Chroma Interval Content as a Key-Independent Harmonic Progression Feature

Marcelo Queiroz
Computer Science Dept, University of São Paulo, Brazil
mqz@ime.usp.br

Rodrigo Borges
Computer Science Dept, University of São Paulo, Brazil
rcborges@ime.usp.br

***Abstract*—This paper introduces a novel chroma-based harmonic feature called Chroma Interval Content (CIC), which extends Directional Interval Content (DIC) vectors to audio data. This feature represents key-independent harmonic progressions, but unlike the Dynamic Chroma feature vector it represents pitch-class energy motions based on a symbolic voice-leading approach, and can be computed more efficiently (in time $\mathcal{O}(N \log N)$ as opposed to $\mathcal{O}(N^2)$). We present theoretical properties of Chroma Interval Content vectors and explore the expressive power of CIC both in representing isolated chord progressions, establishing links to its symbolic counterpart DIC, as well as in specific harmony-related MIR tasks, such as key-independent search for chord progressions and classification of music datasets according to harmonic diversity.**

***Index Terms*—chroma features, chord progressions, key-independent representations, directional interval content**

I. Introduction

Models for representation of harmonic progressions (e.g. chord transitions) are useful for a number of MIR tasks, such as indexing and retrieval of items from music collections [6] and similarity-based classification and identification (e.g. genre classification [1] and version/cover song identification [18]), but also for systematic musicological studies related to harmonic idioms [2], [3] and automatic harmonization [9].

The literature offers some alternatives for representing chord progressions, both in the symbolic and the audio domains. Symbolic chord sequences may be represented using Lehrdahl's Tonal Pitch Space hierarchical model based on the circle-of-fifths and de Haas et al. TPS Distance [6], a function which expresses the changes of chordal distance to the tonic over time for any two given chord sequences through a single integer value. Paiement et al. [14] describe a symbolic feature built from a theoretical model of harmonic energy distribution, and propose to measure chord changes in terms of Euclidean distances between such features. Cambouropoulos et al. Directional Interval Content vectors [2], [3] aim at creating a more detailed map of the harmonic flow between voices in a chord progression, by counting intervals between every possible pair of pitch classes from the antecedent and consequent chords.

Harmonic content in audio data is frequently expressed in terms of *chroma vectors* (also referred to as *pitch class profiles*), which may be understood as energy histograms per pitch class, computed for short audio segments. There are several ways of defining and computing chroma features, both in the symbolic and audio domains (e.g. pitch histograms, CENS, HPCP, see [13], [20]). While chroma vectors describe static harmonic contents, chord progressions are characterized by the changes from one chord to the next; more specifically, chord progressions may be expressed by variational chroma features, such as Delta Chroma and Dynamic Chroma vectors [10]. Delta Chroma is defined simply as the difference between chroma vectors representing consecutive chords, and measures ebbs and tides of pitch class energies as harmony moves from one chord to the next. Dynamic Chroma, on the other hand, considers all possible rotations[1] of the consequent chord, producing a vector of likelihoods that the chord progression would correspond to a certain chord rotation.

The goal of this paper is to introduce Chroma Interval Content (CIC), a novel audio feature based on chroma and DIC for representing chord progressions which is key-independent (or transpositionally-invariant), i.e. one which represents changes in harmonic content in a relative way very much like traditional or functional harmony express all chords as relative to a certain tonal center, and which at the same time is not dependent on the idea of tonality or on specific harmonic idioms. The only other audio feature with similar properties is Dynamic Chroma, which is key-independent, but its idiom-independence remains untested: being motivated by considering differences between chromas of rotated chords, this feature reflects a musical model based on harmonic functions which are obtained by rotation (e.g. as G-major is a +7-semitone rotation of C-major). By contrast, CIC achieves both its key-independence and idiom-independence by extending DIC from the symbolic domain to the chroma domain, through viewing chord progressions as multi-layered displacements of chroma energy in many simultaneous directions, similarly to the harmonic flows in voice-leading. By avoiding explicit rotations and exploring mathematical structure, the new feature also reduces the theoretical complexity from $\mathcal{O}(N^2)$ to $\mathcal{O}(N \log N)$ for a chord progression between N-dimensional chroma vectors.

This paper is structured as follows. Section II formally introduces CIC, comparing it to Dynamic Chroma and discussing

The first author acknowledges the support of FAPESP grant 2018/09373-8 and CNPq grant 309645/2016-6.

[1]The term *rotation* of a chord refers to pitch class *transposition* (i.e. circularly shifting all pitch classes by a fixed number of semitones) and is formally defined in Section II.

some of its mathematical and musical properties. Section III provides experimental evidence of CIC's usefulness for indexing and retrieval of harmonic progressions and for harmony-based classification tasks. Finally, concluding remarks and ideas for future work are presented in Section IV.

II. Chroma Interval Content

Review of DIC: Directional Interval Content (DIC) vectors have been introduced in [2], [3] as an idiom-independent symbolic harmonic feature for characterizing harmonic progressions represented by pitch-class sets, which is inspired by Lewin's interval function [11]. Pc sets are meant to abstract from and simplify general chord representations in terms of pitches, inversions, instruments or voices, by using octave equivalence and keeping the bare minimum that allows us to distinguish chord types, i.e. the set of pitch-classes (integers from 0 to $N-1$, where $N=12$ for the chromatic scale) that belong to a given chord.

For each potential voice-leading motion between pc sets $A, B \subseteq \{0, 1, \ldots, N-1\}$, the DIC harmonic feature counts the corresponding displacement (in pc steps modulo N, or semitones for $N = 12$) and thus form an N-dimensional vector DIC with DIC$[n]$ representing how many potential voice-leading motions of n steps exist in the chord progression $A \longmapsto B$. For example, between a C-major chord $\{0, 4, 7\}$ and a G-major chord $\{2, 7, 11\}$ (see Figure 1) we find possible

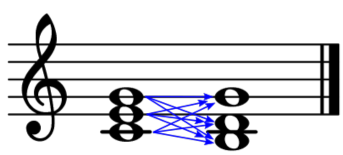

Fig. 1. Potential voice-leading motions between C-major and G-major.

voice motions between $0 \mapsto 2$, $0 \mapsto 7$, $0 \mapsto 11$, $4 \mapsto 2$, $4 \mapsto 7$, $4 \mapsto 11$, $7 \mapsto 2$, $7 \mapsto 7$, $7 \mapsto 11$, and so for this example DIC $= (1, 0, 1, 1, 1, 0, 0, 3, 0, 0, 1, 1)$. We write "potential voice-leading motions" to emphasize that no voices are actually known from pc sets, but actual instrumentations/orchestrations of a given chord progression might include any of those potential voice-leading motions (we are ignoring idiom-dependent stylistic conventions that might forbid some of these motions).

Specifically, starting off from a vector $\mathrm{DIC}_{A\mapsto B} = 0$, each possible pair $(a, b) \in A \times B$ entails a unit increment in the counter $\mathrm{DIC}_{A\mapsto B}[b-a]$. This algorithmic definition (as it is expressed in [2], [3]) can be formalized in many ways, e.g.

$$\mathrm{DIC}_{A\mapsto B}[n] = \sum_{a\in A}\sum_{b\in B} \delta_{n,(b-a)\%N} \tag{1}$$

by using Kronecker's delta $\delta_{i,j} = \begin{cases} 1 & \text{if } i = j, \\ 0 & \text{otherwise,} \end{cases}$ or

$$\mathrm{DIC}_{A\mapsto B}[n] = \sum_{m=0}^{N-1} I_m^A I_{(m+n)\%N}^B \tag{2}$$

by using indicator functions $I_i^X = \begin{cases} 1 & \text{if } i \in X, \\ 0 & \text{otherwise.} \end{cases}$

It can be seen that both definitions are equivalent[2] and transpositionally-invariant[3], and, while the first form seems to reflect more directly the algorithmic definition, it is the second one that lends itself to a generalized and more efficient extension from DIC's symbolic domain to the audio domain of chroma vectors, which instead of indicating which pitch classes belong or not to a certain chord, represent how strongly each pitch class is represented in a corresponding audio signal.

Formal definition of CIC: given two chroma vectors $X, Y \in \mathbb{R}_+^N$ representing subsequent chords in an audio signal, we define

$$\mathrm{CIC}_{X\mapsto Y}[n] = \sum_{m=0}^{N-1} X_m Y_{(m+n)\%N}, \tag{3}$$

which can be recognized as the cross-correlation between X and Y, or the circular convolution between X and $\overleftarrow{Y}$, the pitch-class inversion[4] of Y, with all indexes also inverted, i.e.,

$$\mathrm{CIC}_{X\mapsto Y} = \overleftarrow{X * \overleftarrow{Y}}.$$

This immediately suggests a fast convolutional computation of CIC as

$$\mathrm{CIC}_{X\mapsto Y} = \overleftarrow{\mathcal{F}^{-1}\left(\mathcal{F}(X)\mathcal{F}\left(\overleftarrow{Y}\right)\right)}, \tag{4}$$

where $\mathcal{F}$ denotes the DFT operator. Equation 4 can be computed in time $\mathcal{O}(N \log N)$ by using an FFT implementation of the DFT, as opposed to $\mathcal{O}(N^2)$ for N applications of Equation 3. It is evident that this theoretical advantage only translates into significant computational savings for higher-dimensional chroma vectors (e.g. quarter-tone, cent scale, etc). Although higher-dimensional chroma representations are not popularly used for the analysis of traditional music based on the 12-semitone scale, their interest is evidenced in ethnomusicological studies [8], [17], in the study of tonal stability in western baroque and classical music [7] and also in general MIR tasks [16].

Comparison to Dynamic Chroma: It is useful to compare the above definition with the Dynamic Chroma feature vector as defined in [10]:

$$\mathrm{DC}_{X\mapsto Y}[n] = Z - \|Y^{(n)} - X\|,\ n = 0, 1, \ldots, N-1, \tag{5}$$

where $Y_k^{(n)} = Y_{(k-n)\%N}$ and $Z = \max_n\{\|Y^{(n)} - X\|\}$. According to its authors, their intention is to represent a likelihood that Y would be a circularly-rotated version of X, for each possible rotation $Y^{(n)}$, $n = 0, 1, \ldots, N-1$. This is especially appealing if indeed X and Y share the same chord structure (e.g. if both are major chords), because in this case one of the differences $\|Y^{(n)} - X\|$ might be very close to zero and the corresponding Dynamic Chroma (DC)

[2] $I_m^A I_{(m+n)\%N}^B = 1 \iff a = m \in A \wedge b = (m+n)\%N \in B \iff (b-a)\%N = n \iff \delta_{n,(b-a)\%N} = 1$.

[3] $C \mapsto D$ with $C = (A+n)\%N$ and $D = (B+n)\%N$ implies $\mathrm{DIC}_{C\mapsto D} = \mathrm{DIC}_{A\mapsto B}$.

[4] $\overleftarrow{Y}_n = Y_{(-n)\%N}$, for $n = 0, \ldots, N-1$.

component would be maximum. DC is easily seen to be key-independent (since $\|Y^{(n)} - X\| = \|Y^{(m+n)} - X^{(m)}\|$ for all m) and computable in $\mathcal{O}(N^2)$ time.

Mathematical and musical properties of CIC: Unlike the DC feature vector, DIC's original formulation (Equation 1) is not oriented towards finding possible matches between rotated chroma vectors of similarly-structured chords. CIC's extension to chroma vectors through generalization of Equation 2 is not motivated by measuring componentwise differences between subsequent chroma vectors, but by aggregating information that would represent chroma energy redistribution, i.e. how energy from each pitch class in the preceding chroma vector is redistributed to other pitch classes in the subsequent chroma vector, similarly to how DIC operates on potential voice-leading motions between chords of arbitrary harmonic structures. One way of verifying the plausibility of this interpretation is to consider a simple scenario in which both chroma vectors are binary, i.e. $X, Y \in \{0,1\}^N$. In this very simple case, X and Y can be interpreted as indicator vectors for pc sets $A = \{n|X_n = 1\}$ and $B = \{n|Y_n = 1\}$, and Equations 2 and 3 are actually the same, i.e.

$$\mathrm{CIC}_{X\mapsto Y} = \mathrm{DIC}_{A\mapsto B}$$

(since $X = I^A$ and $Y = I^B$). In this sense, since any chord progression between pc sets $A \mapsto B$ can be equivalently represented by a chord progression between corresponding indicator vectors $X \mapsto Y$, it is seen that the proposed CIC feature is actually an *extension* of the symbolic DIC vectors to general (i.e. non-binary) chroma vectors.

CIC inherits all mathematical properties of convolutions of non-negative vectors, such as commutativity, linearity in both arguments, invariance under identical circular rotations of both arguments, and monotonicity in both arguments. Many musically-relevant properties of CIC, such as key-independence, amplitude dependence and harmonic monotonicity, are also immediately inherited from DIC or translatable from equivalent mathematical properties.

Specifically, identical circular rotations of $X^{(n)}$ and $Y^{(n)}$, an operation musically equivalent to key transposition by $+n$ semitones within a tonal idiom, translates simply into a reordering of the terms in Equation 3, and so

$$\mathrm{CIC}_{X^{(n)}\mapsto Y^{(n)}} = \mathrm{CIC}_{X\mapsto Y},$$

which explains that CIC, like DC, is a *key-independent* harmonic progression feature, but unlike DC this property is not dependent on an explicit enumeration of all rotations of Y (as per Equation 4). A similar property corresponds to the effect of transposing only one of the chords: in this case, illustrated in Figure 2 through the progressions C-major↦G-major and C-major↦D-major, the resulting CIC is rotated by the same amount, i.e.

$$\mathrm{CIC}_{X\mapsto Y^{(n)}} = \mathrm{CIC}^{(n)}_{X\mapsto Y},$$

as can be seen by a simple change of variables in Equation 3.

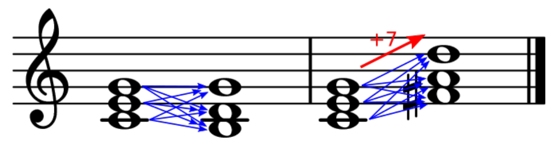

Fig. 2. Effect on CIC of rotation of second chord: all voice-leading motions are warped by the same amount.

Both DC and CIC are affected by differences in amplitude levels between chroma vectors X and Y, which may be circumvented by working with l_∞-normalized chroma vectors[5]. By rewriting Equation 3 as

$$\overline{\mathrm{CIC}}_{X\mapsto Y}[n] = \sum_{m=0}^{N-1} \frac{X_m}{\|X\|_\infty} \frac{Y_{(m+n)\%N}}{\|Y\|_\infty},$$

it is seen that $\overline{\mathrm{CIC}}$ is an *amplitude-independent* feature which also extends DIC (since in the case of non-zero binary vectors $\|X\|_\infty = \|Y\|_\infty = 1$ implies $\overline{\mathrm{CIC}}_{X\mapsto Y} = \mathrm{CIC}_{X\mapsto Y}$). As an audio feature measuring relative levels of displacement of energy among pitch classes, another interesting normalization is

$$\overline{\overline{\mathrm{CIC}}}_{X\mapsto Y} = \frac{\mathrm{CIC}_{X\mapsto Y} - \min(\mathrm{CIC}_{X\mapsto Y})}{\max(\mathrm{CIC}_{X\mapsto Y}) - \min(\mathrm{CIC}_{X\mapsto Y})},$$

which maps all CIC values to the interval $[0,1]$. This type of normalization proved to be useful in applying CIC to MIR tasks and was used in Section III.

Commutativity of convolution corresponds to two musically meaningful properties of CIC, namely

$$\mathrm{CIC}_{X\mapsto Y} = \overleftarrow{\mathrm{CIC}_{Y\mapsto X}}$$

and

$$\mathrm{CIC}_{X\mapsto Y} = \mathrm{CIC}_{\overleftarrow{Y}\mapsto\overleftarrow{X}}.$$

Musically, if X, Y are binary chroma vectors the first property states that the reversed chord progression $Y \mapsto X$ has

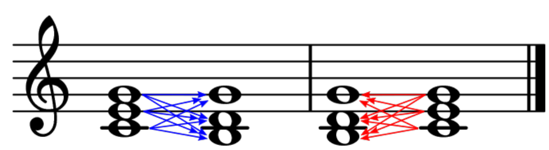

Fig. 3. Effect on CIC of chord reversal: all voice-leading motions are inverted.

the same potential voice-leading motions as $X \mapsto Y$, but each individual voice-leading motion is inverted, as in the example of Figure 3. The second property states that the chord progression from the pitch-class inversion[6] of chord $B = \{n|Y_n = 1\}$ back to the pitch-class inversion of chord

[5] where each component of a vector X is divided by $\|X\|_\infty = \max\{X_n\}$. This is the most commonly used normalization for chroma vectors.

[6] If C is a pc set, then $\overleftarrow{C} = (-C)\%N = \{(-c)\%N \mid \forall c \in C\}$. This is equivalent to reflecting each pitch class using 0 as a mirror: B becomes D♭, E becomes A♭, etc.

$A = \{n | X_n = 1\}$ possesses the exact same set of potential voice-leading motions found in chord progression $A \mapsto B$. Figure 4 shows an example for the progression C-major↦G-

Fig. 4. Effect on CIC of chord inversion and reversal: all voice-leading motions are precisely the same, even though they appear to be mirrored. For instance, all blue arrows entering note B in the first chord progression, corresponding to displacements of -1, -5 and -8 semitones, also appear leaving D♭ in the second chord progression; this corresponds to D♭ being the inversion of B (D♭ $= 1 = -11\%12 = \overleftarrow{B}$). The same can be verified for blue arrows entering D and red arrows leaving B♭ (D = $\overleftarrow{B\flat}$), and also for blue arrows entering G and red arrows leaving F (G = $\overleftarrow{F}$).

major and its pitch-class inversions B♭-minor↦F-minor. The same interpretation is extended through CIC to progressions between arbitrary chroma vectors by considering voice-leading motions as chroma energy redistributions.

Harmonic monotonicity is another musically meaningful property of CIC also satisfied by DIC: if one or both chords in a chord progression are enlarged, i.e. $A \subseteq \hat{A}$ and $B \subseteq \hat{B}$, then

$$\text{DIC}_{A \mapsto B} \leq \text{DIC}_{\hat{A} \mapsto \hat{B}},$$

(meaning component-wise $\leq$) since all potential voice-leading motions of $A \mapsto B$ exist in $\hat{A} \mapsto \hat{B}$. This translates immediately into harmonic monotonicity for CIC if applied to chord progressions between binary chroma vectors $X, Y, \hat{X}, \hat{Y} \in \{0,1\}^N$: i.e. $X \leq \hat{X}$ and $Y \leq \hat{Y}$ implies

$$\text{CIC}_{X \mapsto Y} \leq \text{CIC}_{\hat{X} \mapsto \hat{Y}}.$$

This property extends beyond binary vectors to arbitrary non-negative vectors $X, Y, \hat{X}, \hat{Y} \in \mathbb{R}_+^N$, a fact that can be immediately derived from Equation 3.

III. Experiments and Discussion

In this section we present preliminary experiments which aim to identify potential application cases for the proposed feature. We start with an experiment involving symbolic representations of isolated harmonic progressions within an index/search context, and then address a genre classification experiment on a music dataset. In the following experiments, the [0,1]-normalized form $\overline{\overline{\text{CIC}}}$ is used throughout due to its better empirical performance.

A. Indexing Harmonic Progressions

The first experiment uses a fairly simplified scenario in order to explore the theoretical proximity between the symbolic DIC and the chroma-based CIC features, using ad-hoc queries for retrieving chord progressions from synthesized examples representing a large palette of timbres. The question that motivates this experiment is the following: is CIC a representation feature for chord progressions that allows us to search for specific configurations within an audio file, while being confident that similar chord progressions would be retrieved? In other words, will variations in timbre, which are known to affect chroma, be tolerated by a search mechanism based on symbolic representations of chord progressions, e.g. CIC vectors built from binary chroma templates as search keys, as is done for instance in chord recognition?

We adopted the following method for this experiment:

- a dataset $\mathcal{S}$ of all three-note chords that may be obtained from a given diatonic scale $\{0, 2, 4, 5, 7, 9, 11\}$; due to the key-independence and single-rotation properties of CIC we restricted $\mathcal{S}$ to chords containing pitch-class 0, since any other chord could be circularly shifted towards 0 without affecting the sequence of DIC values (only its particular rotation). There are 15 such chords, including major, minor, diminished and augmented triads, but also clusters and suspended chords such as $\{0, 2, 4\}$, $\{0, 5, 7\}$, as well as other chord types.
- for every chord A in $\mathcal{S}$ a set of pseudo-chroma vectors $\mathcal{C}(A)$ was built based on geometric progressions of amplitudes $\varrho^k,\ k = 0, 1, \ldots, K-1$ with several amplitude decaying factors $\varrho \in \{0.2, 0.4, 0.6, 0.8, 1.0\}$ and number of harmonics $K \in \{1, 2, 5, 10, 15, 20\}$, totaling 30 different pseudo-chroma realizations for each chord.
- for every chord progression $A \mapsto B \in \mathcal{S} \times \mathcal{S}$ a binary archetype $Z = \text{CIC}_{I_A \mapsto I_B} (= \text{DIC}_{A \mapsto B})$ is used as a search item within the set

  $$\mathcal{T} = \{\text{CIC}_{c_a \mapsto c_b} \mid \forall c_a \in \mathcal{C}(A), c_b \in \mathcal{C}(B)\}$$

  of CIC vectors obtained from all pseudo-chroma realizations of the chord progression $A \mapsto B$, using the sets $\mathcal{C}(A)$ and $\mathcal{C}(B)$ defined above. The search retrieves the closest vector $\text{CIC}_{c_a \mapsto c_b}$ to Z according to the cosine distance, and is considered successful if the chord progression $a \mapsto b \in \mathcal{S} \times \mathcal{S}$ is equivalent to the original progression $A \mapsto B$ in the sense that $\text{CIC}_{I_a \mapsto I_b} = \text{CIC}_{I_A \mapsto I_B}$.

For the sake of comparison, the exact same method was applied to DC to check for its adequateness in retrieving chord progressions, represented as binary archetypes $\text{DC}_{I_A \mapsto I_B}$, from the set of DC vectors obtained from pseudo-chroma realizations of all chord progressions

$$\mathcal{T}' = \{\text{DC}_{c_a \mapsto c_b} \mid \forall c_a \in \mathcal{C}(A), c_b \in \mathcal{C}(B)\}.$$

Interestingly enough, both methods did retrieve 100% of correct results for a large number of progressions in $\mathcal{S} \times \mathcal{S}$: 60.4% of all such progressions got a perfect score for CIC, and 67.6% for DC, meaning that recovery from the binary archetypes $\text{CIC}_{I_A \mapsto I_B}$ and $\text{DC}_{I_A \mapsto I_B}$ was successful for all pseudo-chroma realizations of $A \mapsto B$ for these progression types. Taking all searches combined the overall scores were 96.9% for CIC and 97.3% for DC.

Among progression types that didn't achieve a perfect score, some were slightly favored by CIC and some by DC, which almost always produced similar scores, retrieving the correct chord progressions in between 70% and 96.7% of all pseudo-chroma realizations (each feature within 6.7% of the other's score, or equivalently 1 or 2 misclassified pseudo-chroma realizations for each such chord progression). These cases covered 98.7% of all 225 chord progressions in $\mathcal{S} \times \mathcal{S}$.

The remaining 3 cases (1.3% of 225) were chord progressions $\{0, 2, 4\} \mapsto \{0, 2, 4\}$, $\{0, 2, 11\} \mapsto \{0, 7, 9\}$ and $\{0, 7, 9\} \mapsto \{0, 2, 11\}$, for which DIC obtained 80% of successful searches against respectively 100%, 93.3% and 93.3% for DC.

Based on these results it may be safely assumed that both DIC and DC are good representations for chord progressions, in the sense that their binary archetypes $\text{DIC}_{I_A \mapsto I_B}$ and $\text{DC}_{I_A \mapsto I_B}$ built out of binary chroma versions of a chord progression $A \mapsto B$ are good representatives to be used in searching mechanisms, being able to identify accurately over 95% of possible pseudo-chroma realizations of $A \mapsto B$ in the experimental setting described.

B. Genre Classification

In this second experiment, we addressed the following question: how well do CIC vectors capture certain harmonic styles, in the sense of allowing the comparison of different music pieces in terms of diversity of chord progressions? To try to answer this question, a harmony-based genre classification task was set up where both CIC and DC features were employed. It should be mentioned that genre-classification is a difficult task, for which harmony alone is certainly not sufficient to allow precise discrimination, due not only to genre's dependency on all music aspects such as timbre, rhythm and even lyrics, but also due to the very disagreement of genre attribution between listeners [15]. Nevertheless, the goal here is simply to have a preliminary comparison between two alternative representations for chord progressions within the context of a popular MIR task.

Both CIC and DC are calculated from chroma vectors representing *subsequent chords*, and so one requirement of the audio dataset was to provide a reliable segmentation of frames corresponding to the same chord, preferably with chord annotations that would allow double-checking of computed CIC and DC features. One dataset that fulfilled this requirement is described in detail by Clercq and Temperley [5], and is based on a selection made by Rolling Stone magazine entitled "500 Greatest Songs of All Time" of popular pieces from the 1950's through the 1990's. The harmonic transcriptions were manually done by both researchers for 200 of these songs, and are available online[7].

Genre-related information is not available in the original dataset, but can be obtained from Dave Tompkins' Music Database [19]. This repository classifies the same 500 songs into 16 genres, along with a TBD label for unclassified songs. Considering only songs for which harmonic annotations were available and classes with at least 10 songs, we arrived at a reduced dataset comprising 128 songs, distributed among three genres according to Table I.

[7]http://rockcorpus.midside.com/; Accessed December 09th, 2018.

Genre	# of items
Rock	85
Slow	33
Dance	10

TABLE I
GENRES AND CORRESPONDING NUMBER OF SONGS IN THE REDUCED DATASET.

For each song in the reduced dataset, average chroma features were obtained for each segment corresponding to a single chord: individual chromas with $N = 12$ were computed using Librosa [12] over 2048-sample windows. For each pair of average chroma vectors corresponding to adjacent single-chord segments, CIC and DC features were obtained. Each song is thus represented through a CIC-gram or DC-gram of size $12 \times M$, where $M + 1$ is the length of the annotated chord sequence for that song (i.e. there are M chord transitions).

In order to produce a comparable representation among all songs, independently of M, a bag-of-words approach has been used. This representation is based on a clusterization of all possible CIC (or DC) vectors into K clusters with centroids $c_1, \ldots, c_K$, and then each song is represented by a normalized histogram with counters associated to the number of its features closest to the k-th centroid. Clusterization has been carried out using Kmeans, where K has been chosen to optimize the average silhouette coefficient

$$\sum_i \frac{1}{L} \frac{b(i) - a(i)}{\max\{a(i), b(i)\}},$$

where L = total of CIC vectors, i is a CIC vector in cluster I, $a(i) = \max\{d(i, j) \mid j \in I\}$ and $b(i) = \min\{d(i, j) \mid j \notin I\}$. For K between 2 and 50 the optimal silhouette coefficient was obtained for $K = 4$ for CIC and $K = 5$ for DC. After the clusterization step each song is then associated to its K-order histogram, which is used in the classification task.

XGBoost [4] has been used for classifying songs according to the designated genres, due to its performance and ease of use; we used Scikit-learn's XGBoost implementation with parameters γ=0, learning_rate=0.2 and max_depth=15.

For this multi-class classification task a series of binary one-versus-all classifiers are employed, and performance measurements may be combined in several ways:

macro-average is an average over all binary classification tasks with equal weights, which is not suited for unbalanced scenarios where classes have very different sizes;

weighted-average considers the size of each class as the weight applied to the corresponding performance measurement;

micro-average instead of combining isolated performance measurements, all individual labeling results (TP, FP, TN and FN)[8] are combined, and an overall performance measurement is derived from these values.

[8]These are standard metrics for binary classification: TP (True Positives), FP (False Positives), TN (True Negatives), and FN (False Negatives).

Average	CIC performance	DC performance
Macro	0.30 (± 0.10)	0.24 (± 0.04)
Weighted	0.54 (± 0.10)	0.47 (± 0.09)
Micro	0.59 (± 0.09)	0.52 (± 0.12)

TABLE II
F-MEASURES ($\mu \pm \sigma$) FOR TERNARY CLASSIFICATION OF MUSIC GENRES ROCK/SLOW/DANCE IN THE REDUCED DATASET, BASED ON CIC AND DC FEATURES.

Metric	CIC performance	DC performance
F-measure	0.74 (± 0.02)	0.72(± 0.07)
Precision	0.68 (± 0.06)	0.70 (± 0.16)
Recall	0.81 (± 0.05)	0.75 (± 0.09)

TABLE III
F-SCORE, PRECISION AND RECALL FOR BINARY CLASSIFICATION BETWEEN ROCK AND SLOW MUSIC GENRES, BASED ON CIC AND DC FEATURES.

Table II summarizes the results of a multi-class classification task with the classes Rock/Slow/Dance in the reduced dataset; macro/weighted/micro averages for F-measure are presented, using a 5-fold cross-validation.

Macro averages are expectedly worse due to differences in class sizes. Micro and Weighted averages are marginally better for CIC compared to DC. Table III presents F-measures, precision and recall for a binary classification with the 2 largest classes, Rock and Slow. It is apparent that CIC obtains not only slightly better values of F-measure and Recall, but also smaller variations over the folds.

The results above are modest but may be considered as preliminary evidence that the newly proposed feature CIC is worth pursuing further investigation and employment in other harmony-related MIR tasks, especially those dealing with harmonic progressions, harmonic idioms or harmonic diversity.

IV. Conclusion

In this paper we introduced a novel harmonic feature Chroma Interval Content CIC which extends Directional Interval Content (DIC) vectors to chroma features, allowing the representation of chord progressions from audio-extracted data. CIC's motivation comes from a voice-leading perspective applied to chord progressions, generalizing this interpretation to chroma energy displacements. CIC has many interesting computational, mathematical and musical properties, and is potentially applicable to a wide-ranging series of harmony-related MIR tasks.

Being equivalent to a cross-correlation between subsequent chroma vectors, CIC is closely related to Dynamic Chroma vectors, a feature obtained by measuring differences between rotated chords in a chord progression. We have discussed some of their relationships on a theoretical level, and compared both features in similar experimental tasks involving chord progression identification and genre classification. Results have shown that both features provide similarly adequate representations for chord progressions, and CIC may have a slight advantage over DC in a harmony-based genre-classification task.

Future work include a more thorough evaluation of CIC in other harmony-related MIR tasks, such as cover-song identification [10], [18], and also exploration of its application to modeling of harmonic idioms and synthesis of chord progressions within an automatic harmonization context [9].

References

[1] Amélie Anglade, Rafael Ramirez, Simon Dixon, et al. Genre classification using harmony rules induced from automatic chord transcriptions. In *Proceedings of the ISMIR*, pages 669–674, 2009.

[2] Emilios Cambouropoulos. A directional interval class representation of chord transitions. In *Proceedings of the Joint Conference ICMPC-ESCOM 2012*, 2012.

[3] Emilios Cambouropoulos, Andreas Katsiavalos, and Costas Tsougras. Idiom-independent harmonic pattern recognition based on a novel chord transition representation. In *Proceedings of the 3rd International Workshop on Folk Music Analysis (FMA)*, 2013.

[4] Tianqi Chen and Carlos Guestrin. Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 785–794. ACM, 2016.

[5] Trevor Clercq and David Temperley. A corpus analysis of rock harmony. *Popular Music*, 30(1):47–70, 2011.

[6] W Bas De Haas, Remco C Veltkamp, and Frans Wiering. Tonal pitch step distance: a similarity measure for chord progressions. In *Proceedings of the ISMIR*, pages 51–56, 2008.

[7] Agustín Martorell Domínguez. *Tonal stability modeling from audio chroma features*. PhD thesis, Master thesis, Universitat Pompeu Fabra, Barcelona, 2009.

[8] Andre Holzapfel and Yannis Stylianou. Parataxis: Morphological similarity in traditional music. In *Proceedings of the ISMIR*, pages 453–458. ISMIR, 2010.

[9] Maximos Kaliakatsos-Papakostas, Marcelo Queiroz, Costas Tsougras, and Emilios Cambouropoulos. Conceptual blending of harmonic spaces for creative melodic harmonisation. *Journal of New Music Research*, 46(4):305–328, 2017.

[10] Samuel Kim and Shrikanth Narayanan. Dynamic chroma feature vectors with applications to cover song identification. In *Multimedia Signal Processing, 2008 IEEE 10th Workshop on*, pages 984–987. IEEE, 2008.

[11] David Lewin. Re: Intervallic relations between two collections of notes. *Journal of Music Theory*, 3(2):298–301, 1959.

[12] Brian McFee, Colin Raffel, Dawen Liang, Daniel PW Ellis, Matt McVicar, Eric Battenberg, and Oriol Nieto. librosa: Audio and music signal analysis in python. In *Proceedings of the 14th Python in Science Conference*, pages 18–25, 2015.

[13] Meinard Müller, Frank Kurth, and Michael Clausen. Audio matching via chroma-based statistical features. In *Proceedings of the ISMIR*, volume 2005, page 6th, 2005.

[14] Jean-François Paiement, Douglas Eck, and Samy Bengio. A probabilistic model for chord progressions. In *Proceedings of the ISMIR*, number EPFL-CONF-83178, 2005.

[15] Haukur Pálmason, Björn Thór Jónsson, Markus Schedl, and Peter Knees. Music genre classification revisited: An in-depth examination guided by music experts. *Proceedings of CMMR*, 2017.

[16] Maria Panteli, Simon Dixon, et al. On the evaluation of rhythmic and melodic descriptors for music similarity. In *Proceedings of the ISMIR*, 2016.

[17] Sertan Şentürk, Andre Holzapfel, and Xavier Serra. Linking scores and audio recordings in makam music of turkey. *Journal of New Music Research*, 43(1):34–52, 2014.

[18] Joan Serra, Emilia Gómez, and Perfecto Herrera. Audio cover song identification and similarity: background, approaches, evaluation, and beyond. In *Advances in Music Information Retrieval*, pages 307–332. Springer, 2010.

[19] Dave Tompkins. Music Database. https://www.cs.ubc.ca/~davet/music/list/Best9.html. [Online; accessed December 09th, 2018].

[20] George Tzanetakis, Andrey Ermolinskyi, and Perry Cook. Pitch histograms in audio and symbolic music information retrieval. *Journal of New Music Research*, 32(2):143–152, 2003.