EVALUATING RULE- AND EXEMPLAR-BASED COMPUTATIONAL APPROACHES FOR MODELLING HARMONIC FUNCTION IN MUSIC THEORY PEDAGOGY

Johanna Devaney[1] and Daniel Shanahan[2,]

[1]*Ohio State University*

[2]*Louisiana State University*

Correspondence should be addressed to: devaney.12@osu.edu

Abstract: One of the goals of many undergraduate theory curricula is for students to move beyond simply assigning Roman numerals to vertical sonorities toward a higher-level understanding of harmonic function. The "phrase model" provides music theory students with the tools for identifying tonic, predominant, and dominant function at the phrase level and thus acts as a building block for achieving a higher-level understanding of tonal music. It is described in Steve Laitz's textbook *The Complete Musician* through a combination of written rules and labelled musical examples. Computational modelling provides an opportunity to evaluate the relative contributions of these two pedagogical methods; the rules outlined in the texts were converted into functional algorithms and the musical examples were modelled with an exemplar-based probabilistic approach, namely a hidden Markov model (HMM). A third model was also developed, which combined the Rule- and Exemplar-based models. When evaluated on both the textbook examples and exercises from the corresponding workbook, the Rule- and Exemplar-based models exhibited different strengths and the Combined model outperformed both. This indicates that both approaches have unique information that is necessary to produce an appropriate phrase-level analysis from Roman numerals.

1. INTRODUCTION

This paper describes the implementation of a phrase-level function model of Western harmony, specifically the model described by Laitz in his textbook *The Complete Musician* [1]. Musical phrases are built on the harmonic functions of tonic, pre-dominant and dominant, which, when presented in ordered form, generate a complete musical statement. In *The Complete Musician*, Laitz describes his phrase model through written text and annotated musical examples. The goal of this project is to perform a comparative analysis of the relative amount of information that is presented in these two mediums. The information in the text was formalized as a series of rules and the musical examples' annotations were modelled with an exemplar-based approach. The latter was implemented as a single-layer three state (tonic, pre-dominant, dominant) hidden Markov model (HMM), with Roman numerals as observations and transition probabilities learned from a hand-encoded corpus of Laitz's textbook examples.[1] A Combined model, which used the results of the Rule-based model as a prior to guide the Exemplar-based model was also implemented to evaluate whether leveraging both types of information resulted in better performance. The models were tested on both the textbook examples (using five-fold cross-validation in the case of the exemplar-based model) and on the answers to the corresponding workbook, which are available in the instructor manual [2].

This approach was inspired by music theory pedagogy, as educators are often faced with a similar dilemma: should a typical lesson focus primarily on examples from the repertoire, or is it best to enforce clear-cut rules from a text? Ideally, classes would contain both approaches, but a proper examination on the amount of information contained in approach may facilitate future research in music theory pedagogy, and might provide educators with insight into the relative strengths of these two approaches.

[1] The encoded data and code for this project is available for download at http://github.com/jcdevaney/AFURN.

2. PREVIOUS WORK

For decades, symbolic musical notation has been used to facilitate the computational inference of both surface level and deeper level harmonic structures and the analysis of surface-level harmonic sonorities. One significant area of research has been key-finding algorithms, which date back to the early work of Longuet-Higgins & Steedman [3] and continue through today [4]. Our work currently avoids the issue of key finding by working directly from Roman numerals, rather than the musical surface. Although key finding will be an issue in the later phases of this project when the musical surface is used as input, rather than Roman numerals.

The more relevant work for this project is in the area of automatic harmonic analysis of symbolic music notation, which dates back to Winograd's work in the late 1960s [5]. Winograd developed a computational model that provided both a harmonic analysis and a key. His work was later built upon in the rule-based systems for predicting Roman numeral labels from the musical surface, such as those developed by Ulrich [6], Steedman [7], Maxwell [8], Prather [9], Temperley and Sleator [10], Pardo and Birmingham [11] and Terrat [12], and similar examplar-based approaches, such as the work of Raphael and Stoddard [13] and Radicioni and Esposito [14]. Other work has looked more explicitly at phrase-level analysis. While earlier work in this area focused on melodic, rather than harmonic phrases [15, 16], there is now a growing body of work that uses a model phrase-level function in Roman numeral labelling systems, most notably the work by Rohrmeier [17], Granroth-Wilding [18], and Haas and colleagues [19].

Music theory textbooks offer an appealing type of ground truth for computational analysis because they contain expert analyses of graduated musical excerpts. For example, Temperley encoded the analyses from Kostka and Payne's *Tonal Harmony* [20] for key finding [21] and examining statistical properties of harmony [22]. Researchers have also used formalizations available in textbooks, such as Schmuckler use of Piston's "Table of Usual Root Progressions" for an experiment on musical expectation [23].

3. PHRASE MODEL FOR WESTERN ART MUSIC

3.1 General Overview

Phrases are complete musical statements that are built from the ordered presentation of three harmonic functions [1]. Phrases typically contain tonic, pre-dominant, and dominant functions, but sometimes they contain simply tonic and dominant functions. Phrases end with a cadence, either remaining on the dominant function for a half cadence (phrases that end with a V chord) or returning to a tonic function for an authentic cadence (ending with a I or i chord) or a deceptive cadence (ending with a VI or a vi chord). Many ideas about functional harmony can be traced back to the early theoretical works of Rameau [24], although the specification of tonic, pre-dominant, and dominant, appeared later in Riemann's writings [25].

3.2 Functions

The tonic function (T) at the beginning of a phrase serves to establish the tonal centre, and at the end of a phrase to signal its return. The pre-dominant function (PD) prepares for the arrival of the dominant function, although this function may be skipped in shorter phrases. The dominant function (D) sets up an opposition to the tonic, creating a sense of tension that is resolved by the return to the tonic function in phrases that end on an authentic or deceptive cadence.

3.3 Duration

There is a certain amount of flexibility in the amount of time a phrase spends in each function, though typically the tonic function is longer than either the pre-dominant or dominant functions. This is demonstrated in **Table 1**, an example taken from the Laitz textbook [26], where the functions in four different example four-measure phrases are mapped out temporally. In the first example, an equal amount of time is spent in each in-stance of a function. In the fourth example, the tonic function lasts six times longer than either the pre-dominant or dominant functions. This raises the question as to how much the inclusion of duration values influences the model's accuracy, which we address in our comparison of the rule- and exemplar-based models.

Table 1: Examples of typical four-measure phrase examples described by Laitz [26]. Tonic function is represented by T, pre-dominant by PD, and dominant by D. Parentheses indicate the continuation of a function from the previous measure.

	M. 1	M. 2	M.3	M.4	Cadence
1	T	PD	D	T	Authentic
2	T	(T)	PD D	T	Authentic
3	T	(T)	(T) P	D	Half
4	T	(T)	(T)	PD D	Half

4. RULE-BASED MODEL

4.1 Roman Numeral-level Rules

A list of rules was derived from the Laitz text, and was implemented individually over each file using the pattern tool from the Humdrum Toolkit [27]. The rules are detailed in Table 1. Direct labelling (Rules: 1, 5, 7, 8, 12, 14, 16), such as I chords being labelled tonics, IV chords being labelled pre-dominant, or V being labelled dominant, were run first, followed by multiple-chord progressions (Rules: 2, 3, 4, 6, 9, 10, 11, 13, 15, 17, 18), such as a labelling I64 chord facilitate passing motion between a IV and IV6 chord as pre-dominant. Of the fourteen rules, three focused specifically on "I" chords (Rules 1–4), two on "ii" chords (Rules 5–6), one on "iii" (Rule 7), four focused on "IV" chord (Rules 8–11), two focused on "V" chord (Rules 12–13), two focused on "vi" chords (Rules 14-15), and two focused on or "viio" chords (Rules 16–17). One additional rule (Rule 14) addressed the tendency of second inversion chords, which are often part of pedal, passing, and arpeggiated motion, to continue the previous' chord's function. Of these fourteen rules, six of them referred to ascription of tonic function (Rules 1, 5, 6, 8, 10, and 12), five to dominant function (2, 7, 9,11, and 13), and two to pre=dominant function (Rules, 3 and 4). For the sake of simplicity the labels for the chords types refer to the diatonic major chords, but should be considered inclusive of the diatonic minor chords.

4.2 Phrase-level Rules

In a second version of the Rule-based model, each file was run through an rule that ensured that the phrase model was consistently implemented by In order to facilitate this, instances of P or D labels which occurred in between T labels were changed to T, for example: T-T-T-D-T-PD-D-T was changed to T-T-T-T-T-PD-D-T. This was then repeated T or D labels between two PD labels, which were labelled PD, and T or P between two D labels, which were labelled as D. Although many of the original rules (found in Table 2) took such elaborations into account, this served as a way of enforcing such harmonic progressions.

Table 2**:** Function Rules. Note that these rules were executed in the order listed.

	I Chords	
1	Opening and closing I chords	T
2	I^{64} followed by a V chord	D
3	I^{6} chords between a ii and a ii^{6}	PD
4	I^{64} chords between a IV and a IV^{6}	PD
	ii Chords	
5	ii chords	PD
6	ii^{65} chords before or after a I chord	T
	iii Chords	
7	iii chords	T
	IV Chords	
8	IV chords	PD
9	IV chords before or after a I chord	T
10	IV^{6} chords between I and I^{6} chords	T
11	IV^{6} chords between V and V^{6} chords	D
	V Chords	
12	V and V7 chords	D
13	V6 between two I chords	T
	vi chords	
14	vi chords	T
15	vi chords between two V chords	D
	viio Chords	
16	vii^{o} chords	D
17	viio6 between two I chords	T
	Second Inversion Chords	
18	Second inversion chords are assigned the function of the previous sonority	

5. EXEMPLAR-BASED MODEL

The exemplar-based model was implemented with a single-layer hidden-Markov model (HMM) with three states using Kevin Murphy's Toolbox for MATLAB [28].

5.1 States

The three functions of the phrase model (tonic, pre-dominant, and dominant) form the four states of the HMM. As shown in the state space diagram in Figure 1, all phrases begin in the tonic state. The other two states (pre-dominant and dominant) occur in sequence after the initial tonic, although the second state (pre-dominant) may be skipped. A phrase may also end on the dominant state (in the case of a half cadence) or it may end in a return to the tonic (in the case of an authentic or deceptive cadence).

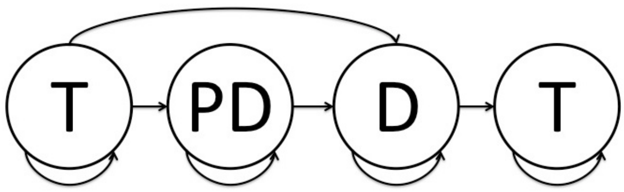

Figure 1: State-space diagram for single-level HMM. Bolded circles indicate possible ending states.

4.2 Observations and Predictions

The observations for the HMM are the Roman numeral labels, while the function labels serve as the predictions. Two different versions of the observations were used, one incorporating duration information and one without. The duration information was incorporated through repetition of the observations; the number of repetitions corresponded to the number of 8th notes that each Roman numeral was held for.

4.3 Textbook Dataset

The textbook dataset used to train the HMM was manually encoded from the Roman numerals and function labels supplied by Laitz for the 126 musical examples that contained phrase-level annotation in chapters 7–14, the section of the book dealing with diatonic harmony. Examples with plagal relationships and dominant anacruses were not included as they are exceptions to the phrase model, as it was defined in the textbook. The data were hand encoded using a modified version of Humdrum's **kern format [27]. Specifically, Roman numerals were annotated in the *chord column using Humdrum's *harm format: standard Roman numerals are used with letters specifying the inversion (a for 1st, b for 2nd, c for 3rd). In the *function column, the functions were labelled as T (tonic), P (pre-dominant) or D (dominant) in the second. Immediately preceding the labels in both columns, were standard **kern rhythmic values. Standard **kern headers, footers and bar lines were also used. The header of the file indicates the composer (COM), the name of the piece (OTL), the page the example occurred on (Laitz) and the meter of the piece. The numeric values in the columns indicate the rhythmic duration of each event (1 for a whole note, 2 for a half note, etc., with a period indicating a dotted note). An example of the encoding is shown in Figure 2. The accuracy of the encoding files were triple-checked by the researchers.

```
!!!COM: Beethoven
!!!OTL: Piano Sonata in D minor, op.31, no.2
!!!Laitz: 230
**kern  **kern
*chord  *function
*M4/4   *M4/4
=-      =-
4i      4T
4Vb     4T
2ia     2T
=       =
1ia     1T
=       =
4iv     4P
4ib     4P
2iva    2P
=       =
1V      1D
==      ==
*-      *-
```

Figure 2: Example encoding of an excerpt from Beethoven's "Piano Sonata in D minor " Op. 31, no. 2.

A distribution of the functions in the non-duration and duration versions of the textbook dataset is shown in Table 3. While the proportions are similar between the functions for each version of the dataset (as further demonstrated by the transition matrices for in Table 4), there is a notable increase in the number of observations with a tonic function in the duration version. This means that simply seeing improvements in the tonic prediction for the duration model over the non-duration model is not sufficient to demonstrate that the duration model is better, but that improvement is the pre-dominant and dominant predictions is necessary.

Table 3: Distribution of the tonic, pre-dominant, and dominant functions in the duration and non-duration versions of the annotated musical examples in textbook dataset.

	T	PD	D
Duration	1537	331	446
(2314)	(67%)	(14%)	(19%)
Non-duration	380	97	125
(602)	(63%)	(16%)	(21%)

Table 4: Transition probabilities for the textbook training set.

Duration	T	PD	D	T
T	0.94	0.05	0.01	0
PD	0	0.80	0.20	0
D	0	0	0.86	0.14
T	0	0	0	1.00

Non-duration	T	PD	D	T
T	0.73	0.20	0.07	0
PD	0	0.33	0.67	0
D	0	0	0.40	0.60
T	0	0	0	1.00

The textbook examples contain 31 unique Roman numerals (shown in Table 5), each a different inversion diatonic major and minor chords. This table also shows the proportional number of times that each Roman numeral occurs in each function, for both the non-duration and duration versions of the data. This table is a useful complement to the rules in Table 2, as it allows for correspondence between the rules derived from and the content of the musical examples. Of the 126 phrases in the complete training set, 97 (77%) end on a tonic function (95 with authentic cadences, ending on I or i, and 2 with deceptive cadences, ending on vi or VI) and 29 (23%) end on a dominant function (all half-cadences, ending on V). The phrase endings were comparable in the reserved training set to those in the complete training set.

Table 5: Roman numerals present in training set, C64 refers to a cadential I64 or i64 chord. The numbers represent a percentage of specified function labels (columns) that a given Roman numeral label (rows) corresponded to, thus each column sums to 100.

	With Duration			**Without Duration**		
	T	***PD***	***D***	***T***	***PD***	***D***
I	43.7	0.0	0.0	36.8	0.0	0.0
I6	7.9	1.8	0.0	9.5	2.1	0.0
I64	0.7	0.0	0.0	1.1	0.0	0.0
C64	0.0	0.0	12.6	0.0	0.0	15.2
i	19.9	0.0	0.0	20.8	0.0	0.0
i6	3.5	0.0	0.0	4.7	0.0	0.0
i64	0.7	2.1	0.0	0.5	4.1	0.0
ii	1.4	10.9	0.0	1.1	10.3	0.0
ii6	1.5	26.0	0.0	2.1	23.7	0.0
ii7	0.0	2.4	0.0	0.0	3.1	0.0
ii65	0.0	3.3	0.0	0.0	3.1	0.0
ii°6	0.7	8.5	0.0	0.5	5.2	0.0
ii°65	0.0	0.6	0.0	0.0	1.0	0.0
III	0.3	0.0	0.0	0.3	0.0	0.0
III6	0.0	0.6	0.0	0.0	1.0	0.0
iii	0.9	0.0	0.0	1.1	0.0	0.0
IV	0.4	24.2	0.0	0.5	21.6	0.0
IV6	0.1	0.0	0.9	0.3	0.0	1.6
iv	0.1	8.8	0.0	0.3	10.3	0.0
iv6	0.0	7.9	0.0	0.0	11.3	0.0
V	3.1	0.0	57.4	3.9	0.0	52.0
V6	0.8	0.0	1.3	1.3	0.0	2.4
V64	1.0	0.0	1.8	1.3	0.0	1.6
v6	0.7	0.0	0.0	1.3	0.0	0.0
V7	3.0	0.0	24.0	2.9	0.0	23.2
V65	2.7	0.0	0.4	2.1	0.0	0.8
V43	2.1	0.0	0.4	2.1	0.0	0.8
V42	1.4	0.0	0.4	1.8	0.0	0.8
VI	0.7	1.2	0.0	0.5	2.1	0.0
vi	2.5	1.8	0.0	2.4	1.0	0.0
vii°6	0.5	0.0	0.7	0.8	0.0	1.6
Total	100	100	100	100	100	100

6. Combined Model

In addition to running the Rules- and Exemplar-based models separately on the data, a combined model was implemented with used the predictions from the Rule-based model as a prior to guide the HMM. This was implemented by adding small values at the locations of the predictions from the Rule-based model to the observation matrix for the HMM that was used for calculating the Viterbi path.

7. Evaluation

As described in section 4, two versions of the Rule-based (R) model were evaluated, one with just the Roman numeral-level rules (R1, see 4.1) and one with both the Roman numeral-level and phrase-level rules (R2, see 4.2). There were also two versions of the Exemplar-based (E) model evaluated, as described in 5, one with duration information (E1) and one without (E2). The possible combination of the R and E models produced four versions of Combined (RE) model (R1+E1=RE1, R1+E2=RE2, R2+E1=RE3, and R2+E2=RE4), resulting in a total of eight models for evaluation (summarized in Table 7).

Table 6: Summary of the eight versions of the Rules- and Exmplar-models evaluated. PME refers to the enforcement of the phrase model in the Rules-model, Dur refers to the inclusions of duration information in the Exemplar-model.

	PME	Dur
R1		
R2	X	
E1		X
E2		
RE1		X
RE2		
RE3	X	X
RE4	X	

7.1 Test Datasets

All of the models were tested on Laitz's textbook examples and workbook exercises. The inclusion of the workbook exercises is a useful proxy for student learning, particularly for the Combined model when the text and musical examples in the textbook were used for training and the workbook is used for testing. For the Exemplar-based and Combined models, five-fold cross validation was used for testing on the textbook dataset, where every fifth example was used for testing and the rest were used for training. This method of sampling was used because the examples in the text are progressive in terms of their complexity and difficulty. Only those textbook examples containing full phrases were included in the test set, 85 of the 126 examples met this criterion. Statistics for the reduced textbook dataset were comparable to those found in Tables Table 3Table 5. The workbook dataset set contains 51 individual phrases from for chapters 7–14 for which Laitz provided Roman numerals and function annotations in the instructor manual. As with the textbook examples, the workbook exercises were encoded using the format described in 4.3 and the encoded were validated in triplicate by the researchers.

A distribution of the functions in the non-duration and duration versions of the work dataset is shown in Table 7, corresponding to the data for the textbook dataset shown in Table 3**.** 32 (63%) phrases ended on a tonic function (all authentic cadences) and 19 (37%) ended on a dominant function (half-cadences). Unlike the textbook dataset, there is no difference in the relative proportions of Roman numerals with tonic, pre-dominant, and dominant functions between the non-duration and duration versions of the workbook dataset, which suggests the results between E1 and E2 should be more similar when tested on this dataset than on the textbook dataset. In comparison to the textbook dataset, the examples in the workbook dataset more of the Roman numeral labels have a tonic function.

Table 7: Distribution of the tonic, pre-dominant, and dominant functions in the duration and non-duration versions of the annotated musical examples in workbook dataset.

	T	PD	D
Duration (1410)	1017 (72%)	146 (10%)	247 (18%)
Non-duration (487)	355 (73%)	52 (11%)	80 (16%)

7.2 Results

The performance of the models (two Rule-based versions, two-Exemplar-based versions, and four Combined versions) are summarized in Table 8, which shows the F-scores by function for each model, and Table 9, which consists of confusion matrices for each of the models' versions. The balanced F-scores provide information about the overall accuracy of the models and the confusion matrices provide detailed information about how misclassification trends.

The inclusion of the phrase-level rules in R2 (detailed in Section 4.2) improved the performance over R1, the version of the Rule-based model that used Roman numeral-level rules (detailed in Section 4.1), for all of the functions in the textbook test set and for the tonic and dominant functions in the workbook test set. There was, however, a reduction in performance for the pre-dominant function in the workbook set, which suggests that the phrase-level rules may be overzealous is re-assigning pre-dominant function labels.

E1, the duration version of the Exemplar-based model, performed better than E2, the non-duration version, across all of the functions for both the textbook and workbook sets. This improvement is likely related to trend related to average duration for how long chords are when they are in their typical function (e.g., a I chord in a tonic function at the beginning of a phrase) versus when they elaborate another function (e.g., a I6 chord between a ii chord and a ii6 chord in a pre-dominant function). This trend also held for the workbook test set.

Both RE1 (duration) and RE3 (phrase-level rules and duration) models performed better on the textbook set than RE2 (no duration) or RE4 (phrase-level rules and no duration), though comparably to each other. This result follows from the better performance of E1 over E2, confirming the importance of duration modelling. RE1 and RE3 also performed better than an either R2 or E1, indicating that there is unique information captured by both the Rule- and Exemplar-based models. On the workbook test set, the inclusion of the phrase-level rules improved the performance of RE3 above that of RE1. This suggests that there may be different phrase-level characteristics between the textbook and the workbook. The phrase-level modelling in the Exemplar-based model substitutes for the phrase-level rules when the training and testing are occurring on different parts of the textbook but it is not as generalized as the phrase-level component of the Rules-based model, and thus performs less robustly on the workbook test set.

Table 8: Balanced F-scores, by function, for the Rule-based (R), Exemplar-based (E), and Combined (RE) models for both the textbook and workbook test sets.

	Textbook			*Workbook*		
	T	PD	D	T	PD	D
R1	91	84	86	86	75	65
R2	95	89	92	93	69	83
E1	95	87	93	91	72	68
E2	92	85	84	89	61	55
RE1	97	91	95	92	74	80
RE2	93	88	86	91	70	68
RE3	96	92	94	95	83	89
RE4	94	92	88	94	81	81

Table 9: Confusion matrices for the Rule-based (R), exemplar-based (E), and Combined (RE) models for the textbook test set. The matrices indicate number of predictions of each function (rows) according to the to the ground truth (columns). The values have been normalized so that each column sums to 100.

Textbook R1

	T	PD	D
T	88	10	5
PD	5	88	2
D	7	2	93

Textbook R2

	T	PD	D
T	97	16	9
PD	1	83	0
D	2	1	91

Textbook E1

	T	PD	D
T	95	9	9
PD	4	91	0
D	1	0	91

Textbook E2

	T	PD	D
T	96	19	25
PD	3	81	0
D	2	0	75

Textbook RE1

	T	PD	D
T	96	5	5
PD	3	95	0
D	1	0	95

Textbook RE2

	T	PD	D
T	96	12	21
PD	3	88	0
D	1	0	79

Textbook RE3

	T	PD	D
T	96	4	8
PD	3	96	0
D	1	0	92

Textbook RE4

	T	PD	D
T	97	8	18
PD	2	92	0
D	1	0	82

Table 10: Confusion matrices for the Rule-based (R), exemplar-based (E), and Combined (RE) models for the workbook test set. The matrices indicate number of predictions of each function (rows) according to the to the ground truth (columns). The values have been normalized so that each column sums to 100.

Workbook R1

	T	PD	D
T	79	13	12
PD	5	83	0
D	16	4	88

Workbook R2

	T	PD	D
T	94	46	10
PD	0	54	0
D	6	1	90

Workbook E1

	T	PD	D
T	99	43	47
PD	0	57	0
D	1	0	53

Workbook E2

	T	PD	D
T	98	54	60
PD	1	46	0
D	1	0	40

Workbook RE1

	T	PD	D
T	96	30	29
PD	3	69	0
D	1	1	71

Workbook RE2

	T	PD	D
T	96	35	44
PD	2	62	0
D	2	4	56

Workbook RE3

	T	PD	D
T	98	26	17
PD	1	74	0
D	1	0	83

Workbook RE4

	T	PD	D
T	98	25	28
PD	1	71	0
D	1	4	72

Overall, all of the models performed better on the textbook test set than workbook test set, particularly for pre-dominant and dominant functions. The confusion matrices provide a means of exploring this discrepancy is more detail.

The confusion matrices for the textbook data in Table 9 show that the inclusion of the phrase-level rules in R2 improved the accurate determination of tonic functions and reduced the confusion between pre-dominant and dominant in both directions. It did, however, increase the confusion of pre-dominant and dominant function as tonic. For the Exemplar-based model, the inclusion of duration information did not improve the correct determination of tonic function but it did reduce the confusion of pre-dominant and dominant functions for tonic. In both cases there was no confusion between pre-dominant and dominant, which did occur within the Rules-based model. The Combined models were able to exploit the robustness of the Exemplar-models for the pre-dominant and dominant distinction, while leveraging the information in both to reduce the confusion with tonic when the duration information was included.

The workbook confusion matrices, shown in Table 10 show that, as with the textbook test, the models are generally robust at distinguishing between pre-dominant and dominant. In contrast, all of the models mislabel a larger number of pre-dominant and dominant functioning chords as tonic on the workbook test set than they did on the textbook test set. Overall, the tendency to over-assign tonic function can be attributed to the larger amount of time that phrases spend in the tonic function. An open question is to how more clearly define when the harmony has moved away from the tonic function to a pre-dominant and dominant function.

8. Conclusions

This paper describes a set of computational implementations of the phrase-level function model of Western harmony (speficially the model described by Laitz in his textbook *The Complete Musician*). The phrase model was implemented as a Rule-based model, derived from the rules explicated in the written text, an Exemplar-based model, trained on the annotated musical examples, and a Combined model, which used the Rule-based model as a prior to guide the Exemplar-based model. Both models were tested on two test sets of individual musical phrases: one containing annotated musical examples in the textbook and the other containing musical exercises from the corresponding workbook for which the author had provided annotation in the instructor manual.

Overall, the Combined model improved on the performance of both the Rule- and Exemplar-based model, with the inclusion of a phrase-level in the Rule-based model and duration information in the Exemplar-based model contributing the more robust performance. The improved performance of the Combined model suggests that both the rules and musical examples contain unique information that is necessary for accurately generating a phrase-level analysis from Roman numerals. Additionally, the differing levels of accuracy for labelling various functions between models suggests that a useful pedagogical strategy might consist of the implementation of a more rule-based approach for certain functions (such as tonic), and a slightly more repertoire-based approach might allow for more difficult cases (such as various types of dominant elaborations). Again, the success of the Combined model demonstrates that neither approach can be used exclusively, but hopefully this work will inform more directed lesson plans for music theory educators in the future.

8.1 Future Work

Going forward, we plan to extend the model to work on more advanced harmonic material, such as applied chords and chromatic harmony, and irregular phrase structures, such as those starting with dominant anacruses or ending with a plagal cadence. We are also going to explore whether the inclusion of metrical position information improves the accuracy of the models, specifically whether it makes them more robust to the over-assignment of tonic function that was prevalent in the workbook test set.

9. ACKNOWLEDGEMENTS

The authors would like to acknowledge the assistance of Kirsten Nisula in the encoding the textbook and workbook datasets. This was supported in part by an award from Google.

REFERENCES

[1] S.G. Laitz, *The Complete Musician*. 3rd ed., Oxford University Press, Oxford, 2011.

[2] S.G. Laitz, *Instructor's Manual to Accompany the Complete Musician*. 3rd ed., Oxford University Press, Oxford, 2011.

[3] H.C. Longuet-Higgins and M.J. Steedman, *On Interpreting Bach.* In *Machine Intelligence*, volume 6: 221–41, 1971.

[4] C. Sapp, *Computational Methods for the Analysis of Musical Structure*. Ph.D. Stanford University, 2011.

[5] T. Winograd, *Linguistics and the Computer Analysis of Tonal Harmony.* In *Journal of Music Theory*, volume 12: 2–49, 1968.

[6] J.W. Ulrich. *The Analysis and Synthesis of Jazz by Computer*. In *Proceedings of the 5th International Joint Conference on Artifcial Intelligence*. Los Altos, CA. 1977.

[7] M. Steedman, *A Generative Grammar for Jazz Chord Sequences.* In *Music Perception*, volume 2(1): 52–77, 1984.

[8] H.J. Maxwell, *An Expert System for Harmonic Analysis of Tonal Music*, in *Understanding Music with Ai.* In M. Balaban, K. Ebcioglu, and O. Laske,(eds.), *Understanding Music with Ai* in volume pages 335–353, AAAI Press: Palo Alto, CA. 1992

[9] R.E. Prather, *Harmonic Analysis from the Computer Representation of a Musical Score.* In *Communications of the ACM*, volume 39(12): 119, 1996.

[10] D. Temperley and D. Sleator, *Modeling Meter and Harmony: A Preference-Rule Approach.* In *Computer Music Journal*, volume 23(1): 10–27, 1999.

[11] B. Pardo and W. Birmingham, *Algorithms for Chordal Analysis.* In *Computer Music Journal*, volume 26(2): 27–49, 2002.

[12] R.G. Terrat, *A Pregroup Grammar for Chord Sequences.* In *Journal of New Music Research*, volume 34(4): 355–60, 2005.

[13] C. Raphael and J. Stoddard, *Functional Harmonic Analysis Using Probabilistic Models.* In *Computer Music Journal*, volume 28(3): 45–52, 2004.

[14] D.P. Radicioni and R. Esposito. *Learning Tonal Harmony from Bach Chorales*. In *Proceedings of the 7th International Conference on Cognitive Modelling*. Trieste, Italy. 2006.

[15] R. Bod, *A Unified Model of Structural Organization in Language and Music.* In *Journal of Artificial Intelligence Research*, volume 17: 289–308, 2002.

[16] D. Temperley, *Cognition of Basic Musical Structures*. MIT Press, Cambridge, MA, 2004.

[17] M. Rohrmeier, *Towards a Generative Syntax of Tonal Harmony.* In *Journal of Mathematics and Music*, volume 5(1): 35–53, 2011.

[18] M. Granroth-Wilding, *Harmonic Analysis of Music Using Combinatory Categorial Grammar*. PhD. University of Edinburgh, 2013.

[19] W.B. Haas, et al., *Automatic Functional Harmonic Analysis.* In *Computer Music Journal*, volume 37(4): 37–53, 2014.

[20] S. Kostka and D. Payne, *Tonal Harmony: With an Introduction to Twentieth Century Music*. McGraw-Hill, New York, NY, 2008.

[21] D. Temperley, *A Bayesian Approach to Key-Finding*, in *Music and Artificial Intelligence*. In C. Anagnostopoulou, M. Ferrand, and A. Smaill (eds.), *Music and Artificial Intelligence* in *Lecture Notes in Computer Science*, volume 2445, pages 195–206, Springer Berlin Heidelberg. 2002

[22] D. Temperley, *A Unified Probabilistic Model for Polyphonic Music Analysis.* In *Journal of New Music Research*, volume 38(1): 3–18, 2009.

[23] M.A. Schmuckler, *Expectation in Music: Investigation of Melodic and Harmonic Processes.* In *Music Perception*, volume: 109–49, 1989.

[24] J.-P. Rameau, *Treatise on Harmony*. Trans, Philip Gossett, (Dover, Toronto, ON). Jean-Baptiste-Christophe-Ballard, Paris, 1722.

[25] H. Riemann, *Harmony Simplified; or, the Theory of the Tonal Functions of Chords.*, Augener, London, 1896.

[26] S.G. Laitz, *Writing and Analysis Workbook to Accompany the Complete Musician*. Oxford University Press, Oxford, 2011.

[27] D. Huron, *The Humdrum Toolkit: Reference Manual*. Center for Computer Assisted Research in the Humanities, Menlo Park, California, 1995.

[28] K. Murphy. *Hidden Markov Model (Hmm) Toolbox for Matlab*, 1998. Available from: http://www.cs.ubc.ca/~murphyk/Software/HMM/hmm.html.