2007-01-01

Key Signature Estimation

Matt Cranitch
Technological University Dublin, matt.cranitch@cit.ie

Derry Fitzgerald
Cork Institute of Technology

Matthew Hart
Cork Institute of Technology

Key Signature Estimation

Matthew Hart, Derry Fitzgerald, Matt Cranitch

Dept. of Electronic Engineering,

Cork Institute of Technology

IRELAND

Matthew.Hart@cit.ie

Abstract— The problem of automatic key signature detection has been the focus of much research in recent years. Previous methods of key estimation have focused on chromagrams and key profiling techniques. This paper presents a remarkably simple but effective method of estimating key signature from musical recordings. The algorithm introduces the 'keyogram', a concept resembling the chromagram, and is aimed for use on traditional Irish music. The keyogram is a measure of the likelihood of each possible major key signature based on a masked scoring system.

***Keywords* – Automatic Key Estimation, Music Information Retrieval, Audio.**

1 Introduction

The topic of automatic key signature detection has been the subject of much recent research [1,2,3,4]. While the subject of key signature recognition is important in its own right, it is only one part of a larger problem, that of automatically generating scores from musical audio signals [5,6]. The method presented herein is intended as a single module of a larger system, designed for the automatic generation of score from audio signals. Key signature estimation is a necessary step in generating a realistic and manageable score that can be read with ease by a musician.

Previous methods of key estimation have used techniques such as Spiral Arrays [3] and Tree based structures [4] as seen in the MIREX competition. More recent methods focus on extracting chromagrams from audio, and comparing them to key profiles [1]. The method presented in this paper more closely resembles the chromagram methods. A chromagram is a restructured spectral representation of frequencies that are mapped onto a limited set of 12 chroma values, and so represents the tonality of a piece of music based on its harmonic content. Frequencies are assigned to bins representing ideal chroma values of the equally tempered scale, where each bin corresponds to the twelve chromas in one octave. The chromagram spectrum *S(f)* is built from the combination of spectral content of the foreground pitch and musical background level *n(f)*,

$$S(f)=\sum_{n=1}^{N} h^{n-1}\, W(nf)\, A(n)(\eta(f)+\delta(np_i-f))$$

where the spectral pitch content is modelled as a scaled impulse train reflecting the interpretation of pitch as a harmonic series; it contains high energy bursts at integral multiples np_i, for harmonic index n. Further, N denotes the number of harmonics, $h(\leq 1)$ denotes the factor controlling peak contribution to the pitch percept, $W(-)$ is an arc-tangent function representing the transfer function of the auditory sensitivity filter, and $A(n)$ is the gain for harmonic n.

To utilise the chromagrams with respect to key signature estimation, many chromagrams are generated from a large set of audio recordings, of which the key is known. Averaging the results for each known key gives a unique template, known as a key profile, for that key signature. The key profiles are then compared against the chromagrams from test audio, and the closest correlation to a profile is the estimated key.

This method is further improved upon by van de Par *et Al.* [2] whereby chromagrams are calculated from three sections of the recordings, start, middle, and end, with each having a specific weight. As the key can change during the course of a piece of music, the accuracy of the method can be improved by weighting the start and end sections higher than the middle section. The start and end sections of most recordings are in the main key signature, with the middle section modulating between other key signatures. The key profiles generated from three chromagrams give better results than those generated from a single chromagram. A single chromagram cannot account for the modulation that occurs midway through the majority of classical piano recordings.

While these methods have proved effective they require the training of the templates on a database of music. Further, they go through an intermediate stage, that of calculating the chromagram prior to key estimation. It should be possible to tackle the problem of key estimation in a more direct manner without the use of training data. One such possible technique is presented in the next section.

2 METHOD

As noted previously, current key estimation methods take an indirect approach to the problem of key signature estimation, and require the use of training data. We propose a method that gives a frame by frame local estimate of the likelihood of each major key, which is then summed across time to obtain a global estimate of the most likely key signature. We term the matrix of these local frame-by-frame key estimates a 'keyogram', in an analogous manner to the use of chromagram. The keyogram is calculated in the following manner:

1. A Short-time Fourier Transform (STFT) is performed on the audio,

   $$X(k,m) = \left| [\sum_{n=0}^{N-1} w(n)x(n+mH)e^{-j2\pi nk/N}] \right|$$

   where *X(k,m)* is the absolute value of the complex STFT, where *m* is the time frame index, H is the hop size between frames, *k* is the frequency bin index, *N* is the FFT window size and where *w(n)* is a suitable window of length *N* also.

2. All harmonics above a threshold, (half the amplitude of the loudest harmonic for that frame), are denoted by a value of 1. This gives the presence or otherwise of a given harmonic, independent of amplitude,

   $$B_{k,n} = \begin{matrix} 1 & Y_{k,n} \geq th \\ 0 & Y_{k,n} < th \end{matrix}$$

   where $Y_{k,n}$ is the harmonic amplitude for index *k* of frame *n*, and *th* is the threshold amplitude for frame *n*. Figure 1 shows a section of a piece of music with the notes above threshold highlighted.

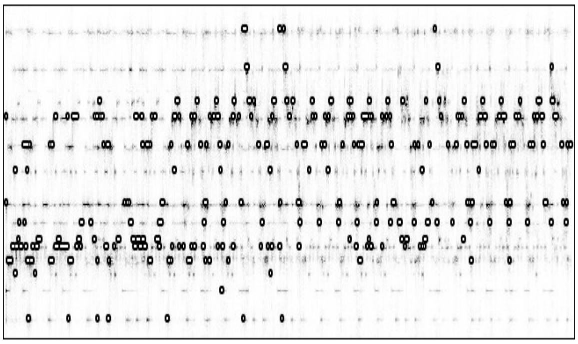

Figure 1: A section of audio highlighting the notes found to be above the threshold.

3. Each frequency bin in the binary matrix is mapped to a note value according to the following formula,

   $$P(k) = [12\log_2(k/NFFT \cdot f_s/f_1)] mod\ 12$$

   where NFFT is the FFT length, f_s is the sampling rate, and f_l is the reference frequency, in this case, A 440Hz

4. For each key signature *i*, the note values of *P(k)* that are members of M_i are summed, yielding $K_{i,n}$ where *K* denotes the keyogram, and *n* denotes the frame number. The set M_i contains the notes associated with key signature *i.*

   $$K_{i,n} = \sum_{P(k) \in M_i} B_{k,n}$$

5. Step 4 is repeated for each key signature *i.* The resulting matrix is a local keyogram for that frame. Figure 2 shows the local keyograms for the spectrum shown in Figure 1.

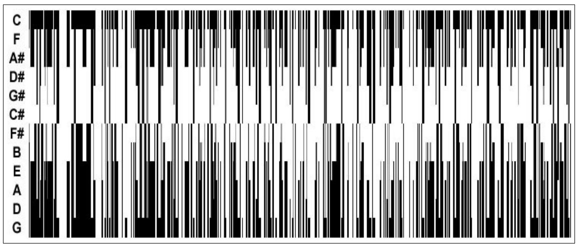

Figure 2: The local keyograms related to Figure 1, plotted over time.

6. The summation of the keyograms across all STFT frames gives a global keyogram representing the likelyhood of all possible major key signatures. The key signature within the keyogram holding the highest value is chosen as the estimated key signature. Figure 3 shows the final global keyogram and Figure 4 shows the contents of the keyogram as it changes over time.

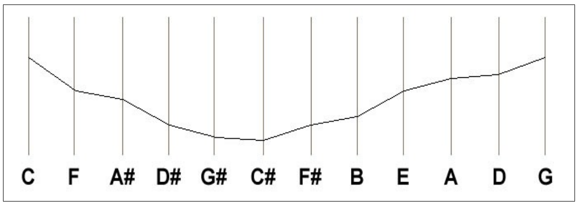

Figure 3: The global keyogram associated with the spectrum shown in Figure 1.

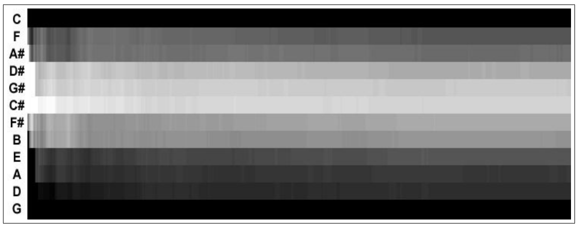

Figure 4: The global keyogram associated with Figure 1 as it changes over time.

This method has the advantage of performing the key signature estimation directly, without the need to first train the system before it can be utilised. No prior data is needed, enabling the method to be music genre independant, unlike the key profiling systems mentioned above. The key profiling systems are trained on large sets of classical piano recordings, and are most effective when presented with such music. The keyogram performs equally well on most genres of music even though it has been aimed at traditional Irish music.

Figures 1 through 4 show the algorithm in action on a section of Track 5 – The Lucky Penny (Table 1). In Figure 1 the most audible notes have been isolated. In many cases more than a single note per frame has been identified providing the algorithm with a more comprehensive set of data to work with. Chords provide the algorithm with more information in a shorter time span allowing it to identify the key signature quickly.

In Figure 2 we can see each local keyogram generated by each note detection frame. The local keyograms vary significantly over time. After the notes are mapped to the global keyogram we can compile Figure 3, showing the cumulative scores of each key signature at the final frame of Figure 1. The keyogram is arranged in the circle of fifths for clarity and shows how similar the complimentary key signatures appear to be, such as C and G.

Figure 4 shows the contents of the global keyogram as it changes over time. The high scoring signatures are tending towards black, and the improbable remain light. It can be seen clearly in the graph that the global keyogram stablises quickly in contrast to the local keyograms. Less obvious is that this global keyogram correctly identified the key signature of the audio. Even though both C and G appear to be black, G has finished with the highest score.

3 EVALUATION

The evaluation of the algorithm consisted of the automatic key signature estimation of 10 traditional Irish music recordings. These recordings were taken from the CD Take A Bow [7]. The correct keys and notation for these pieces are available in The Irish Fiddle Book [8].

Of the 10 recordings, 1 track was estimated as its relative dominant (Track 2 – The Knockaboul Polka) and 1 track was estimated as its perfect fourth (Track 7 – The Broken Pledge) giving an 80% success rate. Using a similar scoring system to the MIREX Key Finding competition we mark the key signatures identified as their relative dominant as half the marks of getting the key fully correct. Using this scoring method, with the other 8 correctly identified recordings, the algorithm gives an 85% accuracy when used on traditional Irish music. It should be noted that many traditional Irish recordings make use of gapped scales and mixed tonality. Track 7 is a prime example of mixed tonality, while the track is written in the key of G major, it contains accidentals of F natural leading the algorithm to identify the key as C major. For the purpose of this algorithm we can assume these key signatures are compatible, as the two key signatures have a difference of only a single note (F, F#) and can be said to be closely related. Giving the perfect fourth a half point also gives a 90% accuracy.

Title	Track	Actual Key	Estimated Key
The Providence Reel	1	D	D
The Knockaboul Polka	2	G	D
The Cúil Aodha Slide	3	D	D
Táimse Im Chodladh	4	D	D
The Lucky Penny	5	G	G
The Stack of Wheat	6	G	G
The Broken Pledge	7	G	C
Madam Bonaparte	8	A	A
Lord Gordon	9	D	D
Tripping Up The Stairs	10	D	D

Table 1: Shows the results of the algorithm applied to 10 traditional Irish recordings.

The algorithm has also been tried on some rock, jazz, and blues recordings, on which it performed remarkably well, best identifying jazz. It is suspected that the loud drums and distorted guitar in rock and blues create excess background noise, reducing the quantity of correctly identified harmonics. The use of drum separation techniques such as Fitzgerald & Barry [9] could greatly aid this issue by removing the percussive signals from the audio prior to processing.

4 Conclusion

The keyogram method extracts prominent harmonics from a raw audio signal, representing the frequencies of musical notes. From the discovered notes a local keyogram is generated, whereby each key in the keyogram gains the sum of all notes contained within its key signature. The local keyograms are combined to form a global keyogram containing the likelihood of each major key.

The algorithm is extremely fast, and can track the key signature in real-time as the recording is playing. Currently however, it does not make any attempt to discern major from minor. Where a recording is in a minor key, it will return the relative major key signature associated with that minor key. This is an area for further research and investigation. A combination of key profiling and keyograms may improve identification and discrimination between major and minor keys and is to be the target of future work.

5 References

[1] Pauws, Steffan; Musical Key Extraction From Audio, *ISMIR 2004 Fifth Int. Conf. on Music Inf. Retr. Proc.*, 2004.

[2] de Par, Steven van; Musical Key Extraction From Audio Using Profile Training, *ISMIR 2006: 328-329*

[3] Mardirossian, Arpi; Chew, Elaine; SKEFIS – A Symbolic Key Finding System, *MIREX Symbolic Key Finding Competition,* 2005.

[4] Rizo, David; Tree Symbolic Music Representation for Key Finding, *MIREX Symbolic Key Finding Competition,* 2005.

[5] Klapuri, Anssi; Automatic Transcription of Music, *SMAC,* 2003.

[6] Bello, Juan Pablo; Techniques for Automatic Music Transcription, *ISMIR,* 2000.

[7] Cranitch, Matt; Take A Bow, *CD – The living Tradition OSSIAN Recordings OSS CD 5,* 1992.

[8] Cranitch, Matt; The Irish Fiddle Book, *ISBN 1-900428-90-3,* 1988 - 2001.

[9] Barry, Dan; FitzGerald, Derry; Drum Source Separation using Percussive Feature Detection and Spectral Modulation, *Proc. IEE Irish Signals and Systems Conference*, 2005.