D7.2
Probabilistic harmonic induction model and idiom-independent harmonic learning

Authors	Maximos Kaliakatsos-Papakostas, Emilios Cambouropoulos, Asterios Zacharakis
Reviewers	Allan Smaill, Ewen Maclean, Kai-Uwe Kuhnberger

Grant agreement no.	611553
Project acronym	COINVENT - Concept Invention Theory
Date	March 22, 2015
Distribution	PU

Disclaimer

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.

Abstract

The application of the COINVENT computational methodology on conceptual blending in music concerns melodic harmonisation, which is the employment of harmony on a given melody. A melodic harmoniser is developed that will facilitate conceptual blending, where user input melodies will be harmonised with blended harmonic characteristics from several diverse idioms. To this end, during the first year of the project a dataset of idioms was created by compiling and harmonically annotating several music pieces. Since it is practically impossible to hand-code the harmonic rules that are able to generate harmonisations for all these idioms, the realistic approach was to induce the harmonic rules through probabilistic models in an idiom-independent way, i.e. the same model works for every idiom but with different statistical features. This report analyses the core methodological framework of the COINVENT melodic harmoniser, which is based on probabilistic harmonic induction on several harmonic characteristics in multiple idioms and the utilisation of the learned information in a generative manner. Based on the idiom-independent General Chord Type (GCT) representation of harmony, the learning/generating modules concern chord progressions, high-level harmonic structure through employing intermediate cadences, voice leading of the bass voice and features regarding the voicing layout of chords. Throughout the extent of this report, short insights about the blending potential of the probabilistic framework are given, through methodologies that are being developed in parallel, e.g. chord blending and chord similarity.
Keyword list: **Melodic Harmonisation**, **Machine Learning**, **Harmonic training**, **Chord Progressions**, **Voice Leading**, **Harmonic Structure**

Contents

1 Introduction

Automated melodic harmonisation discusses the assignment of harmonic material on the notes of a given melody. The harmonic material is described by chord symbols, while the harmonisation is completed if voice leading between the notes of successive chords, is defined. The common approach to test an automatic harmonisation system is to utilise it for harmonising melodies that pertain to a musical idiom with harmonic structure that is well-defined. To this end, some pioneering methodologies that were developed for melodic harmonisation, incorporated human expert knowledge encoded in the form of rules, leading to e that could generate harmonisations with explicit stylistic orientation towards the musical idiom that these rules referred to. For a review in the rule–based systems the reader is referred to [37]. A similar approach to the rule–based methodologies is the one followed by systems that utilise genetic algorithms (GA), like the ones shortly reviewed in the recent paper [8] and, also, in [40]. The similarity between these two approaches is that both rely on a set of harmonic rules intended for a specific musical idiom; in the case of the GAs, the employed fitness function quantifies such rules.

However, the rule–based spectrum of methods has a major drawback when discussing melodic harmonisation in many different idioms: the encoding of rules that describe an idiom is not always a realisable task, since idioms abound in complex and often contradicting interrelations between harmonic elements. To this end, the formulation of *probabilistic* techniques and *statistical learning* has been proposed. Among many proposed methodologies, most of which are discussed in Section 4.1, Bayesian networks [46] and prediction by partial matching [50] have been utilised to construct the bass, tenor and alto voices below a given soprano voice, hidden Markov models (HMMs) for constructing chord sequences for a given melody [42] and probabilistic graphical models for relative tasks [38].

The approach to harmonisation that is pursued in the development of the presented melodic harmoniser, pertains to the wider research context of the COINVENT project, according to which the study of automatic melodic harmonisation includes the blending of harmonic concepts among diverse musical idioms, to produce novel harmonic concepts. At its present form, the presented system can produce harmonies that accurately reflect the characteristics of single idioms. Aim of the melodic harmoniser, however, is to facilitate harmonic blending, allowing the user to select and blend characteristics from more than one idioms. The system has been developed with isolated and discretely separated learning modules of harmonic characteristics, enabling the blending potential of the harmoniser. The blending task, however, is a part of ongoing future work that incorporates many challenges, as further discussed in Section 7. A shadowgraph of the melodic harmoniser's algorithmic structure is illustrated in Figure 1, where the main development philosophy is revealed: the methodological framework is broken down to several parts that are trained on an idiom's features, allowing – as a future work – the employment of blending by combining different parts trained on different idioms.

2 Harmonic training/generation overview

The harmonic learning system is trained on several harmonic aspects, which can be divided in two groups: chord generation and the voicing layout. Figure 2 illustrates this setting, where "GCT generation" on the left block refers to the generation of chords symbols in the General Chord

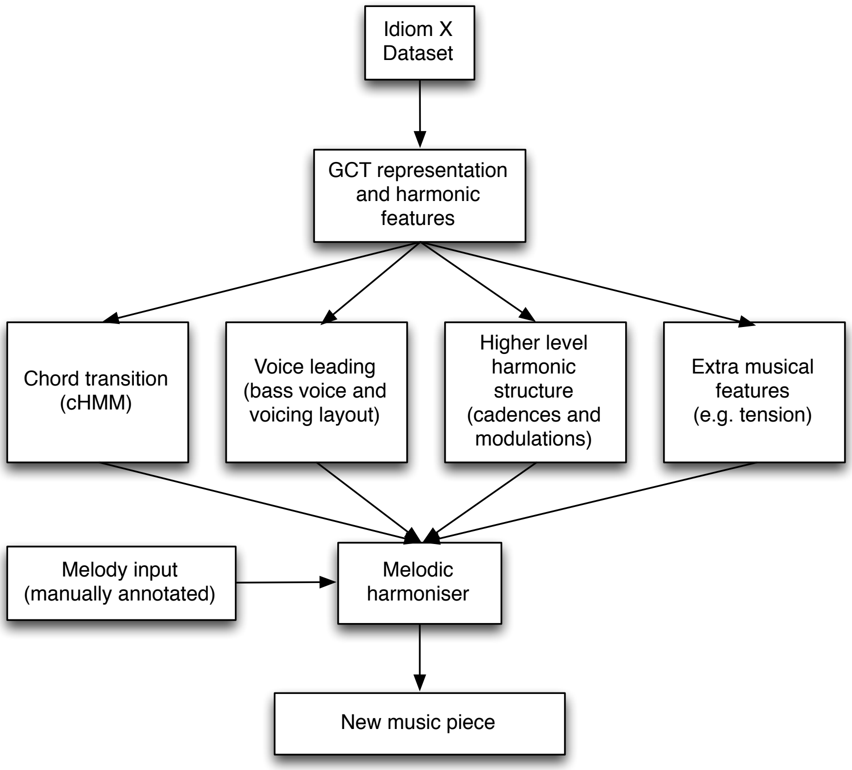


Figure 1: Harmonic learning system overview.

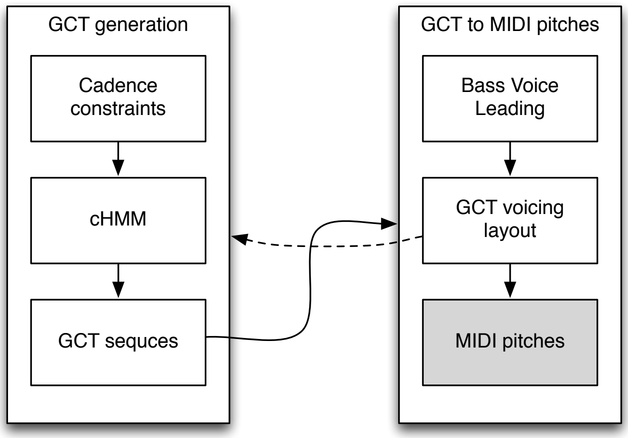


Figure 2: Harmonic learning system information flow overview.

Type (GCT) [3] representation (see Section 3), while the right block refers to the materialisation of GCT chords to music by assigning proper voicing, extracting the final output in MIDI pitches. The arrow leading from the "GCT generation" to the "GCT to MIDI pitches" block indicates the current generative process flow of the melodic harmoniser: first, chord sequences in GCT form are produced and, afterwards, voice leading and voicing layout is applied to the GCT sequences, yielding the finalised output in MIDI pitches. The dashed-lined arrow leading from the "voicing layout" box back to the "GCT generation" block is a part of an ongoing work related to future improvements. According to this future improvement, the selection of a chord will not only depend on the user input melody and the previous chord, but also in the voice leading potential that this chord offers, as discussed in more detail in Section 7.

The system is trained on data that have been annotated by music experts according to their harmonic content. Figure 3 depicts a music score with annotations on the aforementioned harmonic content. On top (ms_0 for "music surface 0"), the music score of the genuine piece is given, while on the bottom staff (ms_1) the harmonic reduction is exported, keeping only the harmonically important notes. The GCT chords are extracted from the reduced version of the score (ms_1), according to the tonality in the current part of the piece, as annotated in the second staff. Finally, the grouping level of the piece's phrases has been annotated in the third score. For more information about tonality and grouping annotations, as well as for the available idioms in the dataset for training the system, the reader is referred to the report of Deliverable 7.1 [26].

Different algorithmic process are combined for training the system, since harmonic learning concerns multiple aspects of harmony. On the chord sequence level (see Section 4), the constrained hidden Markov model (cHMM) [23] algorithm has been developed, which is trained on the GCT representation of chord sequences from the ms_0 score staff. The cHMM is a probabilistic (GCT chord) sequence generator trained on statistics over observable data, extending the typical HMM as it accepts intermediate constraints or "deterministically" defined checkpoints on any position of the sequences it generates. The constraints imposed on the cHMM concern the determination of intermediate and final cadences on places where phrases end, providing the composed harmony with a higher-level structural consistency. Therefore, the harmoniser's cHMM is trained on single

Figure 3: Score information required for training the system.

phrases, as demarcated by the “grouping” staff in the training scores. The cadences are learned by extracting statistics on the final chords of phrases, while they are generated for the user-provided melody through a probabilistic process (see Section 6).

For composing the harmonised output of the input melody, the GCT chords produced by the cHMM and the cadence constraints are materialised into MIDI notes through applying statistically trained models for voice leading and voicing layout. The bass voice leading (BVL) [33] of chord sequences in the ms_0 staff is learned by a HMM that observes the semitone step for the next note of the melody and decides about the step of the bass note, considering also the previous motion of the bass (see Section 5.1). Additional considerations for the voicing layout of the GCT chords regard the probabilities for chord inversions and note doublings (see Section 5.2), as learned by the voicing layout of the ms_0 chords in the training data. These probabilities are taken under consideration, in combination with the BVL requirements, for producing the final voicing layout of the harmony.

After the system is trained, it is able to harmonise a given melody that is accompanied by some additional pointers to its attributes. Figure 4 demonstrates an input protocol for the system, which includes the melody to be harmonised and information regarding some harmonic attributes that cannot be inferred by the system. Initially, the user should accompany the melody with information about the positions where chords should occur (harmonic rhythm), as well as the important notes that should be considered with higher priority when selecting proper chords for each segment. If the user provides no information for these attributes, the system produces default harmonic rhythm and important note selection schemes that might lead to “uninteresting” harmonic

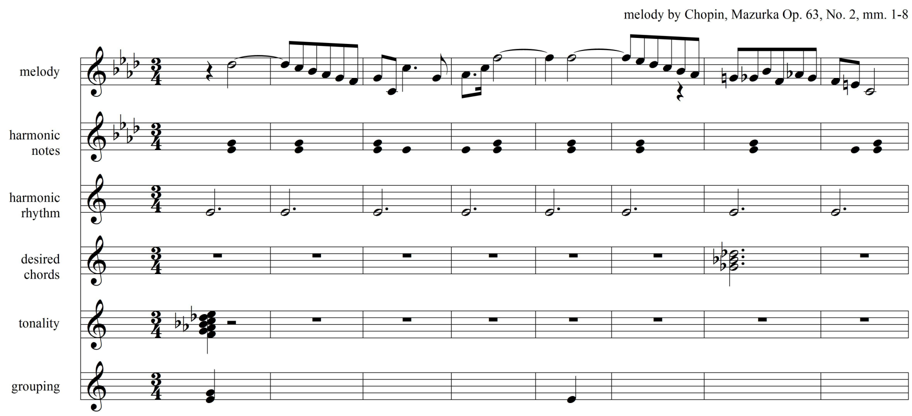


Figure 4: User input requirements to the melodic harmoniser.

results. Additionally, the user has the freedom to choose specific chords at desired locations (desire chords), forcing the system to produce chord sequences that comply with the user-provided constraints, therefore allowing the user to "manually" increase the interestingness of the produced output. Finally, the user should accompany the melody with higher level harmonic information concerning the tonality or tonalities of the piece, as well as with its phrase grouping boundaries.

3 Representing Harmony

There exist different typologies for encoding note simultaneities that embody different levels of harmonic information/abstraction and cover different harmonic idioms. For instance, for tonal musics, chord notations such as the following are commonly used: figured bass (pitch classes denoted above a bass note – no concept of "chord"), popular music guitar style notation or jazz notation (absolute chord), roman numeral encoding (relative chord to a key) [30]. For atonal and other non-tonal systems, pc-set theoretic encodings [12] may be employed.

A question arises: is it possible to devise a "universal" chord representation that adapts to different harmonic idioms? Is it possible to determine a mechanism that, given some fundamental idiom features, such as pitch hierarchy and consonance/dissonance classification, can automatically encode pitch simultaneities in a pertinent manner for the idiom at hand?

Before attempting to answer the above question one could ask: What might such a "universal" encoding system be useful for? Apart from music-theoretic interest and cognitive considerations/implications, a general chord encoding representation may allow developing generic harmonic systems that may be adapted to diverse harmonic idioms, rather than designing ad hoc systems for individual harmonic spaces. This was the primary aim for devising the *general chord type* (GCT) representation. In the case for the project COINVENT, a creative melodic harmonisation system is required that relies on conceptual blending between diverse harmonic spaces in order to generate novel harmonic constructions; mapping between such different spaces is facilitated when the shared generic space is defined with clarity, its generic concepts are expressed in a general and idiom-independent manner, and a common general representation is available.

In recent years, many melodic harmonisation systems have been developed, some rule-based

([9, 37]) or evolutionary approaches that utilise rule based fitness evaluation ([40, 8]), others relying on machine learning techniques like probabilistic approaches ([38, 45]) and neural networks ([19]) , grammars ([15]) or hybrid systems (e.g. [5]). Almost all of these systems model aspects of tonal harmony: from “standard” Bach–like chorale harmonisation ([9, 19] among many others) to tonal systems such as “classic” jazz or pop ([45, 15] among others). Aim of these systems is to produce harmonisations of melodies that reflect the style of the discussed idiom, which is pursued by utilising chords and chord annotations that are characteristic of the idiom. For instance, the chord representation for studies in the Bach chorales include standard Roman numeral symbols while jazz approaches encompass additional information about extensions.

For tonal computational models, Harte’s representation [18] provides a systematic, context-independent syntax for representing chord symbols which can easily be written and understood by musicians , and, at the same time, is simple and unambiguous to parse with computer programs. This chord representation is very useful for annotating manually tonal music – mostly genres such as pop, rock, jazz that use guitar-style notation. However, it cannot be automatically extracted from chord reductions (it is useful for manual annotation) and is not designed to be used in non-tonal musics. In this report, firstly, we present the main concepts behind the General Chord Type (GCT) representation and give an overall description, then, we present an example on a Bach Chorale part that shows the potential of the representation. Some examples of applying statistical learning on such a representation are given in Section 4.

3.1 Analysing and Representing Chords

Harmonic analysis focuses on describing the harmonic content of pitch collections/patterns within a given music context in terms of harmonic labels, classes, functions and so on. Harmonic analysis is a rather complex musical task that involves not only finding roots and labelling chords within a key, but also segmentation (points of harmonic change), identification of non-chord notes, metric information and more generally musical context [48]. In this section, we focus on the core problem of labelling chords within a given pitch hierarchy (e.g. key); thus we assume that a full harmonic reduction is available as input to the model (manually constructed harmonic reductions).

Our intention is to create an analytic system that may label any pitch collection, based on a set of user-defined criteria rather than on standard tonal music theoretic models or fixed psychoacoustic properties of harmonic tones. We intend our representation to be able to cope with chords not only in the tonal system, but any harmonic system (e.g. octatonic, whole-tone, atonal, idiosyncratic traditional harmonic systems, etc.).

3.1.1 Root finding, Consonance and Idiom Independence

Root-finding is a core harmonic problem addressed primarily following two approaches: the standard stack-of-thirds approach and the virtual pitch approach. The first attempts to re-order chord notes such that they are separated by (major or minor) third intervals preserving the most compact ordering of the chord; these stacks of thirds can then be used to identify the possible root of a chord (see, for instance, recent advanced proposal by Sapp [44]). The second approach, is based on Terhard’s virtual pitch theory [49] and Parncutt’s psychoacoustic model of harmony [39]; it maintains that the root of a chord is the pitch most strongly implied by the combined harmonics of all its constituent notes (intervals derived from the first members of the harmonic series are considered as "root supporting intervals").

Both of these approaches rely on a fixed theory of consonance and a fixed (ordered) set of intervals that are considered as building blocks of chords. In the culture-sensitive stack-of-thirds approach, the smallest consonant intervals in tonal music, i.e. the major and minor thirds, are the basis of the system. In the second "universal" psychoacoustic approach, the following intervals, in decreasing order of importance, are employed: unison, perfect fifth, major third, minor seventh, and major second. Both of these approaches are geared towards tonal harmony, each with its strengths and weaknesses (for instance, the second approach has an inherent difficulty with minor harmonies). Neither of them can be readily extended to other harmonic systems.

Harmonic consonance/dissonance has two major components: Sensory-based dissonance (psychoacoustic component) and music-idiom-based dissonance (cultural component) [35]. Due to the music-idiom dependency component, it is not possible to have a fixed universal model of harmonic consonance/dissonance. A classification of intervals into categories across the dissonance-consonance continuum can be made only for a specific idiom. The most elementary classification is into two basic categories: consonant and dissonant. For instance, in the common-practice tonal system, unisons, octaves, perfect fifths/fourths (perfect consonances) and thirds and sixths (imperfect consonances) are considered to be consonances, whereas the rest of the intervals (seconds, sevenths, tritone) are considered to be dissonances; in polyphonic singing from Epirus, major seconds and minor sevenths may additionally be considered "consonant" as they appear in metrically strong positions and require no resolution; in atonal music, all intervals may be considered equally "consonant".

Let's examine the case of tonal and atonal harmony; these are probably as different as two harmonic spaces may be. In the case of tonal and atonal harmony, some concepts are shared, however, actual systematic descriptions of chord-types and categories are drastically different (if not incompatible), rendering any attempt to "align" two input spaces challenging and possibly misleading (FIGURE). On one hand, tonal harmony uses a limited set of basic chord types (major, minor, diminished, augmented) with extensions (7^{ths}, 9^{ths} etc.) that have roots positioned in relation to scale degrees and the tonic, reflecting the hierarchic nature of tonal harmony; on the other hand, atonal harmony employs a flat mathematical formalism that encodes pitches as pitch-class sets leaving aside any notion of pitch hierarchy, tone centres or more abstract chord categories and functions. It seems as if it is two worlds apart having as the only meeting point the fact that tones sound together (physically sounding together or sounding close to one another allowing implied harmony to emerge).

Pc-set theory of course, being a general mathematical formalism, can be applied to tonal music, but, then its descriptive potential is mutilated and most interesting tonal harmonic relations and functions are lost. For instance, the distinction between major and minor chords is lost if Forte's prime form is used (037 for both - these two chord have identical interval content), or a dominant seventh chord is confused with half-diminished seventh (prime form 0258); even, if normal order is used, that is less general, for the dominant seventh (0368), the root of the chord is not the 0 on the left of this ordering (pc 8 is the root). Pitch-class set theory is not adequate for tonal music. At the same time, the tonal music roman-numeral formalism is inadequate for atonal music as major/minor chords and tonal hierarchies are hardly relevant for atonal music.

3.1.2 The General Chord Type Representation

In trying to tackle issues of tonal hierarchy, we have devised a novel chord type representation, namely the General Chord Type representation or GCT representation, that takes as its starting point the common-practice tonal chord representation (for a tonal context, it is equivalent to the standard roman-numeral harmonic encoding), but is more general as it can be applied to other non-standard tonal systems such as modal harmony and, even, atonal harmony. This representation draws on knowledge from the domain of psychoacoustics and music cognition, and, at the same time, "adjusts" to any context of scales, tonal hierarchies and categories of consonance/dissonance.

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 tonality is given, the position within the given scale is automatically calculated.

Input to the algorithm is the following:

- Consonance vector: The user defines which intervals are consonant/dissonant. A 12-point vector, $\vec{v}$, is employed where each vector entry corresponds to a pitch interval in the range $(0-11)$ - in the current version of the algorithm Boolean values are used (i.e., consonant= 1, dissonant= 0). For instance, the vector $\vec{v} = [1,0,0,1,1,1,0,1,1,1,0,0]$ means that the unison, minor and major third, perfect fourth and fifth, minor and major sixth intervals are consonant – dissonant intervals are the seconds, sevenths and the tritone; this vector is referred to in this text as the common–practice consonance vector.
- Pitch Scale Hierarchy: The pitch hierarchy (if any) is given in the form of scale tones and a tonic. For instance, a D major scale is given as: 2, $[0,2,4,5,7,9,11]$, or an A minor pentatonic scale as: 9, $[0,3,5,7,10]$.
- Input chord: list of MIDI pitch numbers (converted to pc-set).

Algorithm 1 GCT computation pseudocode

Input: (**i**) the pitch scale (tonality), (**ii**) a vector of the intervals considered consonant, (**iii**) the pitch class set (pc–set) of a simultaneity

Output: The roots and types of the possible chords describing the simultaneity

```
find all maximal subsets of pairwise consonant tones
for all maximal subsets do
    order the pitch classes of each maximal subset to the most compact form
    create a sequence of maximal subsets (if many) by ordering them so as to have consecutive
    overlapping segments[1]
    keep the maximal subset that appears first in the sequence (chord's type)
    add the remaining pitch classes (chord "extensions") above the highest of the chosen max-
    imal subset's (if necessary, add octave – pitches may exceed the octave range)
    the lowest tone of the chord is the "root"
    transpose the tones of the chord so that the lowest becomes 0
    find position of the "root" in regards to the given pitch scale
end for
```

Table 1: Examples of applying the GCT algorithm.

	example 1	example 2	example 3
tonality	G: $[7,[0,2,4,5,7,9,11]]$	Dm: $[2,[0,2,3,5,7,8,11]]$	C: $[0,[0,2,4,5,7,9,11]]$
cons. vector	$[1,0,0,1,1,1,0,1,1,1,0,0]$	$[1,0,0,1,1,1,0,1,1,1,0,0]$	$[1,0,0,1,1,1,0,1,1,1,0,0]$
input	$[60,62,66,69,74]$	$[50,60,62,65,69]$	$[62,68,77,71]$
to pc–set	$[0,2,6,9]$	$[0,2,5,9]$	$[2,5,8,11]$
maximal subsets	$[2,6,9]$	$[2,5,9]$ and $[0,5,9]$	$[2,5]$, $[5,8]$, $[8,11]$, $[2,11]$
narrowest range	$[2,6,9]$	$[2,5,9]$ and $[5,9,0]$	$[2,5]$, $[5,8]$, $[8,11]$, $[11,2]$
add ext.	$[2,6,9,\mathbf{12}]$	$[2,5,9,\mathbf{12}]$ and $[5,9,0,\mathbf{14}]$	all rotations of $[2,5,8,11]$
lowest is root	2 (note D)	2 and 5	2,5,8, and 11
in root pos.	$[2,[0,4,7,10]]$	$[2,[0,3,7,10]]$ and $[5,[0,4,7,9]]$	$[X,[0,3,6,9]]$, $X \in \{2,5,8,11\}$
relative to ton.	$[7,[0,4,7,10]]$	$[0,[0,3,7,10]]$ and $[3,[0,4,7,9]]$	$X \in \{2,5,8,11\}$
	output after the application of optional steps (see text for explanation)		
	$[7,[0,4,7,10]]$	$[2,[0,3,7,10]]$	$[11,[0,3,6,9]]$

The GCT algorithm encodes most chord types "correctly" in the standard tonal system, however, it is undecided in some cases, and even makes "mistakes" in other cases. In most instances of multiple encodings, it is suggested that these ideally should be resolved by taking into account other harmonic factors (e.g. bass line, harmonic functions, tonal context, etc.). For instance, the algorithm gives two possible encodings for a $[0,2,5,9]$ pc–set, namely minor seventh chord or major chord with sixth (see example 2 in Table 1 above); such ambiguity may be resolved if tonal context is taken into account. Symmetric chords, such as the augmented chord or the diminished seventh chord, are inherently ambiguous; the algorithm suggests multiple encodings which can be resolved only by taking into account the broader harmonic context. Finally, in this version of the algorithm, in the case of two-note chords (dyads) the system prefers a perfect fourth interval to a perfect fifth as the fourth is smaller (narrower range); in tonal music, however, there is usually a preference for the perfect fifth – this is not reflected in this first version of the algorithm.

Since the aim of this algorithm is not to perform sophisticated harmonic analysis, but rather to find a practical and efficient encoding for tone simultaneities (to be used, for instance, in statistical learning and automatic harmonic generation), we decided to extend the algorithm so as to reach in every case a single chord type for each chord (no ambiguity).

GCT algorithm (Extensions) – additional steps:

- If more than one maximal subsets exist, merge them such that a maximally compact ordering occurs (maximal overlapping between subsets) – select as chord type the maximal set at the beginning of the merged list (left–hand side).
- In case of symmetric chords such as augmented triads or a diminished sevenths, prefer permutations in which non-scale tones appear at the end of the list (this is a rather arbitrary rule that works in some cases – alternative rules are currently considered).
- For dyads, prefer perfect fifth over perfect fourth, and prefer sevenths to second intervals.

The additional rules select chord type $[2,[0,3,7,10]]$ in example 2 (maximal overlapping between two maximal subsets), and $[11,[0,3,6,9]]$ in example 3 (last pitch–class is a non-scale degree).

An example harmonic analysis of a Bach Chorale phrase illustrates the proposed GCT chord representation (Figure 5). For a tonal context, chords types are optimised such that pcs at the left

[1]Look at example in Table 1.

J.S.Bach - Chorale 54 (Lobt Gott, ihr Christen, allzugleich) in G major - 2nd phrase

Roman Numeral Analysis:							
	D major	I^6	$vii_o{}^6$	I	ii^6	V^7	I
GCT Analysis (tonal major profile)							
	2,[0,2,4,5,7,9,11]	0,[0,4,7]	11,[0,3,6]	0,[0,4,7]	2,[0,3,7]	7,[0,4,7,10]	0,[0,4,7]
Pc-Set Analysis (chromatic scale):							
	normal orders	[0,4,7]	[0,3,6]	[0,4,7]	[0,3,7]	[0,2,6,9]	[0,4,7]
	prime forms	[0,3,7]	[0,3,6]	[0,3,7]	[0,3,7]	[0,3,6,8]	[0,3,7]
GCT Analysis (atonal profile)							
	[0,1,2,3,4,5,6,7,8,9,10,11]	2,[0,4,7]	1,[0,3,6]	0,[0,4,7]	4,[0,3,7]	7,[0,2,6,9]	2,[0,4,7]

Figure 5: Example of Bach Chorale illustrating the proposed GCT representation.

hand side of chords contain only consonant intervals (i.e. 3[rds] & 6[ths], and Perfect 4[ths] & 5[ths]). For instance, the major 7[th] chord is written as $[0,4,7,10]$ since set $[0,4,7]$ contains only consonant intervals whereas 10 that introduces dissonances is placed on the right-hand side – this way the relationship between major chords and major seventh chords remains rather transparent and is easily detectable. Within the given D major key context it is simple to determine the position of a chord type in respect to the tonic – e.g. $[7,[0,4,7,10]]$ means a major seventh chord whose root is 7 semitones above the tonic, amounting to a dominant seventh. This way we have an encoding that is analogous to the standard roman numeral encoding (Figure 5, top row). If the tonal context is changed, and we have a chromatic scale context (arbitrary "tonic" is 0, i.e. note C) and we consider all intervals equally "consonant", we get the second GCT analysis in Figure 5 which amounts to normal orders (not prime forms) in a standard pc-set analysis – for tonal music this pc-set analysis is weak as it misses out important tonal hierarchical relationships (notice that the relation of the dominant seventh chord type to the plain dominant chord is obscured). Note that relative "roots" to the "tonic" 0 are preserved as they can be used in harmonic generation tasks.

The GCT representation is a first step towards creating a common chord "language" that adapts to different idioms, reflecting their unique harmonic characteristics, and, at the same time, allowing potentially interesting mappings between diverse idioms. How might the GCT encoding behave in an "unknown" harmonic idiom (e.g. in the polyphonic music of Epirus where major second intervals are "consonant" in the sense that they require no resolution and may appear in final chords of phrases – is organising harmonic intervals around major seconds meaningful and does it give any new insights about the style)? What would happen if the GCT encoding is altered for a known idiom in regards to consonance/dissonance or scale hierarchies? Is it possible to "blend" characteristics of different idioms giving rise to "blended" GCT representations? We believe that just being able to ask such questions in the first place and, also, to try out such possibilities is interesting as far as creativity and concept invention are concerned.

4 Learning and Generating Chord Progressions

The exploration of harmonically meaningful chords within musical phrases are considered as distinctively important parts of an idiom. Such parts can be subsequently used as independent blend-able entities, allowing the mechanism of conceptual blending to produce harmonic "checkpoints" that comprise harmonic characteristic from multiple harmonic idioms. An example of structurally important parts are the chords in cadences, as discussed in the literature review presented in Section 4.1. However, the presented approach generalises the notion of "important" chords to a methodology that allows the insertion of fixed–chord constraints in predefined positions of a phrase. Harmonisation with fixed checkpoints is considered a crucial component of the presented work, since it enables the prior definition of important chords in intermediate positions of the melody to be harmonised. The intermediate or "*anchor*" chords of a phrase are considered to be given either from an algorithmic process in a hierarchical level above the "chord progression" level – where chord transitions are defined by the proposed HMM variation – or by a human user. A probabilistic method to incorporate such control is discussed in Section 6, where a method is developed to apply chord constraints on intermediate phrase endings. However, the experimental results in the current section of the report mainly encompass examples where the fixed–chord constraints are provided either by a human expert, or by the chords utilised in the genuine composition of the harmonised melody (from phrases that were not included in the training set). The proposed methodology applies to full reductions of harmonic material, therefore, a phrase is considered to include only the chords and melody notes that encompass harmonic meaning.

An additional fundamental concern of the proposed harmonisation approach is the idiom–independency in the chord symbols, chord relations and melodic considerations. This concern is addressed by utilising the general chord type (GCT) representation, which is briefly discussed in Section 4.2.2. The proposed algorithm acts on a certain level of the harmonic hierarchy, primarily the *phrase* level. Thereby, given some "anchor" chords that remain fixed in a phrase, the aim of the algorithm is to select *"proper"* chord sequences that connect the intermediate parts of the fixed chords, under the conditions introduced by the melodic material to be harmonised. The evaluation of the algorithm incorporates a comparison between the proposed constrained HMM (CHMM) and a "typical" HMM, which incorporates prior probabilities for the beginning and ending chords. The results indicate that CHMMs produce harmonisations that might be completely different to the ones produced by HMMs, depending on the imposed constraints. The results are reported on phrases of a set of J. S. Bach chorales, since they comprise an unofficial "benchmark" dataset for melodic harmonisation methodologies.

4.1 Previous work and motivation

Hidden Markov models (HMMs) have been extensively used for the automatic harmonisation of a given melody, since their formalisation describes the targeted task very well: given a sequence of observed notes (melody), find the most probable (hidden) sequence of chords that is compatible with the observations, according also to a chord transition matrix. In several studies of HMM–based melodic harmonisation methodologies, a straightforward distinction is made on the role that some chords play to the composition – mainly the cadence of the phrase. For instance, the cadences of produced harmonisations by the HMM developed in [2] were utilised to evaluate the system's performance, by comparing the cadence patterns that were produced by the system to the ones observed in the dataset.

Several HMM approaches discuss the utilisation of some methodological tools to amplify the role of the cadence in the harmonisation process. For instance, in [1] and [16] a backwards propagation of the HMM methodology is proposed, i.e. by examining the prior probabilities of the final chord given the final melodic note. The Markov decision process followed in [51], rewards the authentic cadences thus providing higher probabilities to chord sequences that end with an authentic cadence. In [52] the phrases are divided in tonic, subdominant, dominant and parallel tonic chords, allowing a trained HMM to acknowledge the positions of cadences, however the selection of chords is performed through a rule–based process. A commercial application utilising HMM for melodic harmonic is *mySong* [45], which receives the melody by the singing voice of the user, extracts the pitches of the melody and employs an HMM algorithm to provide chords for the melody. The approach followed therein is equivalent to the one described in Section 4.2.1 (and in Equation 1), which is also used as a starting point towards the formalisation of the BCHMM. According to the HMM approach utilised by mySong, prior probabilities are considered not only for the beginning chord of a piece, but also for the ending one, a fact that further biases the choice of solutions towards ones that incorporate first and final chords that are more often met in the training dataset.

The approach developed in the context of COINVENT is motivated by the research in the aforementioned works, but it is different on a fundamental aspect: it allows the *deterministic* (not probabilistic) insertion of chords at any place in the chord sequence. Such an approach is important since it permits the extension of the "learned" transitions with, potentially allowing to build composite harmonisation that comprise characteristics from various idioms. To this end, the isolation of the harmony in "strategic" harmonic positions (e.g. the cadence, the beginning or intermediate parts of a phrase) is expected to contribute to the project's perspective.

4.2 Intermediately–constrained probabilistic harmonisation

The aim of the developed methodology is to allow the probabilistic harmonisation, while allowing prior determination of intermediate chords (also named as *checkpoints* in the literature [5]). The intermediate chords may either be specified by an algorithmic process that determines music structure on a higher hierarchical level, or may be directly inserted by a human annotator. Some examples of algorithm classes on higher hierarchical levels that could be utilised for providing intermediate anchor chords are rule-based approaches, generative grammars, or even Markov models trained with chords on a sparser time scale (e.g. the beginning, the middle and the final chord of phrases). Additionally, the fact that direct human intervention is enabled, allows the presented methodology to be the backbone of a melodic harmonisation assistant, which allows its user to specify a harmonic "spinal chord" of anchor chords that are afterwards connected by chord sequences that give aesthetic reference to a learned idiom.

An abstract example of a melodic harmonisation process that incorporates some fixed anchor points is demonstrated in Table 4.2. Therein, a melodic line denoted by m_i $i \in 1,2,\ldots,8$ (supposed length 8) is harmonised with some given intermediate chords as constraints, namely I_i, $i \in 1,2,3$. The intermediate chords have been applied to specific notes of the melody, i.e. I_1 on m_1, I_2 on m_5 and I_3 on m_8. The first and final notes are harmonised with fixed chords for demonstration purposes, either one or both of them could be harmonised automatically by the variation of the HMM variation discussed in this part of the report. After the intermediate fixed chords have been defined, the *boundary–constrained* variation of the HMM (BCHMM) is utilised for each of the successive parts that begin and/or end with a fixed chord. It has to be highlighted that the BCHMM is an abbreviation signifying an intermediate step of the proposed CHMM methodology. In this step only boundary constraints are considered. In the case where the beginning and ending chords of the phrase are not fixed, the boundary constraints apply only on the fixed edge; the non–fixed edge is harmonised by utilising the typical probabilistic HMM boundary condition, as discussed in the next paragraphs. For the example in Table 4.2, the BCHMM algorithm is applied twice, once of each pair of consecutive anchor chords – namely BCHMM^1 for connecting I_1 with I_2 and BCHMM^2 for connecting I_2 with I_3.

mel.	m_1	m_2	m_3	m_4	m_5	m_6	m_7	m_8
con.	I_1				I_2			I_3
		C_1^1	C_2^1	C_3^1		C_1^2	C_2^2	

BCHMM^1 BCHMM^2

CHMM

Table 2: Abstract example of the proposed harmonisation algorithm. On top (m_i) the melody notes to be harmonised are illustrated, below (I_i) the chord that are given as constraints and in the bottom (C_i^j) the chords produced by the j–th BCHMM method application.

The presented algorithm discusses only the level of chord labelling, i.e. its goal is to attribute a chord symbol – expressed as a GCT structure – disregarding information about harmonic rhythm and voice leading. Harmonic rhythm is a crucial matter that defines a vital part of a harmonisation's character, however, within the context of the prototypical evaluation of the proposed method, a chord is considered to accompany every note of the melody; a similar approach has often been endorsed in past research. Similarly, voice leading is also an important aspect of harmonisation, while one could arguably consider that voice leading is sometimes fundamental in a sense that the movement of each separate voice defines the final vertical shape of the harmonic blocks. Nonetheless, some primitive experimental results on automatic harmonisation with HMMs indicate that the GCT bases and extensions of most probable chord successions, as reflected in a transition matrix, encapsulate the potential of efficient voice leading, allowing the successions of vertical harmonic blocks to be combined in such a way that an efficient voice leading algorithm would potentially interpret some basic horizontal characteristics. Although this argument is clearly supported by the examples presented in the experimental results section, a more elaborate examination is left for future work.

4.2.1 Intermediate anchor chords as boundary constraints

The chords that "connect" two successive fixed–boundary chord segments are defined by the aforementioned variation of HMM, the BCHMM. Throughout the development of the BCHMM, a nomenclature relative to the subject under discussion will be followed, i.e. the dataset will comprise musical pieces (more specifically harmonic reductions of pieces), the states will represent chords and the observations will describe melody notes. To this end, the set of possible states–chords will be denoted by $\mathscr{S}$, while the letters C and c will be used for denoting chords. The set of all possible observations–notes will be denoted as $\mathscr{Y}$, while Y and y will be denoting melody notes. Specifically, the capitalized letters will be used to denote statistical variables, while their instantiation variables will be denoted by lower case letters. For example, $P(C_i = c_i)$ denotes the probability that the chord in the i–th position is a c_i chord (where c_i is a specific chord, e.g. a [7,

$[0,4,7]$, $[10]]$ chord in GCT form, which is a dominant seventh chord).

An initial set of music phrases is considered which will provide the system with the required statistical background, constituting the training set. Through this dataset the statistics that are induced concern three aspects:

1. The probability for each state (chord) to be a beginning chord. This distribution is computed by examining each beginning chord for each phrase in the dataset and is denoted as $\pi(C_1 = c)$, $c \in \mathscr{S}$.
2. The probability for each state (chord) to be an ending chord. This distribution is computed by examining each ending chord for each phrase in the dataset and is denoted as $\tau(C_T = c)$, $c \in \mathscr{S}$.
3. The probability that each state follows another state, denoted as $P(C_i = c_i | C_{i-1} = c_{i-1})$, c_i, $c_{i-1} \in \mathscr{S}$.
4. The probability of a chord being played over a melody note, denoted as $P(C_i = c_i | Y_i = y_i)$.

These probabilities are related during the computation of the *overall* probability that a certain chord sequence ($C_i = c_i$, $i = 1, 2, \ldots, T$) is applied over an observed melody ($Y_i = y_i$, $i = 1, 2, \ldots, T$). This overall probability is computed by

$$P(C_i = c_i | Y_i = y_i) = P_\pi \, P_\mu \, P_\tau, \tag{1}$$

where

$$P_\pi = \pi(C_1 = c_1) \, P(C_1 = c_1 | Y_1 = y_1), \tag{2}$$

$$P_\mu = \prod_{i=2}^{T} P(C_i = c_i | C_{i-1} = c_{i-1}) \, P(C_i = c_i | Y_i = y_i), \tag{3}$$

$$P_\tau = \tau(C_T = c_T) \, P(C_T = c_T | Y_T = y_T). \tag{4}$$

An optimal sequence of chords is one that maximizes the overall probability (in Equation 1)[2], by achieving an optimal path of states that yield a maximal combination for the probabilities in all the counterparts (P_π, P_μ and P_τ), typically through the Viterbi [11] algorithm. The probabilities in P_π promote some chords as better solutions to begin the path of chords: the ones that are more often used in the beginning of pieces in the dataset. Similarly, the probabilities in P_τ advance solutions that are more often met as concluding chords. Although the results reported in past works indicate that P_π and P_τ most probably create satisfactory results, these probabilities do not *guarantee* that the more often met beginning and ending chords will be utilised. A similar comment can be made about some strategies that have been proposed, which focus on constructing satisfactory cadences, by beginning from the end of the phrase to be harmonised and employing the Viterbi algorithm from "right-to-left". Specifically, while the latter approaches have an increased bias towards the

[2]In implementations of HMMs it is usually the negative log–likelihood that is being minimized, i.e. the logarithm of the expression in Equation 1, since the numbers that are yielded by consecutive multiplications of probabilities (quantities ≤ 0) are difficult to be compared by eye because of their small magnitude.

cadence part of the phrase, it is again not guaranteed that the cadence or the beginning chord of the phrase will be satisfactory.

Regarding the probabilistic scheme, the process for computing the probability value in Equation 1, incorporates the extraction of the statistical values for $\pi(C_1 = c_1)$ and $\tau(C_T = c_T)$, according to the number of occurrences of each chord as an initial or final chord respectively. For the BCHMM approach however, no statistics are considered for these boundary points, since they *certainly* (with probability 1) include the chords specified by a higher hierarchical level or by a human annotator. To be compatible with the terminology followed hitherto for the presentation of the HMM model, the latter comment can be expressed by modifying the Equations 2 and 4 so that they indicate the chords selected at temporary boundary points between successive checkpoints as certain, while eliminating the probabilities for any other chords to appear. Specifically, if the beginning and ending chords are selected to be α_1 and α_T respectively, the new probabilities that substitute the ones expressed by Equations 2 and 4 are the respective following:

$$P'_\pi = \begin{cases} 1, \text{ if } C_1 = \alpha_1 \\ 0, \text{ otherwise} \end{cases} \tag{5}$$

$$P'_\tau = \begin{cases} 1, \text{ if } C_T = \alpha_T \\ 0, \text{ otherwise.} \end{cases} \tag{6}$$

The probability that is therefore optimized is the following:

$$P(C_i = c_i | Y_i = y_i) = P'_\pi \, P_\mu \, P'_\tau, \tag{7}$$

where the factor P_μ is the one defined in Equation 3. The employment of the Viterbi algorithm under the constraints imposed by the boundary conditions, as reflected by Equations 5 and 6, assigns zero–value probabilities to all paths, except the ones that begin with α_1 and end with α_T. Figure 6 illustrates the trellis diagram of the Viterbi algorithm under the discussed constraints.

4.2.2 Application of BCHMM in the current harmonisation system

The efficiency of the HMM, and consequently the BCHMM, methodology relies on selecting a proper set of states to represent the chords that are utilised in the training set, which will subsequently be used in the harmonic generation process. The term "proper" indicates that there is a tradeoff in the amount of information of chord representation and the number of states required to delegate each chord in the HMM (and the BCHMM). For instance, by describing the possible chords only as major or minor, the number of states remains small (24 for all 12 pitch classes), however the harmonic description is very poor. Several works in the literature ([5, 45] among others) propose the utilisation of standard chords (e.g major, minor, diminished, augmented and major seventh), applicable to all 12 relative pitch classes of the composition key of the examined pieces. However, by devising such a chord selection scheme it is possible that important harmonic information is excluded, since several pitch class combinations that might appear (rather frequently in some musical idioms) are disregarded.

The chord representation followed in the context of this report is the *general chord type* (GCT) representation, which is able to embody the information of both consonant and dissonant parts of

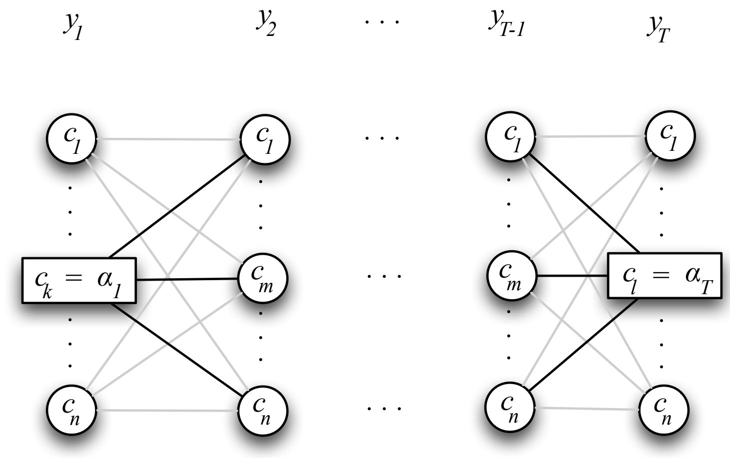


Figure 6: Trellis diagram for the BCHMM. Only transitions from α_1 and to α_T as first and last states respectively are permitted. The intermediate trellis diagram is the same as in a typical HMM.

a pitch class group. The GCT incorporates three parts, the *root*, the *base* and the *extensions* of a chord, denoted with three different entrances in a list of the form [root, [base], [extension]]; for example the pitch class $[7,11,2,5]$ is represented as [7, $[0,4,7]$, [10]], which indicates a dominant seventh chord. These parts are defined for pitch class simultaneities, according to a process that isolates the maximal mutually consonant pitch class combinations of this simultaneity, according to a consonance vector that defines the intervals between pitch classes that are considered consonant. For the chorales of Bach, that constitute the dataset of examination, the consonant intervals are considered to be the major and minor thirds, their inversion–equivalent major and minor sixths and the perfect fifths and fourths. A complete description of the GCT can be found in [3].

The implementation of the HMM incorporated a simple "rule–based" observation–to–state probability assignment ($P(C_i = c_i|Y_i = y_i)$) for defining the probability for each chord to be played with each note of the melody. Specifically, for each note of the melody, this "rule–based" criterion provides a maximum probability for chords that include this note and a minimum for one that does not. Maximum probability is set to 1, while the minimum is set to 10^{-6}. Additionally, the zero entries of the chord transition matrices that are produced by the training simulations, are also assigned a value of 10^{-63}. By removing the zero entries in these matrices, a potential blocking of the algorithm is avoided in situation where zero probabilities occur. Such situations may occur either in the extreme scenario where there is no chord to include a melodic note, or in the even more extreme scenario where there is no probable path connecting two predetermined anchor points.

[3]After the adjustment of the values in either the observation or the transition matrices, these matrices could be normalised to produce a unit sum for each chord. However, since the probability values in the matrix entries are computed only in terms of maximising the total probability (ignoring its magnitude), such a normalisation is not necessary.

4.3 Results

The experimental results demonstrate in a qualitative manner the effectiveness concerning several aspects of the proposed melodic harmonisation approach:

1. The effectiveness of the GCT representation towards capturing the idiom's "chords", providing interpretations that are in agreement with the Roman numeral analysis.
2. The efficient adaptation of the GCT representation to the chord bases and extension characteristics that enable the automatic harmonisation system to be amenable to effective voice leading. Dissonance of extensions, should be treated for special voice leading.
3. The presented methodology's effectiveness in terms of the training data requirements.
4. The increase of interestingness that the insertion of intermediate and/or boundary chords can introduce to the composed harmony.
5. The fact that the HMMs are versatile enough to adapt to "deterministic" harmonic constraints.

During the "unofficial" evaluation of the presented methodology, several test phrases were harmonised, as well as several anchor point insertion setups were examined. The presented results include some indicative harmonisations that have been produced by the system with different anchor point setups. The utilised dataset comprises a selection of phrases from the "benchmark" chorales of J. S. Bach, specifically some chorales in the major mode.

The experimental process aims to provide indications about the fact that the utilisation of the anchor points yield harmonisations that are potentially more "*interesting*" than the ones produced by the typical HMM methodology – depending on the selected anchor points. Therefore, the experimental results expose the ability of the proposed system, as well as the flexibility of the modified HMM scheme towards allowing different – and potentially more interesting – harmonisation alternatives, according to the provided anchor points. To this end, the system's evaluation processes mainly addresses the fact that the proposed methodology is implementable using a relatively small dataset of training pieces.

The presented approach addresses the harmonisation task within the context of a certain key, thus a full harmonic reduction of phrases is considered as input to the system; the term "phrase" will hereby signify the melody notes and their harmonisation, as yielded from the reduction. The phrases of the Bach chorales are divided in two sets according to their key of composition, i.e. in major and minor phrases. Although harmonisations of both modes were tested, the reported results include only major mode phrases. The GCT chords–states that are derived for the major chorales of Bach are 41 and for the minor chorales 38, while many of the major and minor states are overlapping, i.e. exist both in the major and in the minor chorales. Several of these states are redundant since their GCT expression in fact describes chords of the same functionality, e.g. the GCTs [0, $[0,4,7]$, []] and [0, $[0,4]$, []] denote a major chord in the tonic. Additionally, there is a considerable amount of GCT states (around 15 for each mode) that occur only two or three times in the entire dataset. The latter comments indicate that the employment of a GCT clustering technique could group some GCTs according to their harmonic functionality, further reducing the states to approximately 25 for each mode. However, such a grouping methodology is currently under development, while some pointers about this ongoing research are given in Appendices A and B.

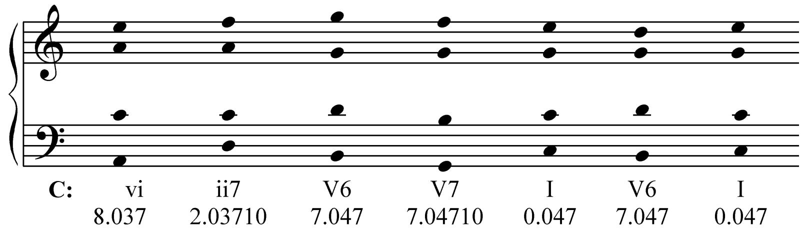
When harmonising a melody with no constraints, the HMM methodology selects the most probable sequence of chords (hidden states) according to probabilities related to the melody's note to be harmonised and to probabilities related to the transitions between pairs of states. The imposition of fixed–chord constraints is intuitively expected to alter the harmonisation "locally", i.e. the CHMM harmonisation is expected to be different than the one provided by the typical HMM a few chords before or after a chord that remains fixed – if the selected chord to be fixed is different than the one provided by the HMM. However, the application of chord constraints in some cases provided different harmonisations throughout the entire length of the phrase. The voice leading in the examples presented below was performed by a music expert; an algorithmic process for voice leading is a future research goal. The score examples that are analyzed in the remaining of this section are produced by HMMs or CHMMs that trained on the same set of 30 random chorale phrases, which did not include the harmonised phrases.

The example in Figure 7 amplifies the role of anchor chords and specifically the beginning and ending chords of a phrase. In this example, a Bach chorale melody is harmonised with the typical HMM methodology (top) and with anchor boundary (beginning and ending) chords denoted by an asterisk. The boundary chords are the ones utilised by Bach in the genuine chorale. An initial comment concerns the fact that the HMM methodology does not "guarantee" that the beginning and ending (boundary) chords of a melody to be harmonised are identical to the ones that would potentially be utilised by a human composer. Additionally, the role of the boundary chords is crucial: the example in Figure 7 demonstrates that different anchor chords provided an entirely different harmonisation. Furthermore, this example shows that that the imposition of constraints "forced" the system to follow more "interesting" and unpredictable chord paths, since, the typical HMM methodology utilised more typical and probable chord progressions between V and I chords. The imposition of constraints on the other hand, forced the HMM methodology to establish temporary secondary tonalities, yielding a richer harmonic interpretation of the melodic sequence.

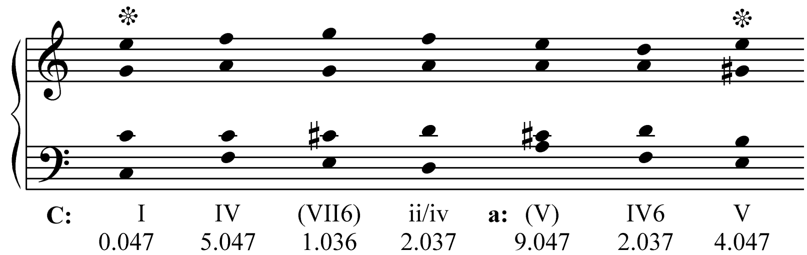
The evidently important role of the beginning and ending chords leads to further inquiries about the ability of the HMM to accurately "predict" the boundary chords of phrases, according to the ones utilised in the genuine compositions. Answers to these inquiries are approached through a statistical comparison between the boundary chords produced by the HMMs and the boundary chords assigned by Bach. Specifically, an intuitively realistic answer is pursued with the utilisation of three different metrics on how "correct" the boundary chords attributed by the HMM are, considering the boundary chords of the genuine Bach chorales phrases as ground–truth. Specifically, when the HMM system harmonises the melody of a phrase, the attributed *first* and *final* GCT chords of the HMM harmonisation are compared (according to the aforementioned three metrics) with the respective GCT chords that exist in the genuine harmonisation of Bach on the same phrase. Therefore, these three metrics are considered to indicate the "*efficiency*" of the HMM harmonisation regarding the beginning and ending GCT chords. These metrics are the following:

1. *Pitch class similarity* ($\texttt{PC}, \in [0,1]$): the percentage of pitch classes (PCs) in the HMM proposed chord that are equal to the pitch classes of the "correct" chord.

2. *Root similarity* ($\texttt{root}, \in \{0,1\}$): 1 if the GCT roots are equal, 0 otherwise.



(a) typical HMM



(b) CHMM with boundary anchor chords

Figure 7: (a) The harmonisation of a Bach chorale melody with the typical HMM methodology and (b) with constraints on the first and final chords (indicated with an asterisk).

3. *Exact similarity* (`exact`, $\in \{0,1\}$): 1 if the GCT chords are completely equal, 0 otherwise.

The `PC` criterion is the most generous one, since it provides a rather positive score to chords that are considered wrong. For example, if the final chord in a phrase is [0, $[0,4,7]$, []] (i.e. I degree) and the HMM proposes an arguably wrong [4, $[0,3,7]$, []] chord (i.e. iii degree), then it receives a score of 0.6667, since the common relative to the root PCs are 4 and 7, while the non–common is only the relative PC 11 (contradicting to 0). The `exact` criterion is the strictest criterion, since it requires that the root, base and extension between chords are the same. The `root` criterion admits that it is an excessive requirement that all the GCT chord characteristics be the same, acknowledging also the fact that potentially different GCT bases and extensions refer to chords of the same functionality, e.g. [0, $[0,4,7]$, []] and [0, $[0,4]$, []]. To this end, the `root` criterion accounts only the similarity of the root GCT part.

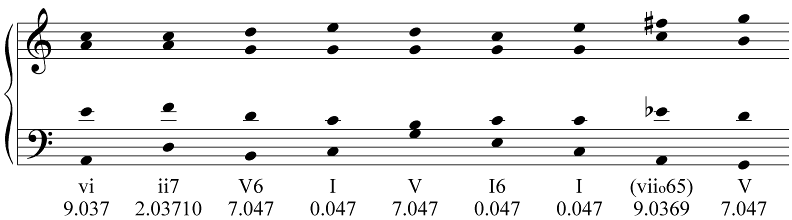
The experimental setup includes four different sets of training excerpts, namely the $\texttt{tr}-5$, $\texttt{tr}-10$, $\texttt{tr}-20$ and $\texttt{tr}-30$ sets. Each of these sets comprises a number of training phrases that is indicated by the numerical part of the name, e.g. the $\texttt{tr}-20$ describes an experimental simulation where 20 phrases are used as training data. Under any training scenario, 10 test melodies are harmonised, which belong to chorale phrases that do not pertain to the training set. The training and testing chorales are randomly selected in 100 random selection–training–harmonising–testing simulations, while different sessions are performed for major and minor mode chorales. Thereby, the statistics that are subsequently presented are extracted from 100 simulations for each setup: major or minor chorale phrases, with different numbers of training phrases (5, 10, 20 and 30) and 10 phrases as harmonising–testing data.

Table 4.3 demonstrates the mean values for the three efficiency measures in the first and final chords of the HMM harmonisations, for the major and the minor chorales and for all training setups (different number of training pieces). A first comment concerns the sensitivity of each metric to the number of training pieces. For instance, the PC metric remains relatively steady regardless of the number of pieces as a training set, while the remaining two metrics increase considerably as the number of training pieces increase. Specifically, for the major pieces the increase is around 10%, while for the minor piece around 4-5%. This fact indicates that the number of coinciding pitch classes is a rather vague measure, incorporating little musical information, since this measure does not reveal the dense impact that the increase of the training data would expectedly have.

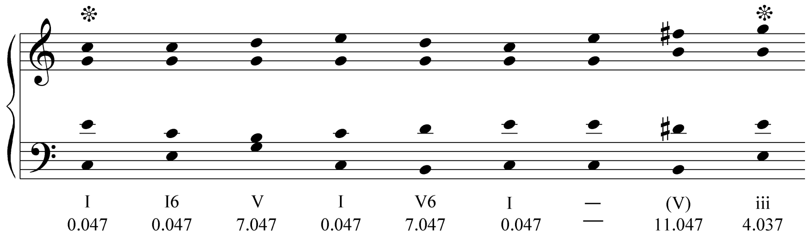
	Beginning		Ending	
	major	minor	major	minor
	pitch class similarity (PC)			
tr-5	0.8635	0.7917	0.9373	0.8777
tr-10	0.8698	0.8014	0.9437	0.8827
tr-20	0.8650	0.8008	0.9358	0.8934
tr-30	0.8670	0.7970	0.9533	0.8884
	root similarity (root)			
tr-5	0.4820	0.4110	0.6860	0.7330
tr-10	0.4940	0.4060	0.7460	0.7720
tr-20	0.4900	0.4420	0.7770	0.8010
tr-30	0.5310	0.4380	0.8230	0.7840
	exact matches (exact)			
tr-5	0.4360	0.3370	0.6530	0.4580
tr-10	0.4530	0.3280	0.7220	0.4990
tr-20	0.4660	0.3740	0.7570	0.5020
tr-30	0.5120	0.3710	0.7920	0.4980

Table 3: Efficiency of the typical HMM harmonisation regarding the first and final chords, according to the three defined metrics.

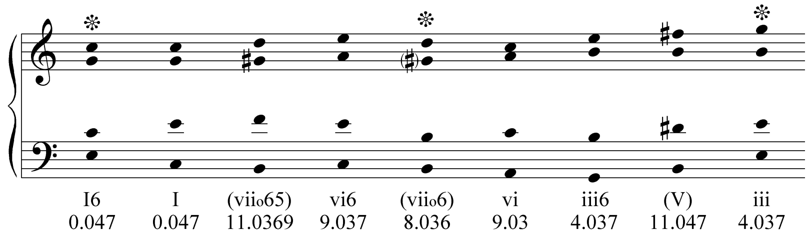
Except from the imposition of boundary chords, the insertion of intermediate chords can also produce interesting results. The example depicted in Figure 8 discusses the harmonisation of a Bach chorale in four different versions. Specifically, Figure 8 (a) demonstrates the harmonisation produced by the typical HMM methodology, while the harmonisation in (b) is produced with constraints on the boundary chords (as indicated by the asterisks). The constraints used in the phrase's boundaries are the ones utilised by Bach in the genuine chorales. The imposition of the boundary constraints does not produce a harmonisation that is entirely different regarding the selection of GCT chords (unlike the example shown in Figure 7), however the voice leading that was assigned by the music expert in both phrases is different. The harmonisation became more interesting when the music expert indicated the insertion of the diminished chord marked with an asterisk in Figure 8 (c) (fifth chord). This anchor chord changed the harmonisation entirely; even when the boundary constraints were alleviated, the harmonisation produced by the CHMM system (Figure 8 (d)) was again completely novel. The fact that different constraint conditions produce diverse harmonisations, amplifies the motivation to utilise a "deterministic" chord selection scheme along with the probabilistic HMM framework.



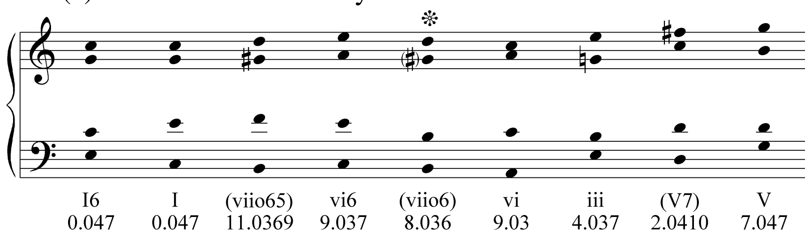
(a) typical HMM



(b) CHMM with boundary anchor chords



(c) CHMM with boundary and intermediate anchor chords



(d) CHMM with an intermediate anchor chord

Figure 8: (a) The harmonisation of a Bach chorale melody with the typical HMM methodology and with constraints on (b) the boundary chords, (c) the boundary and one intermediate chord and (d) only one intermediate chord. The fixed intermediate chords selected by a human annotator are indicated on the score with an asterisk.

4.4 Discussion

This section presented a methodology for performing automatic melodic harmonisation, i.e. providing chords on the notes of a given melody, through a methodology that is based on the hidden Markov model (HMMs), namely the constrained HMM (CHMM), which harnesses the capabilities of the HMMs to perform harmonisations with strictly specific requirements expressed through the employment of certain chords to harmonise certain notes of a melody. Such "anchor" chords would be selected either by an algorithmic process functioning a higher level of the harmonic hierarchy, or by a user. The utilisation of specific chords imminently enhances the automatically produced harmonisations since the proper selection of some key–chords leads the system to interesting harmonic paths. For instance, the selection of the first and final (boundary) chords of a phrase, which chords strongly imply the tonal constitution, is a crucial part for generating harmonisations that provide strong reference to an intended musical idiom – fact that is also highlighted by several works in the automatic harmonisation literature.

According to the experimental results reported herein, the typical HMM approach assigns beginning and ending chords of phrases that are more probable, a fact that potentially contradicts with a composer's choices. Additionally, the imposition of fixed–chord constraints, even only on the boundaries of phrases, force the CHMMs to produce harmonisations that are significantly different to the ones produced without constraints – and often more interesting since they are more "improbable". The chord representation that is employed is the general chord type (GCT) representation, which is a novel technique under development and allows the selection of a relatively small number of chords as states, without disregarding harmonic information from chord extensions.

The proposed technique is a part of an ongoing research in the context of the COINVENT project, according to which the invention of new concepts in automated harmonisation is approached by blending harmonic concepts of several musical idioms. To this end, the determination and utilisation of important harmonic parts of idioms is pursued, e.g. selecting proper fixed–chord constraints ("anchor" chords) and voice leading among others. Therefore, the proposed technique remains to be integrated with an algorithmic "anchor" chord selection mechanism, as well as an algorithmic process that performs idiom–dependent voice leading. The development of the CHMM methodology would potentially be harnessed with even more advanced and abstract harmonic constraints. For example, the user of a system would not only select entire chords to harmonise certain notes of phrases, but also specific notes that should be present along with a note of a harmony, therefore reducing the chord possibilities. Additionally, as the results indicated, by "fixing" the final boundary point it is not expected to lead to a "fixed" cadential pattern, since the absolute similarity in the final chord between the genuine and the artificial harmonies was not followed by an increase to the pre–final chords. The utilisation of longer harmonic segments in places where cadences happen has been previously discussed in the literature [2], providing pointers for future work that would include larger cadential "chunks" as ending boundary points. Finally, the boundary constrained formalisation could be harnessed with a variable order Markov model in the hidden layer, like the predictions suffix trees, producing results by potentially incorporating information over longer harmonic parts for deciding the next chords.

5 Voicing layout

Voicing layout discusses the materialisation of GCTs into MIDI pitches, for producing the final harmonisation result. The voicing layout module is trained on given chord layouts and melodies, inducing statistical rules about the voicing layout attributes of GCTs, as well as the relations of melody notes and bass notes motion. In its generative state, the voicing layout module receives the generated GCT and the user-defined melody as input, while its output is the final harmonisation in MIDI numbers.

5.1 Bass voice motion

Melodic harmonisation systems assign harmonic material to a given melody. Harmony is expressed as a sequence of chords, but the overall essence of harmony is not concerned solely with the selection of chords; an important part of harmony has to do with the relative placement of the notes that comprise successive chords, a problem known as *voice leading*. Voice leading places focus on the horizontal relation of notes between successive chords, roughly considering chord successions as a composition of several mutually dependent voices. Thereby, each note of each chord is considered to belong to a separate melodic stream called a *voice*, while the composition of all voices produces the chord sequence.

Regarding to melodic harmonisation systems, the assignment of *proper* voice leading incorporates the preservation of balanced relations between the melody and all chord-composing voices. The term *proper* is utilised to pinpoint that there are certain sets of "rules" that need to be taken under consideration when evaluating voice leading. However, these "rules" are defined by musical systems, called *idioms*, with many differences. Therefore, different musical idioms potentially employ different sets of rules to evaluate the "appropriateness" of a voice leading scenario.

Such rules have been hand-coded by music experts for the development of rule-based melodic harmonisation systems (see [37] for a review of such methods). Similarly, such hand-coded rules have been utilised as fitness criteria for evolutionary systems (see [8, 40] among others). However, the specification of rules that are embedded within these systems are very complex with many variations and exceptions. Additionally, the formalisation of such rules has not yet been approached for musical idioms that have not hitherto been thoroughly studied.

The work presented in this part of the report is a part of an ongoing research within the context of the COINVENT project, which examines the development of a computationally feasible model for conceptual blending. Thereby, computer systems are given the ability to invent novel concepts through blending two given conceptual input spaces. A part of this project is to apply this methodological framework in melodic harmonisation, creating novel harmonic spaces by blending two given harmonic spaces. Therefore, the construction of harmonic conceptual spaces for many music idioms is pursued. These spaces need to incorporate information on many diverse aspects of harmony, allowing the blending methodology to combine different parts of harmonic information from different music idioms.

The inclusion of many diverse musical idioms in this approach is required for achieving bold results that blend characteristics from different layers of harmony across idioms. However, as previously mentioned, the extraction of hand-coded rules from relatively unstudied idioms demands research resources that go overwhelmingly beyond the scope of the discussed project. Therefore, the approach followed to obtain harmonic information from many diverse music idioms incorporates the extraction of statistics for many aspects of harmony. The aspect of harmony that this part of the report discusses is voice leading of the bass voice. The bass voice leading is an important element of harmony, since it indicates the inversion of the utilised chords, as well as it constitutes a melodic voice by itself.

5.1.1 Previous work

To our best knowledge, no study exists that focuses only on generating voice leading contour of the bass line independently of the actual chord notes (i.e. the actual chord notes that belong to the bass line are determined at a later study). Most of the works that will be hereby presented focus on the generation of harmonies through producing sequences of entire chords, usually starting from an observed (or artificially generated) melody. Experimental evaluation of methodologies that utilise statistical machine learning techniques demonstrated that an efficient way to harmonise a melody is to add the bass line first [50]. This conclusion was made through the information theoretic measure cross-entropy, when the soprano, alto, tenor and bass voice where pairwise compared regarding their statistical relations. To this end, the motivation behind the work presented in this part of the report is further enforced by the findings in the aforementioned paper. There has been a variety of prior work in methodologies and algorithms on probabilistic melodic harmonisation which have direct relevance and usefulness to our own research.

The task of automated melodic harmonisation has been approached from two different perspectives: either to find a satisfactory chord sequence for a given melody (performed by the soprano voice), or to find the remaining three voices that complete the harmony for a given melodic or bass line. The typical form in the latter type of harmonisation is referred to as the “four-part harmony” task, which examines the proper combination of the soprano, alto, tenor, and bass voices. The four-part harmonisation is a traditional part of the theoretical education of Western classical musicians and numerous researches can be found regarding this task [9, 36, 40, 51].

Allan and Williams [1] proposed a four-part harmonisation method based on hidden Markov models (HMM). Therein, two HMMs were utilised to generate chorales in the style of J.S. Bach. The first HMM was employed to yield a sequence of note intervals that accompany each melody beat, while the second produces finer-scale ornamentations. The motivation was to create a model which can be used for the prediction of notes for filling three voices corresponding to the remaining harmonic lines at each time step. Yi and Goldsmith [51] proposed a four-part harmonisation method based on a Markov decision process. A state is represented as a 10-tuple (S1, A1, T1, B1, S2, A2, T2, B2, S3, P), where Si, Ai, Ti, Bi are respectively the soprano, alto, tenor, and bass notes at time i, and P is a temporal position.

Chuan, Ching-Hua, and Chew [5] proposed a hybrid system for generating style-specific accompaniment from a given melody in 3 steps. The first step concerns the determination of chord tones through utilising Support Vector Machines (SVMs) and at a next step the system determines which notes in a given melody need to be harmonised. According to these notes, triads are assigned, first at checkpoints (the bars with all the possible chord solutions are available). The third step is the construction of possible chord progressions using neo-Riemanian transforms.

The utilisation of neural networks has also been examined. Hild, Feulner and Menzel [19] utilised three kinds of neural networks. The first one generated harmonic tree structures from a soprano melody, the second one allocates concrete notes from these skeletons, while the third one is used for ornamentation. Suzuki and Kitahara [47] developed two kinds of computational models, one that contains chord nodes (in the Bayesian networks) and another that does not. Both are capable of generating four-part harmonies using Bayesian networks. They wanted to investigate to what extent the model without chord nodes affects the harmonisation in terms of voice leading compared to the model with chord nodes.

5.1.2 Probabilistic bass voice leading

The proposed methodology aims to derive information from the melody voice in order to calculate the most probable movement for the bass voice, hereby referred to as the *bass voice leading* (BVL). This approach is intended to be harnessed to a larger modular probabilistic framework where the selection of chords (in GCT form [3]) is performed on an other probabilistic module [24]. Therefore, the development of the discussed BVL system is targeted towards providing indicative guidelines to the overall system about possible bass leading movement rather than defining specific notes for the bass voice. A very detailed specification of bass voice notes would result in considerable "over-specification" of harmony, providing information that would overlap with the ones provided by the chord selection module. Additionally, in order to avoid over-specification in the melody's voice, a similar approach is followed to capture melodic movement; descriptions of voice contour is considered rather than melodic notes.

The level of refinement for representing the bass and melody voice movement for the BVL system is also a matter of examination in the current part of the report. It is, however, a central hypothesis that both the bass and the melody voice steps are represented by abstract notions that describe pitch direction, i.e. whether the pitch difference between successive notes in the melody and bass is zero (steady), positive (moving up) or negative (moving down). Several scenarios are examined in Section 5.1.3 about the level of refinement required to have optimal results. Furthermore, non-common refinement scenarios are considered for melody and bass voice, examining whether either one of them should be represented with a more detailed refinement scheme than the other. Table 4 exhibits refinement scale of semitone differences and their level of detail considered for the voice movement. For example, by considering a refinement level 2 for describing the melody voice, the following set of seven descriptors for contour change are considered:

$$\texttt{mel}_2 = \{\texttt{st\_v}, \texttt{s\_up}, \texttt{s\_down}, \texttt{sl\_up}, \texttt{sl\_down}, \texttt{bl\_up}, \texttt{bl\_down},\}$$

while an example of refinement level 0 consideration for the bass voice the set of the following three descriptors are considered:

$$\texttt{bass}_0 = \{\texttt{st\_v}, \texttt{up}, \texttt{down}.\}$$

On the left side of the above equations, the subscript of the melody and the bass voice indicators denotes the level of refinement that is considered. Under this setup, the example chord sequence presented in Figure 9, is given in MIDI pitch numbers as:

$$[67,63,60,48][67,62,65,47],[63,60,65,48],[65,60,60,56],$$

with bass and melody (soprano) voice leading:

$$[-1,0][+1,-4],[+8,+2],$$

while the in terms of the utilised representation it becomes:

$$[\texttt{down}, \texttt{st\_v}\}, \{\texttt{up}, \texttt{bl\_down}\}, \{\texttt{up}, \texttt{sl\_up}].$$

The main assumption for developing the presented BVL methodology is that bass voice is not only a melody itself, but it also depends on the piece's melody. Therefore, the selection of the

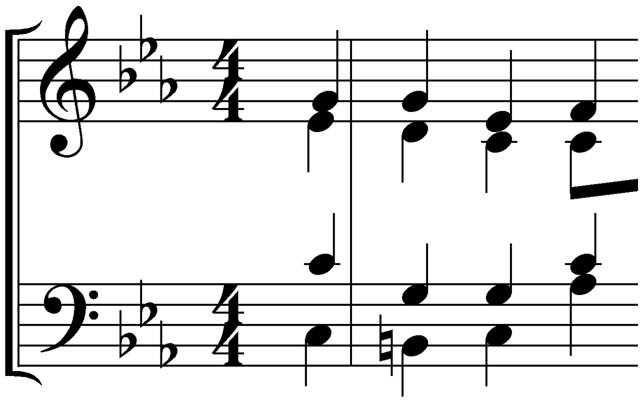


Figure 9: Example of a chord sequence from the Bach Chorale BWV 102.

refinement description	short name	semitone difference range	refinement level
steady voice	`st_v`	0	0, 1, 2
up	`up`	above 0	0
down	`down`	below 0	0
step up	`s_up`	between 1 and 2	1, 2
step down	`s_down`	between −2 and −1	1, 2
leap up	`l_up`	above 2	1
leap down	`l_down`	below −2	1
small leap up	`sl_up`	between 3 and 5	2
small leap down	`sl_down`	between −3 and −5	2
big leap up	`bl_up`	above 5	2
big leap down	`bl_down`	below −5	2

Table 4: The pitch direction refinement scales considered for the development of the proposed BVL system, according to the considered level of refinement.

next bass voice note is dependent both on its previous note(s), as well as on the current interval between the current and the previous notes of the melody. This assumption, based on the fact that a probabilistic framework is required for the harmonisation system, motivates the utilisation of the *hidden Markov model* (HMM) methodology. According to the HMM methodology, a sequence of observed elements is given and a sequence of (hidden) states is produced as output.

The training process of an HMM incorporates the extraction of statistics about the probabilities that a certain state (bass direction descriptor) follows an other state, given the current observation element (melody direction descriptor). These statistics are extracted from a training dataset, while the state sequence that is generated by an HMM system, is produced according to the maximum probability described by the training data statistics – considering a given sequence of observation elements. Concerning the training process, statistics are extracted from a training dataset that incorporate three aspects:

1. The probability for each state (relation between bass voice successive pitches) to be a beginning state. This distribution is computed by examining each state for each piece in the

training dataset and is denoted as $\pi(S_1 = s)$, $S \in \mathscr{S}$.

2. The probability for each state (relation between bass voice successive pitches) to be an ending state. This distribution is computed by examining each ending state for each piece in the dataset and is denoted as $\tau(S_T = s)$, $s \in \mathscr{S}$.
3. The probability that each state follows another state, denoted as $P(S_i = s_i | S_{i-1} = s_{i-1})$, s_i, $s_{i-1} \in \mathscr{S}$.
4. The probability of a state being present over an observation (relation between melody voice successive pitches), denoted as $P(S_i = s_i | Y_i = y_i)$.

A certain sequence of states ($S_i = s_i$, $i = 1, 2, \ldots, T$) is assigned an *overall* probability value, given a sequence of observations ($Y_i = y_i$, $i = 1, 2, \ldots, T$). This overall probability is computed by

$$P(S_i = s_i | Y_i = y_i) = P_\pi \, P_\mu \, P_\tau, \tag{8}$$

where

$$P_\pi = \pi(S_1 = s_1) \, P(S_1 = s_1 | Y_1 = y_1), \tag{9}$$

$$P_\mu = \prod_{i=2}^{T} P(S_i = s_i | S_{i-1} = s_{i-1}) \\ P(S_i = s_i | Y_i = y_i), \tag{10}$$

$$P_\tau = \tau(S_T = s_T) \, P(S_T = s_T | Y_T = y_T). \tag{11}$$

During generation the optimal path of states is pursued, i.e. the one that maximises the overall probability (in Equation 8) – or, equivalently, the path that minimises the negative log–likelihood of the expression in Equation 8. The optimisation of the overall probability is achieved by the (optimal) path of states that yields a maximal combination for the probabilities in all the counterparts (P_π, P_μ and P_τ). This path is typically computed through the Viterbi [11] algorithm.

5.1.3 Experimental results

Aim of the experimental process is to evaluate whether the presented approach composes bass voice leading sequences that capture the intended statistical features regarding BVL from different music idioms. Additionally, it is examined whether there is an optimal level of detail for grouping successive bass note differences in semitones (according to Table 4), regarding BVL generation. To this end, a collection of five datasets has been utilised for training and testing the capabilities of the proposed BVL-HMM, namely:

1. a set of the Bach Chorales,
2. several chorales from the 19th and 20th centuries,
3. polyphonic songs from Epirus,
4. a set of medieval pieces and
5. a set of modal chorales.

These pieces are included in a dataset composed by music pieces (over 400) from many diverse music idioms (seven idioms with sub-categories). This dataset is developed for the purposes of the COINVENT project. For the presented experimental results, each idiom set includes from around 50 to around 150 phrases, each including from 5 to 10 chords.

The Bach Chorales have been extensively utilised in automatic probabilistic melodic harmonisation [1, 21, 38, 34], while the polyphonic songs of Epirus [32, 25] constitute a dataset that has hardly been studies. The utilised datasets have diverse characteristics regarding BVL. The Bach Chorales include very strict rules regarding voice movement, while in medieval pieces the very distinct notion of parallel motion is extensively utilised. The polyphonic songs of Epirus on the other hand, present rather independent shapes of voices for the soprano and the bass, while sometimes the bass is a drone note. However, there are great diversities within some of the aforementioned idioms. For instance, in the chorales from the 19th and 20th centuries many sub-categories can be recognised since, each composer in this dataset utilises very distinctive harmonic tools, a fact that potentially makes training and testing with pieces in this category problematic.

According to the division of consecutive pitch differences presented in Table 4, several refinement level scenarios can be examined for the melody and the bass voices. The scenarios examined during the experimental evaluation process are exhibited in Table 5.

scenario	bass refinement	melody refinement	states × observations
1	1	1	5×5
2	1	2	5×7
3	0	2	3×7
4	0	1	3×5
5	0	0	3×3

Table 5: The examined scenarios concerning bass and melody voice refinement levels. According to Table 4, each refinement level is described a number of states (bass voice steps) and observations (melody voice steps).

Each idiom's dataset is divided in two subsets, a *training* and a *testing* subset, with a proportion of 90% to 10% of the entire idiom's dataset. The training subset is utilised to train a BVL-HMM according to the selected refinement scenario. A model trained with the sequences (bass movement transitions and melody movement observations) of a specific idiom, X, will hereby be symbolised as M_X. For instance, the BVL-HMM trained with the Bach Chorales will be symbolised as M_{Bach}. The testing subset includes the sequences that belong to an idiom but where not utilised for training the model. The testing subset of an idiom X will be hereby denoted as D_X (e.g. the testing pieces taken from the Bach Chorales idiom will be symbolised as D_{Bach}).

Towards evaluating the effectiveness and consistency of the BVL produced by the proposed model, the following question is set: can a model trained by the statistics of a certain idiom (M_X) "describe" the unseen pieces (during training) of this idiom (D_X) *better* than the pieces of any other idiom ($D_Y, Y \neq X$)? This evaluation does not directly aim to provide an evaluation about the quality of the BVL sequences it produces, however, this evaluation process provides indications about whether the examined methodology is able to capture the BVL characteristics of different idioms. An additional inquiry concerns the determination of the refinement level that allows the all trained models to be more effective towards predicting better the testing subsets of their idioms

– the idioms of the subsets they have been trained on.

The evaluation of whether a model M_X predicts a subset D_X better than a subset D_Y is achieved through the cross-entropy measure. The measure of cross-entropy is utilised to provide an entropy value for a sequence from a dataset, $\{S_i, i \in \{1,2,\ldots,n\}\} \in D_X$, according to the context of each sequence element, S_i, denoted as C_i, as evaluated by a model M_Y. The value of cross-entropy under this formalisation is given by

$$-\frac{1}{n}\sum_{1}^{n}\log P_{M_Y}(S_i, C_{i,M_Y}), \tag{12}$$

where $P_{M_Y}(S_i, C_{i,M_Y})$ is the probability value assigned for the respective sequence element and its context from the discussed model.

The magnitude of the cross entropy value for a sequence S taken from a testing set D_X does not reveal much about how well a model M_Y predicts this sequence – or how good is this model for generating sequences that are similar to S. However, by comparing the cross-entropy values of a sequence X as predicted by two models, D_X and D_Y, we can assume which model predicts S better: the model that produces the *smaller* cross entropy value [22]. Smaller cross entropy values indicate that the elements of the sequence S "move on a path" with greater probability values.

Tables 6 to 10 exhibit the cross-entropy values produced by the BVL-HMM models from the systems trained on each available training datasets for each test set's sequences. The presented values are averages across 100 repetitions of the experimental process, with different random divisions in training and testing subsets (preserving a ratio of 90%-10% respectively for all repetitions). The effectiveness of the proposed model is indicated by the fact that most of the minimum values per row are on the main diagonal of the matrices, i.e. where model M_X predicts D_X better than any other D_Y. Additionally, for refinement scenarios 3 and 4, all the diagonal elements are smaller, indicating that a greater refinement for the observed melody's voice than the bass's voice is better. However, an extended musicological interpretation of these results will appear in the final version of our work.

	M_{Bach}	$M_{\text{19th-20th}}$	M_{Epirus}	M_{Medieval}	M_{Modal}
D_{Bach}	3.1387	2.8638	4.3473	3.4858	3.1966
$D_{\text{19th-20th}}$	17.9604	**11.7289**	26.853	16.1352	11.5957
D_{Epirus}	4.4550	3.8403	**3.0906**	3.3136	4.3630
D_{Medieval}	3.6921	3.4480	3.1798	**2.8022**	3.2798
D_{Modal}	3.3813	3.2222	3.7265	3.0456	3.3611

Table 6: Mean values of cross-entropies for all pairs of datasets, according to the refinement scenario 1.

An example application of the proposed BVL system is exhibited in Figure 10, where GCT chords were produced by by the cHMM [23] system. The chordal content of the harmonisation is functionally correct and compatible with Bach's style. The proposed bass line exhibits only two stylistic inconsistencies, namely the two $^6_4$ chords in the first bar. The overall voice leading is correct, except for the parallel octaves (first two chords) and the omitted B (tenor voice) in bar 1.

	M_{Bach}	$M_{\text{19th-20th}}$	M_{Epirus}	M_{Medieval}	M_{Modal}
D_{Bach}	**3.0011**	3.2813	28.7568	16.7499	5.3278
$D_{\text{19th-20th}}$	32.3100	**19.0588**	86.5559	58.4145	30.0222
D_{Epirus}	4.3524	3.9292	**3.1315**	3.8120	4.0281
D_{Medieval}	3.8508	3.5692	3.5483	**3.2555**	3.5563
D_{Modal}	3.5218	3.3549	4.1010	3.4594	3.6161

Table 7: Mean values of cross-entropies for all pairs of datasets, according to the refinement scenario 2.

	M_{Bach}	$M_{\text{19th-20th}}$	M_{Epirus}	M_{Medieval}	M_{Modal}
D_{Bach}	**2.4779**	2.5881	31.0763	16.0368	5.3056
$D_{\text{19th-20th}}$	13.8988	**5.0687**	70.1652	31.6096	15.9747
D_{Epirus}	3.3127	3.1592	**2.8067**	2.9990	3.0378
D_{Medieval}	3.0988	3.0619	3.1845	**2.7684**	2.8539
D_{Modal}	3.0037	2.9028	3.3761	2.9611	**2.7629**

Table 8: Mean values of cross-entropies for all pairs of datasets, according to the refinement scenario 3.

	M_{Bach}	$M_{\text{19th-20th}}$	M_{Epirus}	M_{Medieval}	M_{Modal}
D_{Bach}	**2.4289**	2.4461	3.3213	2.6494	2.4342
$D_{\text{19th-20th}}$	6.4491	**4.3628**	19.0020	7.8220	4.8616
D_{Epirus}	3.1095	3.0163	**2.4523**	2.6928	2.7887
D_{Medieval}	2.9202	2.9144	2.8137	**2.4803**	2.6463
D_{Modal}	2.7970	2.7731	2.9298	2.6107	**2.5569**

Table 9: Mean values of cross-entropies for all pairs of datasets, according to the refinement scenario 4.

	M_{Bach}	$M_{\text{19th-20th}}$	M_{Epirus}	M_{Medieval}	M_{Modal}
D_{Bach}	**1.9500**	2.0307	2.7177	2.0100	1.8438
$D_{\text{19th-20th}}$	2.0896	2.0778	2.3980	2.0164	1.9258
D_{Epirus}	2.2805	2.2837	**2.0177**	2.0314	2.0204
D_{Medieval}	2.3279	2.4342	2.3303	**1.9142**	2.0306
D_{Modal}	2.1305	2.2438	2.3919	1.9469	**1.8999**

Table 10: Mean values of cross-entropies for all pairs of datasets, according to the refinement scenario 5.

5.1.4 Discussion

This part of the report presented a methodology for determining the bass voice leading (BVL) given a melody voice. This work is part of an alternative approach to automated melody harmonisation, which based on the fact that harmony is not solely expressed as a sequence of chords, but in combination with the harmonic movement of each voice that comprise the chords separately. Voice leading concerns the horizontal relations between notes of the harmonising chords, while a sets of rules determine the characteristics of voice leading. Different musical idioms encom-

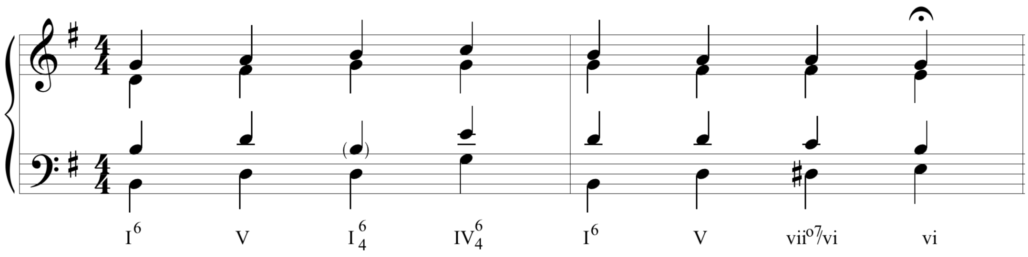


Figure 10: Bach chorale melodic phrase automatically harmonised, with BVL generated by the proposed system (harmonic analysis done manually).

pass different sets of rules, while the specific determination of these rules, especially for relatively unstudied music idioms, is a task that demands overwhelming research resources. To this end, probabilistic methodologies have been extensively utilised for approaching harmonisation rules from a statistical aspect.

The presented work focuses on generating the bass voice directly from a soprano voice; the bass voice is an important part for generating the complete voice leading for all voices in a given chord sequences. This approach is a part of research within the context of the COINVENT project . Purpose of this project is to construct a computationally feasible model for generating novel concepts by blending two given input conceptual spaces. Aim of this study is to generate a method which focuses on the harmonic concept of the bass voice leading. The proposed bass voice leading (BVL) probabilistic model is utilises a hidden Markov model (HMM) to determine the most probable movement for the bass voice (hidden states), by observing the soprano movement (set of observations). Many variations regarding the representation of bass and soprano voice movement have been examined, discussing different levels of representation refinement expressed as different combinations for the number of visible and hidden states. Five diverse music idioms were trained creating the relevant BVLs, while parts of these idioms were used for testing every system separately.

The results indicated low values in term of cross entropy for each trained BVL system with the corresponding testing datasets and high values for examples from different music idioms. Thereby, it is assumed that the proposed methodology is efficient, since some characteristics of voice leading are captured for each idiom. However, not all the combinations regarding the number of hidden and visible states, gave the desired results. In many occasions lowest cross-entropy was observed for musical examples from different music idioms rather the corresponding in which the BVL system was trained for. This abnormal in the results, however, can be investigated as motivation for further research from the musicologists, as well as they can be also used for direct harmonic conceptual blending. The unique harmonic aspects of different music idioms may prove a challenge for automated chord harmonisation using multiple systems providing the bass voice leading generation, as well as relevant systems for the other voices, trained by different musical idioms with common musical characteristics.

For future work, a thorougher musicological examination of the pieces included in the dataset will be pursued, since great difference were observed for the voice leading of pieces in included in the same idiom. Additionally, our aim is the development of the overall harmonisation probabilistic system, where chord selection (based on a separate HMM module) will be also biased by the adequacy of each chord to fulfil the voice leading scenario provided by the voice leading probabilistic module – part of which is presented in this work.

5.2 Chord inversions and note doublings

In Section 5 sequences of GCT chords were produced, encompassing information only about the pitch classes that are present within these chord sequences. The assignment of actual notes requires additional information about the relative pitch height of chords according to the pitch height of the melody's notes (as discussed later in this section). Additional information is required to translate a set of pitch classes (described by GCT chords) to actual notes, namely the chord's *inversion* and its note *doublings*.

All the *inversions* of a chord described by a set of pitch classes are obtained by assigning each of these pitch classes as a bass note. For instance, the chord with pitch classes $[0,4,7]$ has three inversions, with each one having a bass note that corresponds to a different pitch class. In MIDI pitches, these inversions could be $[60,64,67]$, $[64,67,72]$ or $[67,72,76]$, while the pitch height of the bass note in each chord inversion version (60, 64 or 76 respectively) is defined by the bass voice motion module and the distribution of differences between melody and bass note. Note *doublings* concern the notes that are doubled when a pitch class chord is converted to actual notes. In the previously examined example, by considering the inversion prototype $[60,64,67]$ of the $[0,4,7]$ chord, there are four scenarios of single note doublings: $[60,64,67,72]$, $[60,64,67,76]$, $[60,64,67,79]$ and $[60,64,67]$. The first three scenarios concern the respective doubling of each note in the inversion prototype, while the fourth is the no-doubling scenario. It should be noted that it is not impossible for more than one notes to be doubled in a chord expression, however, the studied methodologies are restricted to single note doublings since more than single note doublings are rare.

The voice leading module of the harmonic learning system regarding chord inversions and note doublings, is trained through extracting relevant information from pieces in a training dataset that corresponds to a music idiom. To this end, statistics are extracted regarding the pitch classes of the bass notes and the notes that are doubled for each encountered chord, while the extracted statistics are stored for each GCT chord. Specifically, consider a GCT chord in the form

$$g = [r,\ [\vec{t}],\ [\vec{x}]],$$

where $[\vec{t}]$ and $[\vec{x}]$ are the vectors describing the base and the extensions of the chord. This GCT chord is mapped to a pitch class (PC) chord of the form

$$c = [c_1,\ c_2,\ \dots,\ c_n],$$

where n is the number of pitch classes (c_i, $i \in \{1,2,\dots,n\}$) that describe chord c. The statistics concerning chord inversion are expressed as the probability that each pitch class in c is the bass note of the chord, or

$$p_{\mathrm{i}} = (v_1,\ v_2,\ \dots,\ v_n),$$

where v_i, $i \in \{1,2,\dots,n\}$, is the probability that the pitch class c_i is the bass note. Similarly, probabilities about note doublings are expressed through a probability vector

$$p_{\mathrm{d}} = (d_1,\ d_2,\ \dots,\ d_n,\ s),$$

where d_i, $i \in \{1, 2, \ldots, n\}$, is the probability that the pitch class c_i gets doubled, while there is an additional value, s, that describes the probability that there is no doubling of pitch classes. Table 11 exhibits the extracted statistic for inversions and note doublings for the most often met chords of the major Bach Chorales.

GCT chord	relative PC	inversions	doublings
$[0,[0,4,7]]$	$[0,4,7]$	[0.74, 0.23, 0.02]	[0.68, 0.15, 0.08, 0.09]
$[7,[0,4,7]]$	$[7,11,2]$	[0.78, 0.22, 0.00]	[0.83, 0.02, 0.09, 0.06]
$[5,[0,4,7]]$	$[5,9,0]$	[0.65, 0.34, 0.01]	[0.46, 0.30, 0.11, 0.13]

Table 11: Probabilities for chord inversion (p_i) and note doublings (p_d) in the three most frequently used chords in the major Chorales of Bach.

The probability that a certain GCT chord will be found in a certain state of inversions and doublings is given by multiplying the probability values extracted from the dataset for these states. For example, the probability that the $C = [0,[0,4,7]]$ GCT will be in its second inversion (third as bass) and with the third note doubled is computed as $0.23\ 0.15 = 0.03$. Similarly, the probabilities from the pitch height distribution for the bass note of the chord in relation to the melody note (given by the fitted normal distribution and denoted as $p_h(C)$), along with the probabilities for the bass motion from the previous bass note to the current ($p_m(C)$), contribute to the overall selection of the voicing layout again though multiplication, i.e. the probability value of a chord voicing layout, $l_x(C)$, is given by:

$$l_x(C) = p_{i_x}(C)\ p_{d_x}(C)\ p_{h_x}(C)\ p_{m_x}(C). \tag{13}$$

Therefore, the voicing layout (x_{best}) that is best suited for chord C is found by

$$x_{\text{best}} = \arg\max_x (p_{i_x}(C)\ p_{d_x}(C)\ p_{h_x}(C)\ p_{m_x}(C)). \tag{14}$$

The bass note motion probability is obtained by the BVL module analysed in Section 5.1 and it takes the value 1 if the voicing layout “agrees” with the selected bass motion, and a “penalty” value (near 0) if it “disagrees”.

6 Higher-level Harmonic Structure

The cHMM methodology described in Section 4 has been developed to tackle a shortcoming of the typical HMM model that concerns the lack of identifying harmonic structure on a level higher than chord-to-chord transitions. To this end, the cHMM methodology allows the incorporation of intermediate chord constraints on phrase boundaries, therefore enabling the insertion of intermediate cadential chords that indicate intermediate phrase endings. Since the user input to the system includes information on the phrase structure of the input melody, the intermediate chords are selected to be placed as the final chords in every phrase, defining the cadence end-boundaries of each phrase.

6.1 Cadences, Intermediate Phrase Endings and phrase connection

Musical phrases are hierarchically layered with smaller phrases existing in larger ones. What makes phrases within other phrases discretely perceived is the existence of intermediate cadences that express at some extent a sense of harmonic closure. This section described the statical process followed to capture the characteristics of phrase endings, with a discrimination being made on the type of cadences according to their placement within the piece. Thereby, three types of cadences are assumed:

1. Final cadences: cadences that end a musical piece.

2. Intermediate cadences to same-tonality phrases: ending chords of phrases that lead to phrases with the same tonality.

3. Intermediate cadences to different-tonality phrases: ending chords of phrases that lead to phrases with the different tonality.

Final cadences are important since they signify the end of the piece, reflecting the most prominent characteristics of closure for the idiom that this piece belongs to. The dissociation of intermediate cadences that lead to phrases with same and different tonality is based on the assumption that cadences connecting phrases with different tonalities might incorporate notes that reflect characteristics from both tonalities. Thereby, statistics concerning cadences that lead to different tonalities should be gathered with respect to the incorporated tonalities, e.g. statistics concerning cadences between modulations from one major key to a minor key a fifth up should be accounted as something different from modulations to a minor key a fourth up.

Even though there could be some point in gathering statistics for "all" modulations within an idiom, there are two important reasons why not to do so:

1. *Number of different modulations*: Many idioms in the dataset hardly include any modulations (e.g. modal chorales, polyphonic songs of Epirus), while others include a small number of different modulations (e.g. there specific types of modulations in the Bach Chorales). Therefore, gathering statistics for these modulations would not allow for significant results that inform us about "general" modulation scenarios – since the aim of extracting statistics in the context of the melodic harmoniser is its ability to "generalise" about a given/requested melodic harmonisation scenario.

2. *Melody-relatedness*: The melodic harmoniser produces harmonies in the style of a given idiom for user-given melodies. However, the melody provided by the user might be incorporating tonality modulations that are hardly met within the harmonising idiom and, therefore, constituting modulation-related information irrelevant.

Cadences might be formed of one chord, or incorporate cadential patterns comprised of several chords. The approach followed within the context of the melodic harmonisation system does not require the allocation and utilisation of exact cadential patterns. For the system under development, the utilisation of cadences concerns solely the demarcation of "structural breaks" within the provided melody by providing constraints to the cHMM methodology, a fact that does not require

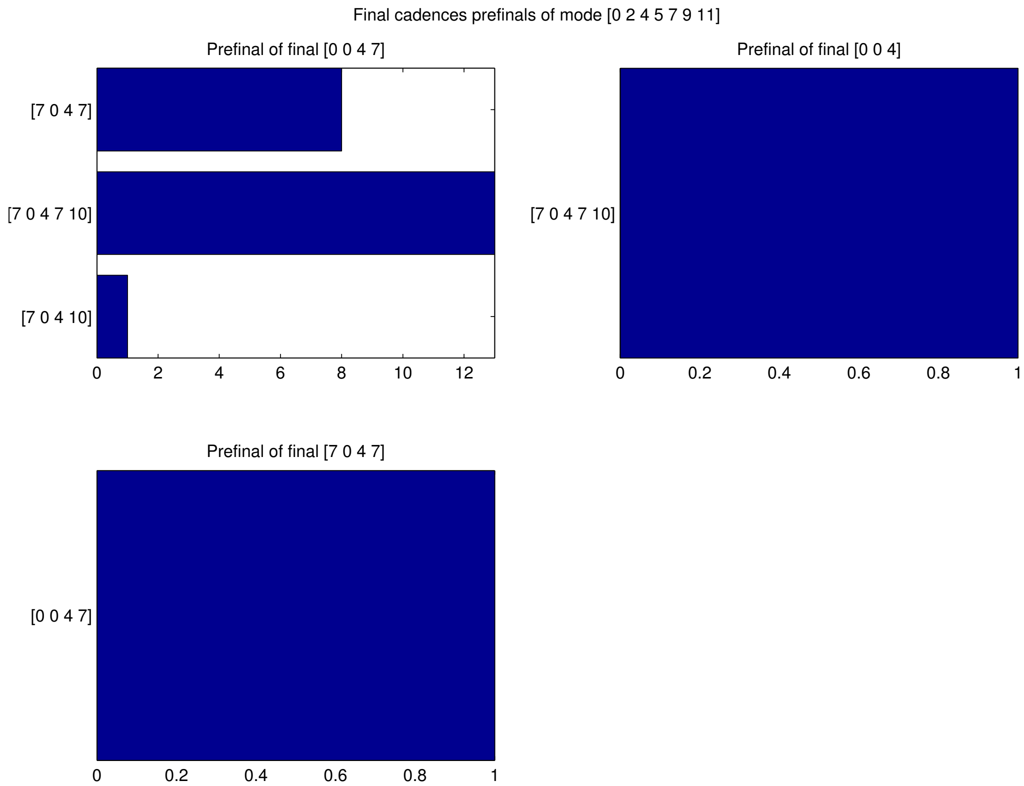


Figure 11: Pre-final chords for each piece-ending chord in the major Bach Chorales.

accurate identification and application of cadential patterns. Thereby, the study of cadences is restricted to measuring statistics concerning the appearances chord pairs in the phrase endings (final cadences, as well as intermediate cadences connecting phrases of same and different tonalities, as described above).

To evaluate the quality of information provided by the cadence statistics, the results gathered from the Bach Chorales were compared to the theoretic aspects of cadences that are known for this idiom. Figures 11 and 12 depict the final and pre-final chord pairs for the ending cadences, where the V-I relation is evidently reflected at piece endings in both modes (major and minor), while in minor mode the picardy third is often used. The fact that different GCTs might express same scale degrees (e.g. $[7,[0,4,7]]$ and $[7,[0,4,7,10]]$ refer to V) will be tackled – in a future work – by a GCT grouping approach shortly described in Appendix B, while further clustering through functional similarity (as shortly discussed in Appendix A) can lead to further clarification of the harmonic relations between different GCTs.

Figures 13 and 14 illustrates the distributions of final and pre-final chord pairs for intermediate phrase cadences that lead to same tonality phrases. Even for intermediate phrases of the major mode, the V-I relation is the most usual, while there are also many instances of half cadences (ending to the V chord) and plagal cadences (IV-I). Through this statistical approach it is revealed that the half cadences most often occur with either the tonic (I) pre-final GCT chord or the secondary dominants of V (V/V), expressed by the $[2,[0,4,7]]$ or the diminished $[6,[0,3,6]]$ GCTs. Additionally, there are many cadences ending in the relative minor scale degree (with either $[9,[0,3,7]]$

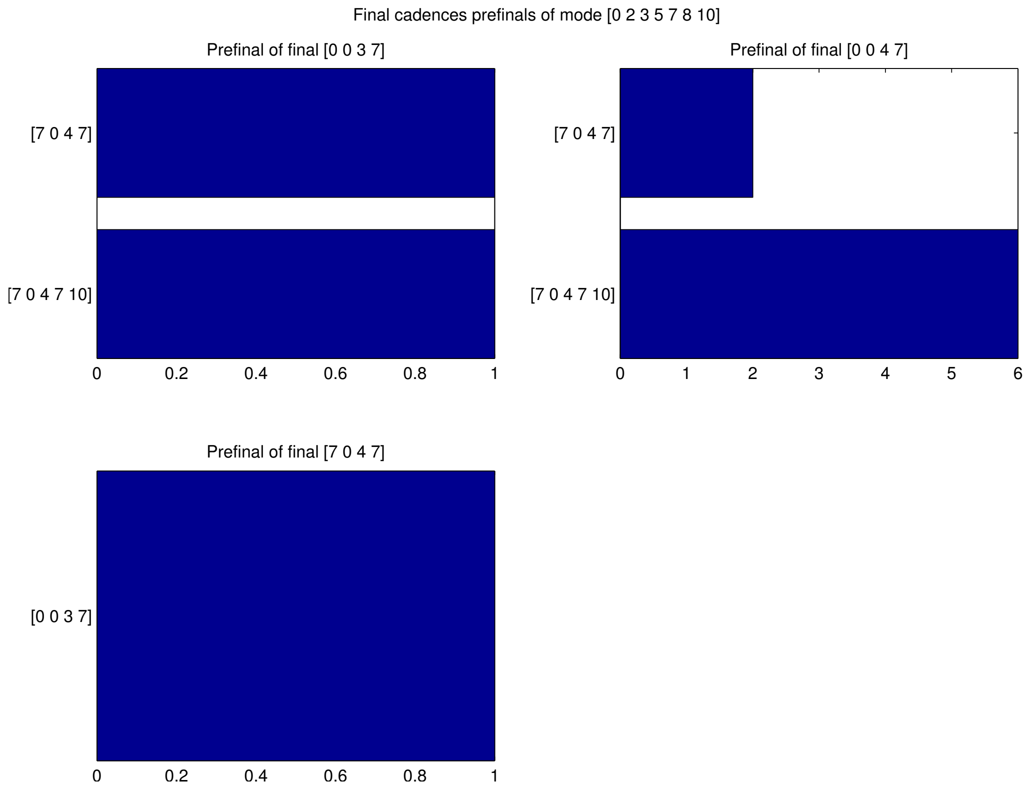


Figure 12: Pre-final chords for each piece-ending chord in the minor Bach Chorales.

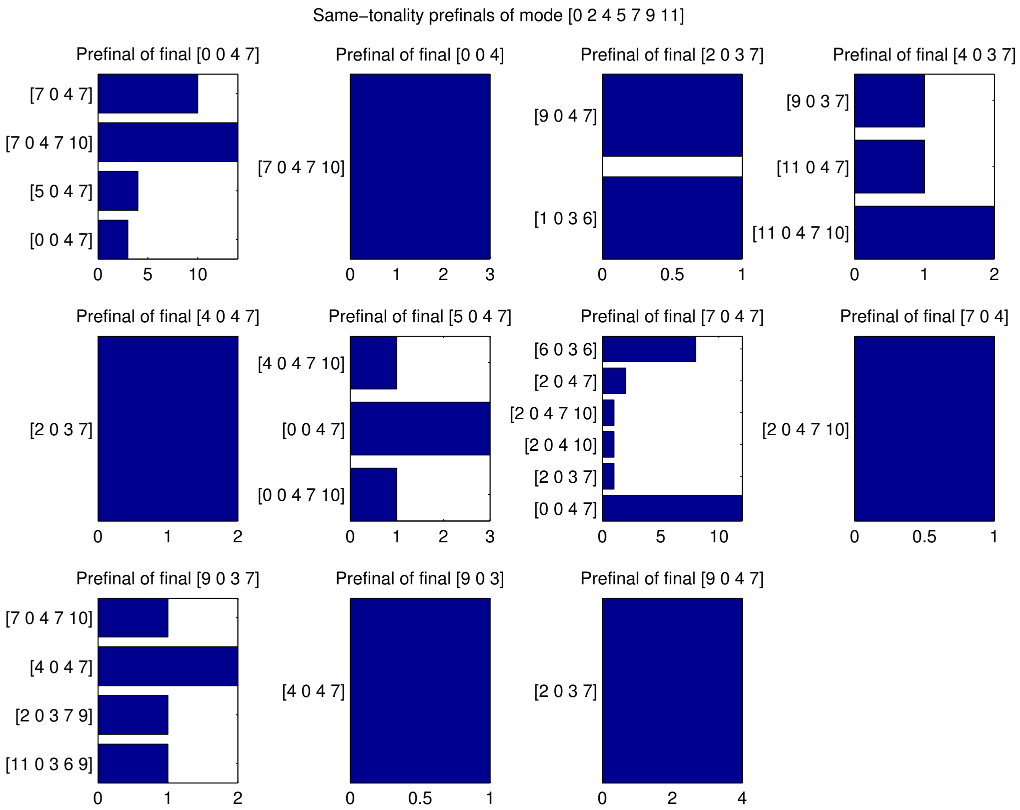


Figure 13: Pre-final chord for each phrase ending chord, with next phrases in the same tonality, for the major Bach Chorales.

or $[9,[0,4,7]]$), a fact that is also predicted by theoretic studies on this idiom. For the minor mode pieces, the perfect cadence is the most often met, while half cadences frequently occur with a minor IV as a pre-final chord – a "popular strategy" for this idiom.

Figures 15 and 16 depicts the distributions of final and pre-final chord pairs for intermediate phrase cadences that lead to different tonality phrases. What could be assumed by these figures is the fact that they have many similarities with the ones in concerning the intermediate cadences to same tonality phrases. This fact reveals the (musically theoretically grounded) fact that tonality modulations in this idiom occur steeply, without hormonic preparation.

The statistical study revealed some harmonic characteristics of the studied idiom that are also musically theoretically grounded, describing the cadences satisfactorily in the examined positions within pieces. Therefore, these statistics can be utilised in a generative fashion, allowing the insertion of intermediate chord constraints – with a roulette process according to the appearance probability of cadence chord pairs – that demarcate harmonic phrase endings in the positions of the melody that the user has indicated.

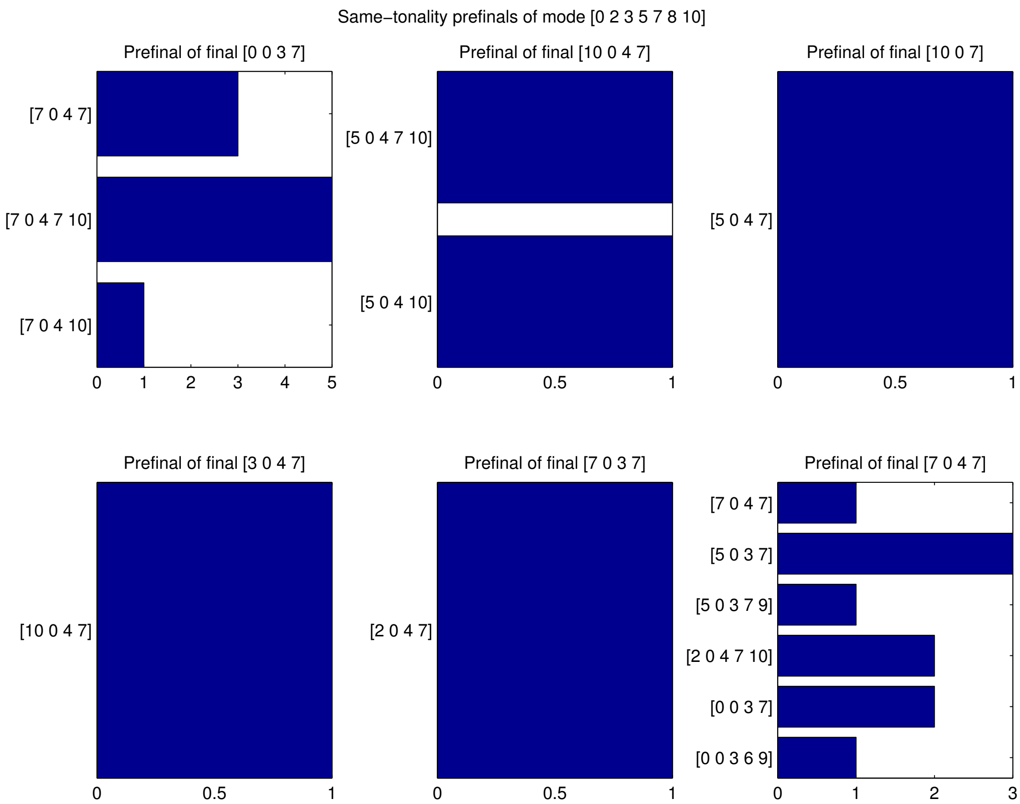


Figure 14: Pre-final chord for each phrase ending chord, with next phrases in the same tonality, for the minor Bach Chorales.

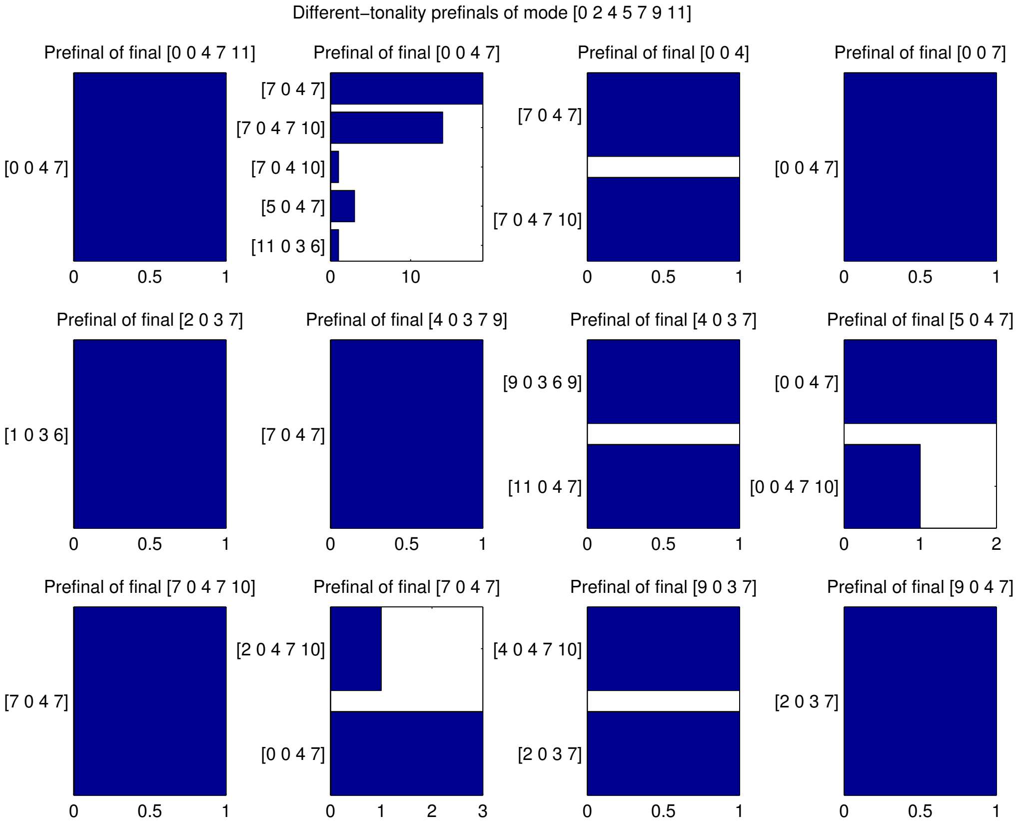


Figure 15: Pre-final chord for each phrase ending chord, with next phrases in different tonality, for the major Bach Chorales.

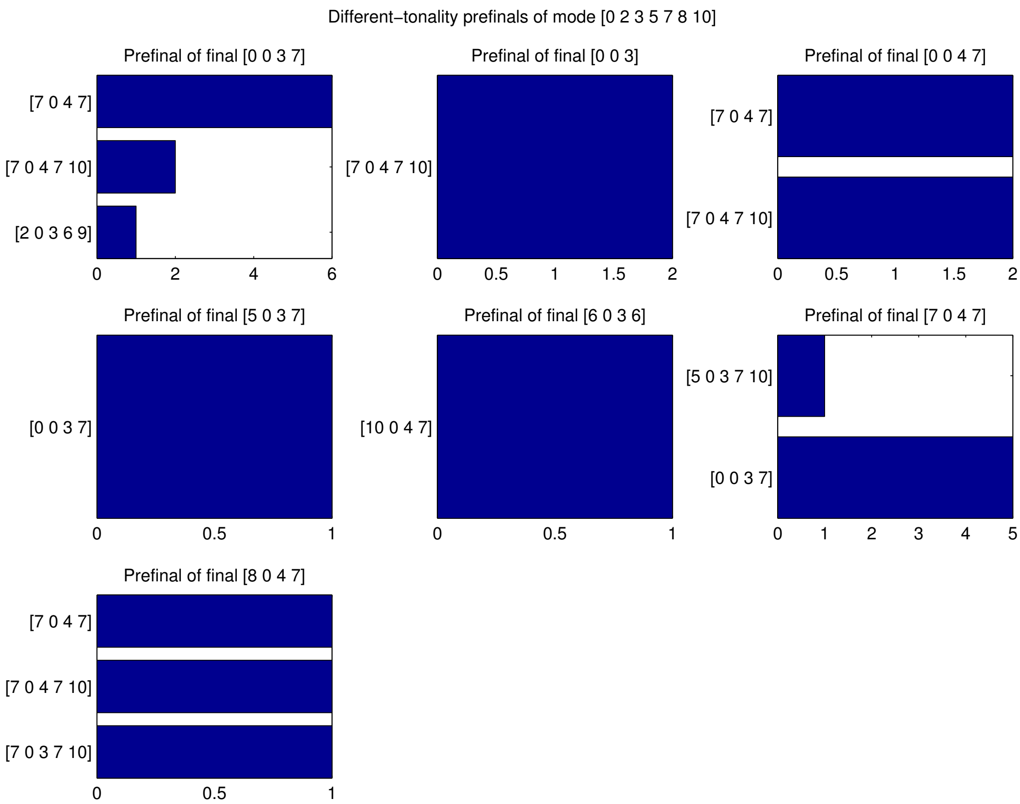


Figure 16: Pre-final chord for each phrase ending chord, with next phrases in the different tonality, for the minor Bach Chorales.

6.2 Tonality Modulations

Besides the melody, several accompanying features are imported to the system, which describe attributes of the input melody. Among these features is the tonality. Although speculations about the local tonality of a melody could be automatically deduced algorithmically [4, 27], manual annotation of tonality changes has been decided for the following reasons:

- Utilisation of non-standard (major/minor) tonalities: The collected dataset include pieces that do not conform to the standard Western Music tonalities, e.g. there are pentatonic or octatonic modes. Additionally, the user is allowed to insert any desirable tonality, which will lead the system to select the proper set of chords to harmonise the given melody.
- Accuracy in tonality-change boundaries: Algorithms that perform melodic segmentation according to tonality are not able to identify the exact location of tonality boundaries. For the COINVENT melodic harmoniser, it is important that the tonality (and phrase level) change locations stay perfectly aligned with the melody segments that a human user perceives.
- The ability to insert "subtle" tonalities: The user is able to introduce short segments of tonality changes in places where an algorithm would "disagree". This ability introduces additional agility and interestingness potential to the system.

Tonality changes are treated differently in different idioms, while, additionally, some idioms do not include, or include very specific tonality modulations between certain – neighbouring in the circle-of-fifths – tonalities. Since tonality modulations are dependent on the melody, and a user input melody might incorporate arbitrary tonality modulations, it is clear that no learning strategy on every idiom could cover the entire spectrum of tonality modulations required by input melodies. For instance, the in the idiom of modal music there are no tonality modulations, since the entire pieces are composed in a certain mode. Therefore, it would be impossible to harmonise a melody that incorporates tonality modulations using the harmony of a modal idiom, since no training paradigms would be available for such a task. For the purposes of the "idiom independent learning" that is required for the presented system, tackling tonality modulations should be done outside the learning framework. However, it is presumed that there are some relations between the chords that are close to tonality modulations. Revealing these relations would enable the development of mechanisms for generating proper chord progressions to harmonise melodies with arbitrary tonality modulations.

For obtaining insights about tonality modulations, statistics where collected on the chords that take part in modulations in the Chorales of Bach. The chords that are considered to take part in a modulation, hereby named also as *"modulation chords"*, are the *"pivot"* (p_M) and the *"modulation"* (M) chords. The pivot chord is the final chord in the first/previous tonality (k_1) and the modulation chord is the first chord of the second/next tonality (k_2). Aim of the study presented in this section is to examine whether either of these chords (p_M and M) are diatonic (all their notes are diatonic) in either keys of the keys (k_1 and k_2) in the modulation. Additionally, it is also examined "how" diatonic these chords are in both tonalities, by measuring the number of notes within these chords that are diatonic to each tonality in the modulation.

To assess statistics regarding relation between modulation chords and their diatonic relation to the modulating tonalities, three sets of chords and pairs of chords are utilised:

1. diat_k: A chord $C \in \text{diat}_k$ if every pitch class of C is diatonic in tonality k. Additionally, a chord $C \in \text{diat}_{k_1 \cap k_2}$ if it is diatonic in both k_1 and k_2 tonalities.

2. comm^n: A pair of two chords C_1 and $C_2 \in \text{comm}^n$ if C_1 and C_2 have exactly n common pitch classes.

3. $\text{diat}^n_{(k_1 \cap k_2)}$: A chord $C \in \text{diat}^n_{(k_1 \cap k_2)}$ if exactly n pitch classes of C are diatonic in both k_1 and k_2 tonalities. (This means that tonalities k_1 and k_2 have at least n common pitch classes.)

Table 12 examines which modulation chords are completely diatonic to each tonality that takes part in a modulation. As expected, the pivot and modulation chords are mostly completely diatonic in their "natural" tonalities (p_M in k_1 and M in k_2 as indicated by the bold percentages), while it is interesting that great percentages (around 80%) of both modulation chords are completely diatonic to both tonalities. Table 13 indicates that in most cases (about 50%) the modulation chords have exactly one common pitch class between them, while in most cases (about 90%) they have at least one common pitch class. Finally, the statistics presented in Table 14 indicate that all modulation chords have at least one pitch class that is diatonic to both modulation tonalities. The cadence schemes to different tonalities illustrated in Figures 15 and 16 indicate that in the Bach Chorales idiom the phrase closure in phrases that lead to tonality modulations is similar to the closure in phrases that lead to same tonalities. However, the statistics presented in this section reveal that the modulation chords are deeply related not one to the other, but also in accordance to the modulating tonalities. This is explained by the fact that the modulating tonalities in the Chorales of Bach are close to the circle of fifths, a fact that constitutes transitions between tonalities "smooth" even if the neighbouring different-tonality phrases are rather independent (i.e. the cadence in the first phrase is not dependent on the second phrase's first chord).

Table 12: Percentages of whether modulation chord are entirely diatonic in the first, second, as well as in both keys.

	$\in \text{diat}_{k_1}$	$\in \text{diat}_{k_2}$	$\in \text{diat}_{k_1 \cap k_2}$
p_M	**0.97**	0.85	0.84
M	0.79	**0.93**	0.79

Table 13: Utilisation of one, two, three, or more than one common pitch classes between modulation chords.

$\in \text{comm}^1$	$\in \text{comm}^2$	$\in \text{comm}^3$	$\in \text{comm}^{\geqslant 1}$
0.54	0.17	0.19	**0.91**

Table 14: Statistics for the number of pitch classes that diatonic in both keys, for pivot and modulation chords.

	$\in \text{diat}^1_{(k_1 \cap k_2)}$	$\in \text{diat}^2_{(k_1 \cap k_2)}$	$\in \text{diat}^3_{(k_1 \cap k_2)}$	$\in \text{diat}^{\geqslant 1}_{(k_1 \cap k_2)}$
p_M	0.00	0.21	0.76	1.00
M	0.00	0.14	0.84	1.00

While the above mentioned statistical relations have little to offer in the context of a trainable generative system, the simple yet statistically grounded conclusions they provided led to the potential of utilising "*conceptual blending*" to resolve tonality modulations between any modulating keys in any idiom. The mechanism that will be utilised pertains to the *chord blending* [10] methodology which has been developed in the context of COINVENT. According to chord blending, attributes of two input chords are combined to create novel ones that inherit characteristics from both input ones. In a specific setting, chord blending facilitates merging two chord sequences with an overlapping segment of chords into a single chord sequence. For instance, by considering two chord sequences, $C_1^a \to \cdots \to C_n^a$ and $C_1^b \to \cdots \to C_m^b$, where the final chord of the first sequence (C_n^a) overlaps with the first chord of the second sequence (C_1^b), a novel chord sequence emerges by blending C_n^a and C_1^b into C^X: $C_1^b \to \cdots \to C^X \to \cdots \to C_m^b$ (see Figure 17). Thereby, the chord yielded by blending plays the role of the pivot chord since it connects the two different modulation parts.

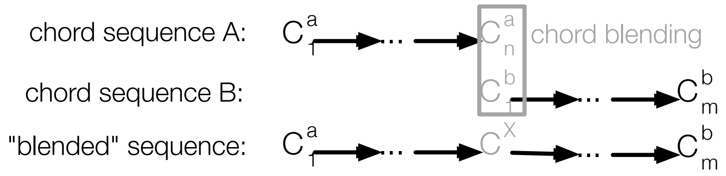


Figure 17: The input chord sequences are pre-composed and incorporate miscellaneous parts, potentially, of different keys.

7 Conclusions and future perspectives

This report presented the research conducted in the context of the COINVENT project, regarding the development of the harmonic learning methodology that will be encompassed in the COINVENT melodic harmoniser. Aim of the "end-product" melodic harmoniser is to facilitate conceptual blending on many levels of harmony, blending harmonic characteristics from many different and diverse idioms. This task is difficult because an accurate description of harmonic characteristics is required, with the additional demand that these descriptions can be utilised for generative practice, i.e. these descriptions/rules should enable the automatic generation of music. As studies on melodic harmonisation have exhibited, only in deeply studied and relatively simple idioms is it possible to utilise "deterministic" rules for rule-based melodic harmonisation. However, the COINVENT melodic harmoniser should be able to compose harmonies form hardly studied and complicated idioms, that do not necessarily comply with the Western Music harmonic dictionary. Furthermore, for the harmonic representation on the most fundamental level, i.e. chords, the General Chord Type (GCT) representation has been developed, which allows the representation of note simultaneities – even simultaneities that cannot be described by the often utilised chord symbolisms – in a chord-like manner.

As a consequence, harmonic learning through statistical methodologies is a reasonable strategy to acquire harmonic characteristics on many levels of harmonies, even for idioms that clearly diverge from the Western Music culture. To this end, harmonic learning is focused on:

- Chord sequence learning: The "constrained hidden Markov model" (cHMM) was developed which extends the typical HMM methodology, but under chord constraints that allow the insertion of chords at specific points in the harmonisation. Thereby, not only the interestingness of the system increases, but also the insertion of chord constraints reflecting higher-level harmonic structure is allowed (e.g. intermediate phrase cadences).

- Learning cadences: The intermediate cadences of phrases, as well as the final cadences of a piece are learned and employed to the locations indicated by the user to the harmonised melody. These cadences act as chord constraints to the cHHM.

- Voice leading of the bass voice: The bass voice leading (BVL) is learned utilising a HMM, by conditioning on the motion of the melody and the previous motion of the bass.

- Voice layout characteristics: Voice layout is learned for each GCT that is met within a dataset. The characteristics that are measured statistically concern the inversion (bass voice) and the note doublings that each chord (and its corresponding GCT) is found.

The probabilities that describe these harmonic aspects are combined to provide a chord progressions, cadences, voice leading and voicing layouts that are highly probable within idioms, allowing the accurate reflection of each idiom's characteristics, according to the currently ongoing pilot studies.

Future work incorporates improvements to the learning/generating module, as well as the utilisation of the blending methodology to the trained parts. Improvements on the one hand concern the consideration of additional harmonic characteristics in the learning/generating process. Such characteristics are be the (statistical) rejection of specific intervals for the BVL, the establishment of metrics regarding drone notes in the bass and middle voices, special voicing configuration on cadences, as well as the determination of distributions for the placement of the inner voices. On the other hand, improvements in the generative part regard the generation of the GCT sequences using the cHMM. In this part of the algorithm the voice leading potential of each selected GCT is not taken under concern, a fact that often leaves the voice leading and voicing layout with little probable choices. To this end, the cHMM methodology could be improved by considering additional probability values that relate to voice leading fitness, for the eligibility of each GCT.

Blending in the context of the melodic harmoniser regards the combination of different harmonic parts, trained on different music idioms. This step, however, requires the advancement of several other methodologies that will allow glueing together harmonic parts that are "inhomogeneous". For instance, as earlier analysed, the utilisation of chord blending allows the combination of entire harmonic parts that pertain to different tonalities or idioms, while this mechanism could also be utilised for inventing novel cadences [10]. Developing proper chord similarity metric (see Appendices A and B), will enable the insertion of chord constraints from alien idioms (e.g. cadences) in the currently utilised cHMM chord generation methodology. Additionally, since voicing layout characteristics are attributed per GCT, similar GCTs among different idioms would potentially "exchange" these characteristics, creating blends on the voicing layout level.

Appendix A: Chord Similarity

For harmonic blending between two different musical idioms (e.g. idiom A and idiom B) the definition of similarity between chords is quite valuable. The way by which harmonic concepts belonging to idiom B are incorporated into idiom A should ideally be governed by principles related to the desired artistic outcome. One such potential principle could be the amount of surprise caused to the listener by the hybrid harmonic progression as a result of introducing unexpected harmonic concepts to an already established harmonic idiom. A harmonic blend with minimal introduction of surprise should include chords that violate expectations of the established idiom as less as possible. This would presuppose the ability of the algorithm to group chords originating from different idioms using a metric of their similarity or dissimilarity as the principal criterion. Thus, surprise would be maximised by replacing a given idiom-A-chord by a totally dissimilar idiom-B-chord, while surprise minimisation would require the replacement of a given idiom-A-chord by its closest counterpart from idiom B. Chord grouping on the basis of similarity is also desired in the intra-idiom level where variations of certain chord categories could be organised in wider chord groups that in turn could be represented by a delegate chord.

However, the evaluation of chord similarity -especially between different harmonic spaces- is not trivial. According to the literature, the overall chord similarity is divided between cognitive similarity and sensory similarity. Psychologists and musicologists have developed cognitive theories that try to model distances between chords within harmonic context, some of which were also backed by a number of empirical experiments. Some notable examples of chord classification systems both in tonal and in atonal context are the work by [20], the classification scheme by [17], pitch-class set (pcset) theory [13, 14], neo-riemannian theory [6, 7], the classic work by [28], tonal pitch space theory by [31] and the work by [41]. Sensory similarity on the other hand, deals with chords in isolation as separate musical entities. [43] has compared the classification schemes by [13, 14], [17], [41] and [20] with empirical data from pairwise dissimilarity listening experiments and found that none of the theories could adequately predict the perceptual distances between chords. Listeners seemed to base their judgments on psychoacoustic factors like spacing type, size of the outermost interval etc. A later study by [29] has largely confirmed Samplaski's findings. From the above, it seems that a hybrid approach -combining both cognitive and sensory factors- is required to adequately capture chord similarity.

This work will additionally utilise information derived from statistical harmonic analysis of various musical idioms in order to achieve chord grouping based on chords' functionality (e.g. frequency of appearance, probabilities of transitions between chords) within each separate idiom. Thus, both intra-idiom and inter-idiom chord similarity and grouping can be achieved on the basis of statistical harmonic analysis. The results of the statistical approach in combination with the cognitive and sensory approaches will enable the formulation of idiom related harmonic structure profiles and facilitate the examination of chord relations between idioms.

Appendix B: GCT Chord Grouping

The GCT algorithm allows the effective determination of a root in a simultaneity of notes, as well as of a base that reflects the basic characteristics. According to the GCT representation further abstraction can be achieved through grouping GCT expressions of simultaneities that "evidently" concern the same chord. For instance, in the major Bach Chorales the following GCTs are met:

$$[0,[0,4,7]],\ [0,[0,4]],\ [0,[0,4,7,11]],,$$

which "evidently" refer to the major chord of the scale's first degree. The fact that all these versions of this chord appear, primarily concerns voice leading aspects of the compositions, where the convergence and divergence of the four incorporated voices lead to the reduced ($[0,[0,4]]$) or the expanded ($[0,[0,4,7,11]]$) version of the "exemplar" I major chord ($[0,[0,4,7]]$).

Further grouping of GCTs has been studied under some basic assumptions about the chord characteristics that are reflected by the root scale degree, the base and the diatonic notes of a GCT expression. Specifically, GCT expressions are grouped into more general GCT categories that contain potentially several GCT members according to the criteria described below: two chords belong to the same group if

- they have the same scale degree root,
- their GCT bases are subset-related and
- they both contain notes that either belong or do not to the given scale context.

Regarding criterion 2, two bases B_1 and B_2 are considered subset-related if $B_1 \subseteq B_2$ or $B_2 \subseteq B_1$, e.g. $[0,4] \subseteq [0,4,7]$ while $[0,4] \not\subseteq [0,3,7]$. Criterion 3 is utilised to identify and group together chords that belong to secondary tonalities within the primary tonality of the piece. For instance, while $c_1 = [0,[0,4,7]]$ and $c_2 = [0,[0,4,7,10]]$ fulfil criteria 1 and 2, according to criterion 3 they are not grouped together since c2 includes the base value 10, which is mapped to the non-diatonic 10 pitch class value (footnote: The scale in the major Bach Chorales is described by the pitch class set: $[0,2,4,5,7,9,11]$.) Thereby, the secondary dominant to the IV (V/IV) is differentiated from the I major chord.

Each GCT group includes the GCTs types that satisfy the aforementioned three criteria. Furthermore, each group is represented by the "exemplar" GCT type, which is the one that is more often met in the datasets under study. Some of the most often met groups in the major scale Bach Chorales are demonstrated in Table 15. Furthermore, this table also includes the functional naming of each group for straightforward comparison of the derived GCT types and the know theoretic music analysis framework.

functional name	exemplar	Group members			
tonic	$[0,[0,4,7]]$	$[0,[0,4,7]]$	$[0,[0,4]]$	$[0,[0,4,7],[11]]$	
dominant	$[7,[0,4,7]]$	$[7,[0,4,7]]$	$[7,[0,4,7],[10]]$	$[7,[0,4],[10]]$	$[7,[0,4]]$
subdominant	$[5,[0,4,7]]$	$[5,[0,4,7]]$	$[5,[0,4]]$	$[5,[0,4,7],[11]]$	
V / IV	$[0,[0,4,7],[10]]$	$[0,[0,4,7],[10]]$	$[0,[0,4],[10]]$		

Table 15: Four often met groups and their exemplar GCTs. Notice how the group of $[0,[0,4,7]]$ has been separated from the group of $[0,[0,4,7],[10]]$, due to the non-diatonic pitch class 10 on the latter.

In some extreme cases, mere utilisation of these three criteria may give rise to ambiguities in grouping some "incomplete" GCT expressions. For instance, if we consider the symmetric scale $[0,1,3,4,6,7,9,10]$ and two "diatonic" chords to this scale, $c_1 = [0,[0,3,7]]$ and $c_2 = [0,[0,4,7]]$, then the $[0,[0,7]]$ type would as well belong the groups of both c_1 and c_2. Additionally, in the less extreme scenario where a major scale pieces is considered ($[0,2,4,5,7,9,11]$), if a $[0,[0,4,7,10]]$ type is followed by a $[0,[0,4]]$, then it would be reasonably assumed that $[0,[0,4]]$ is the same as the $[0,[0,4,7,10]]$ with converging voices to the pitch classes 0 and 4. However, the discussed 3 criteria would categorise $[0,[0,7]]$ in the group of the diatonic $[0,[0,4,7]]$ type, since all tis pitch classes are diatonic in the considered scale – while the $[0,[0,4,7,10]]$ type has the non-diatonic pitch class 10. Therefore, context information are required for further refining the GCT grouping criteria. The results obtained by these criteria, however, are satisfactory regarding the GCT types obtained by the chorales of Bach.

References

[1] ALLAN, M., AND WILLIAMS, C. K. I. Harmonising chorales by probabilistic inference. In *Advances in Neural Information Processing Systems 17* (2004), MIT Press, pp. 25–32.

[2] BORREL-JENSEN, N., AND HJORTGAARD DANIELSEN, A. Computer-assisted music composition – a database-backed algorithmic composition system. B.S. Thesis, Department of Computer Science, University of Copenhagen, Copenhagen, Denmark, 2010.

[3] CAMBOUROPOULOS, E., KALIAKATSOS-PAPAKOSTAS, M., AND TSOUGRAS, C. An idiom-independent representation of chords for computational music analysis and generation. In *Proceeding of the joint 11th Sound and Music Computing Conference (SMC) and 40th International Computer Music Conference (ICMC)* (2014), ICMC–SMC 2014.

[4] CHAI, W. Automated analysis of musical structure, 2005. Thesis (Ph. D.)–Massachusetts Institute of Technology, School of Architecture and Planning, Program in Media Arts and Sciences, 2005.

[5] CHUAN, C.-H., AND CHEW, E. A hybrid system for automatic generation of style-specific accompaniment. In *Proceedings of the 4th International Joint Workshop on Computational Creativity* (2007), Goldsmiths, University of London.

[6] COHN, R. Neo-riemannian operations, parsimonious trichords, and their" tonnetz" representations. *Journal of Music Theory* (1997), 1–66.

[7] COHN, R. Introduction to neo-riemannian theory: a survey and a historical perspective. *Journal of Music Theory* (1998), 167–180.

[8] DONNELLY, P., AND SHEPPARD, J. Evolving four-part harmony using genetic algorithms. In *Proceedings of the 2011 International Conference on Applications of Evolutionary Computation - Volume Part II* (Berlin, Heidelberg, 2011), EvoApplications'11, Springer-Verlag, pp. 273–282.

[9] EBCIOGLU, K. An expert system for harmonizing four-part chorales. *Computer Music Journal 12*, 3 (1988), 43–51.

[10] EPPE, M., CONFALONIER, R., MACLEAN, E., KALIAKATSOS-PAPAKOSTAS, M., CAMBOUROPOULOS, E., SCHORLEMMER, M., CODESCU, M., AND KÜHNBERGER, K.-U.

Computational invention of cadences and chord progressions by conceptual chord-blending. In *International Joint Conference on Artificial Intelligence (IJCAI) 2015* (2015), p. Submitted.

[11] FORNEY, G.D., J. The viterbi algorithm. *Proceedings of the IEEE 61*, 3 (Mar. 1973), 268–278.

[12] FORTE, A. *The structure of atonal music*. Yale University Press, New Haven, 1973.

[13] FORTE, A. Pitch-class set analysis today. *Music analysis* (1985), 29–58.

[14] FORTE, A. Pitch-class set genera and the origin of modern harmonic species. *Journal of Music Theory* (1988), 187–270.

[15] GRANROTH-WILDING, M. T. *Harmonic analysis of music using combinatory categorial grammar*. PhD thesis, Institute for Language, Cognition and Computation School of Informatics University of Edinburgh, Edinburgh, Scotland, Nov. 2013.

[16] HANLON, M., AND LEDLIE, T. Cpu bach: An automatic chorale harmonization system. 2002.

[17] HARRIS, S. J. M. *A proposed classification of chords in early twentieth-century music*. PhD thesis, King's College London (University of London), 1985.

[18] HARTE, C., SANDLER, M., ABDALLAH, S. A., AND GÓMEZ, E. Symbolic representation of musical chords: A proposed syntax for text annotations. In *Proceedings of the 4th International Conference on Music Information Retrieval (ISMIR)* (London, UK, 2005), pp. 66–71.

[19] HILD, H., FEULNER, J., AND MENZEL, W. HARMONET: A neural net for harmonizing chorales in the style of j. s. bach. J. E. Moody, S. J. Hanson, and R. Lippmann, Eds., Morgan Kaufmann, pp. 267–274.

[20] HINDEMITH, P. *The craft of musical composition, vol.1: Theoritical part (Trans. A. Mendel)*. Associated music, New York, 1937/1942.

[21] JORDAN, M. I., GHAHRAMANI, Z., AND SAUL, L. K. Hidden markov decision trees. In *NIPS* (1996), M. Mozer, M. I. Jordan, and T. Petsche, Eds., MIT Press, pp. 501–507.

[22] JURAFSKY, D., AND MARTIN, J. H. *Speech and language processing*. Prentice Hall, New Jersey, USA, 2000.

[23] KALIAKATSOS-PAPAKOSTAS, M., AND CAMBOUROPOULOS, E. Probabilistic harmonisation with fixed intermediate chord constraints. In *Proceeding of the joint 11th Sound and Music Computing Conference (SMC) and 40th International Computer Music Conference (ICMC)* (2014), ICMC–SMC 2014.

[24] KALIAKATSOS-PAPAKOSTAS, M., CAMBOUROPOULOS, E., KÜHNBERGER, K.-U., KUTZ, O., AND SMAILL, A. Concept Invention and Music: Creating Novel Harmonies via Conceptual Blending. In *In Proceedings of the 9th Conference on Interdisciplinary Musicology (CIM2014)* (December 2014), CIM2014.

[25] KALIAKATSOS-PAPAKOSTAS, M., KATSIAVALOS, A., TSOUGRAS, C., AND CAMBOUROPOULOS, E. Harmony in the polyphonic songs of epirus: Representation, statistical analysis and generation. In *4th International Workshop on Folk Music Analysis (FMA) 2014* (June 2011).

[26] KALIAKATSOS-PAPAKOSTAS, M., KATSIAVALOS, A., TSOUGRAS, C., AND CAMBOUROPOULOS, E. D7.1 harmonic training dataset, 2014.

[27] KALIAKATSOS-PAPAKOSTAS, M. A., FLOROS, A., AND VRAHATIS, M. N. A clustering strategy for the key segmentation of musical audio. *Computer Music Journal 37*, 1 (2013), 52–69.

[28] KRUMHANSL, C. L. *Cognitive foundations of musical pitch*, vol. 17. New York: Oxford University Press, 1990.

[29] KUUSI, T. Chord span and other chordal characteristics affecting connections between perceived closeness and set-class similarity. *Journal of New Music Research 34*, 3 (2005), 259–271.

[30] LAITZ, S. G. *The complete musician: an integrated approach to tonal theory, analysis, and listening*. Oxford University Press, New York, 2012.

[31] LERDAHL, F. *Tonal pitch space*. Oxford University Press, 2001.

[32] LIOLIS, K. *To Epirótiko Polyphonikó Tragoúdi (Epirus Polyphonic Song)*. Ioannina, 2006.

[33] MAKRIS, D., KALIAKATSOS-PAPAKOSTAS, M., AND CAMBOUROPOULOS, E. A probabilistic approach to determining bass voice leading in melodic harmonisation. In *Mathematics and Computation in Music: Proceedings of the 5th International Conference, MCM 2015* (London, UK, June 2015), Springer, Berlin, p. Accepted for publication.

[34] MANZARA, L. C., WITTEN, I. H., AND JAMES, M. On the entropy of music: An experiment with bach chorale melodies. *Leonardo Music Journal 2*, 1 (Jan. 1992), 81–88.

[35] OXENHAM, A. *The Perception of Musical Tones*. 2013.

[36] PACHET, F., AND ROY, P. Formulating constraint satisfaction problems on part-whole relations: The case of automatic musical harmonization. In *Proceedings of the 13th European Conference on Artificial Intelligence (ECAI 98)* (1998), Wiley-Blackwell, pp. 1–11.

[37] PACHET, F., AND ROY, P. Musical harmonization with constraints: A survey. *Constraints 6*, 1 (Jan. 2001), 7–19.

[38] PAIEMENT, J.-F., ECK, D., AND BENGIO, S. Probabilistic melodic harmonization. In *Proceedings of the 19th International Conference on Advances in Artificial Intelligence: Canadian Society for Computational Studies of Intelligence* (Berlin, Heidelberg, 2006), AI'06, Springer-Verlag, pp. 218–229.

[39] PARNCUTT, R. *Harmony: A Psychoacoustical Approach*. Springer Berlin Heidelberg, Berlin, Heidelberg, 1989.

[40] PHON-AMNUAISUK, S., AND WIGGINS, G. A. The four-part harmonisation problem: A comparison between genetic algorithms and a rule–based system. In *In proceedings of the AISB'99 symposium on musical creativity* (1999), AISB, pp. 28–34.

[41] QUINN, I. Listening to simmilarity relations. *Perspectives of New Music 39* (2001).

[42] RACZYŃSKI, S. A., FUKAYAMA, S., AND VINCENT, E. Melody harmonization with interpolated probabilistic models. *Journal of New Music Research 42*, 3 (2013), 223–235.

[43] SAMPLASKI, A. G. *A comparison of perceived chord similarity and predictions of selected twentieth-century chord-classification schemes, using multidimensional scaling and cluster analysis.* PhD thesis, Indiana University, 2000.

[44] SAPP, C. S. Computational chord-root identification in symbolic musical data: Rationale, methods, and applications. *Computing in Musicology 15* (2008), 99–119.

[45] SIMON, I., MORRIS, D., AND BASU, S. Mysong: Automatic accompaniment generation for vocal melodies. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (New York, NY, USA, 2008), CHI '08, ACM, pp. 725–734.

[46] SUZUKI, S. Four-part harmonization using probabilistic models: Comparison of models with and without chord nodes, 2013.

[47] SUZUKI, S., AND KITAHARA, T. Four-part harmonization using probabilistic models: Comparison of models with and without chord nodes. In *Proceedings of the Sound and Music Computing Conference 2013 (SMC 2013)* (2013), pp. 628–633.

[48] TEMPERLEY, D. *The Cognition of Basic Musical Structures*. The MIT Press, Sept. 2004.

[49] TERHARDT, E. Pitch, consonance, and harmony. *The Journal of the Acoustical Society of America 55*, 5 (Aug. 2005), 1061–1069.

[50] WHORLEY, R. P., WIGGINS, G. A., RHODES, C., AND PEARCE, M. T. Multiple viewpoint systems: Time complexity and the construction of domains for complex musical viewpoints in the harmonization problem. *Journal of New Music Research 42*, 3 (Sept. 2013), 237–266.

[51] YI, L., AND GOLDSMITH, J. Automatic generation of four-part harmony. In *BMA* (2007), K. B. Laskey, S. M. Mahoney, and J. Goldsmith, Eds., vol. 268 of *CEUR Workshop Proceedings*, CEUR-WS.org.

[52] YOGEV, N., AND LERCH, A. A system for automatic audio harmonization, 2008.