COCHONUT: RECOGNIZING COMPLEX CHORDS FROM MIDI GUITAR SEQUENCES

Ricardo Scholz
Federal University of Pernambuco
reps@cin.ufpe.br

Geber Ramalho
Federal University of Pernambuco
glr@cin.ufpe.br

ABSTRACT

Chord recognition from symbolic data is a complex task, due to its strong context dependency and the large number of possible combinations of the intervals which the chords are made of, specially when dealing with dissonances, such as 7^{ths}, 9^{ths}, 13^{ths} and suspended chords. None of the current approaches deal with such complexity. Most of them consider only simple chord patterns, in the best cases, including sevenths. In addition, when considering symbolic data captured from a MIDI guitar, we need to deal with non quantized and noisy data, which increases the difficulty of the task. The current symbolic approaches deal only with quantized data, with no automatic technique to reduce noise. This paper proposes a new approach to recognize chords, from symbolic MIDI guitar data, called COCHONUT (Complex Chords Nutting). The system uses contextual harmonic information to solve ambiguous cases, integrated with other techniques, such as decision theory, optimization, pattern matching and rule-based recognition. The results are encouraging and provide strong indications that the use of harmonic contextual information, integrated with other techniques, can actually improve the results currently found in literature.

Keywords: Chord recognition, MIDI Guitar, music analysis, bossa nova, music segmenting, music information retrieval.

1. INTRODUCTION

The project "A Country, A Guitar"[1] aims to study the Brazilian interpretation of acoustic guitar when accompanying Brazilian repertoire, in particular bossa nova. It is intuitively recognized by experts that there is a sort of "Brazilian accent" in acoustic guitar interpretation. In order to translate this intuition into formal data, getting a deeper understanding of the phenomena, the use of computer-based retrieval techniques is imperative. In the context of this project, chord recognition has a key role. Given that bossa nova has a jazz-like harmony, favoring high utilization of dissonances, and musicians often carry through re-harmonization, these aspects can be an important information on characterizing a given musician and understanding his/her interpretation. Therefore, for the purpose of this study, requirements on chord recognition are far more rigorous than the requirements considered in current literature. In addition, current approaches for chord recognition from symbolic data do not consider non quantized data [12] or force its quantization to a particular grid [17]. Finally, given that most of the current approaches for chord recognition on audio signals use MIDI synthesis to generate large labeled data sets for supervised learning, a MIDI chord recognition approach that deal with complex chords can be very useful.

In this paper, chord recognition from MIDI sequences is split into three phases: first, a segmentation algorithm is run to identify the possible points where chords change. Then, a utility function is applied to identify the most probable chords for each segment. After that, a graph is built representing the possible chords, and it is split into uncertainty regions, i.e., regions in which there are more than one candidate chord for a given segment, surrounded by certainty regions. A rule base containing common chord sequences patterns in jazz harmony is used to solve ambiguous cases.

2. COMPLEX CHORD RECOGNITION

The data set of the project "A Country, A Guitar" is made of several simultaneous recorded MIDI sequences and audio signals of bossa nova played by musicians on a MIDI guitar. This symbolic input data set is made of several non quantized and noisy MIDI sequences. The ideal output of the process should be the complete information about the chord sequences played, such as chord attack times, types and specially dissonances.

There are two main problems in chord recognition. The first one is segmenting the song in excerpts corresponding to a unique chord, i.e., the identification of chord changes. The second problem is the classification of such segments as representing one of the several chords that a given set of notes may represent. Such tasks can be solved simultaneously or sequentially.

The existence of grace notes is one of the main challenges in segmenting the sequences. Most of the musicians make use of grace notes during execution. Due

[1] **"Um País, Um Violão"**, inspired on the beginning of the lyrics of Corcovado, by Tom Jobim.

to their out-of-chord nature, such notes make chord changes identification and segments classification harder, given the difficulty of telling *a priori* which notes are part of the chord, and which ones are grace notes.

The classification of the segments as chords is a hard task due to the ambiguity and context dependency of chords, i.e., a given set of notes may represent several chords, depending on the context in which they are inserted. A classical example of ambiguous set of notes is the set {E, G, Bb, D}, which can represent a $Em^{7(b5)}$ or a Gm^6. More complex situations may occur when the amount of considered dissonances is larger, as in the set {C, D, F, A}, which can represent a Dm^7, a $C^{sus4(9)(13)}$, a F^6 or even a $A^{sus4(\#9)(b13)}$ depending on its context. In addition, bossa nova execution on acoustic guitars has strong constraints regarding the maximum number of simultaneous notes being played, which very often is limited to four. Therefore, chord shapes on guitar tend to prefer using root and dissonances, and discard notes with low harmonic importance, such as fifths. This also increases ambiguity, since some notes of the chords may not be present at all times.

A good solution to harmony extraction must involve temporal and harmonic correctness requirements. **Temporal correctness** states that all chord changes must be recognized with a minimum time difference to the real chord change time. For most of the musical genres, classification of major, minor and diminished chords may be considered as a good model of the chord recognition problem [10]. However, for bossa nova study, which harmony is predominantly jazzistic, and specially for the study of bossa nova execution on acoustic guitars, it is necessary the complete recognition of chords, including all possible dissonances, given that the way they are used has a key role on characterizing interpretation. Thus, it is interesting that the proposed solution is **harmonically correct**, i.e., the recognition of chords include not only its basic notes (root, third and fifth), but all possible dissonances: sevenths, ninths, thirteenths and suspensions. We call these 4-note, 5-note and even 6-note chords, complex chords.

3. CURRENT CHORD RECOGNITION TECHNIQUES

Chord recognition literature considers two main input data types: audio signals and symbolic data, such as MIDI sequences. There are several techniques for chord recognition on audio signals, which use different approaches, such as pattern matching [8, 10] and many machine learning techniques, including supervised learning with labeled datasets acquired from symbolic data [9], genetic programming [16] and search through a hypothesis space [19]. We will not discuss these approaches here, since our focus is on symbolic data.

Considering chord recognition from symbolic data (typically MIDI), Birmingham and Pardo [12] proposed a chord pattern matching and graph search based approach for partition, segmentation and chord detection. First, partition points are created for every NOTE ON and NOTE OFF events. Then, a directed acyclic graph is built such that each partition point is mapped to a node and linked to all other subsequent nodes. The weights of the edges are calculated through a utility function applied to a set of chord patterns and roots, considering the notes which are being played between the partition points mapped by the starting and ending nodes of the edge. Finally, they use the HarmAn [12] algorithm to find the best path in the graph, i.e., the most plausible chords. This way, some nodes may be discarded along the path finding, allowing neighbor segments to be classified as belonging to the same chord. This technique simultaneously segments and detects chords in the sequence.

Also using symbolic data as input, Melisma Music Analyser [17] is one of the most used systems to build labeled data sets by supervised learning approaches [9]. Melisma uses some policies during harmony extraction, such as the preference for seeking a cycle of fifths between consecutive chords, or consider harmony changes in strong beats. However, Melisma does not actually recognize the chords; it only recognizes the root of each segment and outputs a list of notes which are present in the given segment.

All current approaches, symbolic or audio-based [2, 3, 8, 9, 12, 15, 16, 17, 19], are focused in recognizing only simple chords (basically, root, third and fifth, and, sometimes seventh). None of the approaches is capable of recognizing the complete set of complex chords recurrent in jazzistic harmony, with all their dissonances. In addition, supervised learning approaches [9] need large sets of labeled data for training and validation, suggesting the use of symbolic data approaches to automatically generate such sets. Then, harmonic correctness of such approaches is limited to the amount of chord patterns that symbolic data approaches can identify. Finally, although current symbolic approaches can reach good results, they are not able to deal with non quantized data [12, 17].

4. COMPLEX CHORD RECOGNITION THROUGH COCHONUT

The approach proposed by COCHONUT process splits chord recognition problem into three:

- the segmentation of the sequence through the identification of partition points which correspond to chord changes in the song;
- the identification of a set of more probable chords, according to local information;
- and finally, the choice of the best chord for each

segment, considering contextual information.

Our approach is inspired from Birmingham and Pardo's, but we introduce several technical and conceptual changes/extensions, getting better results. A sequential approach has been chosen, despite of a simultaneous one, because in the later, the use of contextual information would be harder. However, we do agree that a simultaneous approach, using contextual information, could be more appropriate, and we are currently working on improvements with this purpose.

Input data used as training and testing sets for this work have been recorded by two musicians executing bossa nova pieces in a MIDI acoustic guitar, and subsequently hand labeled. Chord grids have been made available; however it has been allowed that musicians made changes in harmony during their interpretations.

It is important to mention that symbolic data generated by MIDI guitars contains noise. We have developed a rule based approach to semi-automatically clean such noises by using local partial harmonic information. Unfortunately, due to space limitations, a more detailed description of the noise reduction process is out of the scope of this paper.

Also due to space limitations, we could not include a complete working example of the technique in the article, although it can be found on [14].

A priori, the method can be applied over any genre which have jazz-like harmony. Moreover, by changing the sets of chord patterns and chord sequences patterns, the method may succeed with other genres.

4.1. Partitioning and Segmentation Technique

As discussed in this section, we introduced several modifications to the Birmingham and Pardo's original algorithm [12] aiming to increase its performance on partitioning the training MIDI acoustic guitars sequences acquired from our guitar players.

The partitioning technique we adopted has a more restrictive policy for partition points's (chord changes) detection, in comparison to Birmingham and Pardo's [12]. The partition algorithm considers as valid partition points only the ones where there are at least three or more simultaneous attacks within a predefined time window.

It is possible to justify such an approach since bossa nova accompaniment recordings on acoustic guitars are basically composed of chord changes, where three or more strings are pulled together. Although this may seem too specific for bossa nova, it can also be applied to non quantized inputs from other popular genres, since chord changes also tend to happen when many notes are played within a small time window.

Experiments have been done considering two, three and four simultaneous notes. The considered time window guarantees a duration between a sixteenth note and an eighth note, depending on the tempo actually used by the musician while recording the song. Given that chord changes very often occur in strong beats, whose granularity is not greater than a quarter note, we believe that this time window is very reasonable.

Best results were obtained when policy considered a threshold of three simultaneous notes. This approach has improved the classification of correct partition points, although it still provided a large number of false-positives. However, most of the incorrectly detected partition points had a particularity: they resulted in consecutive identical state segments. Then, a post processing of the found segments have been defined, aiming to merge consecutive identical segments. This approach has significantly improved results.

4.2. More Probable Chords Identification

Aiming to choose a set of more probable chords for each segment, a set of chord patterns is confronted with the state of the segments, using each of the twelve notes of the chromatic scale as roots. Then, a numeric adaptation value is calculated for each chord (i.e., each pair root/pattern) at each segment.

Pattern matching approaches have been used before [10, 12], however the chord patterns previously defined did not reach good results in jazzistic songs due to the lack of complex chord patterns.

We opted for using a set of chord patterns which included basic chord patterns together with jazzistic harmony [1, 5, 6]. Several experiments have been carried through, from an initial set of chord patterns, built from all partterns on [12] and other more complex ones, aiming to remove patterns not appropriated to jazz harmony and keep or include patterns that leaded to best results on the training set.

The utility function proposed by Birmingham and Pardo [12] was used without any changes to calculate the fitness of each segment to an ordered pair {root, pattern}. Results obtained by Birmingham and Pardo approach using our richer chord patterns set were better than the ones obtained using the initial chord patterns set, although it was not enough to achieve expected results.

Then, all chords which score is lower than 85% of the winner score are discarded, and the remaining chords are considered as candidate chords. Such a threshold was obtained empirically, through the performance on the training set. After that, a graph is built in such a way that each segment is mapped to a layer in the graph, and each layer contains all candidate chords for its segment. Then, every layer is totally and uniquely linked to its next layer. The resulting graph G is a k-partite graph, having as disjoint vertex sets the candidate chords for each segment, where k is the number of segments found during partitioning phase as shown in figure 1.

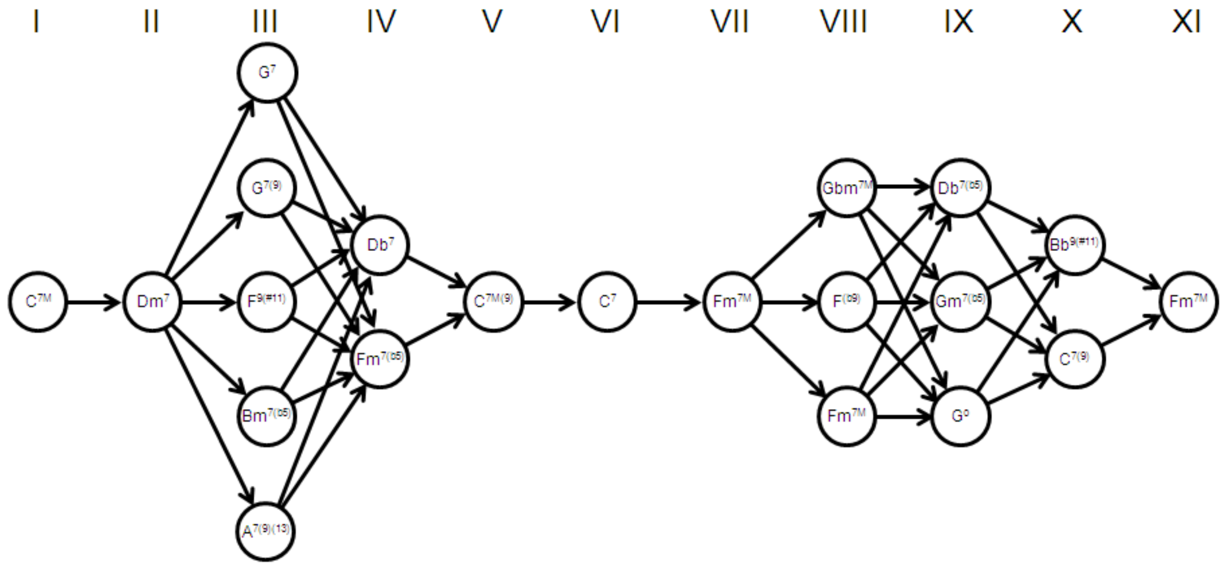

Figure 1. Example of graph built from candidate chords for each segment.

It is important to highlight that this graph, by definition, is much different from Birmingham and Pardo's, since its nodes are connected to all nodes in the immediately subsequent layer, and no other nodes. In addition, Birmingham and Pardo's nodes map to partition points, while nodes in our graph map to segments (defined between two consecutive partition points). And finally, in opposition to Birmingham and Pardo's graph, edges in our graph have no weights.

The problem is then reduced to the choice of a path in the graph G which represents more adequately the harmonic information of the musical piece.

4.3. Contextual Analysis

Local analysis was not enough to resolve ambiguous cases, i.e., it was not possible to choose the best chord for each segment by the application of the utility function, as proposed by Birmingham and Pardo. Then, we assumed that, together with the fitness of the set of notes to a given chord pattern and a given root, we had to make use of the context in which the segment was inserted.

Context is modeled through a set of chord sequence patterns. We defined a set of rules, trying to find recurrent paths over the graph, i.e., the chord sequence patterns. Formally, each sequence pattern defines a set of twelve paths on the graph, one for each possible root of the chromatic scale. In addition, they map recurrent chord sequences in jazzistic harmony, turning the problem of finding the more appropriate path into a problem of matching recurrent paths into the graph. The chord sequence patterns used are a subset of the patterns used in other works related to automatic functional harmonic analysis [13].

Each rule can only apply in layers which nodes had not being used by other rules to fire. After the execution of the inference engine, all non used nodes in each layer are deleted, and the remaining nodes make the chosen path. Priority among rules follows two criterions: the recurrence of the chord sequence and its length, larger chord sequences having greater priority. Results obtained using this approach were very satisfactory.

5. RESULTS

The experiments aimed to measure the quality of the proposed approach according to the requirements defined on section two. The testing dataset contained four songs recorded by two guitar players, in a total of 270 chords. All MIDI sequences have been manually labeled by an expert.

Experiments consisted of the automatic chords recognition through the use of several combinations of COCHONUT and Birmingham and Pardo's approach, varying partitioning algorithm, chord patterns set and graph construction and building technique. The results were then compared with the labeled data set.

The policy for chords comparison aimed to measure the correctness of the whole recognition, and considers the exact comparison with labeled chords, including all dissonances. Differences between enharmonic chords were not considered as errors, given that there are functional harmonic analysis techniques which are able to fix these information very well [13].

Some metrics were defined to obtain more accurate quantitative information about the classifications. These metrics, which appear in the result tables, are: **Total number of label chords (A)**; **Time Correctness Percent (TCP)**, the percentage of the song duration in which the recognized chord is correct; **Strict Correctness Percent (SCP),** percentage of the correctly identified chords, both harmonically (right name) and temporally (within 250ms around the actual chord attack); and **Harmonic Correctness Percent (HCP):** percentage of the correctly identified chords harmonically but whose time occurrence is outside of the window.

The experiments were done for the combinations of techniques described at table 1. Experiment 1 uses only Birmingham and Pardo approach, whereas experiment 5, only COCHONUT's. Experiments 2, 3 and 4 mix both approaches. Table 2 shows the results.

Experi-ment	**Partitioning**	**Chord Patterns**	**Graph**
1	B. & Pardo	B. & Pardo	B. & Pardo
2	B. & Pardo	COCHONUT	B. & Pardo
3	COCHONUT	B. & Pardo	B. & Pardo
4	COCHONUT	COCHONUT	B. & Pardo
5	COCHONUT	COCHONUT	COCHONUT

Table 1. Experiments on chord recognition, considering different combinations of partitioning policy, chord patterns set and graph creation and search algorithm.

Experiment	**A**	**TCP**	**SCP**	**HCP**
1	270	33,70%	21,11%	41,11%
2	270	41,61%	22,22%	50,37%
3	270	53,20%	35,18%	44,81%
4	270	66,94%	48,88%	62,22%
5	270	74,77%	64,81%	74,44%

Table 2. Results of the experiments.

Every piece of the approach substituted on Birmingham and Pardo's has increased results, as shown by experiments 2, 3 and 4. However, the particular difference among results obtained by Birmingham and Pardo and COCHONUT approaches, on experiments 1 and 5, respectively, on table 2, is evident. The amount of time the recognized chords are correct increased about 40%. This provides clear indications of the unfitting of Birmingham and Pardo approach to the harmonic complexity requirements of the problem, as well as the particularities of MIDI guitar data.

We figured out a clear decreasing on the amount of recognized chords when partition technique changed, on experiments 3, 4 and 5. The proposed partitioning algorithm achieved very good results, as it was able to identify a number of partition points very close to the amount of real partition points. Analyzing the amount of identified partition points together with the amount of correct identified chords, and the low number of anticipated and late chords identification, we can conclude that **temporal correctness** has been achieved.

Harmonic correctness can be better verified by the time correctness percents (TCP), as well as the strict correctness percent (SCP) and harmonic correctness percent (HCP). The average strict correctness percent (SCP) is greater than 64%, and considering delays and anticipations in chord classification, it reaches 74,44% (HCP). Comparing to the other experiments, there is a clear improvement on performance. Then, we consider that COCHONUT achieved good results, even when harmonic complexity requirements are rigorous.

Obtained results provide clear indications that the use of an hybrid approach, involving an utility-based chord pattern matching technique to identify more probable chords for each segment, and a contextual harmonic information based approach to choose the best option from ambiguous chords sets, can reach good results for the problem of chord recognition, specially when dealing with complex chords.

6. CONCLUSIONS AND FUTURE WORK

The main computational challenges of chord recognition are the ambiguity and the strong contextual dependency. The greater the harmonic complexity considered is, the greater the existence of ambiguous cases is. Data captured from MIDI acoustic guitars provide additional computational challenges, such as its non quantization and the large amount of noise, inherent to the current MIDI guitar technology.

The proposed process reached very satisfactory results, specially regarding harmonic complexity, i.e., the correct and complete identification of the dissonances on recognized chords. Although it is still possible to improve partitioning and classification, results give clear traces that a hybrid approach, using complete, although ambiguous, chord patterns and information over the harmonic structure can reach good results.

The main contributions of the proposed process are: (1) the innovative use of recurrent harmonic patterns matching aiming to capture harmonic context and to find the best path in chord recognition graphs; (2) a solution that meets more demanding harmonic requirements (i.e., complex chords) working on non-quantized symbolic data, considering that no other work on literature has dealt with such a complex chords set.

As a future work suggestion, we consider the inclusion of more complex contextual information, such as harmonic field and modal loans. Such modifications may allow the use of more chord patterns on the first phase, increasing harmonic completeness.

The use of musical structural information (e.g., sections) could also help to improve results [18], as well as the use of optimization techniques to find the best path over the graph [4], and a different way to build the graph such that partitioning and classification are done simultaneously.

Finally, it would be interesting to propose an adaptation of the approach for audio chord recognition. However, identification of partition points could be hard to adapt. The other phases could be adapted by comparing the Pitch Class Profiles of each segment with the chord patterns, and generating a graph with the high scored chords, in which the search technique proposed by COCHONUT could be applied.

7. REFERENCES

[1] Ferrara, J. Jazz Piano and Harmony: A Fundamental Guide. John Ferrara Music Pub. 2000.

[2] Bello, J. and Pickens, J. "A robust mid-level representation for harmonic content in music signals", *Proceedings of the International Symposium on Music Information Retrieval*, London, UK, 2005.

[3] Cabral, G. Briot, J. and Pachet, F. "Impact of distance in pitch class profile computation", *Proceedings of the 10th Brazilian Symposium on Computer Music*. Belo Horizonte, Brazil, 2005.

[4] Cook, W., Cunningham W., Pulleyblank W. and Schrijver, A. Combinatorial Optimization. John Wiley and Sons. Nova York, USA, 1998.

[5] D'Accord Music Software. Guitar Dictionary – Demo Version. www.daccord.com.br. September, 2007.

[6] Faria, N. *Acordes, Arpejos e Escalas para Violão e Guitarra*. Lumiar Editora. Brazil. 1999.

[7] Figueira Filho, C. and Ramalho, G. *Jeops – the Java embedded object production system*. In M. Monard and J. Sichman (eds). Advances in Artificial Intelligence. IBERAMIA-SBIA 2000. London, United Kingdom, 2000.

[8] Fujishima, T. "Realtime chord recognition of musical sound: a system using common Lisp music", *Proceedings of International Computer Music Conference*. Beijing, China, 1999.

[9] Lee, K. and Slaney, M. "Automatic chord recognition from audio using an HMM with supervised learning", *Proceedings of the 7th International Conference on Music Information Retrieval*. Victoria, PAIS, 2006.

[10] Lee, K. "Automatic chord recognition from audio using enhanced pitch class profile", *Proceedings of the International Computer Music Conference*. New Orleans, USA, 2006.

[11] Lima, E., Madsen, S., Dahia, M., Widmer, G. and Ramalho, G. "Extracting patterns from Brazilian guitar accompaniment data", Proceedings of the 1st European Workshop on Intelligent Technologies for Cultural Heritage Exploitation at the 17th European Conference on Artificial Intelligence. Riva del Garda, Italy, 2006.

[12] Pardo, B. and Birmingham, W. "The chordal analysis of tonal music", Technical report CSE-TR-439-01. The University of Michigan, Dept. of Electrical Engineering and Computer Science. Ann Arbor, PAIS, 2001.

[13] Scholz, R., Dantas, V. and Ramalho, G. "Automating functional harmonic analysis: the Funchal system", *Proceedings of the Seventh IEEE International Symposium on Multimedia*. Irvine, USA, 2005.

[14] Scholz, R. COCHONUT References Home Page. www.cin.ufpe.br/~reps/msc/. September/2008.

[15] Sheh, A. and Ellis, D. "Chord segmentation and recognition using EM-trained Hidden Markov Models", *Proceedings of the International Symposium on Music Information Retrieval*. Baltimore, PAIS, 2003.

[16] SONY Computer Science Lab. Holistic Signal Processing. http://www.csl.sony.fr/items/2004/holistic-signal-processing/. January, 2008.

[17] Temperley, D. and Sleator, D. "The Melisma Music Analyser". http://www.link.cs.cmu.edu/music-analysis/. January, 2008.

[18] Widmer, G. "Learning expressive performance: the structure-level approach", Journal of New Music Research 25(2), 179-205. 1996.

[19] Yoshioka, T. et al. "Automatic chord transcription with concurrent recognition of chord symbols and boundaries", *Proceedings of the 5th International Conference in Music Information Retrieval*. Barcelona, Spain, 2004.