TEMPLATE-BASED CHORD RECOGNITION : INFLUENCE OF THE CHORD TYPES

Laurent Oudre[1], **Yves Grenier**[1], **Cédric Févotte**[2]
[1]Institut TELECOM ; TELECOM ParisTech ; CNRS LTCI
[2]CNRS LTCI ; TELECOM ParisTech
37-39 rue Dareau, 75014 Paris, France
{oudre,grenier,fevotte}@telecom-paristech.fr

ABSTRACT

This paper describes a fast and efficient template-based chord recognition method. We introduce three chord models taking into account one or more harmonics for the notes of the chord. The use of pre-determined chord models enables to consider several types of chords (major, minor, dominant seventh, minor seventh, augmented, diminished...). After extracting a chromagram from the signal, the detected chord over a frame is the one minimizing a measure of fit between the chromagram frame and the chord templates. Several popular measures in the probability and signal processing field are considered for our task. In order to take into account the time persistence, we perform a post-processing filtering over the recognition criteria. The transcription tool is evaluated on the 13 Beatles albums with different chord types and compared to state-of-the-art chord recognition methods. We particularly focus on the influence of the chord types considered over the performances of the system. Experimental results show that our method outperforms the state-of-the-art and more importantly is less computationally demanding than the other evaluated systems.

1. INTRODUCTION

Chord transcription is a compact representation of the harmonic content and structure of a song. Automatic chord transcription finds many applications in the field of Musical Information Retrieval such as song identification, query by similarity or structure analysis.

The features used for chord recognition may differ from a method to another but are in most cases variants of the 12-dimensional *Pitch Class Profiles* [1]. Every component represents the spectral energy of a semi-tone on the chromatic scale regardless of the octave. The succession of these chroma vectors over time is called *chromagram* : the chord recognition task consists in outputting a chord label for every chromagram frame.

The first chord recognition systems consider many chord types. The method proposed by Fujishima [1] considers 27 chord types. The transcription is done either by minimizing the Euclidean distance between *Pitch Class Profiles* and 12-dimensional chord templates constituted by 1's (for the chromas present in the chord) and 0's (for the other chromas) or by maximizing a weighted dot product. Sheh & Ellis [2] use a Hidden Markov Model composed of 147 hidden states each representing a chord (7 types of chords and 21 root notes). All the HMM parameters are learned by a semi-supervised training with an EM algorithm.

These two methods have been improved upon by reducing the number of chord types considered. Fujishima's system is improved in [3] by reducing the number of chords types from 27 to 4 (major, minor, augmented, diminished) and by calculating a more elaborate chromagram including notably a tuning algorithm. Chord transcription is then realized by retaining the chord with larger dot product between the chord templates and the chromagram frames. Sheh & Ellis method is modified in [4] : the number of hidden states is reduced from 147 to 24 by only considering major and minor chords for the 12 semi-tones root notes. Musical knowledge is introduced into the model by initializing the HMMs parameters with values inspired by musical and cognitive theory. Since then, almost all the chord transcription methods [5], [6], [7], [8], [9], only consider major and minor chords.

Our chord recognition system is based on the intuitive idea that for a given 12-dimensional chroma vector, the amplitudes of the chromas present in the chord should be larger than the ones of the non-played chromas. By introducing chord templates for different chord types and roots, the chord present on a frame should therefore be the one whose template is the *closest* to the chroma vector according to a specific measure of fit.

The paper is organized as follows. Section 2 gives a description of our recognition system. Section 3 describes the corpus and the protocol of evaluation. Section 4 presents the results of our system, a study on the influence of the chord types considered, a comparison with the state-of-the-art and an analysis of the frequent errors. Finally the main conclusions of this work are summarized in Section 5.

2. DESCRIPTION OF THE SYSTEM

2.1 General idea

Given N successive chroma vectors $\{\mathbf{c}_n\}_n$, K chord templates $\{\mathbf{p}_k\}_k$ and a measure of fit D, we define :

$$d_{k,n} = D\left(h_{k,n}\,\mathbf{c}_n; \mathbf{p}_k\right). \quad (1)$$

$h_{k,n}$ is a scale parameter whose role is to fit the chroma vector $\mathbf{c}_n$ with the chord template $\mathbf{p}_k$ according to the measure of fit used. In practice, $h_{k,n}$ is calculated such as :

$$h_{k,n} = \underset{h}{\operatorname{argmin}}\, D\left(h\,\mathbf{c}_n; \mathbf{p}_k\right). \quad (2)$$

The detected chord $\hat{k}_n$ for frame n is then the one minimizing the set $\{d_{k,n}\}_k$:

$$\hat{k}_n = \underset{k}{\operatorname{argmin}}\, \{d_{k,n}\}. \quad (3)$$

In our system, the chroma vectors are calculated from the music signal with the same method as Bello & Pickens [4]. The frame length is set to 753 ms and the hop size is set to 93 ms. We use the code kindly provided by these authors.

We have omitted for sake of conciseness the expressions of $d_{k,n}$ and $h_{k,n}$ which are easily obtained by canceling the gradient of (1) wrt $h_{k,n}$.

2.2 Chord models

The intuitive chord model is a simple binary mask constituted of 1's for the chromas present in the chord and 0's for the other chromas [1], [3].

Yet, the information contained in a chromagram captures not only the intensity of every note but a blend of intensities for the harmonics of every note. Like Gomez [10] and Papadopoulos [5], we assume an exponentially decreasing spectral profile for the amplitudes of the partials. An amplitude of 0.6^{i-1} is added for the i^{th} harmonic of every note in the chord.

In our system three chord models are defined, corresponding to 1, 4 or 6 harmonics. Examples for C major and C minor chords are displayed on Figure 1.

From these three chord models we can build chord templates for all types of chords (major, minor, dominant seventh, diminished, augmented,...). By convention in our system, the chord templates are normalized so that the sum of the amplitudes is 1.

2.3 Measures of fit

We consider for our recognition task several measures of fit, popular in the field of signal processing : the **Euclidean distance** (later referred as *EUC*), the **Itakura-Saito divergence** [11] and the **Kullback-Leibler divergence** [12].

Since the Itakura-Saito and Kullback-Leibler divergence are not symmetrical, they can be calculated in two ways. $D\left(h_{k,n}\,\mathbf{c}_n|\mathbf{p}_k\right)$ will respectively define *IS1* and *KL1*, while $D\left(\mathbf{p}_k|h_{k,n}\,\mathbf{c}_n\right)$ will define *IS2* and *KL2*.

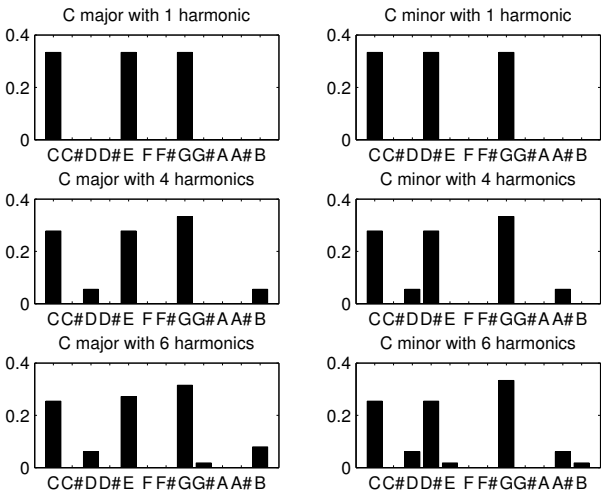

Figure 1. Chord templates for C major / C minor with 1, 4 or 6 harmonics.

2.4 Filtering methods

In order to take into account the time-persistence, we introduce some post processing filtering methods which work upstream on the calculated measures and not on the sequence of detected chords.

The new criterion $\tilde{d}_{k,n}$ is based on L successive values $\{d_{k,n'}\}_{n-\frac{L-1}{2}\leq n'\leq n+\frac{L-1}{2}}$ previously calculated. In our system two types of filtering are used.

The **low-pass filtering** takes the mean of the L values. It tends to smooth the output chord sequence and to reflect the long-term trend in the chord change.

The **median filtering** takes the median of the L values. It has been widely used in image processing and is particularly efficient to correct random errors.

In every case, the detected chord $\hat{k}_n$ on frame n is the one that minimizes the set of values $\left\{\tilde{d}_{k,n}\right\}_k$:

$$\hat{k}_n = \underset{k}{\operatorname{argmin}}\, \left\{\tilde{d}_{k,n}\right\} \quad (4)$$

3. EVALUATION

3.1 Corpus

The evaluation database used in this paper is made of the 13 Beatles albums (180 songs, PCM 44100 Hz, 16 bits, mono). The chord annotations for these 13 Beatles albums are kindly provided by Harte and Sander [13].

In these annotation files, 17 types of chords and one 'no chord' label (N) corresponding to silences or untuned material are present.

The most common chord types in the corpus are major (63.89% of the total duration), minor (16.19%), dominant seventh (7.17%) and 'no chord' states (4.50%). Figure 2 shows the repartition of the chord types among the 13 albums of the Beatles. We can see that the number of major, minor and dominant seventh chords varies much with the album. Yet, the last six albums clearly contain more chord

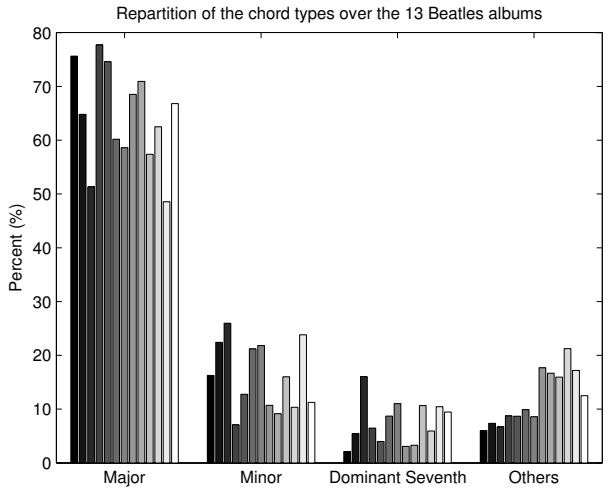

Figure 2. Repartition of the chord types as percentage of the total duration for the 13 Beatles albums.

types (other than major, minor and dominant seventh) than the first seven ones.

3.2 Protocol of evaluation

The evaluation method used in this paper corresponds to the one used in MIREX 08 for the Audio Chord Detection task. [1]

As the evaluation method only takes into account major and minor chords, the 17 types of chords present in the annotation files are first mapped into major and minor types following the rules used in MIREX 08 :

- major : maj, dim, aug, maj7, 7, dim7, hdim7, maj6, 9, maj9, sus4, sus2
- minor : min, min7, minmaj7, min6, min9

For the systems detecting more chord types (dominant seventh, diminished, etc.), once the chords have been detected with their appropriate models, they are then mapped to the major and minor following the same rules than for the annotation files.

A score is calculated for each song as the ratio between the lengths of the correctly analyzed chords and the total length of the song. The final *Average Overlap Score (AOS)* is then obtained by averaging the scores of all the 180 songs. An example of calculation of an Overlap Score is presented on Figure 3.

4. RESULTS

The five previously described measures of fit (*EUC*, *IS1*, *IS2*, *KL1* and *KL2*), three chord models (1, 4 or 6 harmonics) and two filtering methods (low-pass and median) with neighborhood sizes from $L = 1$ to $L = 25$ are tested. For every method we only present the results for the optimal parameters (measure of fit, chord models, filtering method and neighborhood size).

[1] http://www.music-ir.org/mirex/2008/

4.1 Results with major/minor chord types

Considering only major and minor chords (like most of the chord recognition methods of the actual state-of-art), we obtain a *Average Overlap Score* of 0.70 over the 13 Beatles albums. The optimal parameters are the Kullback-Leibler divergence *KL2*, the single harmonic chord model and the median filtering with a neighborhood size of $L = 17$.

4.2 Introduction of other chord types

The simplicity of our method allows to easily introduce chord templates for chord types other than major and minor : we study here the influence of the chord types considered over the performances of our system. The choice of these chord types is guided by the statistics on the corpus previously presented : we introduce in priority the most common chords types of the corpus.

4.2.1 Dominant seventh and minor seventh chords

In the Beatles corpus, the two most common chord types other than major and minor are dominant seventh *(7)* and minor seventh *(min7)* chords. The results for major, minor, dominant seventh and minor seventh chords are presented in Table 1. The score displayed in a case is the best *Average Overlap Score* obtained by considering the chord types of the corresponding row and column.

	min	min7	min & min7
maj	0.70	0.64	0.69
7	0.69	0.63	0.65
maj & 7	0.71	0.66	0.69

Table 1. Average Overlap Scores with major, minor, dominant seventh and minor seventh chords.

The best results are obtained by detecting major, minor and dominant seventh chords, with the Kullback-Leibler divergence *KL2*, the single harmonic chord model and the median filtering with $L = 17$ giving a recognition rate of 71%. Only the introduction of dominant seventh chords, which are very common in the Beatles corpus, enhances the results. The introduction of minor seventh chords, which are less common, degrades the results. Indeed, the structure of minor seventh chords (for example *Cmin7*) leads to confusion between the actual minor chord and the relative major chord (*E♭* in our example).

4.2.2 Augmented and diminished chords

Augmented and diminished chords have been considered in many template-based chord recognition systems [1], [3]. Interestingly, while the augmented and diminished chords are very rare in the Beatles corpus (respectively 0.62% and 0.38% of the total length), the introduction of chord templates for augmented and diminished chords does not degrade the results. We obtain a recognition rate of 69% by considering major, minor, augmented and diminished chords and of 71% by taking into account major, minor, dominant seventh, augmented and diminished chords.

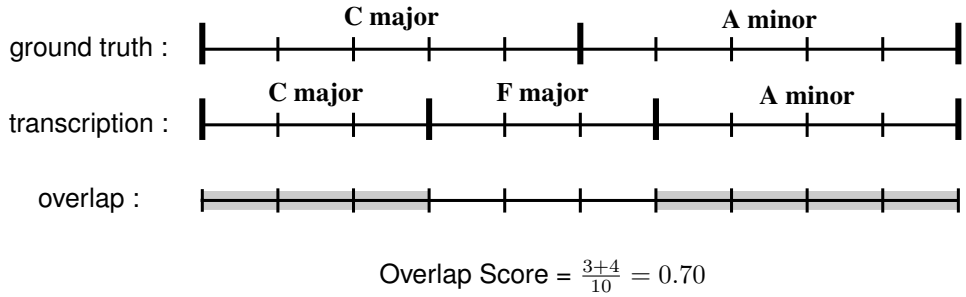

Figure 3. Example of calculation of an Overlap Score.

4.2.3 Other chord types

The introduction of other chord types (ninth, major seventh, sus4, etc.) does not improve the results. This can be explained either by the structures of the chords which can lead to confusions with other chord types or by the low number of chords of these types in the Beatles corpus. Indeed, the introduction of a model for a new chord type gives a better detection for chords of this type but also leads to new errors such as false detections. Therefore only frequent chords types should be introduced, ensuring that the enhancement caused by the better recognition of these chord types is larger than the degradation of the results caused by the false detections.

4.3 Influence of the album

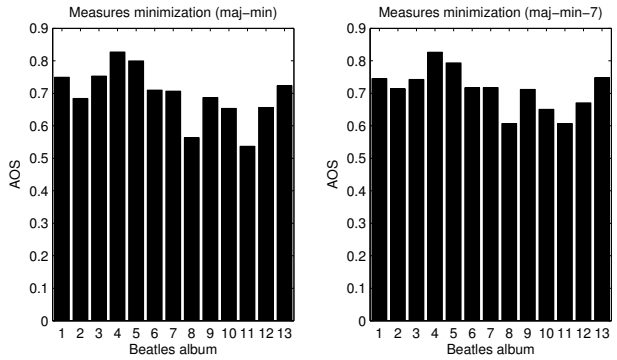

Figure 4. Average Overlap Scores for the 13 Beatles albums (in chronological order) for the major/minor and the major/minor/dominant seventh methods.

We can see on Figure 4 that results are better for the first seven albums : this can be explained by the low number of chords other than major, minor and dominant seventh on these albums (see Figure 2). Surprisingly the introduction of dominant seventh chords tend to improve results not necessarily on albums containing many dominant seventh chords (for example album number 3) but on albums containing many chords other than major, minor and dominant seventh (for example albums number 8 & 11).

4.4 State-of-the-art

Our method is now compared to the following methods that entered MIREX 08.

Bello & Pickens [4] use 24-states HMM with musically inspired initializations, Gaussian observation probability distributions and EM-training for the initial state distribution and the state transition matrix.

Ryynänen & Klapuri [6] use 24-states HMM with observation probability distributions computed by comparing low and high-register profiles with some trained chord profiles. EM-training is used for the initial state distribution and the state transition matrix.

Khadkevich & Omologo [7] use 24 HMMs : one for every chord. The observation probability distributions are Gaussian mixtures and all the parameters are trained through EM.

Pauwels, Verewyck & Martens [8] use a probabilistic framework derived from Lerdahl's tonal distance metric for the joint tasks of chords and key recognition.

These methods have been tested with their original implementations on the same Beatles corpus than before and evaluated with the same protocol (AOS). Results of this comparison with the state-of-the-art are presented on Table 2.

	AOS	Time
Our method (Maj-Min-7)	0.71	796s
Bello & Pickens	0.70	1619s
Our method (Maj-Min)	0.70	790s
Ryynänen & Klapuri	0.69	1080s
Khadkevich & Omologo	0.64	1668s
Pauwels, Varewyck & Martens	0.62	12402s

Table 2. Comparison with the state-of-the-art.

First of all it is noticeable that all the methods give rather close results : there is only a 9% difference between the methods giving the best and worse results. Our method gives the best results, but more importantly with a very low computational time. It is indeed twice as fast as the best state-of-the-art method (Bello and Pickens).

4.5 Analysis of the errors

In most chord transcription systems, the errors are often caused by the structural similarity (common notes) and the harmonic proximity between the real chord and the wrongly detected chord.

Two chords are likely to be mistaken one for another when they *look alike*, that is to say, when they share notes (especially in template-based systems). Given a major or minor chord, there are 3 chords which have 2 notes in common with this chord : the parallel minor/major, the relative minor/major (or submediant) and the mediant chord.

Besides the structural similarity, errors can also be caused by the harmonic proximity between the original and the detected chord. Figure 5 pictures the doubly nested circle of fifths which represents the major chords (capital letters), the minor chords (lower-case letters) and their harmonic relationships. The distance linking two chords on this doubly nested circle of fifths is an indication of their harmonic proximity.

Given a major or minor chord, the 4 closest chords on this circle are the relative (submediant), mediant, subdominant and dominant. One can notice that these 4 chords are also structurally close to the original chord, since they share 1 or 2 notes with it.

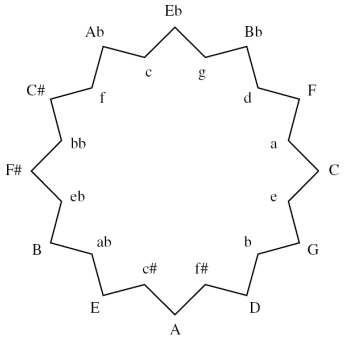

Figure 5. Doubly nested circle of fifths [4].

We have therefore brought out 5 potential sources of errors among the 23 possible ones (i.e., the 23 other wrong candidates for one reference chord). Examples of these potential sources of errors for C major and C minor chords are displayed on Figure 6.

Reference chord	**C**	**Cm**
parallel	Cm	C
relative (submediant)	Am	A♭
mediant	Em	E♭
subdominant	F	Fm
dominant	G	Gm

Figure 6. Particular relationships between chords and potential sources of errors : examples for C major and C minor chords.

Figure 7 displays the repartition of these error types as a percentage of the total number of errors for every evaluated method. Errors due to the bad detection of the 'no chord' states are represented with the 'no chord' label.

The main sources of errors correspond to the situations previously described and to the errors caused by silences ('no chord'). Actually, in most methods, the 5 types of errors previously considered (over the 23 possible ones) represent approximately 60% of the errors.

The introduction of the dominant seventh chords clearly reduces the proportion of the errors due to relative (submediant) and mediant (-9%). Another noteworthy result is that the methods by Ryynänen & Klapuri, Bello & Pickens and our major/minor method approximately have the same error repartition despite the different structures of the methods, which proves that the semantic of the errors is inherent to the task. Pauwels, Varewyck & Martens' system is mostly penalized by the wrong detection of the 'no chord' states, when Khadkevich & Omologo's method produces a wider range of errors.

5. CONCLUSION

Our system offers a novel perspective about chord detection. The joint use of popular measures and filtering methods distinguishes from the predominant HMM-based approaches. The introduction of chord templates allows to easily consider many chord types instead of only major and minor chords. Since our method is only based on the chromagram no information about style, rhythm or instruments is required and thank to the fact that no training or database is needed, the computation time can be kept really low.

6. ACKNOWLEDGMENT

The authors would like to thank J. Bello, M. Khadkevich, J. Pauwels, M. Ryynänen for making their code available. We also wish to thank C. Harte for his very useful annotation files.

This work was realized as part of the Quaero Programme, funded by OSEO, French State agency for innovation.

7. REFERENCES

[1] T. Fujishima. Realtime chord recognition of musical sound: a system using Common Lisp Music. In *Proceedings of the International Computer Music Conference (ICMC)*, pages 464–467, Beijing, China, 1999.

[2] A. Sheh and D.P.W. Ellis. Chord segmentation and recognition using EM-trained hidden Markov models. In *Proceedings of the International Symposium on Music Information Retrieval (ISMIR)*, pages 185–191, Baltimore, MD, 2003.

[3] C.A. Harte and M.B. Sandler. Automatic chord identification using a quantised chromagram. In *Proceedings of the Audio Engineering Society*, Barcelona, Spain, 2005.

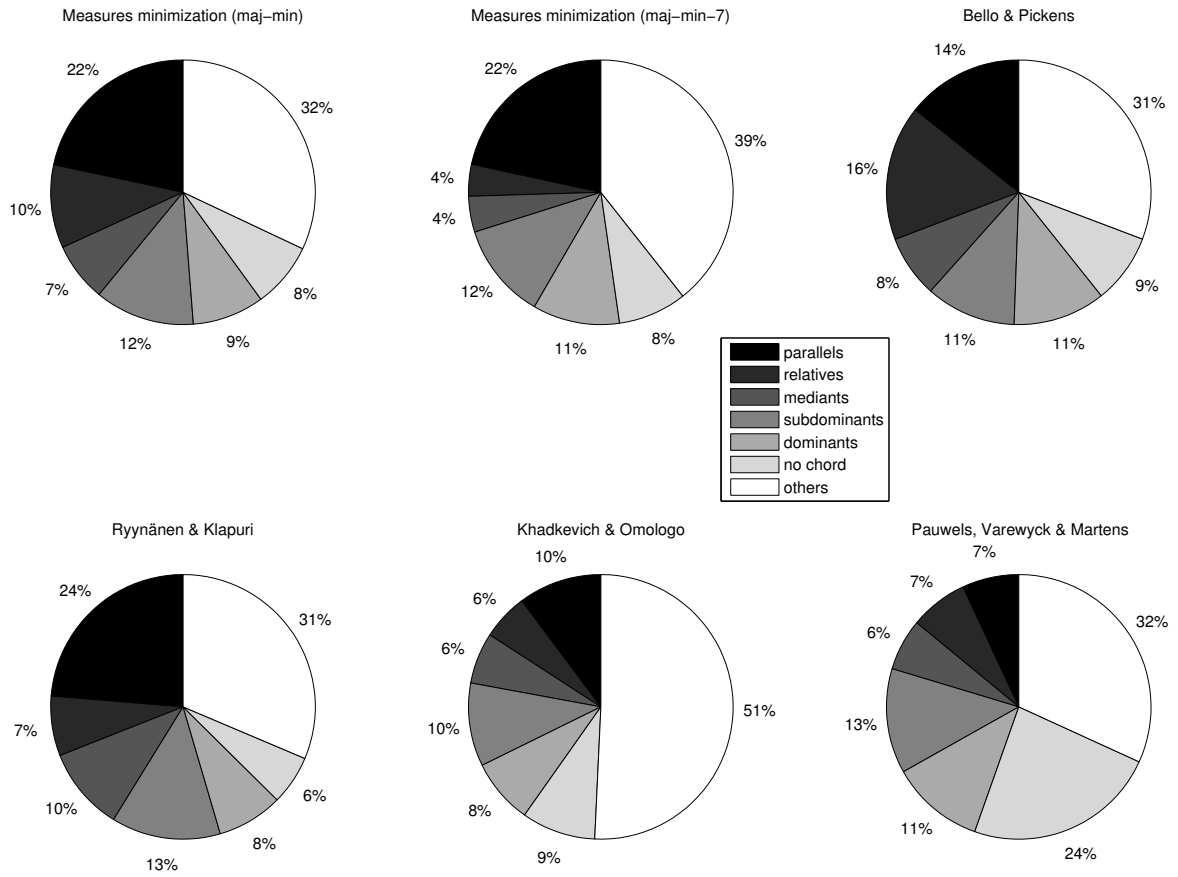

Figure 7. Repartition of the errors as a percentage of the total number of errors.

[4] J.P. Bello and J. Pickens. A robust mid-level representation for harmonic content in music signals. In *Proceedings of the International Symposium on Music Information Retrieval (ISMIR)*, pages 304–311, London, UK, 2005.

[5] H. Papadopoulos and G. Peeters. Large-scale study of chord estimation algorithms based on chroma representation and HMM. In *Proceedings of the International Workshop on Content-Based Multimedia Indexing*, pages 53–60, Bordeaux, France, 2007.

[6] M.P. Ryynänen and A.P. Klapuri. Automatic transcription of melody, bass line, and chords in polyphonic music. *Computer Music Journal*, 32(3):72–86, 2008.

[7] M. Khadkevich and M. Omologo. Mirex audio chord detection. Abstract of the Music Information Retrieval Evaluation Exchange, 2008.

[8] J. Pauwels, M. Varewyck, and J-P. Martens. Audio chord extraction using a probabilistic model. Abstract of the Music Information Retrieval Evaluation Exchange, 2008.

[9] K. Lee and M. Slaney. Acoustic chord transcription and key extraction from audio using key-dependent HMMs trained on synthesized audio. *IEEE Transactions on Audio, Speech and Language Processing*, 16(2):291–301, 2008.

[10] E. Gómez. Tonal description of polyphonic audio for music content processing. In *Proceedings of the INFORMS Computing Society Conference*, volume 18, pages 294–304, Annapolis, MD, 2006.

[11] F. Itakura and S. Saito. Analysis synthesis telephony based on the maximum likelihood method. In *Proceedings of the International Congress on Acoustics*, pages 17–20, Tokyo, Japan, 1968.

[12] S. Kullback and R.A. Leibler. On information and sufficiency. *Annals of Mathematical Statistics*, 22(1):79–86, 1951.

[13] C. Harte, M. Sandler, S. Abdallah, and E. Gomez. Symbolic representation of musical chords: A proposed syntax for text annotations. In *Proceedings of the International Symposium on Music Information Retrieval (ISMIR)*, pages 66–71, London, UK, 2005.