N-grams. The first is whether consecutive repetitions of the same chord should be permitted in a chord N-gram. The second is how to consider the number of beats assigned to each component of an N-gram. The two issues are related to each other and come directly from the fact that an arbitrary number of beats can be allocated to a single chord. For example, considering a 12-bar chord progression with the rhythm signature of 4/4 where the first four measures are `[F - F(♯5) - F6 - F7]`, the next two are `B♭7`, the next two are `F7`, and then during the next two bars the chord moves `G7` `G♭7`, `F7`, `E7`, `E♭7`, `D7`, `D♭7`, and `C7`, resolving to `F7` in the last two measures. If the progression is scanned with a sliding window of two measures, the chromatic descent in measures 9 and 10 are captured in three windows, while the `[I - IV - I]` motion that occurs in measures 4 through 7 can never be captured within such a small window. One may suggest to use a double-sized window for the scan, but then the clone of the progression in which each chord has twice as many number of beats as the original creates exactly the same problem.

To resolve these issues, an N-gram here is considered to be proper. This allows an N-gram to have an arbitrarily large number of beats in it. In the above example, the `[I - IV - I]` motion is captured as a 3-gram `[F7 - B♭7 - F7]` with the weight of 32 beats, and the eight-chord chromatic descent is captured as a collection of six 3-grams `[G7 - G♭7 - F7]`, ..., `[D7 - D♭7 - C7]` with the weight of 3 beats each. Also, after simplification, all the consecutive entries whose chord-part are identical to each other should be merged into a single entry. Once this modification has been done, the simplified chord progression has the property that every neighboring pair of chords are different and thus every one of its subsequences is a proper N-gram. Such a chord progression is called a *proper* chord progression. The reader might be cautioned that application of simplification to a proper chord progression without neighbor merging may produce a non-proper N-gram. For example, in the above example, the original sequence with the inclusion of duration can be represented as:

$$\texttt{[F:4 - F(}\sharp\texttt{5):4 - F6:4 - F7:4 - B}^\flat\texttt{7:8 -}$$
$$\texttt{F7:8 - G7:1 - G}^\flat\texttt{7:1 - F7:1 - E7:1 - E}^\flat\texttt{7:1 -}$$
$$\texttt{D7:1 - D}^\flat\texttt{7:1 - C7:1 - F7:8].}$$

Applying the T_0-simplification and merging identical neighbors yields

$$\texttt{[F:16 - B}^\flat\texttt{:8 - F:8 - G:1 - G}^\flat\texttt{:1 - F:1 -}$$
$$\texttt{E:1 - E}^\flat\texttt{:1 - D:1 - D}^\flat\texttt{:1 - C:1 - F:8].}$$

2.3.2 N-gram Transposition

Since popular songs are transposed to different keys, one might be interested in studying chord changes relative to the first chord of the N-gram. One can thus transpose each N-gram locally, in such a way that each N-gram starts with a code having A as the root. Since A is simply nominal, the Roman numerals I, II, III, and so on, can be used. For example, from a five-chord sequence [FMI7, B$^\flat$7, E$^\flat$MA7, CMI7, B7], three 3-grams can be obtained, [FMI7 - B$^\flat$7 - E$^\flat$MA7], [B$^\flat$7 - E$^\flat$MA7 - CMI7], and [E$^\flat$MA7 - CMI7 - B7], which are then transposed respectively to [IMI7 - IV7 - VII$^\flat$MA7], [I7 - IVMA7 - IIMI], and [IMA7 - VIMI7 - VI$^\flat$7]. This transposition process is called the *A-transpose*.

2.3.3 Chord Sequences and Weight of N-grams

As mentioned earlier, a chord progression is a series of chord names such that each chord name is accompanied by a positive rational that represents the number of beats during which its chord is to be played. For an N-gram of a chord progression, its weight represents the contribution that the N-gram makes to the whole progression. For example, the weight has to be assigned so as to distinguish the contribution of a 4-chord pattern DMI7 - G7 - EMI7 - A7 with one beat assigned to each of the four chords from the contribution of the same 4-chord pattern appearing elsewhere in the same chord progression progress with four beats assigned to each chord. The contribution of an N-chord pattern can be approximated by the total number of beats assigned to the chords. Let $\Pi = [a_1 : \ell_1, \ldots, a_M : \ell_M]$ be a progression; that is, it is a series of M chord names $a_1, \ldots, a_M$ and for each i, $1 \leq i \leq M$, ℓ_i is the number of beats assigned to the chord name a_i. Then, for each i, $1 \leq i \leq M - N + 1$, the contribution of the N-gram at position i, $(a_i, a_{i+1}, \ldots, a_{i+N-1})$, is defined to be $\ell_i + \cdots + \ell_{i+N-1}$.

2.3.4 N-gram Profile of a Chord Progression

Figure 1 shows the melody and the chord progression of "Witch Hunt" composed by a Jazz giant Wayne Shorter. Without any simplification the progression is

[CMI7:32 - $E^\flat$7:16 - CMI7:16 - $G^\flat$7:4 - F7:4 -
E7:4 - $E^\flat$7:4 - $A^\flat$MI7(11):4 - A on $A^\flat$:4 -
$A^\flat$MI7(11):4 - G7($^\flat$5):4].

With the B_0 (remove bass) simplification and the T_1 (no tension notes) simplification, the progression becomes

[CMI7:32 - $E^\flat$7:16 - CMI7:16 - $G^\flat$7:4 - F7:4 -
E7:4 - $E^\flat$7:4 - $A^\flat$MI7:4 - A:4 - $A^\flat$MI7:4 - G7($^\flat$5):4].

Without transpose, the progression has the following 3-grams:

- [CMI7 - $E^\flat$7 - CMI7] (64 beats),
- [$E^\flat$7 - CMI7 - $G^\flat$7] (36 beats),
- [CMI7 - $G^\flat$7 - F7] (24 beats),
- [$G^\flat$7 - F7 - E7] (12 beats),
- [F7 - E7 - $E^\flat$7] (12 beats),
- [E7 - $E^\flat$7 - $A^\flat$MI7] (12 beats),
- [$E^\flat$7 - $A^\flat$MI7 - A] (12 beats),
- [$A^\flat$MI7 - A - $A^\flat$MI7] (12 beats), and
- [A - $A^\flat$MI7 - G7($^\flat$5)] (12 beats).

With transpose, the fourth and the fifth ones become identical, so we have

- [IMI7 - $III^\flat$7 - IMI7] (64 beats),
- [$I^\flat$7 - VIMI7 - $III^\flat$7] (36 beats),
- [IMI7 - $V^\flat$7 - IV7] (24 beats),
- [I7 - VII7 - $VII^\flat$7] (24 beats),
- [I7 - VII7 - IIIMI7] (12 beats),
- [I7 - IVMI7 - $V^\flat$] (12 beats),
- [IMI7 - $II^\flat$ - IMI7] (12 beats), and
- [I - VIIMI7 - $VII^\flat$7($^\flat$5)] (12 beats).

Now we obtain the profile of this composition with respect to the (B_0, T_1)-simplification by dividing the weight in terms of the number of beats by their sum.

2.3.5 An Alternative Weighting Scheme

An alternative to the number-of-beats-based weight is the simple frequency count. Let Π be a proper chord progression generated from a given input progression after a certain simplification. Let $g_1, \ldots, g_k$ be an enumeration of all unique N-grams appearing in Π and for each i, $1 \leq i \leq k$, let m_i be the number of times that the N-grams g_i appears in Π. Then for each i, $1 \leq i \leq k$, we define the weight c_i assigned to g_i to be $m_i/(m_1 + \cdots + m_k)$.

Table 2 shows the weights of the N-grams of "Witch Hunt" in the two schemes.

2.3.6 Mathematical Notation

We view an N-gram profile construction scheme is a triple consisting of the N-gram length N, the choice of whether or not to transpose, and the choice of weighting

Fig. 1 The melody and the chord progression of "Witch Hunt" by Wayne Shorter

Table 2 A comparison between the two weight schemes on "Witch Hunt"

3-gram	Weight Scheme	
	Number of Beats	Frequency
[IMI7 – III$^\flat$7 – IMI7]	0.3265	0.1111
[I$^\flat$7 – VIMI7 – III$^\flat$7]	0.1837	0.1111
[IMI7 – V$^\flat$7 – IV7]	0.1224	0.1111
[I7 – VII7 – VII$^\flat$7]	0.1224	0.2222
[I7 – VII7 – IIIMI7]	0.0662	0.1111
[I7 – IVMI7 – V$^\flat$]	0.0662	0.1111
[IMI7 – II$^\flat$ – IMI7]	0.0662	0.1111
[I – VIIMI7 – VII$^\flat$7($^\flat$5)]	0.0662	0.1111

scheme. For an N-gram profile construction ν and a (simplified) chord progression θ, $\nu(\theta)$ represents the chord progression profile created from θ by applying the N-gram construction scheme ν. For a collection of some k tunes, $\theta_1,\dots,\theta_k$, the collective N-gram profile of the collection with respect to ν is

$$\nu(\{\theta_1,\dots,\theta_k\}) = \frac{1}{k}\nu(\theta_i).$$

We can view an N-gram profile as the set of all pairs $w : c$, where w is a proper N-gram appearing in Π and c is the total contribution of w (since w may appear at more than one place in Π) scaled by the total contribution of all N-grams appearing in Π. Since the N-gram profile is created using a fixed length for N-grams, the total number of N-grams that can appear is finite. By assuming that the weight is 0 for all N-grams not appearing in Π, $\Theta[\tau,N](\Pi)$ can be viewed as a vector of finite dimension, whose entries are all nonnegative and add up to 1.

2.4 Comparison Using Cosine-Based Similarity Measure

Given two vectors with the same number of dimensions, $u = (u_1,\ldots,u_d)$ and $v = (v_1,\ldots,v_d)$, their mutual distance can be measured using various methods. In particular, we test the cosine distance

$$1 - \frac{u_1v_1 + \cdots + u_kv_k}{\sqrt{u_1^2 + \cdots + u_k^2}\sqrt{v_1^2 + \cdots + v_k^2}},$$

and the Hellinger distance,

$$\frac{\sum_{i=1}^{d}(\sqrt{u_i} - \sqrt{v_i})^2}{2}.$$

Both distance measures have the value range of $[0,1]$. Also, both have the property that the value is 0 if and only if $u = v$.

Let π and σ be two chord progressions. Let τ be a simplification and let ν be an N-gram scheme. Let δ be a distance function. Then the distance between π and σ with respect to the triple $\xi = (\tau,\nu,\delta)$ is defined to be

$$\mathrm{dist}[(\tau,\nu,\delta)](\pi,\sigma) = \delta(\nu(\tau(\pi)),\nu(\tau(\sigma))). \quad (1)$$

In other words, it is the distance with respect to δ between the vector representation of the N-gram profile constructed from π by applying τ and ν and the one from σ by applying τ and ν. Given a collection D of such triples, $[(\tau_1,\nu_1,\delta_1),\ldots,(\tau_m,\nu_m,\delta_m)]$, the distance between π and σ with respect to the collection is the average of the m distance values, that is,

$$\mathrm{dist}[\Delta](\pi,\sigma) = \frac{1}{m}\sum_{i=1}^{m}\mathrm{dist}[(\tau_i,\nu_i,\delta_i)](\pi,\sigma) = \frac{1}{m}\delta(\nu(\tau(\pi)),\nu(\tau(\sigma))). \quad (2)$$

3 Proof-of-Concept Analysis

A proof-of-concept analysis has been carried out on a data collected from ten composer groups to test the efficacy of N-gram chord progression profiles.

3.1 Data

A data base of 301 chord progressions is constructed from various sources. The data base consists of the following:

- 218 compositions of composers John Coltrane (28 tunes), Chick Corea (25 tunes), Duke Ellington (25 tunes), Herbie Hancock (16 tunes), Freddie Hubbard (17 tunes), Thelonius Monk (27 tunes), Wayne Shorter (47 tunes), and Horace Silver (33 tunes), collected from Jazz fake books (Real Book 1, 2, and 3; New Real Book 1, 2, and 3; Jazz Limited);
- 63 “standard” tunes from Real Book 1, excluding compositions by modern Jazz musicians and Bossa Nova tunes;
- 20 compositions of The Beatles from the Hal Leonard Publishing “Anthology Volume 3”.

The Beatles compositions are considered to be something very different from standards or Jazz composer tunes. The data can be obtained at the first author’s web page: http://www.cs.miami.edu/∼ogihara/chord-sequence-files.zip.

3.2 Comparison of the Simplification Methods

3.2.1 The choice of N and bass note simplification

To determine the value for N and to choose the bass note simplification, we calculate the cosine-based similarity between the standards and D. Ellington with respect to each of the twelve simplification methods and for $N = 1,2,3,4$. Since D. Ellington played the most prominent role in founding the modern Jazz theory and the chord progressions of the Fakebook standard tunes in some sense summarize the chord sequences resulting from Jazz reharmonization, it is anticipated that the two groups are very similar, in particular, when the tension notes are excluded (namely, T_0 simplification). The similarity values are shown in Table 3. It appears that either $N = 3$ or $N = 4$ will be a good choice.

The choice of the bass note simplification (the B-part) does not seem to affect much the similarity measure, while the choice of the tension note simplification (the T-part) makes a substantial difference, in particular, for 3-grams and 4-grams. The phenomenon that the selection on the bass note simplification does not change much the similarity value can be explained by the fact that only a small fraction (less than 5%) of the chords appearing the data had a bass note. This observation leads us to

Table 3 The cosine-based similarity between the standards and D. Ellington with respect to various simplification methods and for $N = 1,2,3,4$.

Method		N				Method		N			
T	B	1	2	3	4	T	B	1	2	3	4
T_0	B_0	0.990	0.950	0.818	0.579	T_2	B_0	0.950	0.798	0.504	0.197
	B_1	0.990	0.950	0.818	0.579		B_1	0.949	0.797	0.500	0.190
	B_2	0.990	0.950	0.818	0.576		B_2	0.947	0.796	0.497	0.187
T_1	B_0	0.954	0.835	0.628	0.319	T_3	B_0	0.952	0.805	0.502	0.194
	B_1	0.953	0.836	0.630	0.320		B_1	0.951	0.804	0.500	0.189
	B_2	0.952	0.834	0.626	0.310		B_2	0.950	0.804	0.500	0.185

choose B_0 (bass note omission) for the bass note simplification, because it is the simplest operation.

3.2.2 Tension Simplification

We next examine how the similarity values vary depending on the choice of the T-part. It is anticipated that the more aggressive the simplification is, the higher the similarity value becomes. This anticipation is clearly confirmed in Table 3, which shows the similarity values between the standards and the D. Ellington tunes. According to the table, there isn't much difference between the T_2 and T_3 simplifications. Since T_2 is more aggressive than T_3, and thus, the resulting chord notation is generally simpler with T_2 than with T_3, we should choose T_2 over T_3.

We then compare T_0 and T_1 using the songs by The Beatles and those by the others. The similarity values are shown in Table 4. There is a substantial difference in the similarity value between T_0 and T_1. Given that The Beatles is in the Pop/Rock genre and the rest are in Jazz, we feel that T_1 is more appropriate than T_0. Since the

Table 4 Comparison between T_0 and T_1

Composer	1-gram		2-gram	
	T_0	T_1	T_0	T_1
CC	0.933	0.594	0.527	0.250
DE	0.993	0.521	0.715	0.239
FH	0.921	0.570	0.456	0.114
HH	0.827	0.354	0.346	0.078
HS	0.962	0.483	0.621	0.178
JC	0.983	0.562	0.790	0.241
TM	0.998	0.551	0.691	0.243
WS	0.950	0.373	0.500	0.164

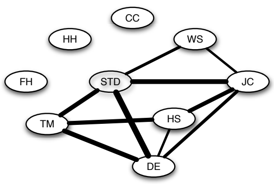
similarity of The Beatles to these composers seems very high for T_0, we consider using T_1 instead of T_0. These observations narrow our choices down to (B_0, T_1) and (B_0, T_2).

Table 5 shows the comparison of the standards against The Beatles, T. Monk, and H. Hancock with respect to the (B_0, T_1)-simplification and the (B_0, T_3)-simplification. We note that as N increases the similarity of the standards more quickly decays with The Beatles and Herbie Hancock than with Thelonius Monk and the decay with respect to the (B_0, T_1) simplification appears to be more dramatic than the decay with resect to the (B_0, T_2) simplification.

Figure 2 shows the cosine-based similarity of the profiles among the Jazz composers with respect to 3-grams and (B_0, T_2)-simplification. Two composers are connected if the similarity is 0.2500 or higher. The thicker the line is, the higher the similarity value is. Since the similarity is symmetric, the upper right portion of the table is left blank and the two $<$'s appearing in the last line indicate that the similarity value is not more than 0.2500.

Table 5 Cosine-distance-based similarity between the standards and each of The Beatles, T. Monk, and H. Hancock

(B_0,T_1)-simplification				(B_0,T_3)-simplification			
	Standards Versus				Standards Versus		
N	The Beatles	T. Monk	H. Hancock	N	The Beatles	T. Monk	H. Hancock
1	0.430	0.922	0.875	1	0.414	0.886	0.829
2	0.163	0.716	0.390	2	0.162	0.676	0.185
3	0.040	0.437	0.114	3	0.040	0.378	0.051
4	0.017	0.199	0.038	4	0.018	0.1580	0.010



(a) The similarity graph of the Jazz composers.

	STD	DE	HS
DE	0.504		
HS	0.349	0.376	
TM	0.379	0.422	0.363
JC	0.402	0.278	0.349
WS	0.267	<	<

(b) The similarity table.

Fig. 2 The composer similarity

This graph seems to reflect well the relations among the composers from the historical perspective. According to the year of the first recording session as a leader, these composers are ordered as follows: Ellington (1924), Monk (1947), Silver (1955), Coltrane (1957), Shorter (1959), Hubbard (1960), Hancock (1962), and Corea (1966). The graph connects among the first five along with the standards and disconnects the remaining three from every one else.

3.3 Artist Clustering Using Profiles

The observation that the 3-gram similarity with respect to the (B_0,T_2) simplification reflects relations among artists from the Jazz historical perspective appears to be more strongly represented in hierarchical clustering of the composers. Figures 3 shows the hierarchical clusters of the composers generated using 3-grams.

4 Exhaustive Analysis Using Top-K Queries

The analysis presented in the previous section shows that among all possible triples of distance measure, simplification method, and N-gram scheme, there exist some combinations that very well reflect the development of Jazz music when they are

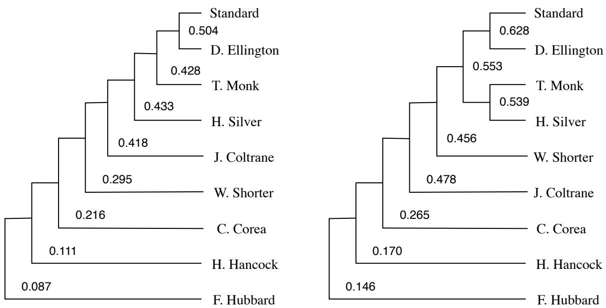


Fig. 3 Hierarchical clustering of the composers. Left panel: with respet to the (B_0, T_1)-simplification. Right panel: with respect to the (B_0, T_2)-simplification.

used for the purpose of comparing different composers. This section considers the problem of distinguishing a composer from the others in a top-k query environment by bringing together more than one triple of distance measure, simplification method, and N-gram scheme.

4.1 Method

Let k be a fixed integer. Suppose that a data base of chord progressions D is given. Let dist$[\Delta]$ be a distance measure. That is, Δ is a series of triples, $\xi_1, \ldots, \xi_r$, where for each i, $1 \leq i \leq r$, δ_i is a triple of distance function, simplification method, and N-gram scheme. We define the precision of dist$[\Delta]$ on a given data set D to be the proportion of compositions Π in D such that the set of k compositions in the data base $D - \{\Pi\}$ that are the closest to Π with respect to dist$[\Delta]$ have at least one composition by the composer of Π.

Let $w_{\max} \geq 1$ be a parameter that bounds from above the N-gram length. Let $r_{\max} \geq 1$ be a parameter that bounds from above the number of triples in Δ. Let $c_{\max} \geq 1$ be a parameter that specifies the number of elements carried over from a stage to the next in the algorithm below. We search for the best distance measure in terms of the aforementioned precision value, in a greedy manner as follows:

Step 1 Set T to the collection of all N-gram schemes where the N-gram length is at most $w_{\max}$. Set U to the collection of all distance functions of interest. Set $\Delta = []$, $C = \{\Delta\}$, and $W = \emptyset$.

Step 2 For $i = 1$ to $r_{\max}$, do the following:

Step 2a For each member Δ in C, for each distance function δ in U, for each simplification method τinS, and for each N-gram scheme $\nu \in T$, do the following:

Step-2a(i) Let Δ' be the series constructed from Δ by appending (τ, ν, δ).

Step-2a(ii) Compute the precision of dist$[\Delta']$ with respect to top k-queries.

Step-2a(iii) If C_0 has less than $c_{\max}$ elements, add $[\Delta']$ to C_0; otherwise, if dist$[\Delta']$ has precision higher than the distance measure in C_0 with the lowest precision among of the group, then replace that distance measure by $[\Delta']$.

Step 2b Set $C = C_0$. Add to W the element in C_0 having the highest precision value.

Step 3 Output the element in W having the highest precision value.

Note that the members of C at the end of each loop body with respect to i has i components each, so the distance measure with the highest precision value produced by the algorithm has at most $r_{\max}$ components. Note also that components in a distance measure may be identical. Since the components are assigned an equal weight, a triple that appears n times receives weight n times as high as the weight a triple appearing only once receives. This in a naive way makes it possible to assign unequal weights to triples.

4.2 Experiments

4.2.1 Data set

A data set consisting of 340 chord progressions is used. The set covers 17 composers and from each composer 20 compositions are selected. The composers are the previous 10 plus seven new: Richie Beirach, Bill Evans (pianist), Keith Jarrett, Pat Metheny, and Steve Swallow; a Brazilian Bossa Nova composer Antonio Carlos Jobim; and an Argentinian "Nuevo Tango" composer Astor Piazzolla.

4.2.2 Parameter Choices

We set $r_{\max}$, the maximum number of rounds, to 10, set $c_{\max}$, the number of distance measures carried over to the next round, to 10, and set $w_{\max}$, the maximum N for the N-gram length N, to 4.

4.3 Results

Figure 4 shows the result of the experiment with respect to top-k queries for $k = 2, \ldots, k = 10$. The precision is the proportion of chord progressions for which at least one of the three closest progressions is composed by the same composer.

In all cases, the precision increases steadily in the first three rounds and then, for a majority of the k-values, the growth tapers off.

Table 6 shows the plotted precision values in a chart. The last row of the table is the baseline precision; that is, the probability that a set of randomly selected pairwise-distinct k compositions from the pool of compositions other than the query contains the composition by the same composer. The query fails when the selected k distinct elements are compositions by someone else. There are 320 compositions composed by someone else and so the number of selections that lead to failure is

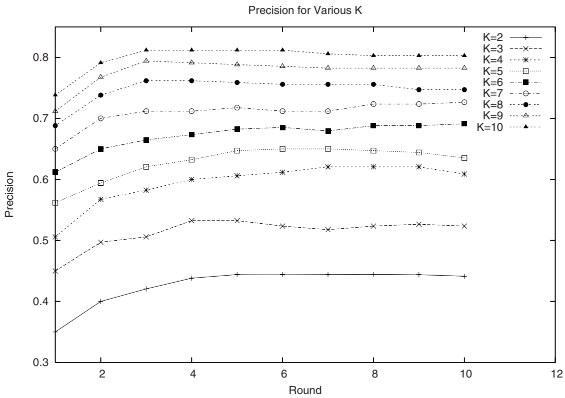


Fig. 4 The accuracy of the best performer in each round for each value of $k, 2 \leq k \leq 10$

Table 6 The precision table

k		2	3	4	5	6	7	8	9	10
Round	1	0.3500	0.4500	0.5059	0.5618	0.6118	0.6500	0.6882	0.7118	0.7382
	2	0.4000	0.4971	0.5676	0.5941	0.6500	0.7000	0.7382	0.7676	0.7912
	3	0.4206	0.5059	0.5824	0.6206	0.6647	0.7118	**0.7618**	**0.7941**	**0.8118**
	4	0.4382	**0.5324**	0.6000	0.6324	0.6735	0.7118	0.7618	0.7912	0.8118
	5	**0.4441**	0.5324	0.6059	0.6471	0.6824	0.7176	0.7588	0.7882	0.8118
	6	0.4441	0.5235	0.6118	**0.6500**	0.6853	0.7118	0.7559	0.7853	0.8118
	7	0.4441	0.5176	**0.6206**	0.6500	0.6794	0.7118	0.7559	0.7824	0.8059
	8	0.4441	0.5235	0.6206	0.6471	0.6882	0.7235	0.7559	0.7824	0.8029
	9	0.4441	0.5265	0.6206	0.6441	0.6882	0.7235	0.7471	0.7824	0.8029
	10	0.4412	0.5235	0.6088	0.6353	**0.6912**	**0.7265**	0.7471	0.7824	0.8029
Best		0.4441	0.5324	0.6206	0.6500	0.6912	0.7265	0.7618	0.7941	0.8118
Baseline		0.1091	0.1593	0.2069	0.2519	0.2944	0.3347	0.3728	0.4088	0.4428
Gap		0.3350	0.3731	0.4137	0.3981	0.3968	0.3918	0.3890	0.3853	0.3690

$\binom{320}{k}$. On the other hand, since there are 339 compositions other than the query itself, the number of possible selections of k distinct elements is $\binom{339}{k}$. Thus, the probability of failure is $\binom{320}{k}/\binom{339}{k}$ and the probability of success is: $1 - \binom{320}{k}/\binom{339}{k}$. Note that the gain from the baseline by the use of chord progression profile ranges from 0.33 to 0.41. This indicates that the chord progression profile can be a highly effective method for identifying compositions by the same composer.

Table 7 shows the summary of precision values of the 17 composers over the ten rounds for top-5 query analysis. At each round, 10 distance measures that achieved the highest precision are selected. For each such measure (there are a total of 100 measures), the precision (or accuracy) is calculated with respect to each artist. The maximum, minimum, the average, and the standard deviation of the 100 values for

Table 7 Composer-wise accuracy distribution for top-5 queries. The composers are presented in the decreasing order of average accuracy.

Composer	Max	Min	Average	StdDev
BEATLES	1.0000	0.8500	0.9335	0.0332
STANDARDS	1.0000	0.7500	0.9025	0.0597
ASTOR PIAZZOLLA	0.9500	0.7500	0.8915	0.0469
THELONIUS MONK	0.8500	0.5000	0.7685	0.0599
WAYNE SHORTER	0.8500	0.2500	0.7360	0.1229
DUKE ELLINGTON	0.8500	0.4000	0.7500	0.0745
BILL EVANS	0.8500	0.2000	0.6595	0.0999
PAT METHENY	0.8000	0.3500	0.6140	0.0645
RICHIE BEIRACH	0.6500	0.4000	0.5930	0.0806
KEITH JARRETT	0.7000	0.2500	0.5625	0.0931
HORACE SILVER	0.6500	0.3500	0.5345	0.0523
ANTONIO CARLOS JOBIM	0.7000	0.3500	0.5275	0.0676
STEVE SWALLOW	0.6000	0.3000	0.4945	0.0509
JOHN COLTRANE	0.7500	0.4000	0.4845	0.0695
HERBIE HANCOCK	0.5000	0.2000	0.4325	0.0694
FREDDIE HUBBARD	0.7500	0.2000	0.3890	0.0780
CHICK COREA	0.5500	0.2000	0.3325	0.0719

each artist are presented. The 10 top composers in this ranking are: The Beatles, Standards, Astor Piazzolla, Thelonius Monk, Wayne Shorter, Duke Ellington, Bill Evans, Pat Metheny, Richie Beirach, and Keith Jarrett. The standard deviation is small for The Beatles, Standards, Astor Piazzolla, Thelonius Monk, Horace Silver, and Steve Swallow. This indicates that for these composers the top-k query makes consistent performance.

Next, for each value of k the distance measure components (that is, the triples of simplification, N-gram scheme, and distance function) are collected from the best performing distance measure. A total of 58 components are collected, which are shown in Table 8. The most frequently occurring simplifications are (B_0, T_1), (B_0, T_2), and (B_2, T_0). They appear 11 times, 9 times, and 7 times, respectively. The N-gram length is 1 for 15 times, 2 for 11 times, 3 for 7 times, and 4 for 15 times. The average length is 2.07.

5 Conclusion

This article explores the question of whether a chord progression can be summarized for music retrieval. Various possibilities for chord progression simplification schemes, N-gram construction schemes, and distance functions are explored. Experiments demonstrate that such profiles can be used for artist grouping and for composition retrieval via top-k queries. The precision of nearly 65% is achieved with top-5 queries involving 17 composers, with a large margin of 40% from the

Table 8 The table of components appearing in the best performing distance measures

Bass	Count	Tension	Count	Transpose	Count	Length	Distance	Count
B_0	29	T_0	4	No	2	4	Cosine Frequency	1
						2	Hellinger Weight	1
				Yes	2	4	Hellinger Frequency	1
						2	Hellinger Weight	1
		T_1	11	No	10	1	Cosine Frequency	1
						1	Hellinger Frequency	2
						1	Hellinger Weight	3
						3	Cosine Weight	1
						4	Cosine Frequency	2
						4	Hellinger Weight	1
						4	Cosine Weight	1
				Yes	1	4	Cosine Weight	1
		T_2	9	No	5	1	Hellinger Frequency	3
						4	Cosine Weight	1
						4	Cosine Frequency	1
				Yes	4	2	Cosine Frequency	4
		T_3	5	No	2	4	Hellinger Weight	1
						4	Cosine Frequency	1
				Yes	3	2	Cosine Frequency	2
						2	Hellinger Frequency	1
B_1	5	T_1	1	No	1	1	Hellinger Frequency	1
		T_2	2	Yes	2	1	Hellinger Frequency	1
						2	Cosine Frequency	1
		T_3	2	Yes	2	1	Cosine Weight	1
						4	Hellinger Weight	1
B_2	12	T_0	7	No	4	3	Hellinger Weight	1
						3	Cosine Frequency	3
				Yes	3	4	Hellinger Weight	1
						4	Cosine Frequency	2
		T_1	4	No	2	3	Hellinger Frequency	1
						4	Cosine Weight	1
				Yes	2	1	Hellinger Frequency	1
						3	Cosine Weight	1
		T_2	1	No	1	4	Cosine Frequency	1
B_3	2	T_1	1	Yes	1	1	Hellinger Frequency	1
		T_1	1	Yes	1	1	Cosine Weight	1

baseline of 25%. This result seems highly promising. An interesting question will be how the performance decays for much larger sets of diverse composers. Another question is whether the N-gram profiles will be effective in identifying composers in terms of composer classification or genre/style classification. Finally, it will be interesting to study extensions of such approaches to include melodic fragments.

References

1. Brandt, C., Roemer, C.: Standardized chord symbol notation: a uniform system for the music profession, 2nd edn. Roerick Music Co., Sherman Oaks (1976)
2. Cahill, M., O'Maidín, D.: Melodic similarity algorithms – using similarity ratings for development and early evaluation. In: Proceedings of the 6th International Conference on Music Information Retrieval, pp. 450–453 (2005)
3. Doraisamy, S.C., Rüger, S.M.: Robust polyphonic music retrieval with n-grams. Journal of Intelligent Information Systems 21(1), 53–70 (2003)
4. Downie, J.S.: Evaluating a simple approach to music information retreival: Conceiving melodic n-grams as text. PhD thesis, University of Western Ontario, London, Ontario, Canada (1999)
5. Jurafsky, D., Martin, J.H.: Speech and Language Processing. Prentice Hall, Upper Saddle River (2000)
6. Kline, R.L., Glinert, E.P.: Approximate matching algorithms for music information retrieval using vocal input. In: Proceedings of the Eleventh ACM International Conference on Multimedia, pp. 130–139 (2003)
7. Mauch, M., Dixon, S., Casey, M., Harte, C., Fields, B.: Discovering chord idioms through Beatles and Real Book songs. In: Proceedings of the International Symposium on Music Information Retrieval, pp. 255–258 (2007)
8. Ogihara, M., Li, T.: N-gram chord profiles for composer style representation. In: Proceedings of 9th International Conference on Music Information Retrieval, pp. 671–676 (2008)
9. Paiement, J.-F., Eck, D., Bengio, S., Barber, D.: A graphical model for chord progressions embedded in a psychoacoustic sapce. In: Proceedings of the 22nd International Conference on Machine Learning, Bonn, Germany (2005)
10. Swanson, R., Chew, E., Gordon, A.: Supporting musical creativity with unsupervised syntactic parsing. In: Creative Intelligent Systems, AAAI Spring Symposium Series (2008)
11. Tojo, S., Oka, Y., Nishida, M.: Analysis of chord progression by HPSG. In: AIA 2006: Proceedings of the 24th IASTED International Conference on Artificial Intelligence and Applications, Anaheim, CA, USA, pp. 305–310. ACTA Press (2006)
12. Uitdenbogerd, A.L., Zobel, J.: Matchng techniques for large music databases. In: Proceedings of the Seventh ACM International Conference on Multimedia, pp. 57–66 (1999)
13. Volk, A., Garbers, J., van Kranenborg, P., Wiering, F., Veltkamp, R.C., Grijp, L.P.: Applying rhythmic similarity based on inner metric analysis to folksong research. In: Proceedings of the Eighth International Symposium on Music Information Retrieval, pp. 293–300 (2007)