Non-chord Tone Identification Using Deep Neural Networks

Yaolong Ju
Schulich School of Music, McGill University
Centre for Interdisciplinary Research in Music Media and Technology (CIRMMT)
Montréal, Canada
yaolong.ju@mail.mcgill.ca

Nathaniel Condit-Schultz
Schulich School of Music, McGill University
Montréal, Canada
nathaniel.condit-schultz@mcgill.ca

Claire Arthur
Schulich School of Music, McGill University
Montréal, Canada
claire.arthur@mcgill.ca

Ichiro Fujinaga
Schulich School of Music, McGill University
Montréal, Canada
ichiro.fujinaga@mcgill.ca

ABSTRACT

This paper addresses the problem of harmonic analysis by proposing a non-chord tone identification model using deep neural network (DNN). By identifying non-chord tones, the task of harmonic analysis is much simplified. Trained and tested on a dataset of 140 Bach chorales, an initial DNN was able to identify non-chord tones with F1-measure of 57.00%, using pitch-class information alone. By adding metric information, a small size contextual window, and fine-tuning DNN, the model's accuracy increased to a F1-measure of 72.19%. These results suggest that DNNs offer an innovative and promising approach to tackling the problem of non-chord tone identification, as well as harmonic analysis.

CCS CONCEPTS

• **Information systems → Probabilistic retrieval models**; • **Computing methodologies → Neural networks**;

KEYWORDS

Non-chord Tone Identification; Harmonic Analysis; Deep Neural Networks; Machine Learning; Bach Chorales

ACM Reference format:
Yaolong Ju, Nathaniel Condit-Schultz, Claire Arthur, and Ichiro Fujinaga. 2017. Non-chord Tone Identification Using Deep Neural Networks. In *Proceedings of the 4th International workshop on Digital Libraries for Musicology, Shanghai, China, October 28, 2017 (DLfM' 17),* 4 pages.
https://doi.org/10.1145/3144749.3144753

1 INTRODUCTION

Non-chord tones are elaborative notes, created by idiomatic stepwise melodic contours, which do not belong to the local structural harmony. Identifying non-chord tones is essential to many music analytic tasks, including melodic analysis [7], polyphonic music retrieval [12], harmonization [1], and harmonic analysis [10, 11, 16, 17]. Raphael and Stoddard [13], and Mearns [10] addressed the theoretical difficulty and importance of non-chord tone identification for harmonic analysis. Pardo and Birmingham [11] and Sapp [16] integrate non-chord tone identification into their harmonic analysis models. Chuan and Chew [1] and Illescas et al. [7] used decision trees to identify non-chord tones, based on extracted features from notes: the former to facilitate the process of harmonization, and the latter for melodic analysis. Still, few scholars have proposed complete, dedicated non-chord tone identification models.

In this paper, we construct a non-chord tone identification model based on deep neural networks (DNNs), trained on chorale music by J.S. Bach. Machine learning of harmonic analysis is difficult due to the large number of chord classes, which require large amounts of training data to learn. In contrast, the relatively simple task of non-chord tone identification requires much less training data. Once non-chord tones are identified, harmonic analysis becomes a relatively simple task, which can be accomplished by a rule-based algorithm. Deep learning has achieved substantial success in numerous complex tasks, often outperforming shallower models [6] and sometimes even surpassing human performance [4]. DNNs are nonetheless relatively simple, suitable for our preliminary research.

2 DATASET

This project draws on a convenient dataset, Rameau[9], consisting of 156 Bach chorales with expert harmonic annotations.[1] This dataset is problematic in several ways: First, the identity and expertise of the "expert" annotators is unclear—indeed, the quality of the Rameau analyses is inconsistent, with poor interpretations of many of the more contrapuntally complex passages.[2] A second, more fundamental, concern regards the musical sample itself: why use Bach chorales? Applying modern harmonic analysis to Bach's music is anachronistic, as Bach and his contemporaries had a contrapuntal, not harmonic, conception of music. However, Bach's music is widely used in modern harmonic pedagogy and is often treated as the archetype of Western tonality. Indeed, Bach's chorales have

DLfM' 17, October 28, 2017, Shanghai, China

ACM ISBN 978-1-4503-5347-2/17/10…$15.00
https://doi.org/10.1145/3144749.3144753

[1] Available at https://github.com/kroger/rameau/tree/master/rameau-deps/genos-corpus. The *music* directory contains 371 Bach chorales, all in Lilypond format, while the *answer-sheets* directory contains text files that contain chord labels for 156 Bach chorales.

[2] Indeed, the Rameau annotation scheme (described below) is not actually complex enough to adequately capture the details of all such passages.

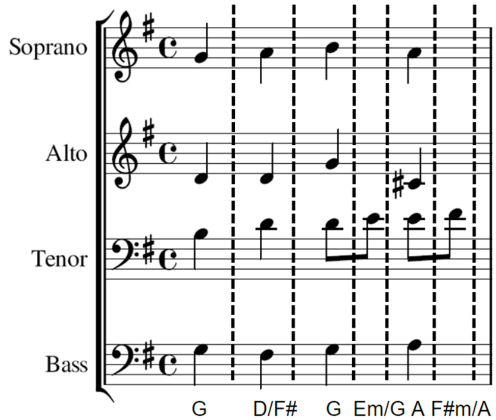

Figure 1: Illustration of *salami slice* temporal parsing of music. Dotted lines indicate boundaries between slices. Harmonic analyses of each slice are shown at the bottom (From BWV 30/6 "Freu dich sehr, o meine Seele").

been used as paradigmatic examples of tonal harmony in a variety of studies [5, 8, 9, 15]. Part of the reason the chorales are widely used is their simple, consistent four-voice musical texture, which affords their usage in empirical research.

The Rameau dataset is only used to train our initial models (reported below). Using our preliminary models, and working closely with music theory experts, we intend to build an improved Bach chorale dataset, including high-quality, harmonic/contrapuntal annotations of all 371 chorales. This dataset will be of interest to many researchers with diverse music theoretic and MIR goals. Nonetheless, models trained on this improved dataset will only serve as a jumping off point for the development of models for more complex music, such as the TAVERN dataset of Mozart and Beethoven piano music[3].

2.1 Dataset Description

Harmonic labels in the Rameau dataset[3] are aligned with the music as *salami-slices* [9, 11]: a "salami-slice" is formed whenever a new note onset occurs in *any* musical voice. Fig. 1 illustrates how the score aligns with the chord labels by slice. As can be seen, each slice in a Bach chorale contains four pitches.

2.2 Dataset Pre-processing

Although the Rameau dataset contains harmonic labels for 156 chorales, 41 analysis files showed misalignment errors with the Lilypond scores, making these analyses unusable.[4] After case-by-case evaluation, we were able to repair alignment issues in 25 files, leaving usable harmonic labels for 140 Bach chorales, totalling 10,564 slices.

In the Rameau dataset most slices are simply labeled with a chord name, while some slices are labeled with []—indicating non-chord

[3] 20 chorales are labeled with Roman numerals (e.g., "I") and chord symbols (e.g., "c") while 136 are labeled with only chord symbols. In this paper, we use chords symbols as labels.

[4] In some cases, errors occurred when annotators used ad-hoc shortcuts, like using "c*2" to refer to two "c" chords in a row.

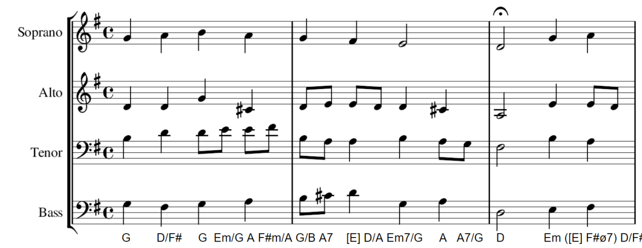

Figure 2: The first three measures of BWV 30/6; Original score and chord labels.

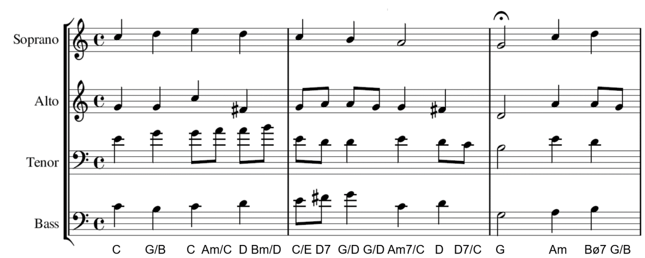

Figure 3: The first three measures of BWV 30/6; Transposed score and translated chord labels.

tones—or ()—indicating alternate interpretations. For example, on the fourth beat of the last measure in Fig. 2, the note (E) could be interpreted as a non-chord (suspension) or as a chord tone (7th of F#ø7). To facilitate processing, and to be consistent with standard analytical labeling, we translated all [] and () labels into appropriate chord labels. To make the tonal relationships between picth-classes consistent across the dataset, we also transposed all the chorales into the same key.

Fig. 3 illustrates the annotations associated with a chorale after pre-processing. Once each slice is consistently labeled with a chord name, chord tones associated with each chord label can be identified using the music21[2] toolbox and subtracted from the slice; any remaining notes are considered as non-chord tones.

3 METHOD

As input to the DNN, each slice is represented by a vector of twelve ones or zeros, representing which pitch classes (C, C#/Db, D, D#/Eb, etc.) are present (1) or absent (0) in the slice. For the DNN's output we use a similar vector of length four, indicating which, if any, of the four voices contains a non-chord tone. These input and output vectors are illustrated in Fig. 4: The third slice in the second measure contains the pitch-classes D, G, A, and B, represented by the input vector [0,0,1,0,0,0,0,1,0,1,0,1]. The corresponding output vector for this slice [0,0,1,0] indicates that the pitch (A), which is the third "1" from the left in the input vector, is a non-chord tone.

Using these input and output representations of the data, we experimented with the number of hidden layers in the DNN (from 2–4) and the number of hidden nodes (from 100–500), finding an optimal performance with 2 hidden layers of 200 nodes each. Other settings were determined empirically (reported in Table 1). We

Figure 4: Non-chord tones in the first three measures of BWV 30/6 (transposed), and the corresponding DNN inputs and outputs.

No. of hidden layers	2
No. of hidden nodes	200
Optimizer	ADAM (Adaptive Moment Estimation)
Loss function	Binary cross-entropy
Training proportion	80%
Validation proportion	10%
Test proportion	10%
Evaluation metric	Precision, recall, F1-measure
Evaluation method	10-fold cross validation

Table 1: The DNN settings.

used an "early stopping" training algorithm to optimize the model's performance.

We tested several modifications to the model's input: First, metric information about each slice was added to each input vector, specifically whether the current slice is on beat (1) or off (0). Second, slices adjacent to the current one were added to the input vector, creating a windowed input that allows the model to consider context. Different window sizes were tested, from a window of three slices (one slice before and one slice after the current slice) up to eleven slices (five before, five after). When the slice is near the beginning or the end, missing values were filled with zeros.

Readers may take issue with the use of pitch-classes (without octave information) and the lack of voice information (bass, tenor, etc.) in the input vector, such that the model is not directly privy to contrapuntal patterns in the music, which are (theoretically) essential to harmonic interpretation. We too expected contrapuntal information to be essential, and experimented with adding this information to the model's input, but found no performance improvement. It seems that, given the simple musical texture of chorales, the model is able to indirectly infer enough information from the pitch-class vectors to achieve adequate performance. As we expand our project to more complex musical datasets (like TAVERN), we expect that detailed octave and voicing information will be needed.

Features	Precision	Recall	F1-measure
PC (D:12)	77.95±2.24%	45.65±9.30%	57.00±7.67%
PC + B (D:14)	78.87±4.97%	46.32±9.60%	57.66±8.01%
PC + B + WS1 (D:42)	**86.02±3.35%**	**63.14±10.81%**	**72.19±7.68%**
PC + B + WS2 (D:70)	85.56±2.55%	59.55±10.58%	69.64±7.54%
PC + B + WS3 (D:98)	82.05±2.41%	57.96±10.51%	67.35±7.47%
PC + B + WS4 (D:126)	81.43±4.22%	54.45±11.03%	64.56±8.48%
PC + B + WS5 (D:154)	79.94±3.24%	53.58±9.08%	63.80±7.39%

Table 2: Model performances with different input features. (PC: pitch-class; B: on/off beat feature; WS: window size; D: number of dimensions). The optimal model is reported in bold text.

4 EVALUATION

Ten-fold cross validation was used for evaluation. However, as noted by Rizo et al. [14], there is a significant imbalance between the number of chord tones and non-chord tones: only about 8% of the dataset are non-chord tones. As a result, raw accuracy is not the best form of evaluation—if the model labeled every note as a chord note, it would still achieved 92% accuracy. Therefore, we report the metrics of precision, recall, and F1-measure. The results are shown in Table 2. As we can see, the baseline system achieved 57.00% F1-measure, while the beat-position input feature added some moderate improvement. When the windowing input was added, the performance of the model improved significantly, with the F1-measure increasing from 57.00% to 72.19% with a window size of one. Larger window sizes did not improve the performance.

Fig. 5 illustrates the output of the optimal model on a single Bach chorale: The top four musical staves are the original chorale music, while the fifth stave is generated by music21's chodify function, summarizing the pitch content of the slice. Chord labels are placed immediately below the chordify stave, while the next two lines of text (lowercase letters) indicate non-chord tones in the ground-truth data and the model's output respectively. As we can see, the model is correct for the first six measures, with some errors in the rest of the chorale. Experienced music analysts will see that many of the "errors" in fact represent plausible analytical choices.

5 CONCLUSION

In this paper, a non-chord tone identification model for Bach chorales using feedforward DNN is proposed. This model is trained and tested using the Rameau dataset. Experimentation with model input features suggests that including on/off beat metric information and a contextual window of size three achieves the optimal model performance. Given a very limited amount of data, where the proportion of non-chord tones is less than 10%, the model still achieved an F1-measure of 72.19%. We hope that better performances will be achieved with more, higher-quality, data. Thus, we intend to complete the whole Bach chorale dataset with 371 chorales fully annotated with harmonic/contrapuntal labels. Not only will this dataset help to train and test our model further, it will be useful for many other music analytical tasks. We also intend to experiment with more sophisticated neural networks, in particular the powerful bi-directional long short-term memory (BLSTM) model, to see

Figure 5: The first 18 measures of BWV 389 "Nun lob, mein Seel, den Herren." The first line of text underneath the score is the original chord labels. The second line is the non-chord-tone ground truth(note names). The third line is the model's predicted non-chord tones.

whether it will improve the performance. All this work will then be applied to more complex musical collections, such as the piano music in the TAVERN dataset.

REFERENCES

[1] Ching-Hua Chuan and Elaine Chew. 2011. Generating and evaluating musical harmonizations that emulate style. *Computer Music Journal* 35, 4 (2011), 64–82.

[2] Michael Scott Cuthbert and Christopher Ariza. 2010. music21: A toolkit for computer-aided musicology and symbolic music data. In *Proceedings of the 11th International Society for Music Information Retrieval Conference*. Utrecht, Netherlands, 637–642.

[3] J. Devaney, Claire Arthur, Nathaniel Condit-Schultz, and Kirsten Nisula. 2015. Theme and Variation Encodings with Roman Numerals: A New Data Set for Symbolic Music Analysis. In *Proceedings of the 16th International Society for Music Information Retrieval (ISMIR) Conference*, Meinard Müller and Frans Wiering (Eds.). Malaga, Spain, 728–734.

[4] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. 2015. Delving deep into rectifiers: surpassing human-level performance on ImageNet classification. In *Proceedings of the IEEE International Conference on Computer Vision*. 1026–1034.

[5] Thomas Hedges and Martin Rohrmeier. 2011. Exploring Rameau and Beyond: A Corpus Study of Root Progression Theories. In *Proceedings of the 1st International Conference on Mathematics and Computation in Music*. Berlin, Heidelberg, 334–337.

[6] G. E. Hinton and R. R. Salakhutdinov. 2006. Reducing the Dimensionality of Data with Neural Networks. *Science* 313, 5786 (July 2006), 504–507.

[7] Plácido R. Illescas, David Rizo Valero, Iñesta Quereda, José Manuel, and Rafael Ramírez. 2011. Learning melodic analysis rules. In *Proceedings of the International Workshop on Music and Machine Learning*.

[8] Nori Jacoby, Naftali Tishby, and Dmitri Tymoczko. 2015. An Information Theoretic Approach to Chord Categorization and Functional Harmony. *Journal of New Music Research* 44, 3 (2015), 219–244.

[9] Pedro Kröger, Alexandre Passos, Marcos Sampaio, and Givaldo De Cidra. 2008. Rameau: a system for automatic harmonic analysis. In *Proceedings of International Computer Music Conference*. Belfast, N. Ireland.

[10] Lesley Mearns. 2013. *The computational analysis of harmony in Western art music*. Ph.D. Dissertation. Queen Mary University of London.

[11] Brian Pardo and William P. Birmingham. 2001. The chordal analysis of tonal music. In *The University of Michigan, Department of Electrical Engineering and Computer Science Technical Report CSE-TR-439-01*.

[12] Jeremy Pickens. 2004. *Harmonic modeling for polyphonic music retrieval*. Ph.D. Dissertation. University of Massachusetts at Amherst.

[13] Christopher Raphael and Joshua Stoddard. 2004. Functional harmonic analysis using probabilistic models. *Computer Music Journal* 28, 3 (2004), 45–52.

[14] David Rizo, Plácido R. Illescas, and José M. Iñesta. 2016. Interactive melodic analysis. In *Computational Music Analysis*, David Meredith (Ed.). Springer International Publishing, 191–219.

[15] Martin Rohrmeier and Ian Cross. 2008. Statistical Properties of Tonal Harmony in Bach's Chorales. In *Proceedings of the 10th international conference on music perception and cognition*. Sapporo, Japan, 619–627.

[16] Craig Stuart Sapp. 2007. Computational chord-root identification in symbolic musical data: Rationale, methods, and applications. *Computing in Musicology* 15 (2007), 99–119.

[17] Timothy Judson Willingham. 2013. *The harmonic implications of the non-harmonic tones in the four-part chorales of Johann Sebastian Bach*. Ph.D. Dissertation. Liberty University.