Local key estimation based on harmonic and metric structures

Hélène Papadopoulos, Geoffroy Peeters

LOCAL KEY ESTIMATION BASED ON HARMONIC AND METRIC STRUCTURES

Hélène Papadopoulos[*]
Sound Analysis/Synthesis Team,
IRCAM / CNRS-STMS
Paris, France
Helene.Papadopoulos@ircam.fr

Geoffroy Peeters
Sound Analysis/Synthesis Team,
IRCAM / CNRS-STMS
Paris, France
Geoffroy.Peeters@ircam.fr

ABSTRACT

In this paper, we present a method for estimating the local keys of an audio signal. We propose to address the problem of local key finding by investigating the possible combination and extension of different previous proposed global key estimation approaches. The specificity of our approach is that we introduce key dependency on the harmonic and the metric structures. In this work, we focus on the relationship between the chord progression and the local key progression in a piece of music. A contribution of our work is that we address the problem of finding a good analysis window length for local key estimation by introducing information related to the metric structure in our model. Key estimation is not performed on empirical-chosen segment length but on segments that are adapted to the analyzed piece and independent from the tempo. We evaluate and analyze our results on a new database composed of classical music pieces.

1. INTRODUCTION

Tonality analysis is one of the most important aspects of Western tonal music. Tonality analysis describes the relationship between the different musical keys present in a piece of music. A musical key implies a tonal center that is the most stable pitch (the tonic) and a mode (major or minor). The elements of the melody and the harmony of a musical fragment are related to each other by the musical key. This aspect of music has interested researchers for a long time because key detection task has many applications in content-based music information retrieval such as classification, segmentation, indexing or summarization. Various approaches have been proposed in previous works for estimating the global key of a piece of music. Some approaches were proposed for symbolic data, using template-based approaches [1], [2], [3], or geometry-based approaches [4]. Others were proposed for audio data, using template-based approaches [5], [6], [7] , [8],[9], geometry-based approaches [10], or HMM-based approaches [11]. Finding the main key of a piece of music is only a little part of tonality analysis. Indeed, even if a piece of music generally starts and ends in a particular key referred to as the main or global key of the piece, it is common that the composer will move between keys, sometimes without definitely establish them. A change in the musical key is called a modulation. In this paper, we are interested in the problem of local key estimation: we aim at segmenting the music piece according to the points of modulation and finding the key of each segment. Little work has been conducted on this topic. In this paper we propose to address the problem of local key finding by investigating the possible combination and extension of different previous proposed global key estimation approaches introducing key dependency on the harmonic and the metric structures. Harmony is directly related to the musical key. In Western tonal music, the chord progression determines the harmonic structure of a piece of music. It is strongly related to the musical key of the piece. A musical scale can be associated with each key. Chords that are specific to the key can be constructed around this scale. Although the idea to use chords to find the key of a musical excerpt has already been explored [12], to our knowledge, no precise study about the relationship between the two attributes has been conducted, in particular in the case of local key estimation. This partly comes from a lack of databases labeled in chords and local key. One contribution of this work is to present such a study on classical music pieces labeled in chords and keys containing many modulations. The problem of finding a good analysis window length for local key estimation has been evoked in the past, without any satisfying answer. Another contribution of our work is that we address this problem by introducing information related to the metric structure in our model. Key estimation is not performed on empirical-chosen segment length but on segments that are adapted to each piece.

The structure of the paper is as follows. First, in section 2, we review some previous works on global and local key estimation. We then present in section 3 our model for local key estimation, which relies on a probabilistic model for simultaneous chord progression and downbeat locations estimation. The local key estimation is based on the harmonic and metric structures of the piece. Eventually, in section 4, the proposed model is evaluated on a set classical music pieces. A conclusions section closes the article.

2. RELATED WORK

In this section, we review some previous works on key estimation. We start by template-based approaches proposed for global key estimation that have inspired our work. We then present previous methods proposed for local key estimation and conclude the section by reviewing key estimation methods based on chord progression. A large part of audio global key finding systems is based on the use of key profiles/templates. Pitch Class Profiles of Chroma features are extracted from the signal and then compared to theoretical templates that indicate the perceptual importance of notes or chords within a key. [1] proposes a method called the *probe tone method* that gives a measure quantifying the hierarchy of notes in a given tonal context. For major and minor keys, 12-dimensional vectors representing the perceptual importance of the 12 semitones of a chromatic scale in the considered key are proposed. These key profiles are used to estimate the key of a MIDI melodic line, by correlating it with a vector containing the

[*] This work was supported by the Quaero Project

relative duration of each of the 12 pitch classes within the MIDI sequence. [13] extends the model proposed in [1] to the case of polyphonic audio files by considering that the profile value for a given pitch class represents also the hierarchy of a chord in a given key. The polyphonic profiles for the 24 different keys are built considering only the three main triads of the keys (tonic, subdominant and dominant). This cognition-inspired method is compared with several machine-learning techniques. The methodologies are evaluated over a large audio database, achieving a 64% of correct overall tonality (mode and key-note) estimation. In this study it is found that the use of machine learning algorithms result in very little improvements over the cognitive-based technique. [11] compares a cognitive-based method similar to the one presented in [13] to an HMM-based approach. Two hidden Markov models are trained on a labeled database in order to learn the characteristics of the major and minor modes. From these two models, 24 hidden Markov models corresponding to the 24 keys are derived. The key of the audio file is then obtained by computing the likelihood of its chroma sequence given each HMM and selecting the one giving the highest value. It was found that the HMM-based approach leads to a lower recognition rate. Note that, in this work, the states in the HMMs have no musical meanings. [6] presents a template-based key finding model. The key is estimated by correlating spectral summary information obtained from audio with precomputed templates. The templates are obtained from real instrument sounds. For this, the spectra of the sounds are weighted by key profiles, which approximate the pitch distribution. Several key profiles are compared: Krumhansl's probe-tone ratings [1], Temperley's profiles [2] and a flat diatonic profile (12-dimensional vectors containing 1 at pith classes that are comprised in the considered diatonic scale, 0 elsewhere). The combination of the Temperley's and diatonic profiles was found to give the best results.

Concerning the problem of local key estimation, even if, compared to the problem of global key estimation, little work has been conducted on this topic, various approaches have already been proposed for this task. [14] presents a method for determining points of modulation in a piece of music in the symbolic domain using a geometric model for tonality called the Spiral Array which incorporates simultaneously pitch, interval, chord and key relations. This method has been extended to the audio case in [10]. [15] presents an approach for detecting multiple keys and locating the key boundaries in the melody of popular songs in MIDI format. Overlapping segments are first extracted form the melody using a diatonic scale model, each one corresponding to a single mode. Segments of unrelated modes are eliminated. Key labels and boundaries are determined by grouping the remaining segments. Another geometric tonality model describing relationship between keys has recently been proposed in [16]. [17] proposes a method for detecting changes in the harmonic content of musical audio signals. A new model for equal tempered tonal space is introduced. Segmentation of audio signal and preprocessing stage for chord recognition and harmonic classification algorithms using HMMs are the main potential applications. [18] presents an approach to derive an appropriate representation of tone centers based on the audio signal using constant Q profiles. The constant Q profiles are 12-dimensional vectors where each component refers to a pitch class. They are derived from sampled cadential chord progressions and small pieces of music. Tonal centers of a music piece are tracked by computing cq-profiles of the piece and matching every given cq-profile with a profile of the reference set using a fuzzy distance. [19] proposes an HMM-based method to segment musical signals according to the key changes and to identify the key of each segment. The front-end of the system is based on the calculation of a chromagram. The key detection task is divided into two steps: first the key is estimated without considering the mode because diatonic scales are assumed and relative modes share the same diatonic scale. The mode (major or minor) is then estimated. Classical piano music is employed to test the performances of the proposed method using three measures: recall, precision and label accuracy. [20] proposes an interesting new model for detecting modulations and labeling local keys using a non-negative matrix factorization method for segmentation. To identify sections that are candidates for unique local keys, groups of contiguous chroma vectors are used as input in the segmentation stage. The length of the window is chosen empirically. The local keys are then found using a correlation model. The method is evaluated on three different data sets: pop songs, classical music and Kosta and Payne corpus.

Because chords and musical keys are musical attributes closely related to each other in Western tonal music, the idea to use the chord progression of a piece to find the keys comes out naturally. In the framework of global key estimation, [21] proposes key-dependent chord HMMs trained on synthesized audio for chord recognition and global key estimation. In this approach, 24 key-dependent HMMs, one for each major and minor keys are built. Key estimation and chord recognition are performed simultaneously selecting the model whose likelihood is highest. It is observed that the proposed method is similar to [11] but, whereas in [11] the states in the HMMs have no musical meanings, in [21], hidden states are treated as chords, which also allows identifying the chord sequence. [22] presents a technique to estimate the predominant key in a symbolic musical excerpt. A HMM is used where the hidden states are the 24 major and minor keys and the observations are pairs of consecutive chords. Human expectation of harmonic relationships is encoded in the model using results from perceptual tests. The parameters of the HMM are trained using hand-annotated chord symbols. This work was extended to the audio case in [12]. Although this model has only been evaluated on the case of global key estimation, it could be used for local key estimation. A recent work [23] proposes a probabilistic framework for simultaneously estimating keys and chords. Novel observation likelihood model and chord/key transition models are proposed that are derived from music theory of Lerdahl.

3. PROPOSED APPROACH

In this paper we are interested in the problem of local key finding in polyphonic audio files. We propose to combine and extend methods proposed for global key finding to the case of local key finding. We rely on the above-mentioned method for global key estimation [13] based on key reference profiles, which are correlated with input pitch class profiles. The underlying idea of this work is that in case of polyphonic music, the chords can be used to estimate the musical key. However, in this previous work, as in [11], there is no estimation of the chords and no investigation of their relationship to keys. We study this relationship in the present work. To integrate the concept of key modulating over time, we propose to use an HMM where the hidden states are the keys which can be observed through observable data that are the chords. The use of the HMM allows us to integrate some musical information about key changes, as proposed in [22]. As shown in the last sec-

tion, HMM have already been used for local key estimation ([12], [19]). However, this was done using a frame-by-frame analysis. A contribution of the present work is that we introduce information related to the metric structure of the audio file in order to make the local key estimation robust. One of the problems when segmenting a piece of music into sections with different keys is to accurately choose the length of the analysis window used for key estimation. In the case of global key estimation, only the first seconds of the piece are used to estimate the key. Several studies have shown that the choice of the duration of the analyzed excerpt has a significant impact on the key estimation results (see for instance [6] or [10]). Concerning local key estimation, the length of the analysis window was found empirically in previous works. After computing chroma vectors on short overlapping frames, [19] or [18] perform a frame-by-frame musical key analysis. An interesting alternative to sliding window key center tracking techniques has been proposed by [20] where a segmentation stage which identifies sections that are candidates for unique local keys is performed prior to local key estimation. Groups of contiguous chroma vectors are used as input. Heavily overlapped groups of chroma vectors are averaged over a span of s seconds. The value of the parameter s is found empirically ($7.4s$) after testing several window sizes. The question of optimal segment length remains an open problem. A too small window size would focus the chromagram on individual chords more than on keys whereas the use of a too large window size would lead to segments containing several keys and key modulations points would become ambiguous. The drawback of using an empirically chosen window size is that, if the testset contains for instance some pieces with a fast tempo compared to the others, the window length will probably be too long and changes of keys will be ignored by the algorithm. For pieces with a slow tempo, chords more than keys will be estimated. Ideally, the window length should be related to the tempo of the piece. We get around this difficulty here by segmenting the piece according to the metric structure. We perform a beat-synchronous analysis. For local key estimation, the temporal unity, which is used here for key analysis, is the musical bar. The analysis window length has thus a musical meaning.

3.1. Model

In this section we present a model that allows estimating the local keys of a musical excerpt using the underlying chord progression, which characterize the harmonic structure, and the downbeat locations, which characterize the metric structure. Metrical level is a hierarchical structure. The beat or the tactus level is the most salient metrical level and corresponds to the foot-tapping rate. Musical signals are divided into units of equal time value called *measures* or *bars*. One important attribute of the metric structure is the *downbeats* or the first beats of each measure. Here, the chords and the downbeats are estimated simultaneously using a "double-states" HMM where a state is a combination of a chord type and a position of the chord in the measure. We consider here a chord lexicon composed of the $I = 24$ Major and minor triads (C Major, ..., B Major, C minor, ..., B minor). The local key estimation model is close to the chord estimation model. The 24 key space is modeled by an ergodic 24-states HMM, where each state represents one of the 24 major and minor keys. In our model, the hidden states (keys) can be observed through observable data that are related to the chord progression of the piece. At each time instant, the chords imply a local key. At each time instant, the key generates an observable 24-dimensional vector representing the probability of each of the 24 chords to have been emitted. Given the observations, we estimate the most likely key sequence over time in a maximum likelihood sense. The flowchart of the system is represented in Figure 1.

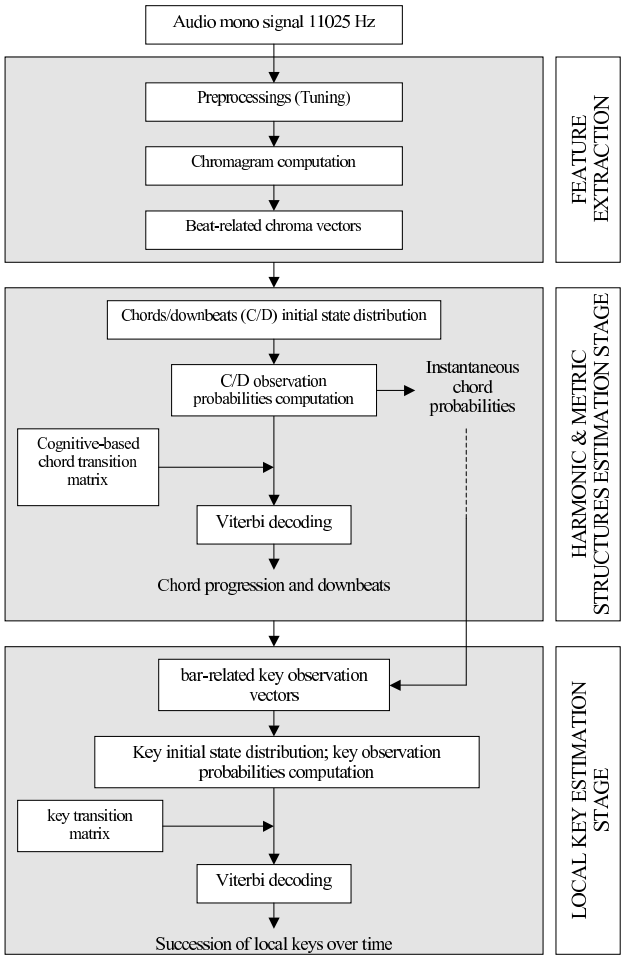

Figure 1: Flowchart of the local key estimation system.

3.2. Feature vectors

As most of chords and key detection systems, the front-end of our system is based on the extraction of a set of feature vectors that represent the audio signal, the Pitch Class Profiles [24] or *chroma* vectors [25]. The succession of chroma vectors over time is known as *chromagram*. The chroma vectors are in general 12-dimensional vectors that represent the spectral energy of the pitch classes of the chromatic scale. For chromagram computation, we use the method we proposed in [26]. To integrate the metric structure of the piece, we built meter-related features by averaging the chroma vectors according to the beat locations so that we obtain one feature vector per beat[1].

[1]This supposes to integrate a beat-tracker as a front-end of the system. In our experiments, the beat locations have been annotated by hand because the testset is composed of classical music pieces containing lots of deviations in tempo that results from the expressivity in classical music.

3.3. Harmonic and metric structures

The harmonic structure is defined by the chord progression and the metric structure is defined by the downbeat locations. These two musical attributes are estimated simultaneously according to the method we proposed in [26]. This method is briefly summarized here. We consider an ergodic $I * K$-states HMM where each state s_{ik} is defined as an occurrence of a chord c_i, $i \in [1 : I]$ at a "beat location in the measure" b_k, $k \in [1; K]$: $s_{ik} = [c_i, b_k]$. In our case $I = 24$ chords and $K = 4$ for a song built on constant four-beats meter, $K = 3$ for a song built on constant three-beats meter. Each state in the model generates with some probability an observation vector $\mathbf{O}(t_m)$ at time t_m defined by the observation probabilities. Given the observations, we estimate the most likely chord sequence over time and the downbeat locations in a maximum likelihood sense.

Initial state distribution: The prior probability π_{ik} for each state is the prior probability to observe a specific chord c_i occurring on a beat location in a measure b_k. Since we do not know *a priori* the chord and the beat location the piece begins with, we initialize π at $\frac{1}{I*K}$ for each of the $I * K$ states.

Observation chord symbol probability distribution: The observation probabilities are computed as:

$$P(\mathbf{O}(t_m)|s_{ik}) = P(\mathbf{O}(t_m)|c_i)P(\mathbf{O}(t_m)|b_k) \qquad (1)$$

where $P(\mathbf{O}(t_m)|c_i)$ corresponds to the chord symbol observation probabilities and $P(\mathbf{O}(t_m)|b_k)$ corresponds to the beat location in the measure observation probabilities. The observation chord symbol probabilities are obtained by computing the correlation between the observation vectors (the chroma vectors) and a set of chord templates which are the theoretical chroma vectors corresponding to the $I = 24$ major and minor triads. In what follows, the succession of these 24-dimensional vectors is referred to as the *chordgram*. The computation of the *chordgram* is detailed below. The beat location in the measure observation probabilities is considered here as uniform.

State transition probability distribution: The transitions between chords result from musical rules which are modeled in the state transition matrix T. These rules are based on the harmonic structure and the metric structure. The transition matrix T used in our HMM takes into account both the chord transitions and their respective locations in the measure. To integrate harmonic rules, we derive the $I * K$-states transition matrix T from a I-states chord type transition matrix T_c based on music-theoretical knowledge about key-relationships. This matrix is the same than the key transition matrix described below. We integrate metric rules in the $I*K$-states transition matrix T relying on the following statement: chords are more likely to change at the beginning of a measure than at other beat locations [27]. To favor chord changes on downbeats, we attribute a lower self-transition probabilities[2] in the state transition matrix T for chords occurring on the K^{th} beat. The transition matrix is for a four-beat meter song represented in Figure 2. The harmony is more likely to change after a chord occurring on the 4^{th} beat of the measure than after a chord occurring on a 3^{rd} beat of the measure.

[2]Here, a self-transition means a transition between two identical chord types, for instance from a CM chord to a CM chord.

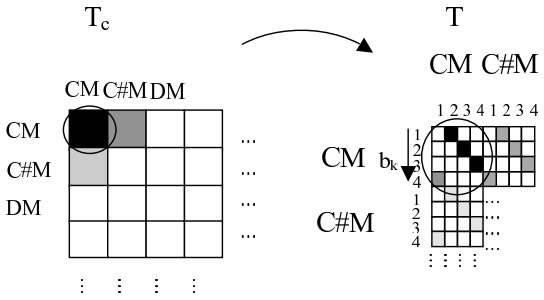

Figure 2: Chord transition matrix for a singles-state HMM [left], transition matrix in the case of a double-states HMM taking into account the position of the chord in the measure [right].

Chord progression and downbeats detection: The optimal succession of states $[c_i, b_k]$ over time is found using the Viterbi decoding algorithm [28] which gives us the most likely path through the HMM states given our sequence of observations. We obtain simultaneously the best sequence of chords over time and the downbeat locations.

3.4. Chordgram

The chordgram is a succession of 24-dimensional vectors representing the probability that each chord has been emitted at each tactus-frame. These instantaneous chord probabilities are obtained by computing the correlation between the chroma vectors and 24 chord templates. Each chord template is a 12-dimensional vector that contains the theoretical amplitude values of the notes and their harmonics composing a specific chord. The chord templates are constructed considering the presence of the higher harmonics of the theoretical notes it consists of, relying on the model presented in [5]: the amplitude contribution of the h^{th} harmonic composing the spectrum of a note is set to 0.6^{h-1}. The chordgram is used for local key estimation.

3.5. Extraction of key observation vectors

The key observation vectors are derived from the chords. In the evaluation part, we will compare two methods. In the first case, the key observation vectors are built from the *chordgram* using the instantaneous chord probabilities. In the second case, they are built directly from the estimated chord progression. In general, the musical key of a music piece changes much less often that the chords and remains the same during several bars. We segment the piece into overlapping segments whose length is related to the measures delimited by the downbeat. The local key is thus estimated on segments that have a musical meaning. Because musical phrase have often length duration of 8 or 4+4 bars, we have chosen to segment the pieces into 2-bars segments with 1-bar overlap. Because key changes occur in general on the first beat of a measure it is important that the analysis starts on a downbeat. In our experiments we have tested the algorithm using other window analysis length and found that the local key estimation results accuracy decreases with longer windows. This is discussed below in section 4.3. The key observation vectors are 24-dimensional vectors obtained by averaging the *chordgram* or the estimated chord progression along the overlapping 2-bars length segments.

3.6. Key estimation from chords using hidden Markov models

From the key observation vectors, we estimate the succession of keys in the track. The method is very similar to the one we proposed for chord estimation. The initial state distribution of keys is uniform ($\frac{1}{24}$ for each of the 24 states) since we have no reason to prefer a key above another. The observation key probabilities $P(k_i|\mathbf{O}(t_m))$ are obtained by computing the correlation between the key observation vectors and a set of key profiles that represent the importance of each triad within a given key.

The key profiles are obtained using a method similar to the one proposed in [13]. In the monophonic case, Krumhansl proposes probe tone ratings [1] that represent the distribution of the pitches according to the musical key. Two types of rating vectors are proposed, one for the major mode and one for the minor mode. Temperley has modified these key profiles in [2]. [6] proposed Temperley-Diatonic pitch-distribution profiles which were extended in [11] to the polyphonic case. In part 4, we will compare all these key templates and propose a new one where all notes have the same weight in the template but the one corresponding to the tonic which has a triple weight. As in [11] and [13] the polyphonic profiles for the 24 different keys are built considering the three main triads of the keys (tonic, subdominant and dominant). For instance, for a C major key, only C major, F major and G major chords are considered. We detail below the key profiles computation for major mode, the minor key profiles are obtained in a similar way. Let $T_i^M, i \in [1, 12]$ denote the monophonic major key templates. The T_i^{Mp} polyphonic major key templates are computed according to the following equation:

$$\begin{cases} T_i^{Mp}(k) & = T_i^M(i), & \text{if } k = i, \\ & = T_i^M((i+5)[12]), & \text{if } k = (i+5)[12], \\ & = T_i^M((i+7)[12]), & \text{if } k = (i+7)[12], \\ & = 0 & \text{otherwise,} \end{cases} \quad (2)$$

where $a[m]$ denotes the mathematical operator *modulo*, the remainder when a is divided by m.

The observation key probabilities $P(k_i|\mathbf{O}(t_m))$ are obtained according to Equation (3) and normalized so that

$$\sum_i P(\mathbf{O}(t_m)|k_i(t_m)) = 1.$$

Let $\mathbf{T}_i^p, i \in [1, 24]$ denote a key template.

$$For \quad i = 1 \dots 24, \quad P(\mathbf{O}(t_m)|k_i(t_m)) = \frac{\mathbf{O}(t_m).\mathbf{T}_i}{\|\mathbf{O}(t_m)\|.\|\mathbf{T}_i\|} \quad (3)$$

Key modulations in a music piece follow musical rules that can be reflected in the state transition matrix. To integrate musical meaning in key transition, we adopt the key transition matrix proposed in [22] already used as a chord transition matrix[3]. In [1], Krumhansl studies the proximity between the various musical keys using correlations between key profiles obtained from perceptual tests. These key profile correlations have been used in [22] to derive a key transition matrix in the context of local key estimation as described below. Krumhansl gives numerical values corresponding to key profile correlations for C major and C minor keys. The values can be circularly shifted to give the transition probabilities for keys other than C major and C minor. In order to have probabilities, all the values are made positive by adding 1, and then normalized to sum to 1 for each key. The size of the final key transition matrix is 24 x 24.

The optimal succession of states over time is found using the Viterbi decoding algorithm that gives us the best sequence of keys over time. The music piece is thus segmented into segments that are labeled by a key.

[3] Chords and key are musical attributes related to the harmonic structure and can be modeled in a similar way.

4. EVALUATION

4.1. Testset

Classical Mozart piano pieces were used to evaluate the algorithm. Trained musicians manually annotated the ground truth for chords and local key by hand. Beat locations have first been annotated using the software *Wavesurfer*. Trained musicians had provided a list of the chords and key with their duration in beats. The list has been then automatically mapped to the beat locations resulting in the ground truth we use[4]. The testset consists in 5 movements of Mozart piano sonatas: KV 283 #1 & 2, KV 309 #1 , KV 310 #1and KV 311 #2 corresponding to about 30 minutes of music. Each piece contains several modulations and this is one of the main reasons why they were selected. It has to be noticed that it is very hard to label Mozart pieces in chords and musical key, even for a well-trained musician because on the one hand, there are a lot of ornamental notes (such as appoggiaturas, suspensions, passing notes etc.) and on the other hand, harmony is frequently incomplete (some notes of the chord are missing). This makes the choice of chords labels very difficult. Changes from one key to another are often ambiguous, in particular when they are very short. Moreover, modulation is very often a smooth process, it can take several bars to establish properly a tonal center. Segments corresponding to transition from one key to another have been labeled as transition parts.

4.2. Evaluation measure

Chord estimation evaluation measure: The result of chord estimation we give corresponds to the mean and standard deviation of correctly identified frames per song. Parts of the pieces where no chord can be labeled (for instance when a chromatic scale is played) have been ignored in the evaluation.

Local key estimation evaluation measure: Concerning local key estimation, we consider, as in [19], two aspects of the results: the label accuracy (how the estimated key is consistent with the ground truth) and the segmentation accuracy (how the detected modulation points are consistent with the actual locations). For evaluating local key label accuracy, we use a measure similar to the one used for evaluating chord label accuracy. For evaluating segmentation accuracy, we use two metrics proposed in [19]. $Precision(P)$ is defined as the ratio of detected transitions that are relevant. $Recall(R)$ is defined as the ratio of relevant transitions detected. We also give the $f-measure(F)$ which combines the two $F = 2RP/(R+P)$. A change of key can take several bars. Two established keys are often separated by a transition part where no key is firmly established. These parts, which have been labeled as transition parts T in the ground truth, need to be taken

[4] The ground-truth chords and keys progression in beats can be obtained by contacting the authors

Table 1: Chords and local keys label accuracy results using a 2-bars length window and the newly proposed templates. Method 1): based on the chordgram. Method 2): based on the chord progression.

	keys method 1)	keys method 2)	chords
label accuracy (%)	80.22	74.11	61.43

into account in the evaluation of segmentation accuracy. For this, a tolerance window w is used in the following way. If a modulation is detected at frame n_1 and close enough to a relevant modulation of the ground truth labeled at frame n_2 such that $|n_1 - n_2| < w$, It is considered as correct. The greatest the value of w is, the highest the precision and recall are. We present below results with w corresponding to 1 or 2 bars.

4.3. Results and discussion

We have carried out several experiments to evaluate the impact of various parameters on the local key estimation results: choice of the key templates, choice of the length of the analysis window, key estimation from the *chordgram* or from the estimated chord progression, influence of the tolerance window.

Relationship between chords and local key: We have evaluated two different methods for local key estimation. In the first one (method 1), the probability of each chord at a given time instant is used to estimate the key. In the second one (method 2), the chords are first estimated using a hidden Markov model and the local key is derived from the estimated chord progression. Label accuracy results are presented in Table 1. It is difficult to select the best between the two presented methods. Indeed, the best label key results are obtained with (method 1) but it can be seen in Table 3 that method 2) slightly outperforms method 2) concerning local key segmentation. Tests on a larger database would be needed to clearly evaluate the performances of the two methods.

The analysis of the results piece by piece shows that there is a correlation between the estimation of the chords and the estimation of the key. We expected that a good estimation of the chords would lead to a good estimation of the keys. This was corroborated when evaluating method 2). A good estimation in the chord estimation resulted in a good estimation in the local keys whereas a poor estimation of the chords resulted in a poor estimation of the local keys. A deeper analysis showed that if the chord estimation errors consisted in confusions with harmonically close chords (such as dominant or subdominant chords), the key was correctly estimated.

Importance of the metric structure: In Table 2, we present the label accuracy results when the key analysis windows are set according to the downbeat locations (OD, on downbeats) and when the starting point is not a downbeat (ND, no downbeats). To investigate the hypothesis of the importance of the metric structure on the local key estimation, we have positioned the starting point of the key analysis windows on a second beat in case of ND (no downbeats). It can be seen that the label accuracy results are better when the starting point is a downbeat. This is because key changes occur in general on downbeats. When the chordgram or the chord progression are averaged against the downbeat locations, some passages with different local keys may be mixed. There is no clear difference in the key segmentation results. This is probably due to the smoothness of the modulations (see below). Considering the metrical structure allows to improve the key estimation results.

Table 2: Local keys results using a 2-bars length window and the newly proposed templates in case of method 1), when the key analysis windows are set according to the downbeat locations (OD, on downbeats) and when the starting point is not a downbeat (ND, no downbeats). The tolerance window is $w = 1$ bar.

	OD	ND
label accuracy (%)	80.21	76.43
segmentation f-measure	0.52	0.50

Effect of the length of the analysis window: In classical music, musical phrases have in general a length of 4 or 8 bars. This is particularly true for Mozart's piano sonatas. Usually, the musical key remains constant within a phrase or at least within half of the phrase (whereas the harmony changes several times). This is why we chose to estimate the local key on segments of length corresponding to musical phrases. We have evaluated the algorithm with different window lengths: 2, 4, 8 and 16 bars. The best results were obtained using a 2-bar length analysis window. This is because, especially in slow movements, some modulations occur after only 2 bars. Passages with different local keys are very likely to be mixed when a longer analysis window is used. The accuracy of the results decreases with the length of the analysis window, as illustrated in Figure 3.

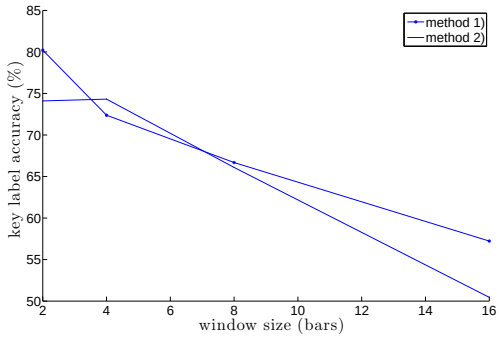

Figure 3: Key estimation results in case of method 1) and 2) according to the length of the key analysis window.

Effect of the choice of the key templates: Several key templates have been proposed for global key estimation based on key templates. We investigated the impact of the type of used key templates on the results. We evaluated the algorithm with 5 types of templates: Krumhansl, Temperley, diatonic, a combination of Temperley and diatonic and finally a newly proposed key template where all notes have the same weight in the template except the

Table 3: Local keys segmentation accuracy (SA) results using a 2-bars length window and the newly proposed templates. Method 1): based on the chordgram. Method 2): based on the chord progression. two tolerance windows: $w = 1$ bar and $w = 2$ bars.

	keys method 1)		keys method 2)	
	$w = 1$	$w = 2$	$w = 1$	$w = 2$
SA precision	0.5723	0.8196	0.4489	0.6805
SA recall	0.4730	0.6874	0.7131	0.8691
SA f-measure	0.5170	0.7327	0.5451	0.7514

one corresponding to the tonic, which has a triple weight. The best results were obtained using the newly proposed templates. The next best results were obtained with the combination of Temperley and diatonic key templates. This corroborated the experimental results obtained in the case of global key estimation in [6] and [11].

Smooth modulations: The key segmentation accuracy results are presented in Table 3 with two tolerance windows: $w = 1$ bar and $w = 2$ bars. It can be seen that the segmentation accuracy results increase a lot when we use a larger tolerance window. This can be explained by the fact that changes in keys are a very smooth process that often takes several bars. It is thus difficult to estimate the precise local keys boundaries. It would be interesting to formulate and add a "local key transition" state in the model. This is left for future works.

5. CONCLUSION AND FUTURE WORKS

In this paper, we have presented a local key finding model that segments an audio file in sections labeled with local keys. The method combines and extends several previous methods proposed for global key estimation. The local key progression over time is modeled according to the harmonic and the metric structures. The local key segmentation has a musical meaning and is independent of the tempo of the piece. Encouraging results are obtained on a set of classical pieces with complex harmony structure and show that the key progression is clearly related to the harmonic and the metric structures. Analysis of the results shows that additional improvement of key segmentation may be achieved in the future using a more complex model that includes key transitions parts. We also plan to introduce chord functional analysis information to improve the results.

6. REFERENCES

[1] C.L. Krumhansl. *Cognitive Foundations of Musical Pitch*. Oxford University Press, New York, 1990.

[2] D. Temperley. *The Cognition of Basic Musical Structures*. Cambridge, MA: MIT Press, 2001.

[3] S. Madsen G. Widmer. "key finding with interval profiles". In *ICMC*, Copenhagen, Denmark, 2007.

[4] A. Mardirossian and E. Chew. "skefis – a symbolic (midi) key-finding system". In *1st Annual Music Information Retrieval Evaluation eXchange, ISMIR*, London, UK, 2005.

[5] Emilia Gómez. Tonal description of polyphonic audio for music content processing. *INFORMS J. on Computing*, 18(3):294–304, 2006.

[6] O. Izmirli. Template based key finding from audio. In *ICMC*, pages 211–214, Barcelona, Spain, 2005.

[7] Steffen Pauws. Musical key extraction from audio. In *ISMIR*, pages 96–99, Barcelona, Spain, 2004.

[8] Y. Zhu and M.S. Kankanhalli. Precise pitch profile feature extraction from musical audio for key detection. *IEEE Transactions on Multimedia*, 8(3):575–584, 2006.

[9] S. van de Par, M.F. McKinney, and A. Redert. Musical key extraction from audio using profile training. In *ISMIR*, pages 328–329, Montreal, Canada, 2006.

[10] Ching-Hua Chuan and Elaine Chew. Audio key finding: considerations in system design and case studies on chopin's 24 preludes. *EURASIP J. Appl. Signal Process.*, 2007(1):156–156, 2007.

[11] G. Peeters. Musical key estimation of audio signal based on hmm modeling of chroma vectors. In *In DAFX, McGill*, pages 127–131, Montreal, Canada, 2006.

[12] K. Noland and M. Sandler. Signal processing parameters for tonality estimation. In *Proceedings of the AES 122nd Convention*, Vienna, Austria, 2007.

[13] E. Gomez and P. Herrera. Estimating the tonality of polyphonic audio files: Cognitive versus machine learning modelling strategies. In *ISMIR*, pages 92–95, Barcelona, Spain, 2004.

[14] E. Chew. The spiral array: An algorithm for determining key boundaries. In *Proceedings of the Second International Conference, ICMAI 2002*, pages 18–31. Springer, 2002.

[15] Y. Zhu, M. Sandler and M. Kankanhalli. Key-based melody segmentation for popular songs. In *ICPR*, Cambridge, UK, 2004.

[16] D. Gatzsche G. Gatzsche, M. Mehnert and K. Brandenburg. A symmetry based approach for musical tonality analysis. In *ISMIR*, Vienna, 2007.

[17] C. Harte, M. Sandler and M. Gasser. Detecting harmonic change in musical audio. In *AMCMM*, Santa Barbara, 2006.

[18] Benjamin Blankertz Hendrik Purwins and Klaus Obermayer. A new method for tracking modulations in tonal music in audio data format. In *International Joint Conference on Neural Networks*, page 2000, 2000.

[19] W. Chai and B. Vercoe. Detection of key change in classical piano music. In *ISMIR*, London, 2005.

[20] O. Izmirli. Localized key finding from audio using non-negative matrix factorization for segmentation. In *ISMIR*, Vienna, 2007.

[21] K. Lee and M. Slaney. A unified system for chord transcription and key extraction using hidden Markov models. In *ISMIR*, Vienna, 2007.

[22] K. Noland and M. Sandler. Key estimation using a hidden Markov model. In *ISMIR*, pages 121–126, Victoria, Canada, 2006.

[23] J.P. Martens B. Catteau and M. Leman. A probabilistic framework for audio-based tonal key and chord recognition. In R. Decker and H.-J. Lenz, editors, *Advances in Data Analysis*, pages 637–644, Berlin, 3 2007. Springer.

[24] T. Fujishima. Real-time chord recognition of musical sound: A system using common lisp music. In *ICMC*, pages 464–467, Beijing, China, 1999.

[25] G.H. Wakefield. Mathematical representation of joint time-chroma distribution. In *SPIE Conf. Advanced Sig. Proc. Algorithms , Architecture and Implementation*, volume 3807, July Denver, Colorado, 1999.

[26] H. Papadopoulos and G. Peeters. Simultaneous estimation of chord progression and downbeats from an audio file. In *ICASSP*, Las Vegas, 2008.

[27] M. Goto. An audio-based real-time beat tracking system for music with or without drum sounds. *Journal of New Music Research*, 30(2):159–171, 2001.

[28] B. Gold and N. Morgan. *Speech and audio Signal Processing: Processing and Perception of Speech and Music*. John Wiley & Sons, Inc., 1999.