Comparing Feature-Based Models of Harmony

Martin Rohrmeier[1] and Thore Graepel[2] [*]

[1] Freie Universität Berlin
[2] Microsoft Research Cambridge
mrohrmeier@cantab.net

Abstract. Predictive processing is a fundamental process in music cognition. While there are a number of predictive models of melodic structure, fewer approaches exist for harmony/chord prediction. This paper compares the predictive performance of n-gram, HMM, autoregressive HMMs as well as feature-based (or multiple-viewpoint) n-gram and Dynamic Bayesian Network Models of harmony, which used a basic set of duration and mode features. The evaluation was performed using a hand-selected corpus of Jazz standards. Multiple-viewpoint n-gram models yield strong results and outperform plain HMM models. However, feature-based DBNs outperform n-gram models and HMMs when incorporating the mode feature, but perform worse when duration is added to the models. Results suggest that the DBNs provide a promising route to modelling tonal harmony.

Keywords: Music; Harmony; Graphical Models; n-gram models; Dynamic Bayesian Networks; Cognitive Modelling; model comparison

1 Introduction

"A mind is fundamentally an anticipator, an expectation-generator." (Dennett, 1996: 57) . Prediction and expectancy formation are fundamental features of our cognitive abilities. The ability to form accurate predictions is important from an evolutionary perspective, ranging from interaction, visual and non-visual perception, synchronisation, complex motor action, or complex communication, be it language or music. Likewise musical expectancy is indispensible for a variety of human musical interactions involving perception, attention and emotion, performance, co-ordination and synchronisation, improvisation, dance. Musical styles ground on established ways to play with patterns of expectancy (such as anticipation, suspense, retardation, revision, garden-path sequences, and deceptive sequences) in order to trigger emotions [20],[11] through the autonomous nervous system based reward and other mechanisms linked with predictive systems [12],[43]. Thus modelling human predictive capacities are fundamental for computational cognitive or interactive models of music.

Harmony is one of the core, and at the same time very complex features of Western tonal music. It is a contingent feature of (only) Western music that emerged as an independent structural feature out of modal polyphony around 1650 and constitutes a fundamental feature across nearly all tonal styles, e.g. Classical, Pop, Rock, Latin, or Jazz music. Harmony is governed by the interaction and sequential organisation of chords as discrete musical building blocks that constitute tonality and reflect the formal structure of a piece (such as frequent schemata like the 12-bar blues form or *rhythm changes* in Jazz). In analogy to linguistic syntax, harmony has been argued to exhibit the organisation of a tree structure [15], [14], [45], [42], [5] and was found to be processed in the same (Broca's) area [16]. Since it governs the mid-level organisation of a piece of music, harmony is one of the cornerstones of Western tonal music, notation and its cognition. Hence a rich model of harmony is fundamental for modelling music cognition as well as related music information retrieval tasks such as piece identity/similarity, segmentation, coversong identification, harmonic analysis, searching, indexing, genre classification or generation.

There are plenty of computational models of harmony in the context of computational musical analysis [46],[47] and music information retrieval. This study addresses specifically the problem of harmonic prediction. While much research was done in cognitive studies of melodic and harmonic prediction [48], [43] as well as in modelling melodic prediction (e.g. [30], [29], [26], [27], [13]), comparably little computational work on predictive cognitive modeling of harmony has been done. One successful step to improve predictive n-gram models of melody was the multiple-viewpoint technique [3], [31], which takes different musical features (like duration, onset, scale degree, etc.) into account in order to enhance melodic prediction. This form of feature-based prediction was, however, mostly used for n-gram models and only preliminarily for models of harmony [50]. An approach using similar feature-based HMM or Dynamic Bayesian network models was employed for the problem of automatic chord transcription [19], [28]. DBNs were also successfully employed for note transcription using chords as predictors [33]. Plain n-gram models were used for modelling harmonic structure in a Bach chorale corpus [37], [40], [49], [39], or composer style representation [23]. While different methods for melodic or harmonic n-gram models and representations were compared by [30] and [44], the performance of different types of predictive harmonic models has not been compared. The contribution of this paper is to evaluate plain and feature-based n-gram and graphical models for the prediction of harmony based on a large Jazz corpus.

2   Methods

2.1   Problem setting

Formally, harmonic structure describes a piece of music as a timeseries of discrete, non-overlapping musical building blocks (chords) drawn from a comparably small finite alphabet that feature timing information (metrical structure, onset and duration). The problem of harmony prediction can be expressed as a

common sequence prediction task for strings of discrete, symbolic events. Given a sequence of events e_i, we let e_a^b denote the subsequence ranging from the indices a to b, employing the notation by [30]. The task is easily specified as modeling the predictive probability distribution $p(e_t|e_1^{t-1})$ of each of the single subsequent events e_t in the chord sequence e_t^T given the sequence of past events e_1^{t-1}. For the current modelling endeavour, we are only focussing on the problem of modelling *which* chord is expected rather than *when* it is expected. As outlined above, the complexity and rich structure of tonal harmony renders it a challenging modeling task which is particularly relevant from a cognitive as well as algorithmic generation perspective.

2.2 Representation

One common problem when modelling harmonic structure is the fact that various forms of harmonic representation are used. In Jazz notation as well as in the formal representation of [9], chords are characterised by root, type, degree attributes and bass note. Such a set of features forms a common denominator shared by different representations (except functional theories, e.g. [36]). According to this representation a chords is represented as specialised sets of pitch classes (roots) with features. A chord symbol such as $E\flat\ aug_G^{7,\ 9}$ indicates a (implied fundamental) root, here $E\flat$, a chord type, here *aug* representing an augmented chord, chord attributes, here 7 9, and a bass note, here G representing the pitch class played in the bass. The richness of this representation and number of chord types is in contrast to frequently employed reduced representations using only 24 major and minor triads (or additional diminished chords), as for instance in the 2008 MIREX task, a fact noted by [19], [44] and others. In the present study a chord was only represented by the Cartesian product of root and type (e.g. $C\sharp m$, $D\flat\ dim$, $G\ aug$) since the other attributes were not sufficiently consistent or correct in the corpus (see below). In addition, we used information about chord duration and mode of the piece.

In total, the data set used here consisted of information drawn from the Cartesian product of chord root, type, duration, and mode (major or minor) of the piece. Due to the considerable computational complexity of the DBN models we refrained from using more additional features in order to motivate this study as a proof-of-concept implementation and a baseline comparison. For the representation of the chord root pitch class, correct enharmonic pitch spelling was used (e.g., $C\sharp \neq D\flat$, $F\flat \neq E\natural$) since this more precise information was available in the corpus and since chord (or pitch) function differs depending on enharmonic pitch spelling and context (for instance, a $E\flat$ chord in $D\ minor$ may function as tritone substitute of the dominant or Neapolitan chord, if it is a sixth chord, while a $D\sharp$ chord is relatively rare and does not fulfill such functions in $Dminor$). For this purpose we were employing the base-40 representation [10]. The root alphabet consisted of $\{C\flat\flat, C\flat, C, C\sharp, C\sharp\sharp, D\flat\flat, ...\}$. The chord type alphabet occuring in the corpus was $\{maj, min, dim, hdim, aug, alt, sus\}$. Two additional padding symbols in the alphabet denoted the beginning and end of a piece. Further, chord duration (*dur*) was represented in beats, and *mode* (of the

Fig. 1. Representation of harmony exemplified by the standard "You must believe in Spring" (above in Real Book style). The top part of the table represents the (relevant) harmony information coded in the corpus. Features used for the model comparison are marked by an asterisk.

Chord	Em7b5	Bb7	A7	Dm	*Dm*♯5	Dm6	D7	Gm7	C7sus	C7	Edim/F	FMaj7
bar	1	1	1	2	2	2	2	3	3	3	4	4
beat	0	2	3	0	1	2	3	0	2	3	0	2
root*	E	Bb	A	D	D	D	D	G	C	C	e	F
type*	hdim	maj	maj	min	min	min	maj	min	sus	maj	dim	maj
bass	E	Bb	A	D	D	D	D	G	C	C	F	F
att	(7b5)	7	7	-	♯5	6	7	7	7	7	-	M7
chord*	E hdim	Bb maj	A maj	D min	*Dmin*	D min	D maj	G min	C sus	C maj	E hdim	F maj
dur*	2	1	1	1	1	1	1	2	1	1	2	2

piece) as binary variable (*major / minor*). Altogether an alphabet of 135 chord symbols occurs in the corpus.

2.3   Models

According to the considerations outlined above, we compared three types of models, (i) multiple viewpoint n-gram models that have been very successfully used (mostly for modelling melody), (ii) HMMs which were also commonly used for musical applications. Further we propose (iii) a novel type of graphical model extending the HMMs by a feature-based approach in analogy to the n-gram methodology.

Multiple Viewpoint n-Gram Models Multiple viewpoint n-gram models (feature-based models) were first suggested for the application to music, and in particular to melodic structure, by [3]. The methodology was extended and extensively evaluated by [29]. It constitutes the heart of the information dynamics of music model (IDyOM, www.idyom.org). The idea behind the multiple-viewpoint technique (MVP) is to combine n-gram models for different structural features, in this case features such as duration, metrical structure or mode (of the piece). Combined viewpoint models (such as $chord \otimes mode$) project the prediction space down to the viewpoint to be predicted by means of marginalising over the other viewpoints (see [29],[3] for details). Such n-gram models have to avoid zero counts and to gain confident predictions between the extremes of overfitting (based on too large contexts) and using overly unspecific information (from too short contexts). A large-scale comparison of different smoothing and interpolation techniques [30] found that Witten-Bell smoothing [51], [17] performed best for the case of melodic prediction. Our implementation of harmonic feature-based n-gram models employed Witten-Bell smoothing and mode, duration, chord, root and type viewpoints.

We use $\kappa_{i,n}$ as a shorthand for the predictive context of the n-gram model, i.e. subsequence $e_{(i-n)+1}^{i-1}$. The probability distribution $\hat{p}(e_i|\kappa_{i,n})$ of the predictive event is modelled by the weighted sum of predictions of all available context-lengths (2). The probability $\alpha(\kappa_{i,n})$ is approximated by the count $c(\kappa_{i,n})$ of n-grams of the given context $\kappa_{i,n}$ and adding a zero-escape approximation to the denominator, which amounts to the number of the encountered symbol types $t(\kappa_{i,n})$. The number of encountered symbol types $t(\kappa_{i,n})$ is further used to adjust the weight of the escape count $\gamma(e_i^j)$ to be approximately proportional to the number of symbol types. ζ denotes the alphabet of surface symbols.

$$\kappa_{i,n} := e_{(i-n)+1}^{i-1} \tag{1}$$

$$\hat{p}(e_i|\kappa_{i,n}) = \alpha(e_i|\kappa_{i,n}) + \gamma(\kappa_{i,n})\ \hat{p}(e_i|\kappa_{i,n-1}) \tag{2}$$

$$\gamma(\kappa_{i,n}) = \frac{t(\kappa_{i,n})}{\sum_{c\in\zeta} c(\kappa_{i,n}) + t(\kappa_{i,n})} \tag{3}$$

$$\alpha(\kappa_{i,n}) = \frac{c(\kappa_{i,n})}{\sum_{e\in\zeta} c(e|\kappa_{i,n}) + t(\kappa_{i,n})} \tag{4}$$

Hidden Markov Models Hidden Markov Models (HMM, [32]) are well-known and do not require a detailed introduction. They are successfully applied across domains, including melodic models, harmonisation, harmonic labelling, transcription or audio alignment problems [35], [34], [1], yet not extensively in contexts that involve harmonic prediction. An HMM models a discrete symbol sequence as a series of symbol emissions the distributions of which are controlled by an underlying Markov process representing a number of hidden states by a single discrete random variable. Inference is performed based on maximum likelihood estimate using the Baum-Welch algorithm. The likelihood of a sequence and prediction is computed based on the forward algorithm [32], [22]. We used the implementation provided by Kevin Murphy's Bayes Net Toolbox (BNT,[21]).

Feature-Based Dynamic Bayesian Networks While HMMs and related graphical models have been employed for the case of music and modelling the relationship of notes and chords [24], [25], [27], [28], we employ a graphical generalisation of the multiple-viewpoint idea that combines and to generalise the idea of viewpoints/ feature-based prediction as applied successfully in n-gram models with the flexibility of greater sequential contexts as available to HMM models. This way the model makes use of the hidden state space as well as principled inference over features compared to the heuristic blending used in n-gram models. We build on *state-space models* and Murphy's research on *Dynamic Bayesian Networks* [22]. For applications of chord transcription, [19] used a harmonic DBN conditioning on key and metrical structure as higher-order model in his transcription system. [28] integrated a representation of metre into the transition matrix of an extended HMM model. In our suggested architecture we make use of *mode* and *duration* features such that the current hidden state depends not

only on the previous state but also on mode and/or previous duration (see Figure 2). We further implemented auto-regressive versions of these models which combine the feature-based prediction with conditioning on the previous chord, and hence incorporate the predictive power of bigram models (cf., [22]). The models were implemented using the unrolled *junction tree* inference algorithm within the BNT framework.
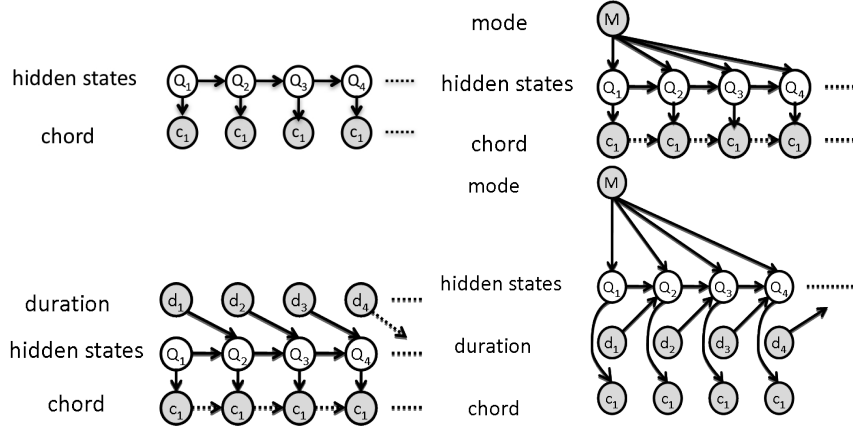
Fig. 2. Architecture of the four types of Dynamic Bayesian Networks, unrolled for 4 time steps. The figure displays the plain Hidden Markov Model (top left), and the structure of its feature-based DBN generalisations using either *mode* (top right), *duration* (bottom left) or both (bottom right). The dotted arrows represent additional auto-regressive versions of these models.

2.4   Evaluation

The corpus consisted of $1,631$ hand-selected Jazz pieces from the Band-in-a-Box (BiaB) corpus [6]. Since the original BiaB corpus contained numerous community-entered pieces which in part contain hundreds of orthographic and syntactic mistakes, the selection was made by a human Jazz expert. The prepared corpus was divided into a *training set* of $1,471$ pieces, which featured $107,505$ chords, and a *testing set* of 160 pieces. The database contained the BiaB metainformation tags, full chord information, chord onset, chord duration as well as song structure. The dataset was preprared from the original BiaB data format based on Mauch's MGU-format reader [18]. The dataset is available online at `www.mus.cam.ac.uk/CMS/people/mr397/`.

For the evaluation, models were trained on the training set and subsequently evaluated on the testing set by predicting each chord of each individual piece. Cross-entropy H_m and perplexity PP_m measures (5, 6) were averaged across the corpus.

$$H_m(p_m, e_1^T) = -\frac{1}{T} \sum_{t=1}^{T} \log_2 p_m(e_t | e_1^{t-1}) \tag{5}$$

$$PP_m(p_m, e_1^T) = 2^{H_m(p_m, e_1^T)} \tag{6}$$

3 Results and Discussion

3.1 Feature-Based n-Gram Models

Figure 3 displays the results. The performance of the simple feature-less n-gram models (operating only on the core chord) illustrates that trigram models perform best, while higher-order models tend to overfit the data. This confirms the findings by [30] for melody in the domain of harmony. In order to have a base-line estimate for the model performance under the impact of knowledge of particular, idiosyncratic sequences (modelling "veridic" tonal knowledge or the partial impact of multiple listening [2], [8]) performance was also computed having the training set in the testing set (also done by [44]). The increase in performance for the veridic evaluations shows that the impact of idiosyncratic sequences is high in 4-grams while 5-gram models are close to be optimal. Perplexity results are similar to/augment results based on cross-entropy. Under all conditions, 3-gram models perform best, while higher-order models tend to overfit. The inclusion of features/viewpoints yields significant improvements over blank n-gram models.[3]

The improvement when adding *mode* is relatively small. However, the duration viewpoint (adding *dur* and *mode* ⊗ *dur*) improves performance remarkably. The mode feature improves performance only slightly. The results thus indicate that n-grams contain mode information to a large extent. This is plausible given that single chords or chord progressions distinguish both modes (e.g. $Dm^{7\flat5}$ is unambiguous $\hat{2}$ of *C minor* while Dm identifies $\hat{2}$ in *C major*, $Fm\ G$ is an unambiguous minor progression). This is consistent with the distinct differences between the top major or minor harmonic n-gram frequencies identified in Bach's chorales [37]. Thus the additional feature does not yield much further enhancement (this explanation is further underpinned by the fact that the improvement of including mode is larger for bigram than for trigram oder higher-order models). In analogy, the combined feature *mode* ⊗ *duration* yields best performance and improves the performance for the duration feature only slightly. The reason why duration improved chord prediction significantly may be that duration is an indicator of chord stability: for instance, shorter chords may occur ornamentally with respect to other chords [42] and, in consequence, behave differently in context than more stable, longer chords.

---

[3] Preliminary tests found (unsurprisingly) that the prediction accuracy is considerably better for the Cartesian product representation of core chords than the viewpoint-based combination of its components (e.g. *root* ⊗ *type*). This confirms the music theoretical notion that the core components root and type are strongly dependent, an effect that cannot be captured by treating them as independent viewpoints.

Fig. 3. N-gram performance results for different plain and feature-based models.
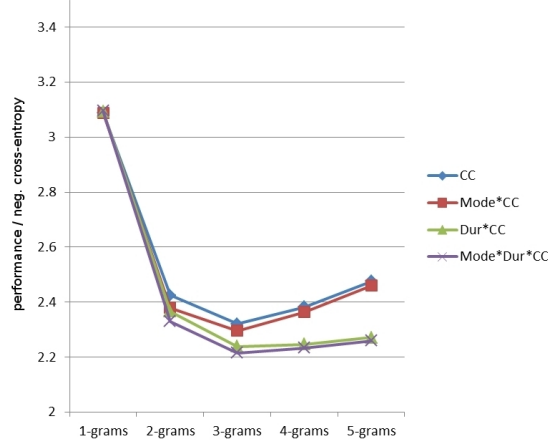

	chord	veridic	chord mode ⊗ chord	dur ⊗ chord	mode ⊗ dur ⊗ chord
1-grams	3.09	3.08	3.09	3.09	3.10
2-grams	2.42	2.36	2.38	2.37	2.33
3-grams	**2.32**	1.97	**2.30**	**2.24**	**2.22**
4-grams	2.38	1.54	2.36	2.25	2.23
5-grams	2.48	1.11	2.46	2.27	2.26

3.2   Hidden Markov Models

Figure 4 displays the performance of the HMM model dependent on the number of hidden states. The HMM performance reaches its best performance at around 2.47 (negative cross-entropy) using 65 hidden states. The mean performance reaching from 25 to 150 hidden states is, however, 2.49. This indicates that a large range of models performes at a similar level and that it infers a comparable amount of information about the chord sequences, independently of whether the inferred information represented in the prior, hidden states, and emission vectors reflects music theoretically established distinctions or not. A large number of hidden states does not improve the performance of the model and extract further harmonic information. With respect to the overfitting problem, HMMs of increasing complexity do not exhibit a comparably strong effect of overfitting as found with n-gram models of increasing context-length: as figure 4 illustrates, the performance decreases comparably slowly for larger numbers of hidden states. Hence HMMs are to some extent more robust with respect to the problem of overfitting. The extent to which the model structure reflects human theoretical knowledge remains to be addressed in further research.

The overall predictive power of HMM models is worse compared with n-gram models. Even the best performing HMMs rank lower than 2-bigram models.
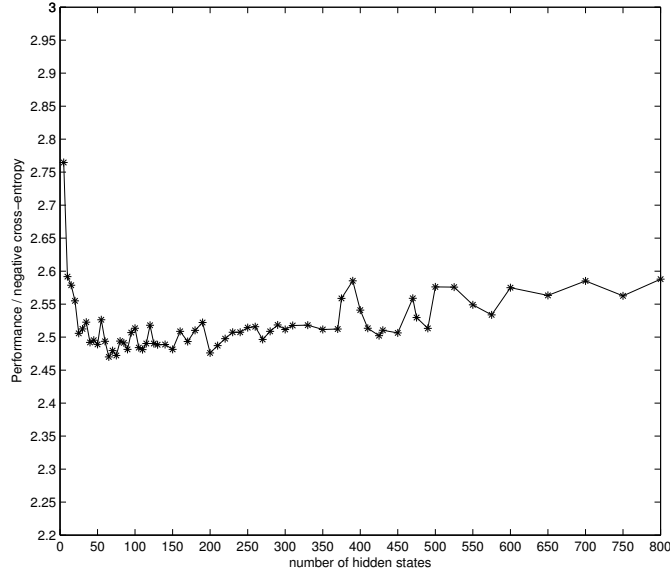
Fig. 4. Performance of plain HMMs.

This finding matches with common results that n-gram models generally tend to outperform other types of models when it comes to prediction accuracy (cf. [17]). The poor performance of HMMs motivates the exploration of enhanced feature-based graphical models in order to improve the predictive power.

3.3 Feature-Based Dynamic Bayesian Networks

The intention behind the feature-based DBNs was to combine the strength of HMM-based sequence modeling with incorporating additional feature information. However, such extensions increase the model complexity considerably and hence, only a small fraction of the design space of these types of models could be explored within reasonable time constraints. Since such graphical viewpoint models have not been evaluated for harmony prediction, the current results constitute a proof of concept and provide estimates of their performance. Table 1 displays the results for the different types of candidate architectures. Every time chord symbols were encoded as Cartesian products of their components because preliminary results showed that models in which these features were separated performed worse.

As expected, auto-regressive HMMs (without viewpoints) perform better than bigram models even with small numbers of states. This confirms that the autoregressive conditioning on the previous chord symbol is (at least) equivalent

Table 1. Results for Hidden Markov Models and Dynamic Bayesian Networks incorporating different feature combinations, numbers of hidden states (hid) and optimal HMM performance for reference. Performance is represented in terms of negative cross-entropy and perplexity.

HMM   **DBNs**

chord			$mode \otimes chord$			$dur \otimes chord$			$dur \otimes mode \otimes chord$		
hid	-CE	PP	hid	-CE	PP	hid	-CE	PP	hid	-CE	PP
30	2.51	6.12	30	2.42	5.79	30	2.36	5.57	30	**2.35**	5.65
90	2.48	5.90	90	2.28	5.34	90	**2.29**	5.55	130	2.41	7.48
130	2.48	5.96	130	**2.26**	5.45						
best											
65	**2.47**	5.91									

Auto-regressive models

chord			$mode \otimes chord$			$dur \otimes chord$		
hid	-CE	PP	hid	-CE	PP	hid	-CE	PP
10	**2.36**	5.53	10	2.28	5.51	15	2.26	5.50
15	2.36	5.53						
25	2.36	5.53						
50	2.36	5.53						

to a bigram model. The better performance indicates that they encode relevant information in the hidden states, yet to a limited extent: the performance does not increase as the number of hidden states increases and they do not reach the performance of 3-gram models. Accordingly, this extension is found not to be advantageous for the HMM.

In contrast, when mode is added as a feature, the DBN outperforms both the mode n-gram models as well as the plain HMM and improves performance with increasing number of states. This indicates that the adding of the mode feature strongly increases the performance and that the Dynamic Bayesian Network extracts further distinctive features of major and minor modes than both other model types. Again this may imply that the model draws information from the longer available context. In contrast to the plain models, the auto-regressive extension of this DBN in turn does not yield additional improvement. Yet conditioning on mode renders the feature-based DBN the best performing model for this feature.

For the present sample computations, the additional duration feature conditioning on the hidden states raises the model complexity drastically but does not yield an improvement of performance compared with the mode feature or the plain HMM (the improvement achieved by auto-regressive conditioning, however, provides a hint that this complex type of model performs as good as the (best) yet simpler mode-featured DBN). Nonetheless none of the DBN models can make comparably efficient use of the additional information embodied in chord durations as multiple-viewpoint n-gram models (reaching a performance of 2.24 and 2.22, see Figure 3). We assume that the comparably weak performance of the

model was due to the fact, that a proportion of the duration values was sparse and that the implementation based on the BNT toolbox lacked specific methods to deal with this problem (to be addressed in future model development). Not surprisingly, the DBN model which combined both mode and duration information performed at the level of the respective multiple-viewpoint bigram models, its performance was, however, much lower than the trigram model performance. This is likely to be due to the same sparsity problem of the distribution of duration values.

Altogether the results show that the inclusion of additional feature information enhances the prediction (and modelling) of harmony. When only chord information was used, n-gram models strongly outperformed HMM models. Autoregressive HMMs, however, achieved a performance improvement over bigram models. When using all available information (chord, mode and duration), n-gram models performed best, which may be explained by the application of enhanced smoothing methods dealing with data sparsity in duration values. Using only mode information, DBN and autoregressive DBN models performed best. This suggests that the combination of combining feature information (without sparsity) and the longer available context yields predictive performance better than n-gram models. This finding suggests that the further development of more advanced feature-based DBN models may provide a promising flexible type of a graphical predictive, cognitive model.

4   Conclusion

The paper presented a comparison of a set of feature-based n-gram, HMM and graphical feature-based Dynamic Bayesian Network models with repsect to predictive modelling of complex harmonic structure in music. These models are important for computational, cognitive and descriptive approaches to music as well as practical applications in terms of generation or real-time interaction. A model comparison showed harmonic prediction improved taking mode or duration information into account. The improvement using mode, however, was comparably small reflecting that differences between both modes are to an extent already reflected in short chord fragments [37]; thus incorporating a dynamic mode feature may be expected to yield only small further improvment. This underpins that harmonic structure is governed not only on mere chord information but also temporal, key or potentially other features. Multiple-viewpoint n-gram models [3],[29] produced best results when duration information is utilised. When only using mode information, however, they are outperformed by feature-based DBNs. These proof-of-concept evaluations illustrate that feature-based DBNs combining the feature approach with the HMM architecture constitutes a promising avenue for further predictive and cognitive modelling of harmony. However, in order to arrive at rich cognitive models further refinements of feature-based model types are required in order to incorporate inference and prediction of metrical structure, dynamic features (like key or mode changes) as well as higher-order features like scale-degree or tonal function (similarly to higher-order melodic viewpoints,

[3], [29]). Moreover, from a theoretical perspective the organisation of harmonic sequences was argued to be hierarchical and exceed simple local Markovian dependencies [45], [38], [40], [42], which in turn may suggest that computational models of similar complexities would yield further improvement.

Ultimately the cognitive aspect of the prediction task requires human baseline measures since the criterion of optimality for the computational models is not most accurate prediction but similar behaviour as human minds – otherwise, intentional musical effects of unpredictable structures like harmonic garden-path or surprise sequences could not be modeled. Future comparison with human experimental results from priming or event-related potential studies may yield further insights into to cognitive adequacy of such probabilistic models and human predictive information processing [43].

References

1. Allan, M., Williams, C. K. I.: Harmonising chorales by probabilistic inference. Advances in Neural Information Processing Systems 17 (2005)
2. Bharucha, J.J., Krumhansl, C.L.: The representation of harmonic structure in music: Hierarchies of stability as a function of context. Cognition 13, 63–102 (1983)
3. Conklin, D., Witten, I.: Multiple viewpoint systems for music prediction. Journal of New Music Research 24(1), 51–73 (1995)
4. Dahlhaus, C.: Untersuchungen über die Entstehung der harmonischen Tonalität. Bärenreiter, Kassel (1967)
5. De Haas, W.B., Rohrmeier, M., Veltkamp, R., Wiering, F.: Modeling harmonic similarity using a generative grammar of tonal harmony. In: 10th International Society for Music Information Retrieval Conference (ISMIR 2009), pp. 549–554 (2009)
6. DeHaas, W.B., Veltkamp, R., Wiering, F.: Tonal pitch step distance: a similarity measure for chord progressions. In: 8th International Conference on Music Information Retrieval (ISMIR) (2008)
7. Dennett, D.: Kinds of Minds. Basic Books, New York (1996)
8. Eerola, T.: The Dynamics of Musical Expectancy. Cross-Cultural and Statistical Approaches to Melodic Expectations. PhD thesis, University of Jyväskylä (2003)
9. Harte, C., Sandler, M., Abdallah, S.A., Gmez, E.: Symbolic representation of musical chords: A proposed syntax for text annotations. In: Proceedings of the 6th International Conference on Music Information Retrieval (ISMIR), pp. 66–71 (2005)
10. Hewlett, W.B.: A base-40 number-line representation of musical pitch notation. Musikometrika 4,1–14 (1992)
11. Huron, D.: Sweet Anticipation: Music and the Psychology of Expectation. MIT Press, Cambridge, Massachusetts (2006)
12. Koelsch, S.: Towards a neural basis of music-evoked emotions. Trends in Cognitive Sciences 14(3), 131–137 (2010)
13. Lavrenko, V., Pickens, J.: Polyphonic music modeling with random fields. In: Proceedings of ACM Multimedia, Berkeley, CA (2003)
14. Lerdahl, F.: Tonal Pitch Space, Oxford University Press, New York (2001)
15. Lerdahl, F. and Jackendoff, R.: A Generative Theory of Tonal Music. MIT Press, Cambridge, MA (1983)
16. Maess, B., Koelsch, S., Gunter, T.C., Friederici, A.D.: Musical syntax is processed in broca's area: An meg study. Nature Neuroscience 4(5):540–545 (2001)

17. Manning, C.D., Schütze, H.: Foundations of Statistical Natural Language Processing. MIT Press, Cambridge, MA:, 1999.
18. Mauch, M., Dixon, S., Harte, S., Casey, M., Fields, B.: Discovering chord idioms through Beatles and Real Book songs. In: 8th International Conference on Music Information Retrieval (ISMIR) (2007)
19. Mauch, M., Dixon, S.: Simultaneous estimation of chords and musical context from audio. IEEE Transactions on Audio, Speech, and Language Processing 18(6),1280–1289 (2010)
20. Meyer, L.B.: Emotion and Meaning in Music. University of Chicago Press, London (1956)
21. Murphy, K.: The bayes net toolbox for matlab. Computing Science and Statistics 33 (2001).
22. Murphy, K.: Dynamic Bayesian Networks: Representation, Inference and Learning. PhD thesis, UC Berkeley (2002)
23. Ogihara, M., Li, T.: N-Gram chord profiles for composer style identification. 9th International Conference on Music Information Retrieval (ISMIR), pp. 671–676 (2008)
24. Paiement, J.F., Eck, D., Bengio, S.: A probabilistic model for chord progressions. In 6th International Conference on Music Information Retrieval (ISMIR) (2005)
25. Paiement, J.F.: Probabilistic Models for Music. PhD thesis, Ecole Polytechnique Fdrale de Lausanne (2008).
26. Paiement, J.F., Bengio, S., Eck, D.: Probabilistic models for melodic prediction. Artificial Intelligence 173(14), 1266–1274 (2009)
27. Paiement, J.F., Grandvalet, Y., Bengio, S.: Predictive models for music. Connection Science 21(2-3), 253–272 (2009)
28. Papadopoulos, H., Peeters, G.: Joint estimation of chords and downbeats from an audio signal. IEEE Transactions on Audio, Speech, and Language Processing 19(1),138–152 (2011)
29. Pearce, M.T.: The construction and evaluation of statistical models of melodic structure in music perception and composition. PhD thesis, City University, London(2005)
30. Pearce, M.T., Wiggins, G.A.: Improved methods for statistical modelling of monophonic music. Journal of New Music Research 33(4), 367–385 (2004)
31. Pearce, M.T., Wiggins, G.A.: Expectation in melody: The influence of context and learning. Music Perception 23(5), 377–405 (2006)
32. Rabiner, L.R.: A tutorial on hidden markov models and selected applications in speech recognition. Proceedings of the IEEE 77:257–286 (1989)
33. Racyński, S., Vincent, E., Bimbot, F., Sagayama, S.: Multiple pitch transcription using DBN-based musicological models. In: 10th International Conference on Music Information Retrieval (ISMIR) (2009)
34. Raphael, C.: Music plus one and machine learning. In: Machine Learning, 27th International Conference (ICML), (2010)
35. Raphael, C., Stoddard, J.: Functional analysis using probabilistic models. Computer Music Journal 28(3), 45–52 (2004)
36. Riemann, H.: Vereinfachte Harmonielehre; oder, die Lehre von den tonalen Funktionen der Akkorde. Augener (1893)
37. Rohrmeier, M.: Towards modelling movement in music: Analysing properties and dynamic aspects of pc set sequences in Bach's chorales. Darwin College Research Reports 04, University of Cambridge, www.darwin.cam.ac.uk/dcrr/ (2006)
38. Rohrmeier, M.: A generative grammar approach to diatonic harmonic structure. In: Spyridis et al. (eds.) 4th Sound and Music Computing Conference, pp. 97–100 (2007)

39. Rohrmeier, M.: Modelling dynamics of key induction in harmony progressions. In: Spyridis et al. (eds.) 4th Sound and Music Computing Conference, pp. 82–89 (2007)
40. Rohrmeier, M., Cross, I.: Statistical Properties of Harmony in Bach's Chorales. 10th International Conference on Music Perception and Cognition (ICMPC), pp. 619–627 (2008)
41. Rohrmeier, M.: Implicit learning of musical structure: Experimental and compuational modelling approaches. PhD thesis, University of Cambridge (2010)
42. Rohrmeier, M.: Towards a generative syntax of tonal harmony. Journal of Mathematics and Music 5(1), 35–53 (2011)
43. Rohrmeier, M., Koelsch, S.: Predictive information processing in music cognition. a critical review. International Journal of Psychophysiology 38(2), 164–175 (2012)
44. Scholz, R., Vincent, E., Bimbot, F.: Robust modelling of musical chord sequences using probabilistic n-grams. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 53–56 (2009)
45. Steedman, M.J.: The blues and the abstract truth: Music and mental models. In: Garnham, A., Oakhill, J. (eds.) Mental models in cognitive science, pp. 305–318. Erlbaum, Mahwah, NJ (1996)
46. Temperley, D.: The Cognition of Basic Musical Structures, MIT Press, Cambridge, MA (2001)
47. Temperley, D.: Music and probability, MIT Press, Cambridge, MA (2007)
48. Tillmann, B.: Implicit investigations of tonal knowledge in nonmusician listeners. Annals of the New York Academy of Science 1060, 100–110 (2005)
49. Tymoczko, D.: Function Theories: A Statistical Approach. Musurgia 10 (3-4), 35–64 (2003)
50. Whorley, R., Wiggins, G.A., Rhodes, C.S., Pearce, M.T.: Development of techniques for the computational modelling of harmony. In: Ventura et al. (eds.) International Conference on Computational Creativity, Lisbon (2010)
51. Witten, I.H., Bell, T.C.: The zero-frequency problem: estimating the probabilities of novel events in adaptive text compression. IEEE Transactions on Information Theory 37(4), 1085–1094 (1991)