

A Transformational Method for Chorale Generation

Raymond Whorley¹ and Darrell Conklin^{1,2}

¹ Department of Computer Science and Artificial Intelligence
University of the Basque Country UPV/EHU, San Sebastián, Spain

² IKERBASQUE, Basque Foundation for Science, Bilbao, Spain
`raymond.whorley@ehu.es, darrell.conklin@ehu.es`

Abstract. Music sampled from a statistical model tends to lack long-term structure. This problem can be ameliorated by transforming an existing piece of music into a new one, such that the structure of the original is retained. A method for transforming Bach chorales is presented here. Sampling is constrained by chord symbols, phrase boundary information, soprano note lengths and key regions from the original music. Two corpora are used: one of hymn tune harmonisations by various composers, and the other of chorale melody harmonisations by J. S. Bach.

1 Introduction

A problem with Markov models for music generation is that music sampled from a model tends to lack long-term structure; that is, the output can seem to wander due to the limited context used. The solution proposed here is to transform an existing piece of music such that the new piece is related to the original in some abstract way. To generate a well-structured four-part harmonic texture, hopefully including a coherent and novel melody, a structural template is created from a chorale melody harmonisation by J. S. Bach. Multiple pieces are then generated in line with the template with the aim of finding low cross-entropy (negative mean log probability) solutions, as these tend to be in better compliance with generally agreed rules of harmony [1].

2 Techniques Employed

Specially selected multiple viewpoint systems [2, 3] are used to model two corpora by machine learning: one of 100 major key hymn tune harmonisations [4] and the other of 50 major and minor key chorales by J. S. Bach (in MIDI format [5], derived from [6]). This affords the opportunity to compare music generated using different corpora. The basic viewpoints (attributes) to be generated are *Duration*, *Pitch* (different enharmonic spellings are not distinguished in the models: this is for future work [3]) and *Cont*, the latter indicating whether a note is freshly sounded or is continuing to sound. Some basic viewpoints which are given by annotation of the template are *Mode* (major or minor key with respect to the

entire piece) and **Phrase** (first in phrase, last in phrase, or neither). Other viewpoints, such as **DurRatio** (sequential duration ratio), **Interval** (sequential pitch interval in semitones) and **ScaleDegree** (number of semitones above tonic), are derived from basic ones; and viewpoints may be linked ($\otimes$, to model conjunctions of attributes) or threaded ($\ominus$, to model longer-term dependencies).

A vertical slice (sonority) comprises four concurrently sounding notes, of which at least one is freshly sounded. The **Pitch** domain (the alphabet of **Pitch** symbols used for statistical modelling purposes) contains vertical pitch slices seen in the corpus, plus transpositions lying wholly within the part ranges. Multiple viewpoint systems create, at all positions in a sequence, a distribution which assigns a probability to all possible predictions. In generating music by random walk, slices are successively sampled until the music reaches a halting point. *Probability thresholds* [7] modify random walk, such that slices having a probability lower than a predetermined fraction of the highest slice probability in the distribution cannot be sampled (e.g. for highest probability of 0.3 and threshold of 0.5, sample slices with a probability of ≥ 0.15). Combining the probability threshold method with iterative random walk [8], it is possible to rapidly find low cross-entropy solutions which are likely to have fairly good voice leading [1].

3 Transformational Method

As a starting point for transformation, a template is constructed by manually analysing the original piece. It comprises tuples containing chord symbol (triad or seventh, ignoring inversion), local key (thus handling key modulations), soprano note **Duration**, key signature, **Mode**, soprano note **Cont**, soprano note onset time, time signature and **Phrase**. **Cont** is usually *false*; but if a chord symbol changes while a soprano note is still sounding, a new tuple is created with a **Cont** value of *true*. Cadences are structurally most important, but the chord symbols in between provide a sensible path through this structure.

Building a statistical model from a corpus by machine learning is a separate exercise. At its most basic level, it consists in automatically gathering n-gram counts. The overall model is hierarchical, however, with different orders of n-gram model combining to give a viewpoint model, and various viewpoint models combining to provide an overall prediction. Multiple viewpoint systems for the separate prediction of **Duration**, **Cont** and **Pitch** are automatically selected on the basis of minimising the cross-entropy of a ten-fold cross-validation of the corpus. The systems contain 3 to 5 viewpoints each, with maximum n-gram order varying from 0 (**Duration**, English Hymnal corpus) to 4 (**Pitch**, EH corpus). For the EH corpus, the best performing viewpoints are **DurRatio** $\otimes$ **Phrase**, **Cont** $\otimes$ **Interval** and **Cont** $\otimes$ **ScaleDegree** for the prediction of **Duration**, **Cont** and **Pitch** respectively.

Transformation takes place by sampling from the statistical model while at the same time being constrained by the template; for example, the original soprano note durations are retained, and generated slices are consistent with the chord symbol sequence, as described below. Key regions are conserved by using notes in the scale of the local key. During transformation all pitches may change,

as may the note lengths of the lower parts: passing notes may occur at any point in the music.

In a chorale harmonisation there is often more than one slice per chord symbol: a mix of pure triads or sevenths and others containing non-chord tones. New viewpoint **PureChord** (not used in modelling) classifies subsets of vertical Pitch slices seen in a corpus as T (pure chord, *true*) or F (pure chord, *false*) with respect to a chord symbol in the template; that is, it constrains the Pitch domain (alphabet) appropriately. T domains of slices are based on complete triads/sevenths, and triads/sevenths with the third, fifth or both missing. F domains comprise slices in which one or two of the notes are a step away from notes in T domain slices. These non-chord tones are restricted to notes of the scale of the local key.

A constraint is implemented requiring *at least one T slice per beat*. Although this does not always happen in music, it ensures that the required chord is always recognisable. A combined T/F prediction probability distribution is used as much as possible. The use of a T distribution only is restricted to the situation in which a **Duration** value is generated which takes us to the end of a beat (or beyond), and a T slice has not yet been generated for that beat. See Figure 1 for examples of problems encountered during preliminary generation runs.

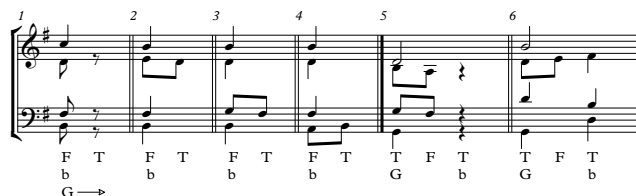


Fig. 1. Examples of problems occurring during generation, assuming a key region of G major. The first example (bars 1 to 4) assumes that a crotchet in the soprano has been harmonised by a B minor quaver slice with one or more non-chord tones in it (F), and that this beat will be completed by a pure B minor slice (T). If the non-chord tone is in the soprano, there is no T slice which can accommodate it (bar 1). In this situation, the soprano note must be a chord tone (bars 2 to 4). The second example (bars 5 and 6) requires a change of chord from G major to B minor beneath a soprano minim. In bar 5, it happens that the soprano D4 generated for the G major chord does not appear in any pure B minor slices seen in the corpus. To avoid such occurrences, slices containing soprano notes not found in the domain constrained by the second chord symbol must be removed from that constrained by the first. In bar 6, the soprano B4 appears in both original domains and therefore also in the constrained domain. These types of domain constraint are handled by software logic.

4 Generated Harmony

For the EH corpus, 1024 transformations were generated for each probability threshold $t = 1.0, 0.9, 0.8, 0.7, 0.6, 0.5, 0.4$ and 0.3 (see Section 2) for Pitch and

Cont (0.0 for Duration) using iterative random walk. Cross-entropies ranged from 5.39 bits/slice ($t = 0.9$) to 7.44 bits/slice ($t = 0.3$). Figure 2 shows a minimal cross-entropy transformation (5.39 bits/slice) of J. S. Bach’s harmonisation of *An Wasserflüssen Babylon*. This has far fewer non-chord tones than the original: indeed, its complexity (as indicated by note density) is subjectively similar to that of typical EH corpus pieces. Voice leading is generally good, but there is a leap of a major 7th in the alto in bar 4. G4 would be better than G3, which results in part crossing with the tenor. There is also a leap of a major 9th in the bass between bars 3 and 4. The identical slices ending bar 6 are better avoided.



Fig. 2. Minimal cross-entropy transformation (5.39 bits/slice) of Bach’s harmonisation of *An Wasserflüssen Babylon*, using the English Hymnal corpus and iterative random walk with $t = 0.0$ for Duration and $t = 0.9$ for Pitch and Cont. First six bars shown.

For the Bach corpus, $t = 1.0, 0.9, 0.8$ and 0.7 were used for all attributes. Cross-entropies ranged from 5.13 bits/slice ($t = 0.8$) to 6.59 bits/slice ($t = 0.7$). Figure 3 shows a minimal cross-entropy transformation (5.13 bits/slice) of J. S. Bach’s harmonisation of *An Wasserflüssen Babylon*. Its complexity subjectively reflects that of typical Bach corpus pieces, and much is good about it; for example, the last two beats of the second phrase contain a well resolved suspension in the alto (of a sort, since the Gs are not tied) leading to a perfect cadence, and the cadence in bar 6 contains a well executed passing note. Conversely, there is a parallel fifth at the end of bar 3 (these occur only very rarely in Bach chorales), and unnecessarily repeated notes occur in the tenor line at the beginning of that bar. Only two pitches are used in the first three bars of the melody, making it rather boring. Of interest with respect to the method is the last beat of bar 6. The template chord is C major, but the F slice generated on the beat forms a pure G major chord. The result is not bad, but not really what is wanted.

5 Conclusions and Future Work

This transformational method can be thought of as creative, in the sense that non-chord tones can appear anywhere in the music; but the output is close to the style of the corpus used, meaning that the Bach corpus is essential if output in the style of Bach is required. The F domain construction method is probably over-prescriptive, so simpler methods will be investigated. In other future work, the length of syllables in the chorale text will be conserved, rather than soprano note lengths, which will allow transformation in the context of conserved lyrics,



Fig. 3. Minimal cross-entropy transformation (5.13 bits/slice) of Bach's harmonisation of *An Wasserflüssen Babylon*, using the Bach corpus and iterative random walk with $t = 0.8$ for all generated attributes. First six bars shown.

and should result in more variety of melody note durations and placement. Chord symbol sequences generated from a statistical model will be used as alternatives to those taken from chorales. The conservation of cadential chord symbols only will also be tried, as will the complete absence of chord symbols (with and without local key region conservation). In addition, harmonisation of existing chorale melodies while conserving chord symbols and so on will be investigated. Finally, one or more objective automatic evaluation techniques, such as one based on general rules of harmony [1], will be employed.

Acknowledgments This research is supported by the Lrn2Cre8 project which is funded by the Future and Emerging Technologies (FET) programme within the Seventh Framework Programme for Research of the European Commission, under FET grant number 610859. We wish to thank our insightful reviewers.

References

1. Whorley, R. P., Conklin, D.: Music Generation from Statistical Models of Harmony. *Journal of New Music Research*, 45(2), 160–183 (2016)
2. Conklin, D., Witten, I. H.: Multiple Viewpoint Systems for Music Prediction. *Journal of New Music Research*, 24(1), 51–73 (1995)
3. Whorley, R. P.: The Construction and Evaluation of Statistical Models of Melody and Harmony. Ph.D. thesis, Department of Computing, Goldsmiths, University of London (2013)
4. Vaughan Williams, R., Ed.: *The English Hymnal*. Oxford University Press, London (1933)
5. MIDI encoding of 185 four-part chorales by J. S. Bach. [Online]. Available: <http://kern.humdrum.org/cgi-bin/browse?l=/185chorales>
6. Bach, J. S.: *Bach-Gesellschaft Ausgabe*, vol. 39, F. Wüllner, Ed. Breitkopf and Härtel, Leipzig (1892)
7. Whorley, R. P., Wiggins, G. A., Rhodes, C., 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), 237–266 (2013)
8. Herremans, D., Sørensen, K., Conklin, D.: Sampling the Extrema from Statistical Models of Music with Variable Neighbourhood Search. In: *Proceedings of the Sound and Music Computing Conference*, pp. 1096–1103. Athens, Greece (2014)