Local Key Estimation from an Audio Signal Relying on Harmonic and Metrical Structures

Hélène Papadopoulos, Geoffroy Peeters

Local Key Estimation from an Audio Signal Relying on Harmonic and Metrical Structures

Hélène Papadopoulos* and Geoffroy Peeters

***Abstract*—In this paper, we present a method for estimating the progression of musical key from an audio signal. We address the problem of local key finding by investigating the possible combination and extension of different previously proposed approaches for global key estimation. In this work, key progression is estimated from the chord progression. Specifically, we introduce key dependency on the harmonic and the metrical structures. A contribution of our work is that we address the problem of finding an analysis window length for local key estimation that is adapted to the intrinsic music content of the analyzed piece by introducing information related to the metrical structure in our model. Key estimation is not performed on empirically chosen segments but on segments that are expressed in relationship with the tempo period. We evaluate and analyze our results on two databases of different styles. We systematically analyze the influence of various parameters to determine factors important to our model, we study the relationships between the various musical attributes that are taken into account in our work, and we provide case study examples.**

***Index Terms*—Acoustic signal analysis, chord, key, downbeat, HMM, chroma, information retrieval, music analysis.**

I. Introduction

TONALITY analysis is one of the most important aspects of Western tonal music. It describes the relationships between the various musical keys present in a piece of music. 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 analysis has interested researchers for a long time because the key detection task finds many applications in content-based music information retrieval such as classification, segmentation, indexing and summarization.

In this article, all musical concepts are formalized in the context of modern Western tonal music, *i.e.* after the 16^{th} century. When an instrument produces a note, the human listener perceives a *pitch* that is a perceptual attribute of sound. In music, the term *note* is a symbol that is used to refer both to the relative duration and to the pitch of a given sound. Pitches are governed by structural principles and music is organized around one or more stable reference pitches. The system of relationships between pitches corresponds to a *key*. A key, as a theoretical concept, implies a tonal center that is the most stable pitch called the *tonic*, and a *mode* (usually major or minor).

A scale is associated with each key. A *scale* is a series of notes arranged in ascending or descending order. Fig. 1 represents a C major scale and its relative A natural minor scale. Two consecutive notes are usually separated either by a *tone* (T) or a *semitone* (S). The position of tones and semitones within a scale associated to a key characterizes its *mode*. The series of seven notes of the scale without any alterations[1] is known as the *diatonic scale*.

In a scale, the *tonic* or first scale degree (I) is the first note and it is the pitch upon which all other pitches of a piece are hierarchically referenced. Among the seven other scale degrees, the 3^{rd} - the *mediant* - and the 5^{th} - the *dominant* - degrees are particularly important since their combination with the tonic results in the most significant chord in a given key.

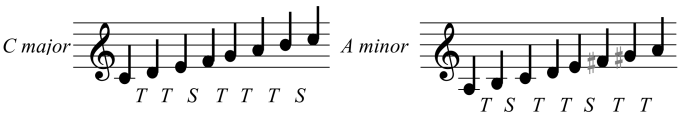

Fig. 1. Example of major and minor scales: C major, A minor. The accidentals that characterize the harmonic and melodic minor scales are represented in grey.

There are two common variations of the natural minor scale: i) The *melodic minor scale*, in which the 6^{th} and the 7^{th} ascending degrees are raised a semitone ($F\#$ and $G\#$ in Fig. 1); ii) The *harmonic minor scale*, in which the 7^{th} degree, both ascending and descending is raised a semitone ($G\#$ in Fig. 1). In this work, we focus on the harmonic minor scale. We consider enharmonic equivalence, *i.e.* notes with different spelling but sounding the same are considered the same (C# is equivalent to Db). In Western tonal music, there are 12 pitches in an octave range. The major and minor scales and the twelve tonics give rise to a total of 24 possible keys.

A set of chords that are specific to the key can be constructed around its scale. A *chord* is defined as a

[1]An alteration is a *sharp* $\#$, which raises the pitch of a note one semitone, or a *flat* b, which lowers it one semitone.

H. Papadopoulos is with the Computer Science Department, University of Victoria, Victoria, BC V8P 5C2, Canada; e-mail: helenepapadopoulos@hotmail.com).

G. Peeters is with the Sound Analysis/Synthesis Team, IRCAM / CNRS-STMS, Paris, FRANCE (e-mail: Geoffroy.Peeters@ircam.fr).

Research partially performed while H.Papadopoulos was a PhD student at IRCAM, Paris and while she was with the Laboratory of Signals and Systems, University Paris 11 - CNRS - Supélec, Orsay, FRANCE. This work was partially supported by the QUAERO project.

combination of three or four notes sounded simultaneously. The succession of chords over time is called a *chord progression*. It is strongly related to the musical key.

Various approaches have been proposed for estimating the *main* or *global* key of a piece of music. Some approaches have been proposed for symbolic data, based on templates [1]–[3], or geometry [4]. Others have been proposed for audio data, based on templates [5]–[9], geometry [10], or hidden Markov models (HMMs) [11]. Finding the main key of a piece of music is only a small 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 moves between keys. A change between different keys is called a *modulation*. Western tonal music can be conceived of as a progression of a sequence of key regions in which pitches are organized around one or more stable tonal center. Such a key region is defined here as a *local key*, as opposed to the *global key*.

Tonality and key are perceptual attributes. Experiments have shown that people with different levels of music training may not perceive equally the stability of a given pitch within a tonal context. Moreover, the sense of key may be more or less strong [1]. At some points, the key may seem ambiguous or unstable and it may not be possible to select one pitch as being more stable than the others. In this case, a tonal center cannot be defined and the key is not *established*. It must be also noticed that, if the notion of key is quite clear in Western "classical" music, it is much more complicated in "popular" music. For instance popular music may use different systems, such as local tonality, apparent tonality, no tonality etc., within a single song. All these considerations make the key estimation task challenging.

The purpose of this work is to investigate the problem of local key estimation, which is more complex than the problem of global key estimation: we aim at segmenting the music according to the points of modulation and finding the key of each segment. Little work has been conducted on this topic. We propose to address the problem of local key finding by investigating the possible combination and extension of various previously proposed approaches for global key estimation and by introducing key dependency on the harmonic and metrical structures.

To our knowledge, although the idea of using chords to find the key of a musical excerpt has already been explored [12]–[16] no precise analysis 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 with chords and local key. One contribution of this work is to present such a study on real recordings of classical and popular music pieces labeled with 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 [6], [10], [16]–[19]. Another contribution of our work is that we address this problem by introducing information related to the metrical structure in our model. Key estimation is not performed on empirically chosen segment length but on segments that are adapted to the musical content of each piece.

The structure of the paper is as follows. First, in Section II, we review some previous works on global and local key estimation. We then present in Section III our model for local key estimation, which relies on a probabilistic model for simultaneous estimation of the chord progression and the downbeat positions. The local key estimation is based on the harmonic and metrical structures of the piece. Eventually, in Section IV, the proposed model is evaluated on two sets of music pieces of various styles. The first one contains classical music pieces and the second one consists in a set of popular music pieces. A conclusion section closes the article.

II. Related work

In this section, we review some previous work 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. For a detailed review of key estimation from audio signals, we refer the reader to [20].

A. Global Key

The problem of automatically estimating the key of a piece of music was first addressed in the context of symbolic music (*e.g.* MIDI format). In what follows, we will review two of the most popular techniques that have been extended later to the case of audio music. For a detailed review of key finding in symbolic music and more generally on tonality induction, we refer the reader to [21] or [22]. Most of the algorithms that extract key from audio start by computing a set of features that represent the signal, typically *chroma* features [23] or *Pitch Class Profiles* (PCP) features [24], which are then used as an input to a tonality induction model. The chroma vectors are in general 12-dimensional vectors that represent the spectral energy of the pitch classes of the chromatic scale. The succession of chroma vectors over time is known as the *chromagram*.

A large part of audio global key finding systems is based on the use of key profiles/templates. *Pitch Class Profiles* or *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. This idea was first proposed by Krumhansl and Schmuckler in [1]. In this work, a method called the *probe tone method* is presented. It gives a measure quantifying the hierarchy of notes in a given tonal context. The algorithm, known as the Krumhansl & Schmuckler (K-S) algorithm, computes the correlation between a vector of pitch-class durations obtained from a musical passage and a set of major and (harmonic) minor key-profiles corresponding to each key. The key profile that provides the maximum correlation is taken as the most probable key of the musical excerpt.

Gómez & Herrera [25] extend 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 expectation of a chord in a given key. The model uses as input features the Harmonic Pitch Class Profiles (HPCP) that account for higher harmonics of the notes. 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 results in little improvements over the cognitive-based technique.

Peeters [11] compares a cognitive-based method similar to the one presented in [25] to an approach based on hidden Markov models. Two HMMs are trained on a labeled database in order to learn the characteristics of the major and minor modes. From these two models, 24 HMMs 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.

Izmirli [6] presents a template-based key finding model. The key is estimated by correlating spectral summary information obtained from audio with pre-computed 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 pitch classes that are comprised in the considered diatonic scale, 0 elsewhere). The combination of Temperley's and the diatonic profiles was found to give the best results.

B. Local Key

Chew [26] 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 is extended to the audio case by Chuan & Chew [10]. In this work, a basic system that generates pitch-class information using a fuzzy analysis and calculates key results using the CEG algorithm is introduced. Three key determination policies are investigated (nearest-neighbor (NN), relative distance (RD), and average distance (AD)). Experiments are conducted on 410 classical music pieces by various composers across different time and stylistic periods (from Baroque to Contemporary). It is found that the AD policy gives the best main key estimation results (79%). Three extensions to the basic key finding system are then proposed (the modified spiral array (mSA), fundamental frequency ($f0$) identification, and post-weight balancing (PWB)) and evaluated on Chopin's 24 *Préludes*. Quantitative evaluation of main key estimation is proposed. The problem of local key finding is only considered on some examples.

Another geometric tonality model describing relationships between keys has recently been proposed in [27]. It is derived from the cognition-based model proposed in [28]. Tones are organized so that tonal symmetries within Western tonal music become apparent.

Some approaches rely on a frame-by-frame analysis or use a sliding analysis window. Purwins *et al*. [18] present an approach to derive an appropriate representation of tone centers based on the audio signal using constant-Q (CQ) profiles. The constant-Q profiles are 12-dimensional vectors where each component is an estimate of 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. The performances of the model are demonstrated over a Chopin's *Prélude* (op.28 no 20), with profiles trained on the 24 Chopin's *Préludes*.

Zhu & Kankanhalli [29] present 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 from the melody using a diatonic scale model, each one corresponding to a single mode. A modality (key style) analysis then determines the center mode of the melody of each segment. Segments of unrelated modes are eliminated. Key labels and boundaries are determined by grouping the remaining segments. The effectiveness of the method is qualitatively measured by analyzing the output of the model while listening to 50 English, Japanese and Chinese popular songs. The ground truth is unknown but it is claimed that key changes can be well perceived.

In order to study the evolution of the tonal center of a piece of music, Gómez and Bonada [30] present a tool to visualize the tonal content of polyphonic audio signals. The tonal content of an audio file of music is represented by the instantaneous evolution of the tonality and its strength. The tool enables the measuring of the effect of the length of the sliding window used for key tracking.

Harte *et al*. [31] propose 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.

In some other approaches, the segmentation stage (segmentation of the analyzed piece into segments that correspond to unique keys) is more elaborate. Temperley [32], [33] proposes a Bayesian key-finding model. The analyzed piece is divided into short segments. The model then searches for the most probable "key structure", where a key structure is a labeling of each segment with a key. Each segment can be expressed as a series of pitch-class sets. Given a segment, the fit between the corresponding key and the pitch-classes composing the segment is measured using "key-profiles" derived from the

Kostka and Payne corpus [34]. The model searches for the most probable key structure using dynamic programming, favoring minimum key changes between segments. It was evaluated for main key estimation (defined by the key of the first of segment of the structure) during the *Music Information Retrieval Evaluation eXchange* (MIREX) 2005 key detection contest [2] and received a weighted score of 91.4% correctly estimated keys on 1252 excerpts from classical pieces.

Chai & Vercoe [17] propose a 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 root 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. Ten classical piano music pieces are employed to test the performances of the proposed method yielding a label accuracy of 84%.

Izmirli [19] proposes a model for detecting modulations and labeling local keys using a non-negative matrix factorization (NMF) 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 containing at least one modulation, classical music annotated with a single key, and excerpts from the Kosta and Payne corpus [34]. Local key estimation achieves up to 82.4% on popular music and 72.5% on the Kosta and Payne corpus.

C. Key Estimation Methods Based on the Chord Progression

In Western tonal music, chords and keys are musical attributes that are closely related to each other. It is thus natural to rely on the chord progression for estimating the key progression.

Hierarchical frameworks based on rule-based approaches have been proposed. For instance, Shenoy *et al*. [13] present a rule-based approach for determining the key of acoustic musical signals from the chord progression. The succession of chords is estimated from beat-synchronous chroma features, on which symbolic inference is applied. Only major and minor chords are considered. The chords, detected across all the frames, are then collected into a single 24-dimensional vector. For each key, a reference 24-dimensional reference vector that corresponds to the theoretical distribution of major and minor chords within the considered key is constructed. For instance, the major and minor chords that can be constructed around the CM scale using the notes of this scale are respectively CM, FM, GM and Dm, Em, Am. The pattern that returns the highest rank is selected as the one being the key of the song. It is found that analysis over 16 bars (64 beats) of audio is sufficient to determine the key of the song. The results obtained on a set of 20 popular English songs spanning four decades of music lead to a key estimation accuracy of 90%. However, chord recognition accuracy is not sufficient to provide usable chord transcription.

An alternative approach is statistical frameworks. Raphael & Stoddard [14], [35] present an approach to functional harmonic analysis based on pitch and rhythm relying on symbolic data. A MIDI representation of a music composition is partitioned into sequences of one-measure length. The goal of this work is to associate a label composed of three variables to each sequence: the tonic (*e.g*. C, C#) and the mode (major or minor) that give the musical key, and the chord characterized by its harmonic function (scale degree, *e.g*. tonic, dominant). The functional analysis of the chord progression is supposed to guide the choice of the key when it is ambiguous. The analysis is performed with a HMM that allows the simultaneous estimation of chord and key. The success of the model is demonstrated over some examples but a quantitative evaluation is not presented.

In the framework of global key estimation, several HMM-based works that estimate the chords and keys have been proposed. Lee & Slaney [36], [37] propose key-dependent chord HMMs trained on synthesized audio for chord recognition and global key estimation. In these approaches, 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 the 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 [36], hidden states are treated as chords, which also allows identifying the chord sequence.

Some works that address simultaneously the problem of estimating chords and keys using HMMs have also been proposed. Burgoyne & Saul [38] present a HMM-based model that tracks key simultaneously with chords. It is claimed that transitions between chords are dependent on their tonal context. Contrary to [37], they do not assume that music remains in a single key from start to end. The model considers chord and key to be inseparable properties of any given harmony. The model is restricted to major and minor triads. Each state of the HMM represents a chord in a possible key (C major in the key of A minor for instance). Simplified rules of tonal harmony are encoded in the transition matrix. The traditional Gaussian emission distribution is replaced with a Dirichlet distribution. The model is trained in an unsupervised manner with the EM algorithm on five Mozart symphonies (K.134, K.162, K.181, K.182 and K.183) and tested on the Minuet of Mozart symphony K.550. The results reveal that a more advanced harmonic model is needed to improve the results.

Noland & Sandler [39] present a HMM technique for estimating the predominant key in a symbolic musical excerpt. 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 is extended to the audio case in [12] and evaluated on 110 Beatles songs,

[2]http://www.mirex.org

yielding a global key accuracy of 68%. Although this model has only been evaluated on the case of global key estimation, it could be used for local key estimation.

Catteau *et al*. [15] propose a probabilistic framework for simultaneously estimating keys and chords. Observation likelihood and chord/key transition models are derived from music theory of Lerdahl & Jackendoff [40]. Parameter tuning and system evaluation are performed using four databases: some cadences and modulations, a set of 10 polyphonic audio fragments of a duration of 60s and a set of 96 MIDI-to-wave synthesized fragments from the MIREX 2005 key detection contest. In this work, the segmentation of audio is based on a fixed frame-by-frame analysis.

Rocher *et al*. [16] describe a method for estimating simultaneously local keys and chords from audio signals using a template-based approach combined with music theory knowledge. Using chord/key pairs as vertices, they build a weighted acyclic harmonic graph. Chord and key progressions are obtained by finding the best path in the graph. A multi-scale analysis is proposed in order to take into account different kind of harmonic information. Evaluation on 174 Beatles songs shows that the chords and the local keys accuracy (respectively 74.9% and 62.4%) are increased considering the mutual dependency between the two musical attributes. Here again, the analysis window size for key estimation is empirically set to a fixed value that does not differ from one piece to another.

D. Summary of Related Works

It is not easy to get an idea of the performances of existing works on local key estimation because the evaluation material and protocol differs a lot from one work to another. In order to provide a clearer idea of existing approaches, especially in local key estimation, we have summarized part of the above-mentioned works on key estimation in Table I.

III. Proposed approach

In this paper we are interested in the problem of local key finding in polyphonic audio files. For this, 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 [25] 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], the chords are integrated into the key profiles, but not directly used to estimate the key. Moreover, their relationship to keys is not explicitly investigated. We study this relationship in the present work. To integrate the concept of key modulating over time, we propose to use a 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 [39].

As underlined in Section II-B, HMMs have already been used for local key estimation [12], [17]. However, this was done using a frame-by-frame analysis. A contribution of the present work is that we introduce information related to the metrical 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 in previous work, the length of the analysis window was found empirically. After computing chroma vectors on short overlapping frames, a frame-by-frame musical key analysis is performed in [17] or [18]. Rocher *et al*. [16] also use a fix-length window size (set to 30s) in their work. An alternative to sliding window key center tracking techniques is proposed in [19] 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 σ seconds. The value of the parameter σ 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 modulation points would become ambiguous. The drawback of using an empirically chosen window size is that key changes may be undetected by the algorithm for pieces with a fast tempo and that, for pieces with a slow tempo, chords may be estimated rather than keys. 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 metrical 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 is related to the structure of the piece and fits its harmonic content. It is thus musically meaningful and not arbitrary.

A. Model

In this section we present a model that allows the estimation of the key progression of a musical excerpt using both the underlying chord progression, which characterizes the harmonic structure, and the downbeat positions, which characterize the metrical 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 metrical structure is the *downbeats* or the first

TABLE I
SUMMARY OF RELATED WORKS ON KEY ESTIMATION.

Reference	Key	Approach	Evaluation Material	Key Label Accuracy
Gómez & Herrera [25]	Global	Correlation between HPCP and (K-K) profiles extended to polyphonic audio comparison with various machine-learning methods.	878 excerpts of classical music (661 for training and 217 for testing).	64%
Peeters [11]	Global	Highest likelihood of a chromagram given 24 HMM trained according to the 24 major and minor keys key.	302 European Baroque, classical and romantic music extracts.	81%
Izmirli [6]	Global	Correlation between spectral summary information and weighted KS-Temperley's key templates.	85 classical music pieces.	86%
Chuan & Chew [10]	Local	CEG algorithm and the Spiral Array Model with various key determination policies (NN, RD, AD)	410 classical music pieces ranging from Baroque to Contemporary, Chopin's 24 *Préludes*.	70%
Purwins *et al.* [18]	Local	Frame-by-frame correlation between input and trained CQ-profiles using a fuzzy distance.	training on the 24 Chopin's *Préludes*, evaluation on Chopin's *Préludes* op.28 no 20.	/
Zhu & Kankanhalli [29]	Local	Melody modality analysis by segmentation of a MIDI melody using a diatonic scale model.	50 popular songs.	qualitative evaluation
Temperley [33]	Local	Bayesian approach that favors minimum key changes between segments of MIDI data.	Main key evaluation on 1252 excerpts from classical pieces.	91.4%
Chai & Vercoe [17]	Local	Key detection divided into two steps: root and mode using two HMMs and a chromagram as input feature.	10 classical piano pieces.	84%
Izmirli [19]	Local	Uses a NMF method for segmentation and correlation between groups of contiguous chroma vectors and templates for labeling	Local key evaluation on 17 pop songs and 17 excerpts from the *Kosta and Payne corpus*.	82.4%
Burgoyne & Saul [38]	Local	Chords and keys simultaneously estimated from audio using a HMM based on Dirichlet distribution and PCP as input.	5 Mozart symphonies (K. 134, K. 162, K. 181, K.182 and K.183) for training and Mozart Symphony K. 550, Minuet for testing.	/
Noland & Sandler [12]	Local/ Global	HMM that uses pairs of consecutive chords as input features and that encodes human expectation of harmonic relationships.	Main key evaluated on 110 Beatles songs.	91%
Catteau *et al.* [15]	Local	Simultaneous estimation of keys and chords using a probabilistic framework in which chord/key transition models are derived from music theory of Lerdahl's.	10 polyphonic audio fragments of 60s and 96 MIDI-to-wave synthesized fragments.	51.2%
Rocher *et al.* [16]	Local	Simultaneous estimation of keys and chords using chroma as input features and a template-based approach combined with music theory knowledge encoded in a weighted acyclic harmonic graph.	174 Beatles songs.	62.4%

beats of each measure. Here, the chords and the downbeats are estimated simultaneously using a "double-state" HMM where a state s_n (n denotes the time index) is a combination of a chord type and a position of the chord in the measure according to the beat positions. We consider here a chord lexicon composed of the $I = 24$ Major and minor triads (C Major, ..., B Major, C minor, ..., B minor). We consider here pieces predominantly in $3/4$ or predominantly in $4/4$ meters. In both cases, the transition matrix will allow $K = 4$ beat positions in the measure.

The 24-key space is modeled by an ergodic 24-state HMM, where each state s_n^{key} represents one of the 24 major and minor keys. The emission probability of each state (each key) is a 24-dimensional vector representing the probability to observe each of the 24 chords in this specific key. They are obtained either directly from the chord progression or using all chord probabilities. Given the observations, we estimate the most likely key sequence over time in a maximum likelihood sense. The flowchart of the model is represented in Fig. 2 and detailed below.

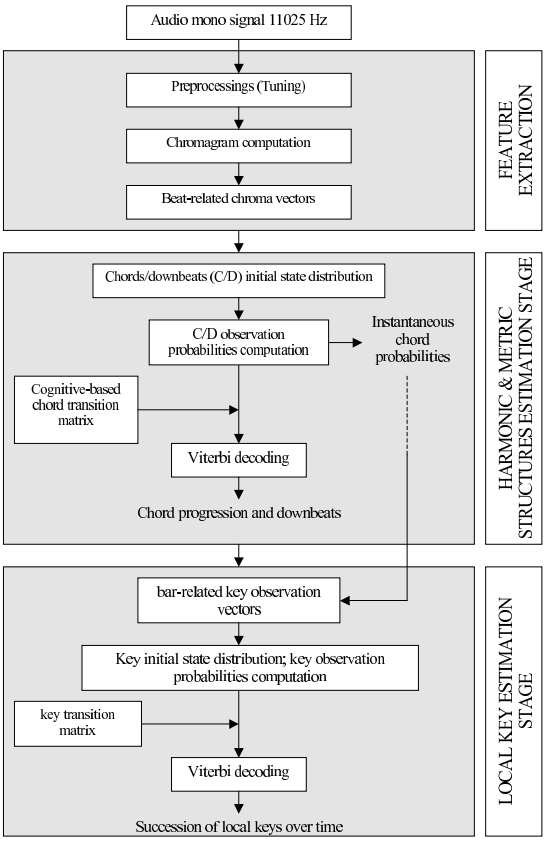

Fig. 2. Flowchart of the local key estimation model.

B. Feature Vectors

As most chord and key detection models, the front-end of our model is based on the extraction of a chromagram that represents the audio signal. For chromagram computation, we use the method we proposed in [41], briefly described here.

The audio signal is first down-sampled to 11025 Hz, converted to mono by mixing both channels and converted to the frequency domain using a constant-Q transform (CQT) [42], which is a time/frequency transform in which the frequency domain channels are geometrically spaced so that the frequency-resolution ratio remains constant. The tuning of the track is estimated to take into account possible deviation with the standard reference pitch $A4 = 440$ Hz. We use the method proposed in [11]. This step is important because the chroma vectors are built by mapping the energy peaks in the CQT to the 12 semitones pitch classes. The center

frequencies of the CQT are geometrically spaced according to the frequencies of the equal-tempered scale that we set according to the estimated tuning. The CQT is then mapped to the 12-dimensional chroma vectors.

To integrate the metrical structure of the piece, we built meter-related features by averaging the chroma vectors according to the beat positions so that we obtain one feature vector per beat. This assumes integration of a beat-tracker as a front-end of the system. In our experiments, we use the beat tracker proposed by Peeters in [43] for the *Quaero test-set*. In the case of the *Piano Mozart test-set*, we use beat positions that have been annotated by hand because the test-set is composed of classical music pieces containing lots of deviations in tempo that result from the expressivity in classical music. The performance of the beat tracker was too poor in this case to be used as an input of our model.

In order to provide robustness against variations of dynamics, the chromagram is normalized so that the components of each chroma vector sums to unity. The feature extraction stage is represented in Fig. 3.

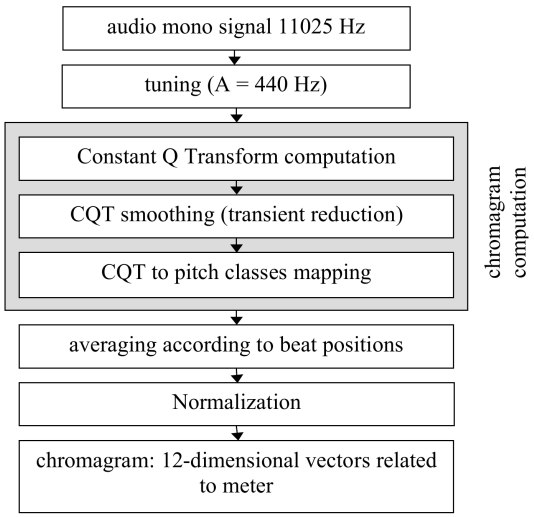

Fig. 3. Chroma features extraction.

C. Harmonic and Metrical Structures

The harmonic structure is defined by the chord progression and the metrical structure is defined by the downbeat positions. These two musical attributes are estimated simultaneously according to the method we proposed in [41], which we briefly summarize here. We propose a specific topology of HMMs that allows us to model chord dependency on metrical structure. This model can handle pieces with complex metrical structures such as beat addition, beat deletion or changes in the meter.

We consider an ergodic $I * K$-state HMM that describes the simultaneous evolution of three processes. Here n denotes the time index. The observable variable is the chroma vector, denoted O_n in the following. The unobservable processes and the corresponding hidden variables are the chord symbols, denoted c_n and the "beat position in the measure", denoted b_n. Each double-state $s_n = [i, k]$ is defined as an occurrence of a chord $c_n = i$, $i \in [1 : I]$ at a "beat position in the measure" $b_n = k$, $k \in [1 : K]$. Here, we consider a chord lexicon composed of $I = 24$ major and minor triads and $K = 4$. Given the observations, we estimate the most likely chord sequence over time and the downbeat positions in a maximum likelihood sense.

1) The Initial State Distribution: is uniform since we have no reason to prefer a state over another.

2) The State-Conditional Observation Likelihoods $P(O_n|s_n)$: are computed as:

$$P(O_n|s_n) = P(O_n|c_n)P(O_n|b_n) \quad (1)$$

where $P(O_n|c_n)$ corresponds to the chord symbol observation probabilities and $P(O_n|b_n)$ corresponds to the "beat position 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 in Subsection III-D. The second term in Eq. (1), $P(O_n|b_n)$, is a constant multiplication that has no effect on the observation probability for state s_n. The state-conditional observation likelihoods $P(O_n|s_n)$ depend actually only on the chord type.

3) The State Transition Matrix T: models the musical rules from which the transitions between chords result. These rules are based on the harmonic and the metrical structures. In our HMM, T takes into account both the chord transitions and their respective positions in the measure. To integrate harmonic rules, we derive the $I * K$-state transition matrix T from a I-state chord type transition matrix T_c based on music-theoretical knowledge about key-relationships. This matrix is the same as the key transition matrix described below.

We integrate metrical rules in T relying on the assumption that chords are more likely to change at the beginning of a measure than at other positions [44]. In order to take into account several cases of metrical structure, two different transition matrices are built. The first one corresponds to the case of songs in 4/4 meter with ternary passages. In this case, we favor 4-beat measures but transitions to 3-beat measures are allowed. The second transition matrix corresponds to the case of songs in 3/4 meter with passages in 4/4. In this case, we favor 3-beat measures but transitions to 4-beat measures are allowed.

To favor chord changes on downbeats, we attribute in the state transition matrix T a lower self-transition probability[3] for chords occurring on the last beat of a measure (*i.e.* on $b_n = K$) than on other beat positions. This is illustrated in Fig. 4.

This model allows us to consider pieces with complex metrical structures including changes in the meter from 3/4

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

to 4/4 time-signature but also various exceptional situations such as the insertions of a measure in 1/4 in a 4/4 meter passage. Our model also allows us to handle errors in the beat tracking stage such as beat insertion or beat deletion due in general to tempo deviation (*e.g.* music tempo speed up or slow down) not detected by the beat tracker.

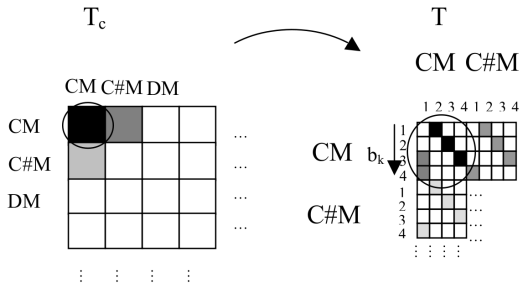

Fig. 4. Chord transition matrix for a single-state HMM [left], transition matrix in the case of a double-state HMM taking into account the position of the chord in the measure [right]. In this figure, the darker the color, the higher the value in the transition matrix.

4) Chord Progression and Downbeats Detection: The optimal succession of states $s_n = [i, k]$ over time is found using the Viterbi decoding algorithm [45] which provides the most likely path through the HMM states $S = (s_1, s_2, \ldots, s_N)$ given the sequence of observations $O = (O_1, O_2, \ldots, O_N)$:

$$\hat{S} = \underset{S}{\operatorname{argmax}}\ p(S, O). \tag{2}$$

where $p(S, O)$denotes the joint probability density of the hidden and observed variables. We estimate simultaneously the most likely chord sequence over time and the corresponding "beat position in the measure", hence the downbeats.

D. Chordgram

We define the *chordgram* as the succession over time of the 24-dimensional vectors representing the probability to observe each of the 24 chords at each beat-synchronous 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}.

E. Extraction of Key Observation Vectors

The key observation vectors are derived from the chords. In the evaluation part, we will compare two methods for local key estimation. They differ from each other in the way the 24-dimensional key observation vectors O^{key} are derived from the chords.

1) *Method 1*: The key observation vectors are built from the *chordgram* using the instantaneous chord probabilities $P(O_n|c_n)$, where O_n corresponds to the chroma vector and $c_n = i, i \in [1:24]$ corresponds to the chords.

$$O_n^{key}(m) = P(O_n|c_n = m) \tag{3}$$

2) *Method 2*: The key observation vectors are built directly from the estimated chord progression $\hat{c} = (\hat{c_1}, \hat{c_2}, \ldots, \hat{c_n})$.

$$O_n^{key}(m) = \begin{cases} 1 & \text{if} \quad \hat{c_n} = m \\ 0 & \text{if} \quad \text{otherwise} \end{cases} \tag{4}$$

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 which are delimited by the downbeats. The local key is thus estimated on segments that are related to the structure and thus fit the intrinsic music content of the piece. Because key changes occur in general on the first beat of a measure it is important that the analysis starts on a downbeat. Musical phrases often have duration of 8 or 4+4 bars. In Section IV-C, we present key estimation results segmenting pieces into 2-bar segments with 1-bar overlap, or into consecutive 1-bar segments. In our experiments we have tested the algorithm using other analysis window lengths and found that the local key accuracy results decrease with longer windows. This is discussed below. The key observation vectors are 24-dimensional vectors obtained by averaging the *chordgram* or the estimated chord progression along the segments. These 24-dimensional vectors represent the probability to observe each of the 24 keys at a given time instant.

F. Key Estimation From Chords and Downbeats 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.

1) 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 over another.

2) The Key Observation Probabilities $P(O_n^{key}|s_n^{key})$: 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 predefined key templates are 24-dimensional vectors with each bin corresponding to one of the 24 major and minor triads. We have tested our model using four key templates as described below.

The first three of them are derived from the knowledge that the most important triads in a given key are those built on the tonic, the subdominant and the dominant [1], [25]. For instance, for a CM key, these chords correspond to CM (C-E-G), FM (F-A-C) and GM (G-B-D).

1) In the first pre-defined key template, we attribute a value of 1 to each of the three main triads. It will be referred to as "main chords" (MC) key template in the following.
2) The second key template is similar to the first one, except that we attribute a higher value $k > 1$ to the chord built on the tonic of the key. In our experiments, we used $k = 3$. It will be referred to as "weighted main chords" (WMC) key template in the following.
3) The third key template is similar to the second one, except that we attribute a value of one to the chord relative to the one built on the tonic (for instance Am chord in a C major key). We consider this case because we know from music theory that this chord has an important function given a key. This key template will be referred to as "weighted main chords relative" (WMCR) in the following.

These three key templates corresponding to the C major (top) and C minor (bottom) keys are represented in Fig. 5.

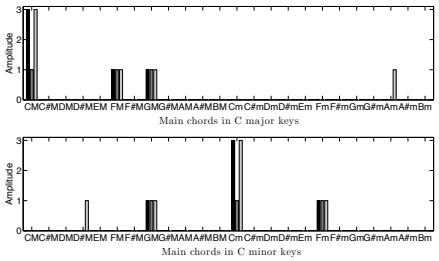

Fig. 5. Pre-defined 24-dimensional key templates based on the three main triads. Dark grey: "main chords" (MC), black: "weighted main chords" (WMC), light grey: "weighted main chords relative" (WMCR).

The 4^{th} pre-defined key-template is built relying on a cognitive experiment conducted by Krumhansl [1] that gives values corresponding to the rating of chords in harmonic-hierarchy experiments. In this experiment, the perceived relative structural significance of chords in tonal context is measured. For this, several trials consisting of a strong key-defining context followed by a single chord are presented to listeners. The listeners are asked to rate how well the final chord fit with the preceding key-defining context. In the experiments, three types of chords are considered: major, minor and diminished. However, since we consider only major and minor chords in our model, the diminished chords were ignored. The cognitive-based key templates corresponding to the C major (top) and C minor (bottom) keys are represented in Fig. 6.

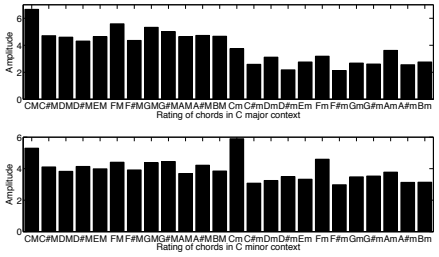

Fig. 6. Pre-defined 24-dimensional cognitive-based key templates [1].

The templates corresponding to the various major and minor keys are obtained by circular permutation from the one corresponding to the C major and C minor keys.

Let $T_i, i \in [1:24]$ denote a 24-dimensional key template. The observation key probabilities $P(O_n^{key}|s_n^{key})$ are obtained according to Eq. (5):

$$For \quad i = 1 \dots 24, P(O_n^{key}|s_n^{key} = i) = \frac{O_n^{key} \cdot T_i}{\|O_{key}\| \cdot \|T_i\|} \quad (5)$$

They are normalized so that $\sum_{i=1}^{24} P(O_n^{key}|s_n^{key} = i) = 1$.

G. State Transition Probability Distribution

Key modulations in a music piece follow musical rules that can be reflected in the state transition matrix. To integrate this information in key transition, we adopt the key transition matrix proposed in [39] already used as a chord transition matrix[4]. 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 [39] 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.

H. Local Key Estimation

The optimal succession of states over time is found using the Viterbi decoding algorithm that gives us the best sequence of keys over time $S^{key} = (s_1^{key}, s_2^{key}, \dots, s_{N^{key}}^{key})$ given the sequence of observations $O^{key} = (O_1^{key}, O_2^{key}, \dots, O_{N^{key}}^{key})$.

$$\hat{S^{key}} = \underset{S^{key}}{\operatorname{argmax}}\ p(S^{key}, O^{key}). \quad (6)$$

The music piece is thus segmented into segments that are labeled by a key.

IV. Evaluation

The aim of this section is to analyze the performances of the proposed approach for local key estimation on two different test-sets of different music styles. The impact of the various parameters will be studied, as well as the relationship between the various musical elements that are taken into account (key, chords, downbeats, musical style).

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

A. Test-sets

The first test-set consists in 5 movements of Mozart piano sonatas listed in Table II corresponding to about 30 minutes of audio music. In what follows, it will be referred to as the *Piano Mozart test-set*. Trained musicians from the Musichorshule of Karlsruhe (Germany) have manually annotated the chord and key progressions ground truth. First, a list of the chords and key with their duration in beats has been provided. Beat positions have then been annotated using the software *Wavesurfer*. Finally, the list has been automatically mapped to the annotated beat positions, resulting in the ground truth we use. The pieces were annotated in part by ear but also relying on the scores when ambiguities were found.

TABLE II
THE *Piano Mozart test-set*.

Reference of the piano sonata	movement
KV 283	1
KV 283	2
KV 309	1
KV 310	1
KV 311	2

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 keys, 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 chord 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 properly establish a tonal center. Segments corresponding to transitions from one key to another have been labeled as transition parts.

The second corpus was provided during the QUAERO project[5] 2009 evaluation campaign. It consists in 16 real audio songs annotated by IRCAM. All the songs were annotated by listening to the audio using the IRCAM QIMA annotation software. Note that one additional difficulty in the case of popular music annotation is that, in general, there is no score. Effects blur musical objects and sometimes, inexistent notes may be perceived while existing notes may be undetected. The songs are listed in Table III and correspond to various artists and styles. In what follows, it will be referred to as the *Quaero 2009 test-set*. For the sake of simplicity, we will refer to this test-set as "popular" music, although it covers various styles of music that include pop, rock, electro and salsa.

The characteristics of the test-sets in terms of number of chords, chord changes, keys and key changes are shown in Fig. 7 and Fig. 8. The total number of chord changes and key segments is respectively 1688 and 90 for the *Piano Mozart test-set* and 1871 and 92 for the *Quaero 2009 test-set*. Note that some pieces are annotated in a single key but present some key changes because they present regions with ambiguous keys (denoted by "T").

[5]http://www.quaero.org

TABLE III
THE *Quaero 2009 test-set*.

Artist	Album	Song
Pink Floyd	Dark Side of the Moon	Breathe
Pink Floyd	Dark Side of the Moon	Brain Damage
Buenavista Social Club	Buenavista Social Club	Chan Chan
Buenavista Social Club	Buenavista Social Club	De camino a la Vereda
Dusty Springfield	Dusty in Memphis	Son of a Preacher Man
Aerosmith	Get A Grip	Cryin
Shack	HMS Fable	Pull Together
UB40	Labour of Love II	Kingston Town
Fall out boy	Infinity on High	This Ain't a Scene it's An Arms Race
Abba	Waterloo	Waterloo
Cher	Believe	Believe
Phil Collins	Single	Another Day in Paradise
Santa Esmeralda	Don't Let Me Be Misunderstood	Don't Let Me Be Misunderstood
Sweet	Desolation Boulevard	Fox on the Run
FR David	Single	Words
Enya	Watermark	Orinoco Flow

Concerning the meter, all pieces in the *Piano Mozart test-set* have a constant $3/4$ or $4/4$ meter while 4 pieces in the *Quaero 2009 test-set* have a variable meter.

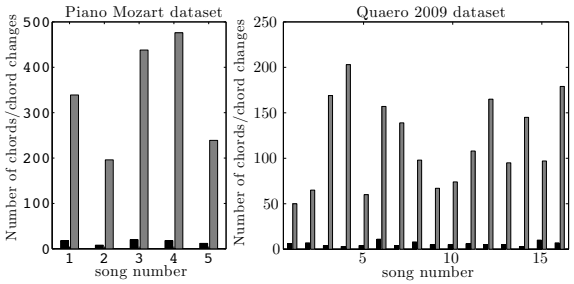

Fig. 7. characteristics of the datasets : number of different keys (black) and number of key changes (grey) per song.

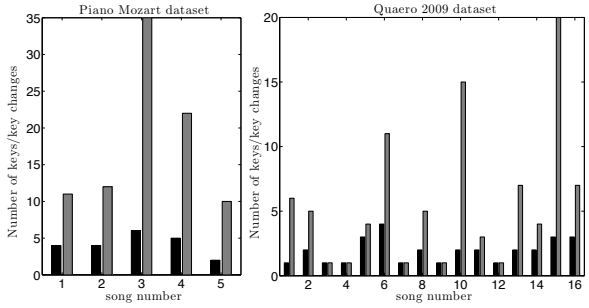

Fig. 8. characteristics of the datasets : number of different keys (black) and number of key changes (grey) per song.

B. Evaluation Measures

1) Chord/key Label Accuracy: For chords and key evaluation, we consider *label accuracy LA*, which measures how the estimated chord/key is consistent with the ground truth. *LA* results correspond to the mean and standard deviation of correctly identified chords/keys per song. Parts of the pieces where no key can be labeled (for instance when a chromatic scale is played) have been ignored in the evaluation. "Non-existing chords", noted "N" in the annotation denote noise, silent parts or non-harmonic sounds. They are unconditionally counted as errors in the evaluation.

Two scores are considered: *Exact Estimation EE* corresponds to the rate of keys/chords correctly detected; *Exact + Neighbor E+N* corresponds to the rate of correctly detected keys/chords including neighboring keys/chords. Neighboring key/chord considered here are harmonically close triads: parallel Major/ minor (EM being confused with Em), relative (Am being confused with CM), dominant (CM being confused with GM) or subdominant (CM being confused with FM).

For local key label accuracy, in addition to the *EE* and *E+N* scores, we consider the *MIREX Estimation* score *ME* , that gives the estimation rate according to the score proposed for the MIREX 2007 key estimation task[6].

2) Local key Segmentation Accuracy: We also consider, as in [17], another aspect of local key estimation: the *key segmentation accuracy SA* indicates how the detected modulation points are consistent with the actual positions. It is expressed with the Precision, Recall and F-measure. $Precision(P)$ 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 $F = 2RP/(R + P)$. Key changes are not abrupt. Two established keys are often separated by a transition part (labeled as T in the ground truth) where no key is firmly established. The transition parts are taken into account in segmentation accuracy by the use of a window w. 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. A high value of w favors high precision and recall. We present below results with w corresponding to 1 or 2 bars.

3) Statistical Significance Testing: During our experiments, we use *paired samples t-test* (or *dependent samples t-test*) at the 5% significance level to measure whether the difference in the results from one method to another are statistically significant or not.

C. 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 analysis window length, key estimation from the *chordgram* or from the estimated chord progression, influence of the tolerance window.

Label and segmentation accuracy results are presented in Tables IV to VII. For each test-set, we present the results obtained with the parameters that were found to perform the best : 2-bar window length / WMC key templates for the *Piano Mozart test-set*, and 1-bar window length / Krumhansl key templates for the *Quaero 2009 test-set*. We discuss below the choice of these parameters (see Sections IV-C5 and IV-C6).

[6]The score is obtained using the following weights: 1 for correct key estimation, 0.5 for perfect fifth relationship between estimated and ground-truth key, 0.3 if detection of relative major/minor key, 0.2 if detection of parallel major/minor key. Fore more details, see http://www.mirex.org.

1) Local Key and Musical Styles: The local key estimation results are different for the two test-sets. The key estimation results are much higher in the case of classical music (label accuracy 80.21%, segmentation accuracy 0.5170) than in the case of popular music (label accuracy 61.31%, segmentation accuracy 0.3410). Also, for popular music, the standard deviation is high. However, most of the errors correspond to neighboring keys, as indicated by the mirex (*ME*, 73.18%) and exact + neighbors (*E+N*, 89.84%) scores.

The WMC key templates and transition matrix we propose (based on cognitive experiments or music knowledge) better reflect the tonal content and the relationships between keys of the *Piano Mozart test-set* than the *Quaero 2009 test-set*. Key changes in the *Piano Mozart test-set* are well-modeled in the key transition matrix, whereas it is not the case for every piece in the *Quaero 2009 test-set*. For instance the song *Words* from FR David mainly consists in key transitions from C# Major to D# minor. However, this transition is only ranked as the 8^{th} most likely one in the transition matrix. Moreover, as discussed in Section IV-C6, the newly proposed key templates do not always reflect accurately the tonal content of pieces in the *Quaero 2009 test-set* .

The *Quaero 2009 test-set* contains 5 pieces with constant key whereas our algorithm favors segmentation into several keys. Removing these pieces from the evaluation (*NC* results) increases the segmentation F-measure to 0.5055. The standard deviation of the label accuracy results is also much lower. As discussed in Section IV-C5, it is difficult to make a compromise between favoring key changes and favoring constant key.

2) Comparison Between Chordgram and Chord Progression: We have proposed two methods for local key estimation:

1) In the first case (*method 1*), the probability of each chord at a given time instant is used to estimate the key.
2) In the second case (*method 2*), the chords are first estimated using the 2-state HMM described in Section III-C. The local key is then derived from the estimated chord progression.

Tables IV and VI show that *method 1* outperforms *method 2*. Indeed, the best key label accuracy results are obtained with *method 1*, both for classical and popular music.

Table VII shows that for popular music, segmentation accuracy is also significantly higher with *method 1*. Moreover, even if *method 2* slightly outperforms *method 1* on the *Piano Mozart test-set* (see Table V), a paired sample t-test shows that the difference is not statistically significant and tests on a larger database are needed. One drawback of *method 2* is that errors in the estimation of the chord progression are propagated to the key estimation step. It thus seems better to rely on the *chordgram* rather than directly on the chord progression for estimating the local key.

It can be noticed that recall segmentation accuracy is higher with *method 2* than with *method 1*. Our explanation is that chord changes are favored on downbeats in the case of *method 2* whereas changes of harmony in the *chordgram* of *method 1* are smoother. As a result, the key changes are

TABLE IV
CHORDS AND LOCAL KEYS LABEL ACCURACY (*LA*) RESULTS ON THE *Piano Mozart test-set*, USING A 2-BAR LENGTH WINDOW AND THE WMC KEY TEMPLATE. *EE*: EXACT ESTIMATION RATE. *E+N*: ESTIMATION RATE INCLUDING NEIGHBOR KEYS/CHORDS. *ME*: MIREX ESTIMATION RATE. *method 1*: BASED ON THE CHORDGRAM. *method 2*: BASED ON THE CHORD PROGRESSION.

Piano Mozart test-set		
key LA method 1 (%)	*EE*	$\mathbf{80.21 \pm 13.56}$
	ME	$\mathbf{84.81 \pm 11.86}$
	E+N	$\mathbf{93.36 \pm 10.08}$
key LA method 2 (%)	*EE*	74.11 ± 18.92
	ME	80.08 ± 18.64
	E+N	91.19 ± 10.55
chord LA (%)	*EE*	61.43 ± 5.50
	E+N	74.11 ± 18.92

TABLE V
LOCAL KEY SEGMENTATION ACCURACY (*SA*) RESULTS USING A 2-BAR LENGTH WINDOW AND THE WMC TEMPLATES. *method 1*: BASED ON THE CHORDGRAM. *method 2*: BASED ON THE CHORD PROGRESSION. TWO TOLERANCE WINDOWS: $w = 1$ BAR AND $w = 2$ BARS.

Piano Mozart test-set			
		method 1	*method 2*
SA precision	$w = 1$	0.5723	0.4489
	$w = 2$	0.8196	0.6805
SA recall	$w = 1$	0.4730	0.7131
	$w = 2$	0.6874	0.8691
SA F-measure	$w = 1$	**0.5170**	0.5451
	$w = 2$	0.7327	0.7514

TABLE VI
CHORDS AND LOCAL KEYS LABEL ACCURACY (*LA*) RESULTS ON THE *Quaero 2009 test-set*, USING A 1-BARS LENGTH WINDOW AND THE KRUMHANSL KEY TEMPLATE. *EE*: EXACT ESTIMATION RATE. *E+N*: ESTIMATION RATE INCLUDING NEIGHBOR KEYS/CHORDS. *ME*: MIREX ESTIMATION RATE. *method 1*: BASED ON THE CHORDGRAM. *method 2*: BASED ON THE CHORD PROGRESSION. *WC*/*NC*: WITH/NO PIECES WITH CONSTANT KEY.

Quaero 2009 test-set			
method		*WC*	*NC*
key LA method 1 (%)	*EE*	$\mathbf{61.31 \pm 36.50}$	67.61 ± 26.43
	ME	$\mathbf{73.18 \pm 27.56}$	78.60 ± 16.67
	E+N	$\mathbf{89.84 \pm 24.70}$	93.74 ± 10.11
key LA method 2 (%)	*EE*	52.14 ± 21.17	54.02 ± 14.20
	ME	62.84 ± 20.71	65.80 ± 13.71
	E+N	80.80 ± 22.64	82.65 ± 13.06
chord LA (%)	*EE*	72.67 ± 16.84	75.11 ± 11.82
	E+N	93.78 ± 5.56	94.78 ± 4.90

TABLE VII
LOCAL KEY SEGMENTATION ACCURACY (*SA*) RESULTS USING A 1-BAR LENGTH WINDOW AND THE KRUMHANSL TEMPLATES. *method 1*: BASED ON THE CHORDGRAM. *method 2*: BASED ON THE CHORD PROGRESSION. TWO TOLERANCE WINDOWS: $w = 1$ BAR AND $w = 2$ BARS. *WC*/*NC*: WITH/NO PIECES WITH CONSTANT KEY.

Quaero 2009 test-set					
		method 1		*method 2*	
		WC	*NC*	*WC*	*NC*
SA precision	$w = 1$	0.2955	0.4394	0.1620	0.2420
	$w = 2$	0.3883	0.5713	0.2682	0.3919
SA recall	$w = 1$	0.5166	0.7765	0.5506	0.8167
	$w = 2$	0.6148	0.9123	0.6875	1.0000
SA F-measure	$w = 1$	**0.3410**	**0.5055**	0.2377	0.3532
	$w = 2$	0.4432	0.6504	0.3632	0.5269

blurred.

3) Relationship Between Chords and Key: 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 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 of confusions with harmonically close chords (such as dominant or subdominant chords), the estimated key was nevertheless either correct or a neighboring key.

4) Importance of the Metrical Structure: Musical elements are highly organized and, when listening to a piece of music, we can feel in general a structure and separate the piece into several segments, as for instance verse/chorus in a popular music song. These segments are in general related to the metrical structure (measures or groups of measures) but also to the key because pitches within a section are organized around a tonal center that is characteristic of the section. It thus seems useful to rely on the metrical structure in order to estimate key progression. As explained above, we propose here to use bar-related key analysis segments. In a fully-automatic analysis, the beats and downbeats are directly estimated from the audio. Beat and downbeat tracking results evaluated on the *Quaero test-set* are presented in Table VIII. Beat tracking is not perfect. However, some errors in the beat tracking do not affect downbeat estimation [41] and downbeat tracking results are fair enough to be useful for key estimation.

TABLE VIII
BEAT AND DOWNBEAT POSITION ESTIMATION RESULTS. PRECISION (PREC), RECALL (REC), F-MEASURE (F-M).

	Precision	Recall	F-measure
Beat	0.53 ± 0.37	0.72 ± 0.32	0.57 ± 0.35
Downbeat	0.86 ± 0.34	0.79 ± 0.40	0.79 ± 0.40

To investigate the hypothesis of the importance of the metrical structure on the local key estimation, we run our model without the metrical information, i.e. we perform an analysis with a constant key analysis window. We present results with of a duration of $30s$ and $10s$, with $0.8s$ overlap[7]. Results are presented in Table IX.

It can be seen that, for both test-sets, key estimation is better when taking into account the metrical structure. The difference in the results between the NM case and WM case is statistically significant for segmentation accuracy for both methods and for label accuracy in the case of *method 2*. Example in Fig. 9 shows an excerpt of the first movement of the Mozart piano sonata KV 283. The size of the window is an essential point in key estimation. If it is too long, key

[7]The window must be long enough to get the sense of the key. Note that other durations have been tested without showing any significative differences

TABLE IX
COMPARISON OF LOCAL KEY RESULTS USING ADAPTED WINDOW LENGTH (WITH METER, WM) OR USING A FIX ANALYSIS WINDOW LENGTH (NO METER, NM) OF 10 OR 30S. THE TOLERANCE WINDOW IS $w = 1$ BAR.

		Piano Mozart test-set			*Quaero 2009 test-set*		
		WM	NM30s	NM10s	WM	NM30s	NM10s
m1	*LA* (%)	80.21	72.63	69.73	61.31	57.32	58.71
	SA F-m	0.52	0.06	0.19	0.34	0.03	0.15
m2	*LA* (%)	74.11	66.25	70.88	52.14	40.25	38.20
	SA F-m	0.55	0.16	0.26	0.24	0.07	0.12

changes may be overlooked by the algorithm. For instance, in the *NM-30s* case, key changes between 2:50 and 3:10 are not detected. When a smaller fixed window is used in the *NM-10s* case, segmentation is better. However, if the analysis window is too small, the algorithm will analyze the chordal structure of the piece instead of the key structure. For instance, in the *NM-10s* case, a G major segment is estimated around 2:40 instead of remaining in D major (see the grey oval in Fig. 9). This is probably due to the presence within 2 bars of a G major chord built on the IV^{th} degree of D major key. Key segmentation is also better when positioning the starting point of the key analysis window on downbeats, since key changes very often occur on downbeats. It helps avoiding mixing some passages with different local keys. This is underlined in Fig. 9 by the dashed rectangles.

5) Effect of the Length of the Analysis Window: We have evaluated the algorithm with different window lengths: 1, 2, 4, 8 and 16 bars. Results are provided in Fig. 10.

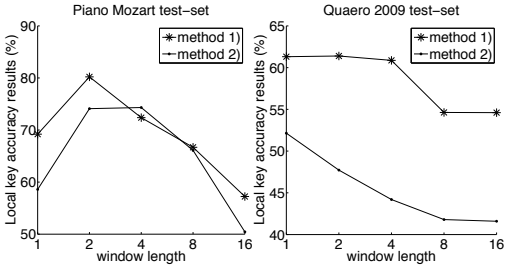

Fig. 10. Key estimation results in case of methods 1 and 2 according to the length of the key analysis window.

In classical music, the length of musical phrases is very often 4 or 8 bars. This is particularly true for Mozart's piano sonatas. Usually, the musical key remains constant within a phrase (whereas the harmony changes several times). This is why we chose to estimate the local key on segments of length related to musical phrases. A 1-bar analysis window length is too short because it captures the harmony (the chords) rather than the local key. The best results were obtained using a 2-bar length analysis window. This may be due to the fact that, 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.

For pieces belonging to the *Quaero 2009 test-set*, the best results were obtained using a 1-bar analysis window length. However, for *method 1*, the difference with a 2-bar and a 3-bar analysis window length is not statistically significant. This is because the structure of the local key progression in pieces belonging to the *Quaero 2009 test-set* is in general quite more complex than in the *Piano Mozart test-set*. Popular music may use many different musical systems inside a song (local tonality, apparent tonality, no tonality etc.). Moreover, the number of key changes within a song varies a lot between one song to another. For instance, the song *Words* from FR David has 19 key changes and the song *Pull Together* from Shack has a constant key, whereas the two songs have the same time duration. Note that, as for the *Piano Mozart test-set*, a too long analysis window results in low key estimation scores because passages with different local keys are very likely to be mixed.

6) Effect of the Choice of the Key Templates: We evaluated the algorithm with 4 templates, as illustrated in Table X.

TABLE X
LOCAL KEY ACCURACY (*EE* IN %)ON USING A 2-BAR LENGTH WINDOW COMPARING VARIOUS KEY TEMPLATES, FOR *method 1 m1* AND *method 2 m2*.

		WMC	MC	WMCR	Krumhansl
Mozart	*m1*	80.2 ± 13.6	79.1 ± 9.0	80.1 ± 11.8	50.0 ± 21.1
	m2	74.1 ± 18.9	71.3 ± 18.4	73.5 ± 19.1	75.7 ± 19.9
Quaero	*m1*	52.2 ± 28.9	52.7 ± 33.8	51.0 ± 30.9	61.3 ± 36.5
	m2	37.7 ± 12.0	41.0 ± 27.8	38.8 ± 13.9	52.1 ± 21.2

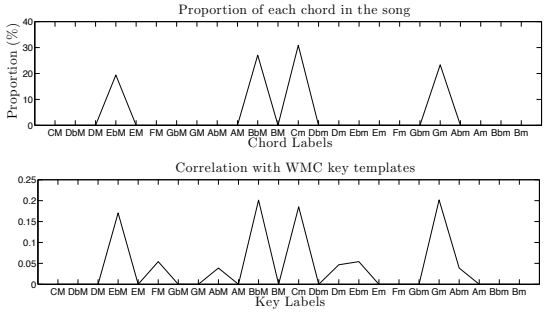

Fig. 11. Up: Proportion of chords in the song *Pull Together*. Bottom: Correlation with the WMC key templates.

For the *Piano Mozart test-set*, the best results are obtained with the weighted main chords WMC templates for *method 1*. In the case of *method 2*, the cognitive-based templates slightly outperform the WMC templates. However, statistical tests indicate that the difference in the results is not statistically significant.

For the *Quaero 2009 test-set*, the results are significantly better with the cognitive-based key templates than with the newly proposed key templates. The WMC key templates are built on the assumption that the proportion of chords built on the tonic of the local key is the highest. This is not always the case in pieces from the *Quaero 2009 test-set*. For instance, the song *Pull Together* from Shack is in constant C minor key. Its chord progression consists in a loop Cm - Fm - BbM - EbM. In the upper part of Fig. 11, we show a 24-dimensional vector that corresponds the mean duration of each chord in the piece. In the lower part of Fig. 11, we show

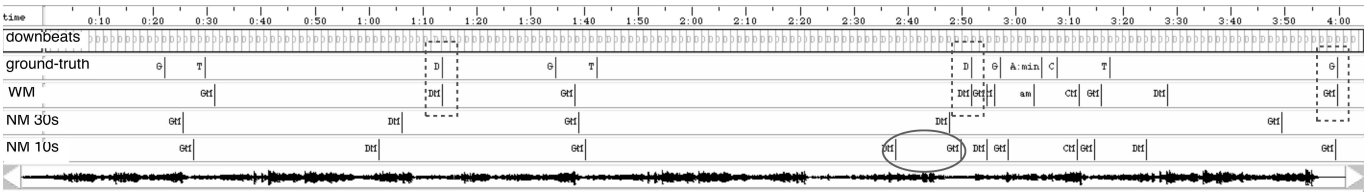

Fig. 9. Estimated key progression of an excerpt of the Mozart piano sonata KV 283. From top to bottom : downbeat positions, key ground truth, estimated key considering metrical structure, results with a 30s fixed window and with a10s fixed window. The image has been obtained using the Open Source tool *Wavesurfer*

the results of the correlation between this vector and each of the 24 chord templates. It can be seen that the template corresponding to Cm key is not the one that best reflects the tonal content of the piece: BbM is much closer.

7) Smooth Modulations: The key segmentation accuracy results are presented in Tables V and VII, in which we consider 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 2-bar tolerance window. This can be explained by the fact that key change is a very smooth process that often takes several bars. Key changes are in general annotated in the ground truth only when a musical element clearly indicates that the key is established (for instance when the tonic of the key is played). Between two key changes, there may be several bars without a precisely established key. It would be interesting to formulate and add a "local key transition" state in the model. This is a direction for our future works.

8) Comparison with a Direct Template-Based Method: In order to have a better idea of the performances of the proposed method for local key estimation, we follow [16] and compare our results to a *direct template-based method* (DTBM). This can be viewed as applying the Krumhansl-Schmuckler key-finding algorithm to successive overlapping frames. This method is used by Sapp in [46] for displaying a visual content of the musical key structure of a composition in a single picture. The signal is divided into overlapping frames of 30s and 0.8s overlap and we compute one chroma feature per frame[8]. For each frame, we compute the correlation with the chroma feature with the 24 Krumhansl's key templates. The estimated key is selected as the one that gives the highest value. Results are presented in Table XI.

For both test-sets, our model performs significantly better than the DTBM method. Both label accuracy and segmentation accuracy are higher. This is illustrated in Fig. 12 which shows the four first minutes of the first movement of the Mozart piano sonata KV 309. All the key segments are correctly detected within a precision window of one measure. Key labels are correct in general, except that short segments

[8]In order to make a fair comparison with our model, we compute the chromagram using the same parameters than for our method (beat-synchronous chroma features) and the average the chroma vectors so that we obtain one feature per 30s-length frame. A length of 30s was found to be a good compromise for local key estimation in [16].

TABLE XI
COMPARISON OF LOCAL KEY RESULTS OBTAINED WITH THE PROPOSED METHOD (*method 1*) WITH RESULTS OBTAINED USING A DTBM METHOD. THE TOLERANCE WINDOW IS $w = 1$ BAR.

		Piano Mozart test-set	*Quaero 2009 test-set*
meth 1	*LA* (%)	80.21 ± 13.56	61.31 ± 36.50
	SA F-measure	0.52	0.34
DTBM	*LA* (%)	62.41 ± 21.52	42.10 ± 31.30
	SA F-measure	0.24	0.08

of related keys are sometimes inserted, as F major in the first C major segment. This is due to the presence of long-duration chords built on the IV^{th} or V^{th}. With the DTBM method, key segmentation precision is much lower (see the dashed rectangles). Moreover, the algorithm gets very confused when there are many successive modulations (see segment from 2:50 to 3:50 in the grey oval). The use of a fixed window that may not fit the structure of the piece results in mixing passages with different keys.

9) Analysis of Errors: For both test-sets, as indicated by the $E + N$ scores in Tables IV and VI , most of the errors correspond to confusions with neighboring keys (perfect fifth relationship between estimated and ground-truth key, relative major/minor key, parallel major/minor key). Details are shown in Fig. 13.

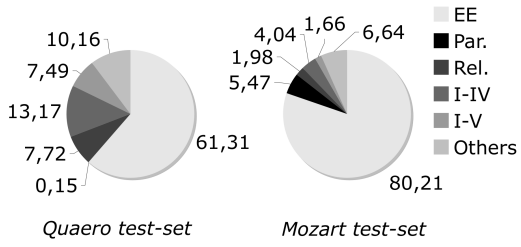

Fig. 13. Repartition of key estimation errors.

The errors may be due to various causes, as discussed above (key template not fitting the harmonic content of the song, length of the analysis window, etc. Note that there is a notable predominance of sub-dominant errors in the results. This may be due to the high value given to transitions between subdominant chords in the cognitive-based transition matrix. If the model does not recognize exactly a key but makes confusion with a neighboring key, the result can still be useful for higher-level structural analysis such as segmentation.

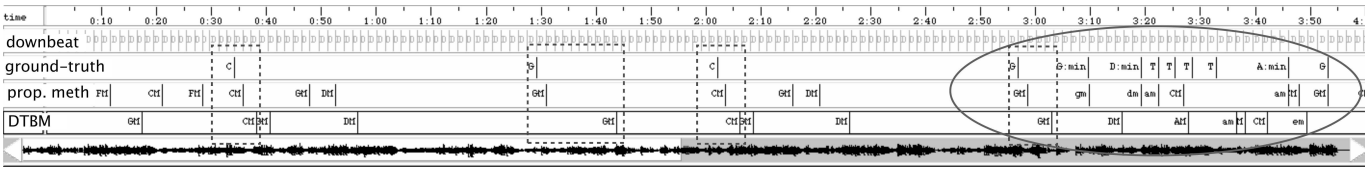

Fig. 12. Estimated key progression of an excerpt of the Mozart piano sonata KV 309. From top to bottom : downbeat positions, key ground truth, estimated key with the proposed method and the DTBM method. The image has been obtained using the Open Source tool *Wavesurfer*

V. Conclusion and Future Works

In this paper, we have presented a model for the estimation of key progression of Western tonal polyphonic excerpts, which segments an audio file into sections labeled with local keys. The method combines and extends several previous methods proposed for global key estimation. Our approach to local key estimation is based on a HMM that uses predefined key templates. In Western tonal music, the chord progression is directly related to the key and, conversely, the sense of key arises from the chords. We thus rely on the chord progression to model the key progression. Previous works use an empirically-chosen key analysis window size. A too small window size would focus the chromagram on individual chords more than on keys whereas a too large window size would result in mixing passages with different keys and some modulations may be overlooked. We rely here on the metrical structure and bar-related key analysis windows. Key estimation is not performed on empirically chosen segments but on segments that are expressed in relationship with the tempo period and thus that are adapted to the intrinsic music content of the piece. Our method takes advantage of the structure of music.

We have evaluated our model on two test-sets of various styles (classical and popular music) that contain pieces with complex tonal structures. Our results, illustrated by case-study examples, show that the key progression can be estimated relying on the harmonic structure and that its robustness is increased when taking into account the metrical structure.

We have studied the impact of the various parameters (key template, analysis window length, distribution of observation key probabilities). As it could be expected, we have found that their choice should depend on the music style. Most of the errors correspond to confusions with neighboring keys. Confusions with keys in perfect-fifth relationship are particularly common errors. Introducing chord functional analysis information would help reducing these kinds of errors.

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. However, the prior results we obtain with our current model may already be useful to some music-content applications such as music mood detection for instance.

Our test-set is quite small, however, we believe it is large enough to draw interesting observations on local key estimation. It is reliable since it has been annotated by trained musicians and it contains more than 180 key segments to be estimated. Moreover, the results we obtain and the comparison between the various methods are in general statistically significant. We plan to explore other music styles in order to see if our hypothesis can be generalized to a wider range of music styles. Popular music offers a wide range of investigations because it may use different systems, such as local tonality, apparent tonality, no tonality, etc. within a single song. Moreover, many other modes than major and harmonic minor are commonly used, such as aeolian and mixolydian modes, which opens additional challenges to the problem of key estimation.

The notion of key relationships is very complex. Our review of previous works on the subject of local key estimation shows that there is currently no unified evaluation methodology. Various evaluation materials and evaluation measures for key estimation have been used and it is not easy to get an idea of the different methods. This partly comes from a lack of pieces annotated in local keys. We plan to concentrate on this point in the future. Annotation is a very important step. Besides providing evaluation material for the designed algorithms, the process of key annotation should help getting a better understanding of the notion of key as a perceptual attribute, and a better understanding of the ability of human beings of getting a sense of how pitches are organized around a tonal center.

VI. Acknowledgment

The authors would like to thank M-C. Papadopoulos and A. Vay for their help in annotating the piano Mozart pieces, and G. Tzanetakis and the reviewers for their very helpful comments.

References

[1] C. Krumhansl, *Cognitive foundations of musical pitch*. New York, NY, USA: Oxford University Press, 1990.

[2] D. Temperley, *The cognition of basic musical structures*. Cambridge, MA, USA: MIT Press, 2001.

[3] G. Widmer and S. Madsen, "Key finding with interval profiles," in *Proceedings of the International Computer Music Conference (ICMC)*, Copenhagen, Denmark, August 27-31 2007.

[4] A. Mardirossian and E. Chew, "SKeFiS – a symbolic (MIDI) key-finding system," in *MIREX*, London, UK, September 11-15 2005.

[5] E. Gómez, "Tonal description of polyphonic audio for music content processing," *INFORMS Journal on Computing*, vol. 18, no. 3, pp. 294–304, 2006.

[6] O. Izmirli, "Template based key finding from audio," in *Proceedings of the International Computer Music Conference (ICMC)*, Barcelona, Spain, September 5-9 2005, pp. 211–214.

[7] S. Pauws, "Musical key extraction from audio," in *Proceedings of the International Conference on Music Information Retrieval (ISMIR)*, Barcelona, Spain, October 10-14 2004, pp. 96–99.

[8] Y. Zhu and M. Kankanhalli, "Precise pitch profile feature extraction from musical audio for key detection," *IEEE Transactions on Multimedia*, vol. 8, no. 3, pp. 575–584, 2006.

[9] S. van de Par, M. McKinney, and A. Redert, "Musical key extraction from audio using profile training," in *Proceedings of the International Conference on Music Information Retrieval (ISMIR)*, Victoria, BC, Canada, October 8-12 2006, pp. 328–329.

[10] C.-H. Chuan and E. Chew, "Audio key finding: considerations in system design and case studies on Chopin's 24 preludes," *EURASIP Journal on Advances in Signal Processing*, vol. 2007, pp. 156–156, January 2007.

[11] G. Peeters, "Musical key estimation of audio signal based on HMM modeling of chroma vectors," in *Proceedings of the International Conference on Digital Audio Effects*, Montreal, QC, Canada, September 18-20 2006, pp. 127–131.

[12] K. Noland and M. Sandler, "Signal processing parameters for tonality estimation," in *Proceedings of the Convention Audio Engineering Society (AES)*, Vienna, Austria, May 5-8 2007.

[13] A. Shenoy, R. Mohapatra, and Y. Wang, "Key determination of acoustic musical signals," in *Proceedings of the IEEE International Conference on Multimedia & Expo (ICME)*, vol. 3, Taipei, Taiwan, June 27-30 2004, pp. 1771–1774.

[14] C. Raphael and J. Stoddard, "Harmonic analysis with probabilistic graphical models," in *Proceedings of the International Conference on Music Information Retrieval (ISMIR)*, Baltimore, MD, USA, October 26-30 2003, pp. 177–181.

[15] B. Catteau, J. Martens, and M. Leman, "A probabilistic framework for audio-based tonal key and chord recognition," in *Advances in data analysis - Proceedings of the 30th Annual Conference of the Gesellschaft für Klassifikation*, R. Decker and H.-J. Lenz, Eds. Berlin, Germany: Springer, March 8-10 2007, pp. 637–644.

[16] T. Rocher, M. Robine, P. Hanna, and L. Oudre, "Concurrent Estimation of Chords and Keys From Audio," in *Proceedings of the International Conference on Music Information Retrieval (ISMIR)*, Utrecht, Netherlands, August 9-13 2010.

[17] W. Chai and B. Vercoe, "Detection of key change in classical piano music," in *Proceedings of the International Conference on Music Information Retrieval (ISMIR)*, London, UK, September 11-15 2005.

[18] H. Purwins, B. Blankertz, and K. Obermayer, "A new method for tracking modulations in tonal music in audio data format," in *Proceedings of the International Joint Conference on Neural Networks (IJCNN)*, Como, Italy, July 24-27 2000.

[19] O. Izmirli, "Localized key finding from audio using non-negative matrix factorization for segmentation," in *Proceedings of the International Conference on Music Information Retrieval (ISMIR)*, Vienna, Austria, September 23-27 2007.

[20] H. Papadopoulos, "Joint Estimation of musical Content Information From an Audio Signal," Ph.D. dissertation, Paris 6 University, Paris, France, 2010.

[21] E. Chew, "Towards a mathematical model of tonality," Ph.D. dissertation, Department of Operations Research, Massachussets Institute of Technology, Cambridge, MA, USA, 2000.

[22] E. Gómez, "Tonal description of music audio signals," Ph.D. dissertation, Universitat Pompeu Fabra, Barcelona, Spain, 2006.

[23] G. Wakefield, "Mathematical representation of joint time-chroma distribution," in *Proceedings of the SPIE Conference on Advanced Signal Processing Algorithms, Architecture and Implementation*, Denver, CO, USA, July 19-21 1999, pp. 637–645.

[24] T. Fujishima, "Real-time chord recognition of musical sound: a system using common lisp music," in *Proceedings of the International Computer Music Conference (ICMC)*, Beijing, China, October 22-28 1999, pp. 464–467.

[25] E. Gómez and P. Herrera, "Estimating the tonality of polyphonic audio files: cognitive versus machine learning modelling strategies," in *Proceedings of the International Conference on Music Information Retrieval (ISMIR)*, Barcelona, Spain, October 10-14 2004, pp. 92–95.

[26] E. Chew, "The spiral array: an algorithm for determining key boundaries," in *Proceedings of the International Conference on Music and Artificial Intelligence (ICMAI)*, Edinburgh, Scotland, September 12-14 2002, pp. 18–31.

[27] G. Gatzsche, M. Mehnert, D. Gatzsche, and K. Brandenburg, "A symmetry based approach for musical tonality analysis," in *Proceedings of the International Conference on Music Information Retrieval (ISMIR)*, Vienna, Austria, September 23-27 2007.

[28] C. Krumhansl and E. Kessler, "Tracing the dynamic changes in perceived tonal organization in a spatial representation of musical keys," *Psychological Review*, no. 89, pp. 334–368, 1982.

[29] Y. Zhu and M. Kankanhalli, "Key-based melody segmentation for popular songs," in *Proceedings of the International Conference on Pattern Recognition (ICPR)*, Cambridge, UK, August 23-26 2004.

[30] E. Gómez and J. Bonada, "Tonality visualization of polyphonic audio," in *Proceedings of the International Computer Music Conference (ICMC)*, Barcelona, Spain, September 5-9 2005.

[31] C. Harte, M. Sandler, and M. Gasser, "Detecting harmonic change in musical audio," in *Proceedings of the Audio and Music Computing for Multimedia Workshop (AMCMM)*, Santa Barbara, CA, USA, October 27 2006.

[32] D. Temperley, *Bayesian approach to key-finding*, ser. Lecture Notes in Computer Science. Springer Berlin / Heidelberg, 2002, no. 2445, ch. Music and Artificial Intelligence, pp. 149–155.

[33] ——, "A Bayesian key-finding model," in *MIREX*, London, UK, September 11-15 2005.

[34] S. Kostka and D. Payne, *Tonal harmony*. New York, NY, USA: McGraw Hill, 1995.

[35] C. Raphael and J. Stoddard, "Functional harmonic analysis using probabilistic models," *Computer Music Journal*, vol. 28, no. 3, pp. 45–52, 2004.

[36] K. Lee and M. Slaney, "A unified system for chord transcription and key extraction using hidden Markov models," in *Proceedings of the International Conference on Music Information Retrieval (ISMIR)*, Vienna, Austria, September 23-27 2007.

[37] ——, "Acoustic chord transcription and key extraction from audio using key-dependent HMMs trained on synthesized audio," *IEEE Transactions on Audio, Speech and Language Processing*, vol. 16, no. 2, pp. 291–301, Feb. 2008.

[38] J. Burgoyne and L. Saul, "Learning harmonic relationships in digital audio with Dirichlet-based hidden Markov models," in *Proceedings of the International Conference on Music Information Retrieval (ISMIR)*, London, UK, September 11-15 2005.

[39] K. Noland and M. Sandler, "Key estimation using a hidden Markov model," in *Proceedings of the International Conference on Music Information Retrieval (ISMIR)*, Victoria, BC, Canada, October 8-12 2006, pp. 121–126.

[40] F. Lerdahl and R. Jackendoff, *A Generative Theory of Tonal Music*. MIT press, 1983.

[41] H. Papadopoulos and G. Peeters, "Joint estimation of chords and downbeats," *IEEE Transactions on Audio, Speech and Language Processing*, vol. 19, no. 1, pp. 138–152, January 2011.

[42] J. Brown, "Calculation of a constant Q spectral transform," *Journal of the Acoustical Society of America*, vol. 89, no. 1, pp. 425–434, 1991.

[43] G. Peeters, "Beat-marker location using a probabilistic framework and linear discriminant analysis," in *Proceedings of the International Conference on Digital Audio Effects*, Como, Italy, September 1-4 2009.

[44] M. Goto, "An audio-based real-time beat tracking system for music with or without drum sounds," *Journal of New Music Research*, vol. 30, no. 2, pp. 159–171, 2001.

[45] B. Gold and N. Morgan, *Speech and audio signal processing: processing and perception of speech and music*. John Wiley & Sons Inc., 1999.

[46] C. Sapp, "Visual Hierarchical Key Analysis," *Computers in Entertainment*, vol. 3, no. 4, pp. 1–19, October 2005.