Generating Subjects for Pieces in the Style of Bach's Two-Part Inventions

Raymond P. Whorley and Robin Laney

The Open University
Raymond.Whorley@open.ac.uk

Abstract. We describe how to begin the computational process of composing a piece in the style of J.S. Bach's two-part inventions by automatically generating plausible musical *subjects*. A generate and test protocol is proposed, whereby subjects would be generated by a modified random sampling technique and then tested against information theoretic measures, with the aim of filtering out unsuitable subjects. The statistical models to be sampled were constructed by machine learning from a corpus comprising the first few bars of the upper parts of the fifteen two-part inventions, using *multiple viewpoint systems*. Using *information content,* we were able to correctly classify subjects as suitable or unsuitable from a pitch structure point of view in 115 out of 120 cases. Mean information content based on a simple short-term model (i.e., the statistics were taken from the generated subject, not from the corpus) was partially successful in filtering out unduly repetitive subjects.

Keywords: machine learning, probability threshold, multiple viewpoint system, statistical model, information theory

1 Introduction

Our longer-term aim is to computationally investigate a theory of how J.S. Bach composed his two-part inventions (Dreyfus, 1996). The ultimate goal is to demonstrate the theory in action by automatically generating music in this style, and by creating software that will allow users to guide the process of generating such music according to their preferences. In brief, an invention is a piece of music that employs smaller building blocks, also called inventions, which can be transformed in largely deterministic ways. The smaller-scale inventions and their transformations are laid out in a process called *disposition* to form the backbone of the piece. A less formal process of *elaboration* fills in the gaps. This paper describes how to begin the compositional process by automatically generating plausible *subjects* (focusing on pitch structure), which are even smaller building blocks used in the construction of the smaller-scale inventions. The subject of the C major invention, for example, comprises the first seven notes of the upper part. Importantly, we analyse information theoretic measures, based on selected statistical models, as a means of filtering out subjects deemed unsuitable on the basis of pitch structure and over-repetitiveness.

2 Methods Employed in This Work

We use random sampling of statistical models as the means of generating musical motifs that may be considered as subjects for pieces in the required style. The models are constructed by machine learning from a small corpus consisting of the first few bars of the upper part of each of the fifteen two-part inventions (complete upper parts may be considered for comparison in future work). The machine learning employs *multiple viewpoint systems* (Conklin & Cleary, 1988; Conklin, 1990; Conklin & Witten, 1995) and *Prediction by Partial Match* (PPM, Cleary & Witten, 1984). A concise description of the modelling of melody by these and associated techniques is given in Section 3.2 of Whorley & Conklin (2016); but an overview is also given below.

Table 1. Basic, test and derived primitive viewpoints relevant to this research are informally defined here.

Viewpoint	Meaning
`Contour`	Falling, equal or rising pitch.
`Duration`	Note duration (e.g., crotchet = 24).
`DurRatio`	Ratio of adjacent note durations.
`ExtendedScaleDegree`	Pitch interval (semitones) above lowest tonic MIDI value.
`FirstInBar`	First beat in bar, or not.
`FirstInPhrase`	First note in phrase, or not.
`FirstInPiece`	First note in piece, or not.
`InScale`	Note within scale defined by `KeySig` and `Mode`, or not.
`Interval`	Pitch interval between adjacent notes.
`IntFirstInBar`	Pitch interval from note on first beat in bar.
`IntFirstInPiece`	Pitch interval from first note in piece.
`IOI`	Onset time difference between adjacent notes.
`KeySig`	Number of sharps or flats (negative for flats).
`Mode`	Key is either major or minor.
`Pitch`	Note pitch MIDI value (e.g., middle C = 60).
`ScaleDegree`	Pitch interval (semitones) above nearest lower tonic.
`Tactus`	On a main beat of a bar, or not.
`TactusDuration`	Note duration as number of main beats (or fraction of beat).
`Tessitura`	Note pitch is high, low or midrange.

2.1 Modelling of Melody

Viewpoints are representations of the data in a corpus in terms of various different alphabets. Pearce (2005) and Whorley (2013) introduced new viewpoints, and viewpoints `ExtendedScaleDegree` and `TactusDuration` are reported for the first time in this paper. For melodic music, the basic viewpoints (to be predicted or generated) are `Duration` and `Pitch`. Data for these viewpoints is readily available in the corpus representation as integers for note lengths (e.g., crotchet = 24) and MIDI pitch values (e.g., middle C = 60). Data for viewpoint `Interval`, for example, must be derived from `Pitch` data of adjacent notes. Boolean test viewpoints indicate rhythmically or structurally important positions in a melody; for example, `Tactus` is true on main beats in a bar, but false elsewhere. Table 1 informally defines the primitive basic, test

and derived viewpoints relevant to this paper. These may be linked (e.g., `Duration` ⊗ `Tactus`) to model combinations of attributes. In this particular case, tuples such as $\langle 12, \text{true} \rangle$ and $\langle 6, false \rangle$ are modelled. They may also be threaded, for example `Pitch` ⊖ `Tactus`. This viewpoint uses the `Pitch` alphabet, but is only defined at positions where `Tactus` is true (thereby modelling longer-term dependencies).

Viewpoint models are variable-order n-gram models with a defined maximum order (i.e., context size). Say that a melody has been partially generated, that the duration of the next note has been generated, and the pitch of that note is about to be generated. Prediction by Partial Match takes the immediately preceding viewpoint context of maximum size, and compares it with contexts of the same size seen in the corpus. On finding a match, associated prediction counts (again from the corpus, and matching the generated duration where applicable) are used to calculate prediction probabilities. We back off to a slightly smaller context size (a partial match, taking with us a proportion of the probability mass, the *escape probability*) to find more predictions. This procedure is followed until all possible predictions have been found.

Having constructed probability distributions for each viewpoint in the system, they must all now be combined. First, they are all converted to `Pitch` distributions, since we wish to generate the pitch attribute of a note. Next, a weighted geometric (multiplicative) combination technique is employed (Pearce, Conklin, & Wiggins, 2005), which gives the highest weighting to the least uniform distribution (i.e., the most certain one). A bias, which can be optimised, is used in the calculation of the weights.

So far only corpus statistics have been considered, leading to the construction of a *long-term model* (LTM). It is also possible to take account of statistics from the piece of music being predicted or generated, resulting in a *short-term model* (STM) or an updated long-term model (LTM+). The LTM(+) and STM probability distributions can be combined by the same geometric technique, but using a separately optimised bias. In this research, however, non-updated LTMs will be used to generate the musical subjects, while LTMs and STMs will be considered for the purpose of evaluating the suitability of subjects (ultimately to filter candidate subjects during generation).

2.2 Viewpoint Selection

The first step of this work was to select viewpoints for multiple viewpoint systems capable of generating note durations and pitches. This was done automatically, following Pearce (2005), on the basis of minimising the *cross-entropy* (an information theoretic measure: see e.g. Jurafsky & Martin, 2000; Manning & Schütze, 1999) of a leave-one-out cross-validation of the corpus. The lower the cross-entropy, the better the model; and cross-validation ensures that over-fitting to the corpus is avoided. Various maximum model orders were tried in order to find the best performing model, in each case optimising the bias after viewpoint selection.

The multiple viewpoint system shown in Table 2 (Appendix A) was selected for the generation of the musical attribute *duration*. Prior to the generation of this attribute for each note, probability distributions are constructed for each viewpoint, converted into `Duration` distributions, and then combined as described in Section 2.1. This system achieves a cross-validation cross-entropy of 0.77 bits/prediction, using a maximum first-order LTM with a bias of 65.3.

A maximum fourth-order LTM system as listed in Table 3 (Appendix A), achieving a cross-validation cross-entropy of 2.79 bits/prediction with a bias of 1.9, was selected for the generation of the musical attribute *pitch*.

2.3 Information Theoretic Measures

The *information content* (MacKay, 2003) of an event, otherwise known as its *pointwise entropy* (Manning & Schütze, 1999), is defined as $h = -\log_2 p$, which is the minimum number of bits required to encode this event. The mean information content of a sequence of n events is therefore $\overline{h} = -\frac{1}{n}\sum_{i=1}^{n} \log_2 p_i$. As the sequence length tends to infinity, so this measure tends to cross-entropy. Laney, Samuels, & Capulet (2015) use the *Information Dynamics of Music* (IDyOM, Pearce & Wiggins, 2012) framework to investigate cross-entropy as a measure of musical contrast.

The first appearance of a musical subject should be solidly within the key of the piece as a whole. Notes that are not within the key can be perceived as surprising in the context of a subject. Information content h is an indicator of surprise or unexpectedness: the higher it is, the greater the surprise. It was thought, therefore, that this measure would afford a way of identifying motifs that are unsuitable in this way, especially in the light of evidence from earlier work (Conklin & Witten, 1995; Potter, Wiggins, & Pearce, 2007).

A motif that is overly repetitive is unlikely to make a suitable subject for an invention. It was thought that a short-term model comprising the single viewpoint `Pitch` would result in a particularly low mean information content $\overline{h}$ if such repetition occurred. This would provide a means of filtering out such motifs. Viewpoint `Interval` was ruled out because a scale is probably acceptable as a subject, but would have low $\overline{h}$ because it consists only (or mainly) of tone and semitone steps in one direction.

2.4 Probability Thresholds

As stated above, musical motifs are generated by random sampling of the overall prediction probability distribution. A large proportion of generated motifs, however, are atypical of music in the corpus: indeed, some motifs may verge on the chaotic. Mean information content $\overline{h}$ correlates well with degree of organisation or chaos. Whorley & Conklin (2016) demonstrated that generating multiple harmonisations of five melodies by random sampling resulted in generally high $\overline{h}$, but that there was a trend towards better adherence to some general rules of harmony as $\overline{h}$ reduced. It was possible to show that this trend continued by modifying the random sampling technique by the introduction of *probability thresholds* (Whorley et al., 2013). Essentially, predictions are ignored if they have a probability lower than a prescribed fraction of the highest probability in a distribution. For example, if the probability threshold is 0.1 and the highest probability in a given distribution is 0.3, then predictions with probabilities lower than 0.03 are overlooked during the sampling process. It should be noted that minimisation of $\overline{h}$ is not the aim, since music with values lower than those found in the corpus can tend to be less interesting.

3 Results

As a starting point, sixteen seven-note motifs were generated using random sampling (i.e., with a probability threshold of zero). We forced them to begin on the second semiquaver of the bar, assuming C major and common time, for direct comparison with the seven-note subject of invention 1. They were all different, with mean information content $\bar{h}$ ranging from 2.83 to 6.09 bits/note. Such large values indicate a tendency towards disorder, and indeed there were many inappropriate chromatically altered notes and large intervals: see Fig. 1a and Fig. 1b for examples. See also Table 4 (Appendix A) for details of all sets of generation runs.

3.1 Probability Thresholds

As indicated in Section 2.4, random sampling can be modified by the use of probability thresholds to generate a greater proportion of suitable subjects. 16 motifs were generated for each of the probability thresholds 1.0, 0.9, 0.8, 0.7 and 0.6 (in all other respects, generation was the same as for the initial sixteen). It was obvious that the motifs were well organised, and that $\bar{h}$ was much lower: 1.54 bits/note (Fig. 1c) for a threshold of 1.0, and 1.54 to 2.11 (Fig. 1d) bits/note for a threshold of 0.6. The disadvantage of such high thresholds, however, is that not many distinct motifs are generated. In fact, for thresholds of 0.8 and above, only one motif is generated in 16 runs (that identified as the subject in invention 1: see Fig. 1c). A threshold of 0.7 results in an additional three motifs; while many more are generated with a threshold of 0.6.

A compromise is required such that a large proportion of inappropriate motifs are filtered out, while still allowing a large variety of motifs to be generated. We found that thresholds of 0.1 and 0.3 for the generation of note duration and pitch respectively fitted the bill. 48 motifs were generated in this way, with $\bar{h}$ from 1.54 to 3.46 bits/note. Some were not suitable as subjects, however, because of the appearance of notes not in the specified key (Fig. 1e), uncharacteristic rhythmic structure (Fig. 1f) or excessive repetition (Fig. 1g); therefore ways and means of further filtering out unsuitable motifs were investigated in this paper, focusing on pitch-related issues.

3.2 Information Content

High information content h can possibly identify inappropriate pitches in motifs ($\bar{h}$ is not used because the effect of a note with high h is diluted, especially in longer motifs). Eight ten-note motifs were generated in B minor and $\frac{6}{8}$ time, starting on the first beat of the bar, using probability thresholds of 0.1 and 0.3 for the generation of duration and pitch respectively. Fig. 1h and Fig. 1i contain a D sharp, which is not in the specified key; so they are not suitable as subjects from a pitch structure point of view.

To begin with, h was calculated based on overall (combined) LTM pitch probability distributions. The highest h associated with Fig. 1h is 3.11 bits; but this is for the seventh note, which is an acceptable F sharp. The D sharp is assigned a value of 1.77 bits. On the other hand, the D sharp in Fig. 1i is assigned 3.40 bits, the highest value in the motif, as we might have hoped. Amongst the motifs with an acceptable pitch

structure, the highest h is 3.96 bits. This value is associated with the fourth note of Fig. 1j, which is an acceptable A sharp. It is clear, then, that information contents based on overall LTM distributions are not able to identify unsuitable notes in motifs.

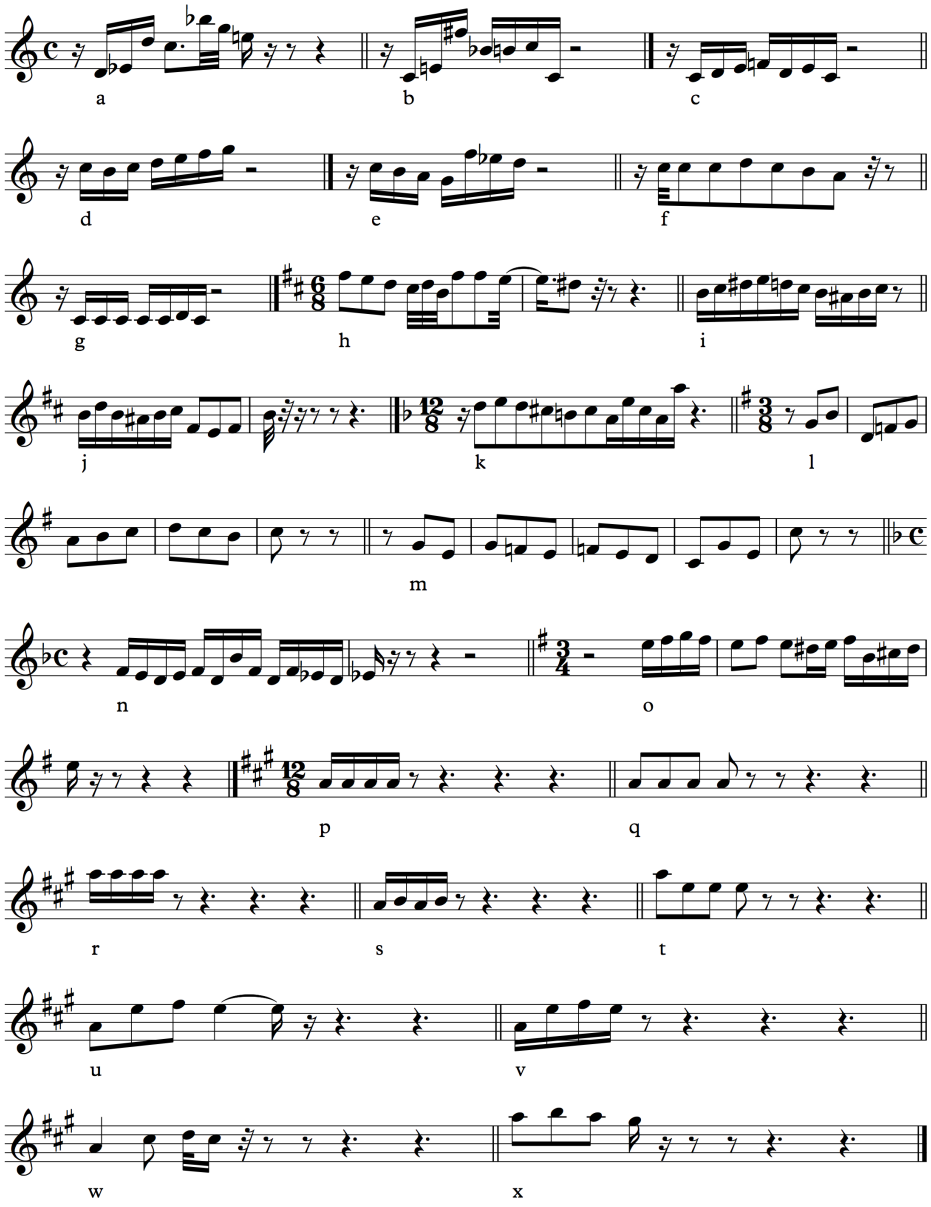

Fig. 1. Examples of generated motifs, illustrating the effect of probability thresholds, information content and mean information content as filters. The significance of individual examples is made clear in the text.

Next, the utility of individual viewpoint LTM distributions was tested on the same 8 motifs. Only viewpoints that had been selected for the pitch-predicting system were investigated, considering that an enormous number of possible primitive and linked viewpoints could be tried. Of the fourteen viewpoints in the system, the one found best able to distinguish between acceptable and unacceptable motifs was `Interval` ⊗ `InScale`. The unsuitable motifs had highest h of 5.95 and 7.25 bits, corresponding to the offending D sharps, whereas those for the suitable motifs ranged from 2.81 to 5.39 bits. It was encouraging to see these non-overlapping ranges; but a much larger sample was required to be confident of the efficacy of `Interval` ⊗ `InScale` h.

A further 112 motifs of 10 to 14 notes, with various key and time signatures, were generated to bring the total sample to 120. For this larger sample size the normal range of highest h for unsuitable motifs was 5.46 to 7.86 bits, and that for suitable motifs 2.10 to 5.39 bits. Although most motifs could be successfully classified as suitable or not, there were some exceptions. Fig. 1k was specified as D minor, but is perceived as D major: the minor third of the scale is absent, and there is a B natural but no B flat. A highest h of 2.58 bits falsely suggests that the motif is acceptable.

Fig. 1l is in G major, but contains an F natural. This minor seventh of the scale was not picked up by the h method, the highest value being 4.56 bits. Similarly, Fig. 1m (also in G major) has two F naturals, while Fig. 1n (in F major) has two E flats. They have highest h of 3.06 and 4.17 bits respectively.

One acceptable motif is rejected on the basis of a very high h. Fig. 1o displays a textbook ascending form of the melodic minor at its conclusion, but the h assigned to the C sharp is 8.31 bits. Other than these five motifs, 92 were correctly classified as suitable (highest $h \leq 5.39$ bits, all notes within the specified key) and 23 correctly classified as unsuitable (highest $h > 5.39$ bits, not all notes within the specified key). For the purposes of this paper, classification was done manually from computationally calculated values of h in trace output.

3.3 Mean Information Content from a Simple STM

It was considered possible that a short-term model comprising the single viewpoint `Pitch` could act as a filter for overly repetitive motifs, by assigning them very low $\overline{h}$. A further twenty-four motifs for each of lengths four to nine notes were generated. The 24 motifs in each set of 4 to 14 notes were arranged in order of $\overline{h}$ (based on an STM for `Pitch`), and two adjacent motifs were selected as being on either side of acceptability from the point of view of repetition. Note that the generation of unison intervals is largely due to the opening of invention 12 (ornaments are ignored in the corpus). Nine of the four-note motifs at the lower end of the $\overline{h}$ range are shown as Fig. 1p to Fig. 1x, where the acceptability boundary is chosen to be between Fig. 1t and Fig. 1u (3.28 and 4.54 bits/note respectively). The relevant values were plotted as 'Definitely retain' and 'Definitely reject' (it should be noted that in a few instances there were no unacceptable motifs in a generated sample), and then a regression curve was added for midway points: see Fig. 2. This curve passes between 'Definitely retain' and 'Definitely reject' except in the case of the nine-note subject, and it can be implemented in software as a hard boundary between retention and rejection. The curve is a coarse filter only: repetitive motifs occur above the line, such as a seven-

note one containing five Ds, four of which appear consecutively at the end. A general reduction in $\bar{h}$ with increasing motif length is expected, since the STM improves as more notes are seen; but the reversal of the curve above 10 notes is probably due to repetitive notes tending to occupy a smaller proportion of a motif.

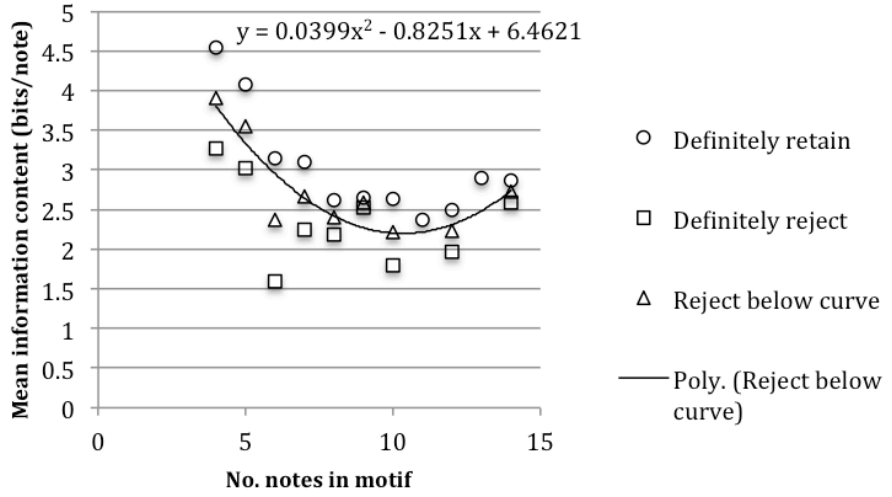

Fig. 2. Plot of *mean information content* (bits/note) against *no. notes in motif*. The circles and squares are for motifs on either side of the retain/reject boundary, on the basis of repetition. The regression curve is for points (triangles) midway between these values, and can serve as a filter to remove repetitive motifs.

4 Conclusions

We have shown that a generate and test approach, using multiple viewpoint systems, can create suitable subjects for pieces in the style of Bach's two-part inventions. Random sampling of the probability distributions, modified by the use of carefully chosen probability thresholds, means that a lot of unsuitable high mean information content motifs are never generated; while at the same time a large variety of motifs are. Information theoretic techniques are then able to filter out a fair proportion of remaining unsuitable motifs: `Interval` ⊗ `InScale` information contents (using LTM) have been demonstrated to be a good way of filtering out motifs containing notes that are not within the specified scale; while `Pitch` mean information content (using STM), is able to identify some repetitive motifs. In the end, the software user will choose from the remaining motifs on the basis of suitability, musicality, personal preference, and so on. This is just the start of the process of creating pieces in the specified style, and much work remains to be done. The work can be adapted to produce musical material for other styles and genres by using different corpora and selecting the best sets of viewpoints to model them.

References

Cleary, J.G. & Witten, I.H. (1984). Data compression using adaptive coding and partial string matching. *IEEE Transactions on Communications, COM-32* (4), 396–402.

Conklin, D. (1990). *Prediction and entropy of music* (Master's thesis). Department of Computer Science, University of Calgary, Canada.

Conklin, D. & Cleary, J.G. (1988). Modelling and generating music using multiple viewpoints. In *Proceedings of the First Workshop on AI and Music* (pp. 125–137). St. Paul, MN.

Conklin, D. & Witten, I.H. (1995). Multiple viewpoint systems for music prediction. *Journal of New Music Research, 24* (1), 51–73.

Dreyfus, L. (1996). *Bach and the patterns of invention.* Cambridge, MA and London, England: Harvard University Press.

Jurafsky, D. & Martin, J.H. (2000). *Speech and language processing.* Upper Saddle River, NJ: Prentice-Hall.

Laney, R., Samuels, R. & Capulet, E. (2015). Cross entropy as a measure of musical contrast. In T. Collins, D. Meredith & A. Volk (Eds.), *Mathematics and computation in music* (LNAI 9110, pp. 193–198). Berlin, Germany: Springer.

MacKay, D.J.C. (2003). *Information theory, inference, and learning algorithms.* Cambridge, England: Cambridge University Press.

Manning, C.D. & Schütze, H. (1999). *Foundations of statistical natural language processing.* Cambridge, MA: MIT Press.

Pearce, M.T. (2005). *The construction and evaluation of statistical models of melodic structure in music perception and composition* (PhD thesis). Department of Computing, City University, London, UK.

Pearce, M.T., Conklin, D. & Wiggins, G.A. (2005). Methods for combining statistical models of music. In U.K. Wiil (Ed.), *Computer music modelling and retrieval* (pp. 295–312). Berlin, Germany: Springer.

Pearce, M.T. & Wiggins, G.A. (2012). Auditory expectation: The information dynamics of music perception and cognition. *Topics in Cognitive Science, 4,* 625–652.

Potter, K., Wiggins, G.A. & Pearce, M.T. (2007). Towards greater objectivity in music theory: Information-dynamic analysis of minimalist music. *Musicae Scientiae, 11* (2), 295–322.

Whorley, R.P. (2013). *The construction and evaluation of statistical models of melody and harmony* (PhD thesis). Department of Computing, Goldsmiths, University of London, UK.

Whorley, R.P. & Conklin, D. (2016). Music generation from statistical models of harmony. *Journal of New Music Research, 45* (2), 160–183.

Whorley, R.P., Wiggins, G.A., Rhodes, C. & Pearce, M.T. (2013). 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.

Appendix A Multiple Viewpoint Systems and Run Details

Table 2. Multiple viewpoint system automatically selected for the generation of basic attribute duration.

Duration ⊗ Tactus
DurRatio ⊗ (IOI ⊖ Tactus)
DurRatio ⊗ (Interval ⊖ Tactus)
TactusDuration ⊗ (ScaleDegree ⊖ FirstInBar)

Table 3. Multiple viewpoint system automatically selected for the generation of basic attribute pitch.

Interval ⊗ InScale
ExtendedScaleDegree ⊗ FirstInPiece
IntFirstInBar ⊗ InScale
Interval ⊗ Tactus
IntFirstInPiece ⊗ InScale
Mode ⊗ (ScaleDegree ⊖ FirstInPhrase)
Interval ⊗ IntFirstInBar
(Contour ⊖ FirstInBar) ⊗ KeySig
InScale ⊗ Tessitura
DurRatio ⊗ Interval
ScaleDegree ⊖ FirstInPhrase
Interval ⊗ ScaleDegree
(Contour ⊖ FirstInBar) ⊗ IOI
(Pitch ⊖ Tactus) ⊗ (ScaleDegree ⊖ FirstInPhrase)

Table 4. Listed here for each set of subject generation runs are: probability thresholds for the generation of basic attributes duration and pitch; key and time signature; number of notes; starting position (e.g., the second semiquaver of the first beat); total number of generated samples and number of distinct subjects; and filter(s) investigated.

run set	1	2	3	4	5	6	7	8	9
duration pt	0.0	1.0	0.9	0.8	0.7	0.6	0.1	0.1	0.1
pitch pt	0.0	1.0	0.9	0.8	0.7	0.6	0.3	0.3	0.3
key	C maj	C maj	C maj	C maj	C maj	C maj	C maj	B min	D min
time sig.	4/4	4/4	4/4	4/4	4/4	4/4	4/4	6/8	12/8
no. notes	7	7	7	7	7	7	7	10	11
start on semiq	2	2	2	2	2	2	2	1	2
of beat	1	1	1	1	1	1	1	1	1
no. samples	16	16	16	16	16	16	48	24	24
no. distinct	16	1	1	1	4	9	47	24	24
filter	n/a	n/a	n/a	n/a	n/a	n/a	n/a	$h, \overline{h}$	$h, \overline{h}$
run set	10	11	12	13	14	15	16	17	18
duration pt	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1
pitch pt	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3
key	G maj	F maj	E min	A maj	B♭ maj	D maj	C maj	G min	F♯ min
time sig.	3/8	4/4	4/4	12/8	4/4	3/8	4/4	4/4	4/4
no. notes	12	13	14	4	5	6	7	8	9
start on semiq	1	1	1	1	2	1	1	1	1
of beat	2	2	3	1	1	2	2	3	3
no. samples	24	24	24	24	24	24	24	24	24
no. distinct	24	24	24	19	21	23	24	23	23
filter	$h, \overline{h}$	$h, \overline{h}$	$h, \overline{h}$	$\overline{h}$	$\overline{h}$	$\overline{h}$	$\overline{h}$	$\overline{h}$	$\overline{h}$