Improving predictions of derived viewpoints in multiple viewpoint systems

HEDGES, TW; Wiggins, GA; ISMIR

To be published by http://wp.nyu.edu/ismir2016/

For additional information about this publication click this link.
http://qmro.qmul.ac.uk/xmlui/handle/123456789/13047

IMPROVING PREDICTIONS OF DERIVED VIEWPOINTS IN MULTIPLE VIEWPOINT SYSTEMS

Thomas Hedges
Queen Mary University of London
`t.w.hedges@qmul.ac.uk`

Geraint Wiggins
Queen Mary University of London
`geraint.wiggins@qmul.ac.uk`

ABSTRACT

This paper presents and tests a method for improving the predictive power of derived viewpoints in multiple viewpoints systems. Multiple viewpoint systems are a well established method for the statistical modelling of sequential symbolic musical data. A useful class of viewpoints known as derived viewpoints map symbols from a basic event space to a viewpoint-specific domain. Probability estimates are calculated in the derived viewpoint domain before an inverse function maps back to the basic event space to complete the model. Since an element in the derived viewpoint domain can potentially map onto multiple basic elements, probability mass is distributed between the basic elements with a uniform distribution. As an alternative, this paper proposes a distribution weighted by zero-order frequencies of the basic elements to inform this probability mapping. Results show this improves the predictive performance for certain derived viewpoints, allowing them to be selected in viewpoint selection.

1. INTRODUCTION

Multiple viewpoint systems [7] are an established statistical learning approach to modelling multidimensional sequences of symbolic musical data. Music is presented as a series of events comprising of basic attributes (e.g. pitch, duration) modelled by a collection of *viewpoints*. For example, pitch may be modelled by pitch interval, pitch class, or even pitch itself. Statistical structure for each viewpoint is captured with a Markovian approach, usually in the form of a Prediction by Partial Match (PPM) [2] suffix tree. Predictions from different viewpoints modelling the same basic attribute are combined, weighting towards viewpoints with lower uncertainty in terms of Shannon entropy [24]. The system can be viewed as a mixture of experts, or ensemble method machine learning approach to symbolic music, dynamically using specialised models which are able to generalise data in order to find structure.

The current research explores a problem associated with a collection of viewpoints known as *derived viewpoints*. Derived viewpoints apply some function to basic attributes aiming to capture some relational structure between basic attributes (e.g. pitch interval), or to generalise sparse data (e.g. pitch class). During training, elements from the basic attribute domain are mapped onto the derived viewpoint domain with a surjective function. Viewpoint models must be combined over a shared alphabet in order to calculate probability estimates, therefore, an inverse function maps from the derived viewpoint domain to the basic attribute domain. Where a derived element maps onto several basic elements, probability mass from the derived element is distributed uniformly between the basic elements [17]. This can be problematic for derived viewpoints with small domains mapping onto large basic attribute domains as the derived elements could refer to many basic elements. Such viewpoints may generalise sparse data and find useful statistical structure, but this information is lost when mapping back to the basic attribute domain. This is especially prevalent where the zero-order (or unigram) distribution of the basic attribute domain is of low entropy, such that a few elements are very frequent and the rest relatively infrequent.

This paper proposes a method for improving predictions from derived viewpoints. The basic premise behind the method is to use the zero-order distribution of the basic attribute to weight the probabilities from the derived viewpoint when mapping back to the basic attribute. This enables the derived viewpoint to take advantage of the zero order statistics of basic attributes in a way which is not possible if the basic and derived viewpoints are modelled separately. After a review of research using multiple viewpoint systems (Section 2), the system used in the current paper is presented (Section 3), and a detailed description of the proposed method given (Section 4). The method is tested on individual derived viewpoints (Section 5.1) before being applied to various full multiple viewpoint systems, including viewpoint selection (Section 5.2).

2. RELATED RESEARCH

Multiple viewpoint systems have become an important tool for statistical learning of music since their inception over twenty-five years ago [3]. This section reviews their uses and applications to both musical and non-musical domains.

Early multiple viewpoint systems [3, 7, 16] focussed on monophonic melodic music, namely chorale and folksong melodies. The seminal paper [7] uses hand-constructed multiple viewpoint systems with a corpus of 100 Bach chorales. Results show that a system of four viewpoints

capturing pitch, sequential pitch interval, scale degree, durational, and metrical information performs best. The system can be used as a generative tool, using a random walk process to generate a chorale in the style of the training corpus. Further work with monophonic melodic music can be seen with the Information Dynamics of Music (IDyOM) model [16], which is developed as a cognitive model of melodic expectation. The PPM* algorithm is refined [20] with a thorough evaluation of smoothing methods, as well as the methods for combining predictions from various individual models, and the method for constructing viewpoint systems [17]. IDyOM is found to closely correlate with experimental data of melodic expectation from human participants, accounting for 78% of variance when predicting notes in English hymns [19], and 83% of variance for British folksongs [21]. Multiple viewpoint systems have also been applied successfully to Northern Indian raags [25], Turkish folk music [23], and Greek folk tunes [6], strengthening their position as a general, domain-independent statistical learning model for music.

Multiple viewpoint systems can be applied to polyphonic musical data, modelling some of the harmonic aspects of music. Musical data with multiple voices is divided into vertical slices [4] representing collections of simultaneous notes, i.e. chords. Relationships between voices can be captured with the use of *linked viewpoints* between voices. This approach has been utilised extensively for the harmonisation of four-part chorales [27, 28]. Harmonic structure can also be modelled directly from chord symbols [5, 10, 22], removing the problems of sparsity and equivalence associated with chord voicing.

Strong probabilistic models of expectation for sequential data can be used for segmentation and chunking. IDyOM is compared to rule-based models for boundary detection in monophonic melodic music in [18], with the statistical model performing comparably rule-based systems. Similar methods have been applied to segmenting natural language at the phoneme and morpheme level [9]. These segmentation studies utilise the fact that certain information theoretic properties, namely information content, can be used to predict boundaries in sequences. The ability for multiple viewpoint systems to model the information theoretic properties of sequences, as well as their general approach to statistical learning, makes them an attractive basis for cognitive architectures capable of general learning, finding higher order structure, and computational creativity [29].

3. A MULTIPLE VIEWPOINT SYSTEM FOR CHORD SEQUENCES

This section presents a brief technical description of the multiple viewpoint system and corpus used in the current research. The corpus consists of 348 chord sequences from jazz standards in lead sheet format from *The Real Book* [11] compiled by [15]. This gives a suitably large corpus of 15,197 chord events, represented as chord symbols (e.g. Dm^7, $Bdim$, $G7$). *The Real Book* is core jazz repertoire comprising of a range of composers and styles, indicating

it is a good candidate for studying tonal jazz harmony. The viewpoint pool is derived from similar multiple viewpoint systems dealing with chord symbol sequences [5, 10].

3.1 Harmonic Viewpoints

Three basic attributes, `Root`, `ChordType`, and `PosInBar`, are used to represent chord labels. `Root` is the functional root of the chord as a pitch class assuming enharmonic equivalence. `ChordType` represents the quality of the chord (e.g. major, minor seventh) and are simplified to a set of 13 (*7, M, m7, m, 6, m6, halfdim, dim, aug, sus, alt, no3rd, NC*) for practical reasons.[1] NC represents the special case where no harmonic instruments are instructed to play in the score. `PosInBar` represents the metrical position in the current bar measured in quavers. Since, by definition, a chord must be stated at the start of each bar, this is a sufficient basic attribute to represent any durational or temporal information in the chord sequence.

The following viewpoints are derived from `Root`. `RootInt` is the root interval in semitones modulo-12 between two adjacent chords, returning the symbol -1 if either is NC. `MeeusInt` categorises root movement (`RootInt`) using root progression theories [14]. The symbol 1 represents dominant root progressions (`RootInt` = 1,2,5,8,9), -1 for subdominant progressions (`RootInt` = 3,4,7,10,11), 0 for no root movement (`RootInt` = 0), -2 for a diminished fifth (`RootInt` = 6), and -3 when either root is NC. Since tonal harmony progresses predominantly in perfect fifths, the `ChromaDist` viewpoint simply represents the minimum number of perfect fifths required to get from one root to the next, or the smallest distance around a cycle of fifths, with -1 representing the NC case. All of these viewpoints return the undefined symbol, $\perp$, for the first event of a piece when the previous event does not exist. `RootIntFiP`, `MeeusIntFiP`, and `ChromaDistFiP`, apply `RootInt`, `MeeusInt` and `ChromaDist` to the current event and the first event of the piece instead of the previous event. Finally, a *threaded viewpoint* (see [7]), `RootInt` ⊖ `FiB`, measures `RootInt` between chords on the first beats of successive bars.

Three viewpoints are derived from `ChordType`, allowing chord types to be categorised in a number of ways. `MajType` assigns a 1 to all chords where the third is major, a 2 to all chords where the third is minor and a 0 to all chords without a third. `7Type` assigns a 1 to all chords with a minor 7th, and a 0 to all other chords, (except a *NC* which is given a -1 symbol.) `FunctionType` assigns all chords with a major third and minor seventh a 0 (dominant chords), all other chords with a major third a 1 (major tonics), all chords with a minor third and minor seventh a 2 (pre-dominant), all other minor chords a 3 (minor tonic), and *NC* a -1. Table 1 summarises all of the harmonic viewpoints presented in this section over a sample chord sequence.

---

[1] See [10] for a detailed explanation of chord type simplification.

	Bm^7	D^7	NC	GM^7
Root	11	2	-1	7
ChordType	min	7	NC	maj
PosInBar	0	0	2	0
RootInt	⊥	3	-1	-1
MeeusInt	⊥	-1	-3	-3
ChromaDist	⊥	3	-1	-1
RootIntFiP	⊥	3	-1	8
MeeusIntFiP	⊥	-1	-3	1
ChromaDistFiP	⊥	3	-1	4
RootInt ⊖ FiB	⊥	3	⊥	5
MajType	0	1	-1	1
7Type	1	1	-1	0
FunctionType	2	0	-1	1

Table 1. Sample chord sequence with basic and derived viewpoints.

3.2 System Description

A fully detailed model description is beyond the scope of this paper, however, broadly the system follows the IDyOM model [16], branching from the publicly available LISP implementation [1]. The system estimates probabilities of sequences of events in a basic event space ξ with viewpoints, τ, operating over sequences formed from elements of a viewpoint alphabet $[\tau]$. Formally, a viewpoint modelling a type τ comprises of a partial function $\Psi_\tau : \xi^* \rightharpoonup [\tau]$, a *type set* $\langle \tau \rangle$ specifying the basic types the viewpoint is capable of predicting, and a PPM* model trained from sequences in $[\tau]$. In order to make predictions over the basic event space ξ, symbols are converted back from $[\tau]$ with the inverse function $\Psi' : \xi^* \times [\tau] \to 2^{[\tau_b]}$ where τ_b is the basic type associated by τ. This many-to-one mapping means that a single derived sequence can represent multiple basic event sequences.

Long-term (LTM) and short-term (STM) models [7] are used to capture both the general trends of the style modelled and the internal statistical structure of the piece being processed. An LTM consists of the full training set, whilst the STM is built incrementally from the current piece and is discarded after it has been processed. Predictions from all viewpoints within the LTM/STM are combined first, before combining the LTM and STM predictions. Prediction combination is achieved with a weighted geometric mean [17], favouring the least uncertain models according to their Shannon entropy.[2] Various smoothing methods are employed, allowing novel symbols to be predicted and predictions from different length contexts to be combined in a meaningful way without assuming a fixed order bound [20].

Multiple viewpoint systems are assessed quantitatively with methods from information theory [13]. The main performance measure is mean information content $\bar{h}$, representing the number of bits required on average to represent each symbol in the sequence of length J (1).

---

[2] For reference, all model combinations in this paper are achieved with an LTM-STM bias of 7 and a viewpoint bias of 2 see [17] for details.

$$\bar{h}(e_1^J) = -\frac{1}{J} \sum_{i=1}^{J} log_2\, p\big(e_i \mid e_{i-n+1}^{i-1}\big) \qquad (1)$$

4. USING ZERO-ORDER STATISTICS TO WEIGHT Ψ'

The focus of this paper is to improve predictions from derived viewpoints by weighting probabilities after the inverse mapping function Ψ' has been applied. Firstly, it is useful to show in detail cases where certain derived viewpoints would be poor predictors for a basic attribute.

Where a derived viewpoint maps an element onto a large number of basic elements, a certain amount of information is lost by dividing the probability mass uniformly. Suppose a prediction from MajType returns a high probability for a major chord, mapping onto a *'7'*, *'M7'*, *'6'*, *'alt'* or *'aug'* ChordType. *'7'* and *'M7'* chords are very common, whilst *'alt'* and *'aug'* chords are comparatively rare. Since MajType must distribute probability mass equally to all five of these basic elements, a considerable amount of information is lost and it remains a poor predictor of ChordType. The predictive strength of these kinds of viewpoints are to generalise data which will become sparse, specifically in sequence prediction when matching contexts in the PPM* model. This strength is likely to be reduced by the uniform distribution of probability mass and could make these viewpoints poor predictors; returning high mean information content estimates and remaining unselected in viewpoint selection.

A general approach to counter this loss of information is to weight probabilities with the zero-order (unigram) frequencies when distributing probability mass from a derived element to the relevant basic elements. For reference, (2) shows a probability estimate of a basic element, $p(t_{\tau_b})$, calculated by uniformly distributing the probability mass of a derived element, $p(t_\tau)$, following [17]. B represents the set of basic elements that are mapped onto from the derived element t_τ. The proposed alternative, shown in (3), uses probabilities from the zero-order model $p_0(t_{\tau_b})$ to weight the distribution of probability mass from t_τ to t_{τ_b}. As with PPM* predictions, probability mass must be reserved for unseen symbols in the basic element alphabet, so a smoothing method and -1th order distribution is utilised. Using an established smoothing framework [20], (4) shows an interpolated smoothing method with escape method C, an order bound of 0 and with no update exclusion. $c(t_{\tau_b})$ is the number of times the symbol t_{τ_b} occurs the training set, J is the length of the training set, $[\tau_b]$ is the alphabet of the basic viewpoint, and $[\tau_b]^s$ the observed alphabet of the basic viewpoint.

$$p(t_{\tau_b}) = \frac{p(t_\tau)}{|B|} \qquad (2)$$

$$p_w(t_{\tau_b}) = p(t_\tau)\frac{p_0(t_{\tau_b})}{\sum_{i \in B} p_0(i)} \qquad (3)$$

$$p_0(t_{\tau_b}) = \frac{c(t_{\tau_b})}{J + |[\tau_b]^s|} + ...$$

$$\frac{|[\tau_b]^s|}{J + |[\tau_b]^s|} \cdot \frac{1}{|[\tau_b]| + 1 - |[\tau_b]^s|} \quad (4)$$

A demonstration of this process is shown in Figure 1. `FunctionType` is used to predict the next `ChordType` symbol with an LTM model given the context *Am7, D7, Bm7, Bbm7*. The top chart shows a strong expectation of a pre-dominant chord which could map onto a *m7*, *halfdim*, or *dim* `ChordType`. With an unweighted Ψ' (2) from `FunctionType` to `ChordType`, these three basic elements are all given equal probability (middle chart). However, since *m7* is far more common than *halfdim* and *dim*, a more accurate probability distribution could be one weighted (3) by the zero-order frequencies (bottom chart), assigning a high probability to *m7*. This approach allows the powerful generalisation of derived viewpoint models to be combined efficiently with more specific predictions from the basic viewpoint.
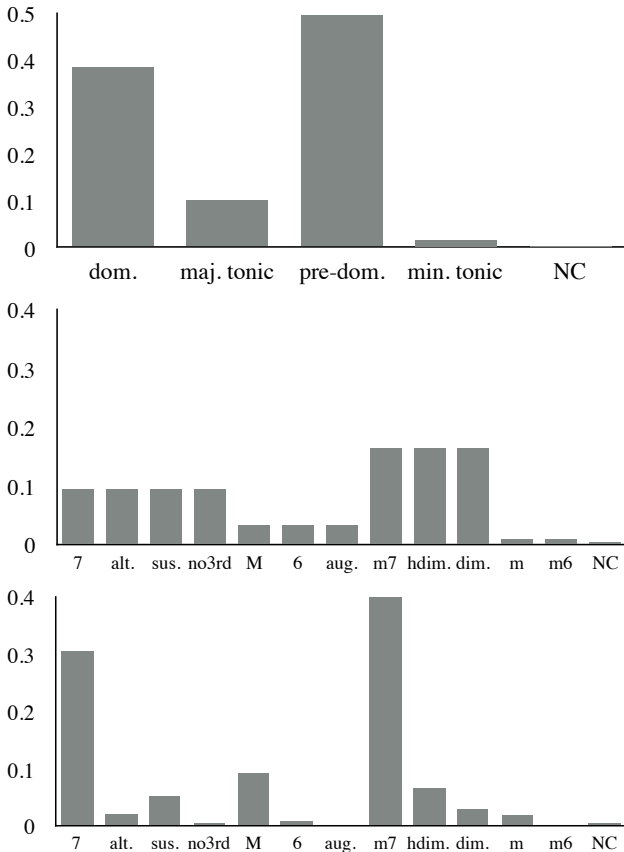
Figure 1. Top: probability distribution of `FunctionType` following the context *Am7, D7, Bm7, Bbm7*. Middle and bottom: probability distributions for `ChordType` predicted by `FunctionType` with an unweighted (middle) and zero-order weighted Ψ' (bottom).

5. TESTING THE IMPACT OF WEIGHTING Ψ'

To investigate the effect of weighting Ψ'_τ with a zero order model, the mean information content, $\bar{h}$ (1), is used as a performance metric to compare predictions with the weighted and unweighted inverse mapping function. In all cases, $\bar{h}$ is calculated with a 10-fold cross-validation of the corpus. The effect of the weighting on individual derived viewpoints is observed first (Section 5.1) before comparing the impact on full multiple viewpoint systems (Section 5.2). The STM is an unbounded interpolated smoothing model with escape method D using update exclusion, and the LTM an unbounded interpolated smoothing model with escape method C without update exclusion [20]. These parameters have been found to be optimal for the current corpus [10].

For the individual viewpoints, it is expected that derived viewpoints which abstract heavily from their basic viewpoint will benefit most from weighting Ψ'. Typically, these are viewpoints derived from `ChordType`, for example, `MajType` reduces the alphabet of `ChordType` from 13 down to 3. By contrast, it is expected that the impact of weighting Ψ' will be far smaller for derived viewpoints with a close to one-to-one mapping between alphabets (e.g. `RootInt`), if significant at all. When constructing a full multiple viewpoint system it is hoped that weighting Ψ' will help more derived viewpoints to be selected over basic viewpoints. Not only should this give a lower mean information content, but also produce a more compact viewpoint model. Successful derived viewpoints should abstract information away from basic viewpoints onto smaller alphabets without a loss in performance.

5.1 Individual Viewpoints Results

Six derived viewpoints for predicting `Root` and `ChordType` are chosen for testing, as well as the basic viewpoints themselves for reference. Table 2 shows the mean information content calculated using both weighted and unweighted Ψ' functions. Effect size measured by Cohen's $d = \frac{\bar{h}_1 - \bar{h}_2}{\sigma_{pooled}}$ across all pieces ($n = 348$) is used to quantify the relative performance for each viewpoint. A one-sided paired *t*-test across pieces assesses statistical significance between the means at the $p < .001$ level, marked with a *.

Strikingly, the derived viewpoints predicting `ChordType` benefit most from the weighting method, all with effect sizes greater than 1.7 and an absolute improvement of around 0.9 bit/symbol. By contrast, the impact of the weighting on the viewpoints derived from `Root` is small and inconsistent, with effect sizes of around 0.1 or less. Indeed, weighting Ψ' has a marginally negative impact on `RootInt`, although only by 0.016 bits/symbol. It is likely that this is because in the majority of cases `RootInt` has a one-to-one mapping with `Root`, except for the *NC* case where a `RootInt` symbol of -1 maps onto the full alphabet of `Root`. It is interesting to note that none of the individual derived viewpoints are able to predict their basic viewpoint better than the

Derived Viewpoint	Unweighted Ψ'	Weighted Ψ'	d
ChordType	1.807	1.807	.000
MajType	3.270	2.315	1.977*
7Type	3.249	2.371	1.766*
FunctionType	3.060	2.080	1.731*
Root	2.259	2.259	.000
RootInt	2.297	2.313	-.030
MeeusInt	3.152	3.076	.129*
ChromaDist	2.688	2.681	.009

Table 2. Predicting `ChordType` (top) and `Root` (bottom) with weighted and unweighted Ψ'. Performance difference is measured by Cohen's $d = \frac{h_1 - h_2}{\sigma_{pooled}}$. * marks differences which are statistically significant at the $p < .001$ level according to a one-sided paired t-test.

basic viewpoint itself, even with a weighted Ψ'. At this point their impact on full multiple viewpoint systems is unknown and must be tested with a viewpoint selection algorithm.

5.2 Viewpoint Selection Results

A viewpoint selection algorithm is a search algorithm to find the locally optimal multiple viewpoint system given a set of candidate viewpoints. Following [17], the current research uses a forward stepwise algorithm which, starting from the empty set of viewpoints, alternately attempting to add and then delete viewpoints from the current system, greedily selecting the best system according to $\bar{h}$ at each iteration. For this study a stopping criteria is imposed such that the new viewpoint system must improve $\bar{h}$ by at least an effect size of $d > .005$, or more than 0.5% of a standard deviation.

Predicting the `Root` and `ChordType` together, given the metrical position in the bar (`PosInBar`), is chosen as a cognitively tangible task for the multiple viewpoint system to perform. In order to predict the two basic attributes simultaneously they are considered as the merged attribute `Root⊗ChordType`. Merged attributes are simply a cross product of basic attributes, equivalent to *linked viewpoints* [7], and have been found to be an effective method for predicting multiple highly correlated basic attributes [10]. An unbounded interpolated smoothing model with escape method C for both STM and LTM is found to be optimal for predicting merged attributes in the current corpus [10], with update exclusion used in the STM only. Using all of the basic and derived viewpoints specified in Section 3.1 and allowing linked viewpoints consisting of up to two constituent viewpoints, or three if one is `PosInBar`, a pool of 64 candidate viewpoints for selection is formed.

The unweighted Ψ' system goes through five iterations of viewpoint addition (without deletion) before termination returning $\bar{h} = 3.037$ (Figure 2). By contrast, the weighted Ψ' system terminates after seven viewpoint additions with a lower $\bar{h}$ of 3.012 (Figure 3). The difference

between these results is found to be statistically significant with a paired one-sided *t*-test at the .001 level ($df = 347$ $t = 5.422$ $p < .001$). However, more importantly, the effect size is found to be small, $d = .026$, owing to the absolute different of .025 bits/symbols between the means. Since the termination criteria is somewhat arbitrary (an appropriate value for d is hand-selected), the unweighted system was allowed to continue up to seven iterations to match the weighted system. This returns $\bar{h} = 3.025$, which is still found to be significantly outperformed by the weighted model ($df = 347$ $t = 3.725$ $p < .001$, effect size $d = .017$).

In the context of the current study the viewpoints chosen from both viewpoint selection runs is highly relevant. The unweighted Ψ' selects only basic viewpoints and viewpoints derived from `Root`. No viewpoints derived from `ChordType` are selected, nor `MeeusInt` or `ChromaDist`. This is to be expected given the findings in Section 5.1, where derived viewpoints with an unweighted Ψ' are found to be poor predictors of `ChordType`. By contrast, during viewpoint selection with a weighted Ψ', linked viewpoints containing `FunctionType` are added on the third and sixth iterations and `MeeusInt` on the fourth iteration. This means that not only does the weighted Ψ' model perform slightly better in terms of $\bar{h}$, but is also more compact since the average viewpoint alphabet size of the seven linked viewpoints selected is 124.4, as opposed to 169 for the unweighted Ψ' model. [3]

6. CONCLUSIONS AND DISCUSSION

This paper has presented a new method for improving predictions from derived viewpoints by weighting Ψ' (the function which maps from the derived to basic alphabet of a viewpoint) with the zero-order frequencies of the basic attribute. Results show that such a weighting significantly improves the performance of derived viewpoints which abstract heavily away from their basic viewpoint, notably `MajType`, `7Type`, and `FunctionType`. On the other hand, viewpoints derived from `Root`, such as `RootInt`, `MeeusInt`, and `ChromaDist`, see only marginal improvements or slight decreases in performance. It has been shown that weighting Ψ' allows more derived viewpoints to be chosen in viewpoint selection. This produces a model which returns a slightly lower mean information content than its unweighted counterpart. This model is also slightly more computationally efficient owing to the smaller alphabet sizes of the selected viewpoints. In practical terms, this creates a model that has a closer fit to the training data whilst taking slightly less time to run for any of the tasks outlined in Section 2 (computational modelling of expectation, segmentation, and automatic music generation).

This paper studied weighting only by zero-order frequency. Useful future research might explore alternative weighting schemes beyond the zero-order frequencies, such as first-order Markov, or even more aggressive, exponential weighting schemes. Furthermore applying the

---

[3] Note that `PosInBar` is a given attribute and so contributes an alphabet size of only 1 during the prediction phase.
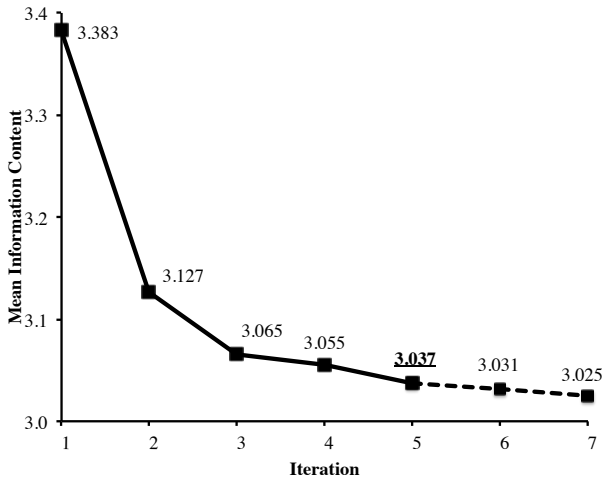
1 + Root ⊗ ChordType ⊗ PosInBar
2 + RootInt ⊗ ChordType
3 + RootIntFiP ⊗ ChordType ⊗ PosInBar
4 + Root ⊗ ChordType
5 + RootInt ⊗ ChordType ⊗ PosInBar
(6 + RootIntFiP ⊗ ChordType)
(7 + RootInt ⊖ FiB ⊗ ChordType)

Figure 2. Viewpoint selection for multiple viewpoint models using unweighted Ψ'. Viewpoints added at each iteration are shown below the graph. Bracketed viewpoints and the dotted line indicate viewpoints added after termination.
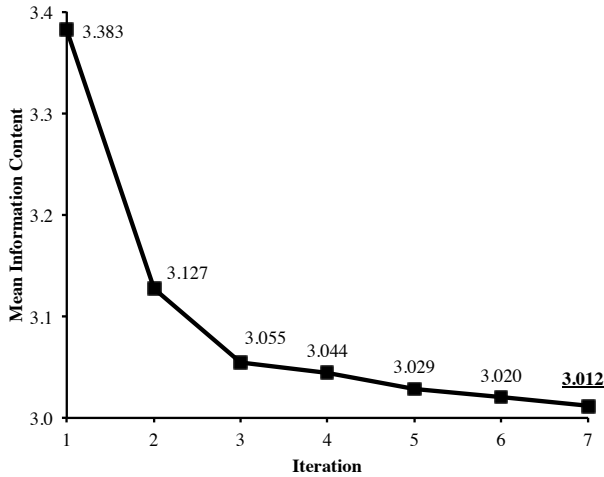
1 + Root ⊗ ChordType ⊗ PosInBar
2 + RootInt ⊗ ChordType
3 + RootIntFiP ⊗ FunctionType ⊗ PosInBar
4 + MeeusInt ⊗ ChordType ⊗ PosInBar
5 + RootIntFiP ⊗ ChordType
6 + RootInt ⊗ FunctionType ⊗ PosInBar
7 + Root ⊗ ChordType

Figure 3. Viewpoint selection for multiple viewpoint models using weighted Ψ'. Viewpoints added at each iteration are shown below the graph.

weighting schemes to a range of domains, genres, and corpora beyond jazz harmony is necessary to prove the methods presented in this paper can be universally applied.

The weighting of Ψ' for derived viewpoints appears to be successful as it combines a more general, abstracted model capable of finding statistical regularities with the more fine-grained model of the basic viewpoint. It could be argued that this is already achieved by multiple viewpoint systems in that they combine predictions from multiple models at various levels of abstraction in an information-theoretically informed manner. However, if the effect of weighting Ψ' with a zero-order model was entirely subsumed by viewpoint combination then almost identical viewpoints would be chosen during the viewpoint selection process, which is not the case (Section 5.2). As the results stand, the weighted Ψ' model selects more derived viewpoints, forming a more compact model and performs slightly better in terms of mean information content.

The compactness of multiple viewpoint systems is relevant both to computational complexity and their relationship with cognitive representations. Searching a suffix tree for the PPM* algorithm with the current implementation using Ukkonen's algorithm [26] is achieved in linear time (to the size of the training data J), but must be done $|[\tau]|$ times to return a complete prediction set over the viewpoint alphabet $[\tau]$, giving a time complexity of $O(J|[\tau]|)$. Selecting viewpoints with a smaller alphabet size has, therefore, a substantial impact on the time complexity for the system. As a model for human cognition [19], selecting viewpoints with smaller alphabets without a loss of performance is equivalent to building levels of abstraction when learning cognitive representations [29].

Additionally, the weighted Ψ' model constructs more convincing viewpoint systems from a musicological perspective. Chord function is an important aspect of jazz music [12] and tonal harmony in general, where common cadences progress in *pre-dominant, dominant, tonic,* patterns. Therefore, the fact that ChordType is selected over MajType and 7Type suggests that chord function as signified by the third and seventh of the chord together is more important than the quality of the third (modelled by MajType) or seventh (modelled by 7type) separately. Similarly, the selection of MeeusInt in the model suggests that functional theories for root progressions may be useful descriptors of tonal harmony. On the other hand, ChromaDist, which considers rising and falling progressions by a perfect fifth equivalent, is not selected. This supports the notion that harmonic progressions in tonal harmony are goal-oriented and strongly directional [8].

7. ACKNOWLEDGEMENTS

8. REFERENCES

[1] https://code.soundsoftware.ac.uk/projects/idyom-project. Accessed: 23-03-2016.

[2] J. Cleary and I. Witten. Data compression using adaptive coding and partial string matching. *Communications, IEEE Transactions on*, 32(4):396–402, 1984.

[3] D. Conklin. *Prediction and Entropy of Music*. PhD thesis, Department of Computer Science, University of Calgary, 1990.

[4] D. Conklin. Representation and discovery of vertical patterns in music. In *IMCAI*, pages 32–42, Edinburgh, Scotland, 2002. Springer.

[5] D. Conklin. Discovery of distinctive patterns in music. *Intelligent Data Analysis*, 14(5):547–554, 2010.

[6] D. Conklin and C. Anagnostopoulou. Comparative Pattern Analysis of Cretan Folk Songs. In *3rd International Workshop on Machine Learning and Music*, pages 33–36, Florence, Italy, 2010.

[7] D. Conklin and I. Witten. Multiple viewpoint systems for music prediction. *Journal of New Music Research*, 24(1):51–73, 1995.

[8] C. Dahlhaus. *Studies on the Origin of Harmonic Tonality*. Princeton University Press, Princetown, NJ, 1990.

[9] S. Griffiths, M. Purver, and G. Wiggins. From phoneme to morpheme: A computational model. In *6th Conference on Quantitative Investigations in Theoretical Linguistics*, Tübingen, Germany, 2015.

[10] T. Hedges and G. Wiggins. The prediction of merged attributes with multiple viewpoint systems. *Journal of New Music Research*, accepted.

[11] H. Leonard. *The Real Book: Volume I, II, III, IV and V*. Hal Leonard, Winoa, MN, 2012.

[12] M. Levine. *The Jazz Theory Book*. Sher Music Co., Petaluma, CA, 1995.

[13] D. Mackay. *Information Theory, Inference and Learning Algorithms*. Cambridge University Press, Cambridge, UK, 2003.

[14] N. Meeus. Toward a post-schoenbergian grammar of tonal and pre-tonal harmonic progressions. *Music Theory Online*, 6(1):1–8, 2000.

[15] F. Pachet, J. Suzda, and D. Martín. A comprehensive online database of machine-readable leadsheets for jazz standards. In *14th International Society for Music Information Retrieval Conference*, pages 275–280, Curitiba, Brazil, 2013.

[16] M. Pearce. *The Construction and Evaluation of Statistical Models of Melodic Structure in Music Perception and Composition*. PhD thesis, City University, London, 2005.

[17] M. Pearce, D. Conklin, and G. Wiggins. Methods for combining statistical models of music. In *CMMR'04: Proceedings of the Second International Conference on Computer Music Modeling and Retrieval*, pages 295–312. Springer-Verlag, 2005.

[18] M. Pearce, D. Mullensiefen, and G. Wiggins. The role of expectation and probabilistic learning in auditory boundary perception: A model comparison. *Perception*, 39(10):1365–1389, 2010.

[19] M. Pearce, M. Ruiz, S. Kapasi, G. Wiggins, and J. Bhattacharya. Unsupervised statistical learning underpins computational, behavioural, and neural nanifestations of musical expectation. *NeuroImage*, 50(1):302–313, 2010.

[20] M. Pearce and G. Wiggins. Improved methods for statistical modelling of monophonic music. *Journal of New Music Research*, 33(4):367–385, 2004.

[21] M. Pearce and G. Wiggins. Expectation in melody: the influence of context and learning. *Music Perception: An Interdisciplinary Journal*, 23(5):377–405, 2006.

[22] M. Rohrmeier and T. Graepel. Comparing feature-based models of harmony. In *9th International Symposium on Computer Music Modeling and Retrieval (CMMR 2012)*, pages 357–370, London, UK, 2012.

[23] S. Sertan and P. Chordia. Modeling Melodic Improvisation in Turkish Folk Music Using Variable-Length Markov Models. In *12th International Society for Music Information Retrieval Conference*, pages 269–274, Miami, FL, 2011.

[24] C. Shannon. A Mathematical theory of communication. *The Bell System Technical Journal*, 27(3):379–423, 1948.

[25] A. Srinivasamurthy and P. Chordia. Multiple viewpoint modeling of north Indian classical vocal compositions. In *International Symposium on Computer Music Modeling and Retrieval*, pages 344–356, London, 2012.

[26] E. Ukkonen. On-line construction of suffix trees. *Algorithmica*, 14(3):249–260, 1995.

[27] R. Whorley. *The Construction and Evaluation of Statistical Models of Melody and Harmony*. PhD thesis, Goldsmiths, University of London, London, 2013.

[28] R. Whorley, G. Wiggins, C. Rhodes, and M. Pearce. 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.

[29] G. Wiggins and J. Forth. IDyOT: A computational theory of creativity as everyday reasoning from learned information. In *Computational Creativity Research: Towards Creative Machines*, pages 127–148. Atlantis Press, 2015.