HARMONY IN THE POLYPHONIC SONGS OF EPIRUS: REPRESENTATION, STATISTICAL ANALYSIS AND GENERATION
Maximos Kaliakatsos-Papakostas, Andreas Katsiavalos, Costas Tsougras, Emilios Cambouropoulos
School of Music Studies, Aristotle University of Thessaloniki, Greece
fmaxk, akatsiav, tsougras, emilios@mus.auth.gr

ABSTRACT
This paper examines a previously unstudied musical corpus derived from the polyphonic singing tradition of Epirus employing statistical methods. This analysis will mainly focus on unique harmonic aspects of these songs, which feature, for instance, unresolved dissonances (major second and minor seventh intervals) at structurally stable positions of the pieces (e.g. cadences). Traditional triadic chord types are inadequate for this corpus’ unconventional harmonic language; pc-set theoretic tools are too general/abstract. A novel chord representation has been devised that adapts to different non-standard tonal harmonic spaces. In the General Chord Type (GCT) representation, the notes of a harmonic simultaneity are re-arranged, depending on a given classification of intervals (that reflects culturally dependent notions of consonance/dissonance), such that abstract idiom-specific types of chords may be derived. Based on these harmonic representations, statistical analyses are performed that provide insights regarding underlying harmony and, especially, on the idiosyncratic use of consonance/dissonance within this idiom. Then, characteristics of harmonic successions are examined via statistical analysis of the common chord transitions in the idiom. Finally, the learned statistical features are used to generate new harmonisations in the Epirus-song style for unseen Epirus-song melodies or for melodies from other distant idioms.

1. INTRODUCTION
The paper at hand paper examines a previously unstudied musical corpus derived from the polyphonic singing tradition of Epirus employing computational statistical methods. This analysis will mainly focus on unique harmonic aspects of these songs, which feature, for instance, unresolved dissonances (M2 and m7 intervals) at structurally stable positions of the pieces (e.g. cadences).

The statistical analysis over a corpus (Gjerdingen, 2014), mainly involves the definition of elements that are believed to occur within the corpus and the subsequent calculation of statistics on these elements, the norms of which indicate frequent relations and patterns within the corpus. Several studies with diverse orientations have been based on analysing the statistics on corpora. These studies can be divided in two categories: the “analytical” and the “compositional” approaches. Roughly, the analytical approaches aim to elucidate some facts about a studied corpus that remain ill-defined or unexplored in the domain of musical analysis. The compositional approaches on the other hand, are targeted towards utilizing the statistical values yielded by the corpus analysis and combine them with machine learning or other generative techniques, to compose novel music that complies with the yielded statistical properties of the corpus.

Among many other studies, the statistical analysis of musical corpora has been performed for tracing the patterns of occurrence of specific musical phenomena, like the “cadenza doppia” (Gjerdingen, 2014) and Koch’s metrical theory (Ito, 2014), or for providing additional insights about well-studied and popular idioms like rock music (de Clercq & Temperley, 2011; Temperley & de Clercq, 2013). Furthermore, through similar analytical techniques more detailed investigations on the functionality of perceptual mechanisms are allowed, like tracking the evolution of the major and minor keys in tonal music (Albrecht & Huron, 2014), measuring the historical development of musical style in relation to the cognition of key (White, 2014), analysing the perceptual functionality of tonality and meter in music (Prince & Schmuckler, 2014), or examining the insertion of preformed structures in improvisation (Norgard, 2014). Statistical analysis on a corpus of a capella flamenco songs (Cabrera et al., 2008) has also been utilized to provide important tonal insights about this previously unstudied folk musical idiom, allowing thus a first ethnomusicological exploration of this idiom. Regarding the analytical part, the paper at hand contributes to the field of ethnomusicology by providing initial insights about statistical facts that concern the harmony in a corpus comprising several polyphonic songs of Epirus.

The polyphonic singing of Epirus, a region of northwestern Greece/southern Albania, is an intrinsically polyphonic Greek-language Balkan folk idiom, sharing features with similar Albanian, Aromanian and Vlachian idioms, and differing with them in language, structure and other vernacular aspects (Samson, 2013, p. 47-48; Ahmedaja, 2008; Kokkonis, 2008, p. 42-44). It is based on anhemitonic pentatonic modes and performed by 2-voice to 4-voice groups, with each voice having a specific musical and narrative role (Lolis, 2006; Kokkonis, 2008): leading solo voice (partēs [taker]), subsidiary solo voice (gyristēs [turner] or klōstēs [spinner]), drone group (isokratēs) and optionally, in 4-part polyphonic songs, the drone elaborator (richtēs [dropper]).

The present analysis focuses on the unique harmonic aspects of this polyphonic idiom, namely the structure and statistical distribution of the non-triadic sonorities and the use of unresolved dissonances (major second and minor
seventh intervals), especially at structurally stable positions of the pieces (e.g. cadences). The pentatonic pitch collection - described as pc set (0,2,4,7,9) - functions as source for both the melodic and harmonic content of the music. This analysis allows further musical hypotheses to be formed and conclusions to be drawn regarding harmonic mechanisms characteristic of the idiom.

Statistical analysis may provide confirmation of musicological hypotheses, comparing results with established musicological ground truth. For instance, in a statistical study on the harmony of the Bach Chorales (Prince & Schmuckler, 2014), the statistical analysis revealed chords with specific harmonic functionality; these results were aligned with the analytic musicological analysis encoded with roman numerals. However, for representing the harmonic elements of the non-standard examined idiom, the roman numerals or standard chord symbols cannot be employed, since chord simultaneities rarely comply with the standard major–minor categories. One can always resort to pc-set theory; however, in this study, the General Chord Type (GCT) representation (Cambouropoulos et al, 2014) is utilized to represent this idiom’s chord simultaneities that is more expressive than pc-sets – see brief description in next section.

There is extensive literature over the topic of automatic music composition by utilizing statistical quantities of a given corpus. Regarding these approaches, the statistical values are used either to straightforwardly provide estimations about the most probable future events (e.g. with variable length Markov models (Dubnov et al., 2003) or multiple view–points (Whorley et al., 2013) among others), or they are used as targeted fitness values to drive evolutionary processes (e.g. with genetic programming (Manaris et al., 2007) among others). The compositional part of the presented study discusses automatic melodic harmonisation through a machine learning technique which is based on the Hidden Markov Models (HMMs), namely the constrained HMMs (cHMMs) (Kaliakatsos-Papakostas et al., 2014). The presented compositional approach, utilizes in a straight forward manner the yielded statistical values to compose novel harmonies on given melodies, even if these melodies are not extracted from the studied idiom. Specifically, the harmonisations provided on melodies from the Bach chorales indicate that the “blended” musical result retains some important harmonic characteristics.

The present study examines only the vertical sonorities incorporated in the idiom, attempting to address the following questions: How can an unconventional musical style be encoded/represented so that meaningful computational analysis may be conducted? What are the most pronounced features of the harmonic language of polyphonic songs of Epirus? Which are the dominant pitches of the scale? Which is the frequency distribution of the sonorities? What constitutes a cadence phenomenon in this idiom? What is the dissonance's role, how freely is it employed and why/where? Finally, how might the statistical data obtained from the analysis be used creatively to generate epirus-style harmonisations of new melodies?

In the next section, the set of polyphonic songs used in this study is described and a novel chord representation, namely the General Chord Type representation, is employed to encode the song’s verticalities. Then, the statistical methodology and results are presented along with musicological interpretations that unveil characteristic harmonic facts about the polyphonic style of Epirus. Finally, a constrained-HMM is used to generate novel melodic harmonisations in the style of the analysed songs.

The paper concludes with remarks about the particular musical idiom and future directions of research.

## 2. DATASET AND GCT REPRESENTATION

### 2.1 The Epirus polyphonic song dataset

A collection of songs from the discussed idiom, transcribed from field recordings and notated in standard staff notation, is available in (Lolis 2006). From this collection, a dataset consisting of 22 songs (10 with 3 voices, 12 with 4) in minor pentatonic scale were selected; the minor pentatonic scale consists of pitch classes: [0,3,5,7,10] (for instance \{A,C,D,E,G\}). The 22 songs were manually segmented into 102 polyphonic phrases, while 37 monophonic phrases where excluded since the study revolves around the harmonic aspects of the idiom.

The songs were encoded at two reductional levels, corresponding to two adjacent levels of the metrical/time-span hierarchy: level \(ms0\) closely describes the musical surface by including embellishing figures, neighbour notes, etc. and corresponds to the metrical level of sixteenth or eighth notes (depending on the metrical tactus and beat level) of the transcription, while level \(ms1\) describes a deeper structure by omitting most elaborations and corresponds to the eighth or fourth notes of the transcription – see example in Figure 1. The lowest level encoding \(ms0\) consists of 1467 chords whereas the next reduction level \(ms1\) of 945 pitch class sets (chords); the pcs reduction ratio (ms1/ms0) of the polyphonic phrases is 0.64 (945 /1467 chord states). The reduction was deemed analytically necessary in order to disclose the idiom's harmonic functions and cadence patterns.

The songs were notated into music xml files; in addition to the original content, each xml file is annotated\(^1\) with structural information using music notation that follows a specific formalism. More specifically, songs in the dataset consist of six staves: two for the original song/content preserving voicing information (ms0), one for tonality annotations where the scale is written as a

\(^1\) All annotations have been prepared by two of the authors.
Figure 1. *Ammo ammo pigena*, polyphonic song from Epirus: Annotated xml file containing original song transcription (ms0), time-span reduction of the original content (ms1) as well as tonality and grouping information (see text). Symbols P, G, R and D denote the different voice groups (partês, gyristês, richtês, drone).

The representations of the polyphonic songs were carried out without any transposition with the use of the GCT representation model. Afterwards, traditional note names or pc sets or other descriptions were employed when necessary. For ease of comprehension, the results of the statistical analysis and relevant figures/examples assume that the A minor pentatonic is used, i.e. A,C,D,E,G.

2.2 The GCT representation

The General Chord Type (GCT) representation (Cambouropoulos et al., 2014), allows the re-arrangement of the notes of a harmonic simultaneity such that abstract idiom-specific types of chords may be derived; this encoding is inspired by the standard roman numeral chord type labeling, but is more general and flexible. Given a classification of intervals into consonant/dissonant (binary values) and an appropriate scale background (i.e. scale with tonic), the *GCT algorithm* computes, for a given multi-tone simultaneity, the ‘optimal’ ordering of pitches such that a maximal subset of consonant intervals appears at the ‘base’ of the ordering (left-hand side) in the most compact form. Since a tonal centre (key) is given, the position within the given scale is automatically calculated. For instance, the note simultaneity [C, D, F#] or [0,2,6,9] in a G major key is interpreted as [7,0,4,7,10] (appearing also in the figures for matter of space as 7.04710), i.e. as a dominant seventh chord.

The proposed representation is ideal for hierarchic harmonic systems such as the tonal system and its many variations, but adjusts to any other harmonic system such as the Epirus polyphonic system. In terms of the current parameter, we have applied the GCT encoding for two different consonance vectors. The first is the standard vector for tonal music, i.e. consonant thirds/sixths and perfect fourths/fifths. The second vector, which is more adequate for ‘atonal’ music, includes additionally major seconds and minor sevenths (i.e., major seconds and minor sevenths are considered ‘consonant’ following the fact that in the Epirus songs these intervals are stable and require no resolution). In the example in Figure 2, we see the GCT encodings arising for the two different consonance vectors.

As can be seen in Figure 2, the two encodings are often different. Using the first consonance vector, the tonic is the root for almost all the GCTs (for A minor pentatonic we have tonic A: 85.39%). Using the second consonance vector, the chord’s ‘root’ is more diverse (A: 61.58%, D: 12.27%, G: 22.85%) and additionally, since the consonance vector is less strict, the GCT chord encodings are more compact. Perhaps the most striking difference between the two encodings is this: the tonal-consonance-based labelling determines as the ‘root’ of almost all the chord types the *tonic* of the minor pentatonic scale which is actually the drone of each piece (notice that almost all
chord types on the upper row of Figure 2 have a 0 on the right before the period; the atonal-consonance based labelling gives ‘roots’ of chords that shift between the tonic (0) and the subtonic (10) (this is interesting as the dissonances that occur between the tonic and subtonic have a special role as will be discussed in the next section).

3. STATISTICAL ANALYSIS AND RESULTS

In this section a number of interesting statistical observations on the polyphonic Epirus dataset are presented. First, plain distributions of pitches and chord types are given and, then, conditional probability results that reflect transition regularities between chords are discussed. For ease of comprehension, whenever the GCT representation is not depicted, we assume that the A minor pentatonic is used, i.e. A,C,D,E,G (for readers not acquainted with the GCT, the traditional note names are easier to follow).

3.1 Pitch and chord distributions

The power set of the minor pentatonic scale has 30 different valid combinations (the empty set and the whole pentatonic set are excluded). The frequency of appearance of chords according to cardinality (number of pcs) is depicted in Figure 3.

![Figure 3. Frequency of chords according to cardinality](image)

When single notes (i.e., unisons) appear in the polyphonic sections, 97.10% of these are the tonic (A). Unisons on other scale degrees are rare (<1%). From all 10 combinations with cardinality 2, dyads: AC, AD, AG are the most frequent (>25% each), AE is not so common, CE, GD, CG are rare (<1%) and DE, DG, EG are absent. For cardinality 3 (10 combinations), ACG is the most frequent (36.91%), ACE, ACD are common, ADG is less common, EAD and EGA are rare and CDE, CDG, CEG, DEG are absent. For cardinality 4 (5 combinations), ADGC is the most frequent of the cardinality class (67.16%) but rare in general (4.76%), ACEG and ACDE are rare and ADEG and CDEG are absent. Considering that from all the polyphonic phrases, almost half are taken from songs with 3 voices and also that it is common for one voice to double the drone tone, the cardinality percentages seem reasonable. Perhaps more 4-voice songs need to be added in the dataset to increase the frequency of higher cardinalities.

Since the drone tone (i.e., the tonic) is found in every chord, the absence of all the chords that don’t contain it is also justified. However, from the 15 combinations of the power set of the scale that contain the tonic, only 4 from 6 chords with cardinality 3 are found (ADE and GAE don’t exist) and only 1 from 4 with cardinality 4 (ACDE, ACEG, AEDG not found).

Apart from the quantitative difference in the frequency of appearance, the chord types found in the original surface (ms0) and the reduced musical surface (ms1) are identical (except chord AEDG that is found only 3 times in ms0 – not found in ms1). The statistical distribution of the dissonant sonorities (containing a major second) is almost identical at reductional levels ms0 and ms1, a feature indicating that dissonance is an integral/structural part of the harmonic idiom, and it is not reduced out as an elaborative event at the deeper level. On the contrary, this occurs in the tonal idiom, where most dissonant chords are the outcome of unstable non-chord tones and they are present mostly at the musical surface, not in reductions.

The appearance of the note E is rather rare. Considering that from the general percentage of all the existing chords that contain this note (19.68%), 8.99% corresponds to the only minor chord of the scale (ACE) and that another 5.18% corresponds to the perfect fifth interval (AE) from the tonal centre, it is clear that this note has relatively low usage.

Considering the pitch content of the chords, we can assume chord subsets and categories depending on various groupings relations. For example, from all the possible intervals that may appear in a chord in this pentatonic dataset (2, 3, 4, 5, 7, 9, 10 semitones) the most psychoacoustically dissonant is that of 2 and its inversion 10. Using this information we can create two groups of the most frequent chords depending on the presence of a 2-semitone interval: Consonant (A, AC, AD, AE, ACE) and Dissonant (AG, ACG, ADG, ADGC, ACD). Furthermore, we can split the Dissonant group according to the quantity of the dissonant intervals present in the chord and the quality of the pitch classes that comprise it. The Dissonant set may be split in three subsets: {ACD}, that contains the CD dissonant interval not involving the tonic A, {ACG, ADG, AG} that contain dissonant interval GA involving the tonic and {ADGC} that contains both 2-semitone intervals GA & CD.

![Figure 4. Frequency of appearance of the chords in the dataset](image)
with tonic) appear mostly on weak beats (not the first beat in the measure), whereas dissonant chord ADC containing interval CD (dissonance not with tonic) appears mostly on the strong beat. This fact shows a structural difference in the use of the two types of dissonant chords. The dissonant interval appearing between the tonic A and subtonic G is very common but seems to be a kind of ‘colouring’ of standard consonances that appear on strong beats (see below dissonance used as cadence extension). On the contrary, the CD dissonance is considered probably milder as it does not clash with the tonic and appears more often on strong beats.

Concerning the phrase positioning of chords, the most common final chords of final phrases are unison A, AG, AC, ACG, and AE (GCT: 0.0, 0.010, 0.03, 0.0310, 0.07) – see Figure 5. In the next section the use of dissonance at cadences is discussed more extensively. There is no significant preference for beginning chords.

**3.2 Chord transition statistics**

A first order transition table reveals the general and immediate chord transition preferences of the chords. Utilizing this information we can track previous and next chords for single chords and for groups as well. The generated first-order transition matrix is presented in Appendix I. In Figure 6, the transition matrix of the Appendix I is converted in a different form that is simpler to read. The current chord is placed in the middle column; on the left we see the previous chords and on the right the following chord; distance from the middle column represents strength of transition (i.e. higher probability).

Variable-length higher-order transition probabilities are additionally being calculated using Prediction Suffix Trees (Ron et al., 1996). Such trees reveal common patterns occurring in the dataset. We have used such a PST to examine cadences in more detail. In Table 1, the most common chords appearing on phrase endings are depicted. Even though dissonances GA and GAC appear as last chords of phrases, we consider them as chord extensions assuming that the actual cadence end point is the unison on the tonic (A) or tonic plus third (AC); this assumption is supported by the fact that the last word of the songs’ lyrics usually ends on a strong metric position consonance whereas the dissonance is introduced immediately after on a weaker beat as a kind of ‘tension colouring’ that requires no preparation or resolution. Approximately 30% of cadences ending in unison A or AC are extended with the addition of G creating a dissonance. More than half of the chords preceding the final chord in the cadence embody dissonances.

<table>
<thead>
<tr>
<th>Semi-final chord</th>
<th>Final chord</th>
<th>Extension</th>
</tr>
</thead>
<tbody>
<tr>
<td>GAC (12+2)</td>
<td>A (33+14+1)</td>
<td>GA (14)</td>
</tr>
<tr>
<td>GA (8+3)</td>
<td>ACG (1)</td>
<td>GAC (4+2)</td>
</tr>
<tr>
<td>AC (3+6)</td>
<td>AD (5+3)</td>
<td></td>
</tr>
<tr>
<td>AE (2+1)</td>
<td>GAC (6)</td>
<td></td>
</tr>
<tr>
<td>AD (3+1)</td>
<td>ACD (3+1)</td>
<td></td>
</tr>
<tr>
<td>ACDG (3)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 1. Most common chords appearing on phrase endings (chords appearing less than three times have been omitted). See text for details.

**Figure 5.** Frequency of appearance of the last chords of all (102), middle (non-final) & final phrases (22).

In Figure 7 two prototypical cadential patterns appearing in this idiom are illustrated. Dissonance usually appears...
just after the consonant downbeat final chord as a kind of

cadence extension that adds a final brushstroke of tension.

\[
\begin{align*}
\text{Figure 7. Prototypical cadences in polyphonic songs from Epirus.}
\end{align*}
\]

Also, the high occurrence rate of the cadential pattern in
this idiom and the fact that pitch class E is mostly absent
in it possibly explains this pc's relatively low statistical
appearance (see 3.1).

4. MELODIC HARMONISATION

The statistical results hitherto discussed provide some in-
dications about harmonic regularities that are encountered
in the examined idiom. The utilization of these statistics
for melodic harmonisation are expected to produce har-
monisations – in a statistical manner – that preserve these
characteristics, even if the melody to be harmonised is
not characteristic of this idiom. The statistical informa-
tion that is utilized in the short examples presented in
this section concern the chord–to–chord transitions, as
reflected by the first order transition maps discussed in
Section 3. To this end, a modification of the first order
hidden Markov model (HMM) is utilized, namely the
constrained-HMM (cHMM), which was presented in
(Kaliakatsos-Papakostas et al., 2014). The cHMM meth-
odology follows the typical HMM methodology, with the
difference that specific chord constraints can be set by the
user or by another algorithmic process at any point of the
melody to be harmonised. From a user perspective, the
cHMM methodology allows the user to choose specific
beginning, intermediate or ending chords or fixed chord
progressions (e.g. cadences), while the remaining har-
monic content is filled with a probabilistic framework
based on the HMM methodology.

The cHMM methodology requires statistical information
from the corpus and deterministic information from the
melody to be harmonised. The statistical information
from the corpus concerns chord (or states in the HMM
nomenclature) transitions, specifically the first order tra-
sitions under the Markov assumption (that the appear-
ance of a chord depends only on its previous chord). The
deterministic information, on one hand, involves chord-to-
ote (observation) rules1 and, on the other, the fixed-
chord constraints. Regarding the chord-to-note rules, a
chord is considered “probable” to harmonise a given note
of the melody if the melodic note’s pitch class is a mem-
er of the chord’s pitch class. In the presented application
of the cHMM, the fixed-chord constraints are provided by

1 The typical HMM methodology incorporate probabilistic rules that
relate observations (melody notes) with states (chords).

the user (one of the authors), according to the harmonic
cHECKPOINTS that are desired – primarily regarding the
preservation of cadence characteristics. For a more de-
tailed overview of the cHMM specifications, the interest-
ed reader is referred to (Kaliakatsos-Papakostas et al.,
2014).

The results presented below discuss the harmonisation of
some melodies, utilizing the corpus statistics for the first
order Markov transitions between chords. Specifically,
the transition tables are extracted from 30 random phrases
of the pieces in the Epirus song corpus, in order to be
aligned with scalability issues that demand efficient train-
ing with limited numbers of training data. Automated me-
loidic harmonisation is performed on melodies of Epirus
songs as well as on melodies from completely different
idioms. It has to be noted that the Epirus songs that are
re-harmonised, are not included in the 30 random phrases
that comprised the training set. The aim of the results is
to demonstrate the adequacy of these statistics regarding
melodic harmonisation, not only in the context of melo-
dies that are typical of the idiom, but also for less typical
melodies.

The harmonisation in the polyphonic style of Epirus of
three melodies is illustrated in Figures 8, 9 & 10. The
first melody is the melody of the Epirus polyphonic song
depicted in Figure 1, whereas the remaining two from dif-
ferent idioms, namely, J.S.Bach’s ‘Chorale nr. 110’ (“Va-
ter unser im Himmelreich”) melody and the second part
of G. Gershwin’s ‘Summertime’ melody; all three are built
on the minor pentatonic scale. In each of these figures,
the melody is illustrated on the top, accompanied by the
chords (in GCT encoding) generated by the aforemen-
tioned cHMM model that was trained on the polyphonic
songs of Epirus. Below each melody, a full harmonisation
following the idiom’s voice leading is presented (created
manually by one of the authors).

The harmonisation of the Epirus melody is very close to
the original transcription depicted in Figure 1; this shows
that the cHMM model is quite good at generating chords
in the learned style (NB., this melody was not included in
the training dataset). The generated harmonisations of the
other two melodies, preserve characteristics of the poly-
phonic style of Epirus and have a unique harmonic fla-
vour which shows that blending between diverse musical
materials may give interesting original outcomes.

The cHMM constraints imposed to all three pieces related
to chords of cadences; namely, for all pieces the final two
chords were respectively: [0,[0,3]] and [0,[0,3],[10]] in
GCT notation. When the cadence constraints were omit-
ted, the generated cadence in most cases included two
repetitions of the tonic unison which is also very common
in the idiom (expressed with the GCT [0,[0]]). As can be
seen, many characteristics of the idiom are preserved
such as the used chord types, the progression of chords,
the interchange of consonance-dissonance and the drone tone (common root in almost all GCTs); there are however, characteristics that are missed out in the current model such as the appearance of specific types of chords (e.g. chord with dissonances) on particular metric positions. The generation model requires embedding in a more general compositional framework that takes into account voice-leading, metric structure, bass-line motion and so on. It is encouraging that such a simple learning model (cHMM) coupled with the GCT representation captures significant components of the harmonic language of the idiom.

5. CONCLUSIONS

In this paper a study has been presented of a musical corpus derived from the polyphonic singing tradition of Epirus that employs computational statistical methods. The analysis focused primarily on harmonic aspects of these songs. Representing a non-tonal idiom in terms of harmonic content is non-trivial. A novel chord representation, namely the General Chord Type (GCT) representation, has been presented that adapts to different non-standard tonal harmonic spaces depending on a given classification of intervals that reflects culturally-dependent notions of consonance/dissonance. It has been shown that the GCT is appropriate for encoding note simultaneities in the given idiom and that different consonance vectors give rise to different labelings of chords that embody alternative facets of core harmonic concepts in the idiom (e.g. drone or tonic-subtonic relation).

Using the GCT as basis, statistical analytic tools have been employed that highlight characteristic distributional and transitional properties of chords in the idiom. It has been shown that some note simultaneities are more common than others, that ‘dissonances’ (major 2nds and minor 7ths) are structurally relatively ‘stable’ as they appear...
in higher reductional levels and also on relatively strong metrical positions, and that certain cadential patterns seem to emerge that describe harmonic content at phrase endings. The transitional patterns learned from a section of the dataset (using the GCT representation and the cHHM methodology) are used to generate new harmonisations for melodies in the style of the studied idiom but also on melodies drawn from other foreign styles (e.g. Bach chorale melody and Gershwin’s Summertime melody); the new harmonisations preserve qualities of the analysed idiom creating novel musical creations.

6. ACKNOWLEDGMENTS

The project COINVENT acknowledges the financial support of the Future and Emerging Technologies (FET) programme within the Seventh Framework Programme for Research of the European Commission, under FET-Open grant number: 611553.

7. REFERENCES


Appendix I. First-order transition matrix of GCT simultaneities for the Epirus polyphonic song dataset