# AUTOMATIC CHORD RECOGNITION USING QUANTISED CHROMA AND HARMONIC CHANGE SEGMENTATION

**Christopher Harte and Mark Sandler**
Centre for Digital Music
Queen Mary, University of London
Mile End Road, London, UK
`christopher.harte@elec.qmul.ac.uk`

## ABSTRACT

This extended abstract describes an entry to the MIREX'09 (Music Information Retrieval eXchange) chord detection competition. The system described here uses a combination of two algorithms previously presented by the authors.

The system first calculates a quantised chromagram from the audio recording. It then uses a harmonic change detection function to segment the chroma features in time. The average chroma values for each segment are found and the results are then analysed by a simple chord recognition algorithm.

**Keywords:** Chord recognition, harmonic change

## 1. SYSTEM OVERVIEW

The chord recognition system comprises several steps. In the first step, the audio data is downsampled to 5512.5Hz. A short time fourier transform (STFT) with window length 4096 samples and frame overlap of 512 samples (1/8th of a frame) is then performed producing a series of linear frequency spectra with a time resolution of approximately 11 frames per second. These linear frequency spectra are then converted to log frequency using the constant-Q (CQ) transform technique described by Brown and Puckette [1]. The CQ transform is calculated over four octaves between frequencies $f_{\text{min}} = 110$Hz (note A2 on the piano keyboard) and $f_{\text{max}} = 1760$Hz (note A6) with 36 bins per octave. The $k^th$ bin centre frequency is therefore:

$$f_k = (2^{1/36})^k f_{\text{min}} \tag{1}$$

The bins of the CQ log frequency spectra are summed across octaves to produce a 36-element vector called a Harmonic Pitch Class Profile (HPCP). For a CQ spectrum $C$ with $M$ bins, the value of the $b^{th}$ bin of the HPCP $H$ is given by equation 2.

$$H_b = \sum_{m=0}^{M} |C_{(b+36m)}| \qquad 1 \leq b \leq 36 \tag{2}$$

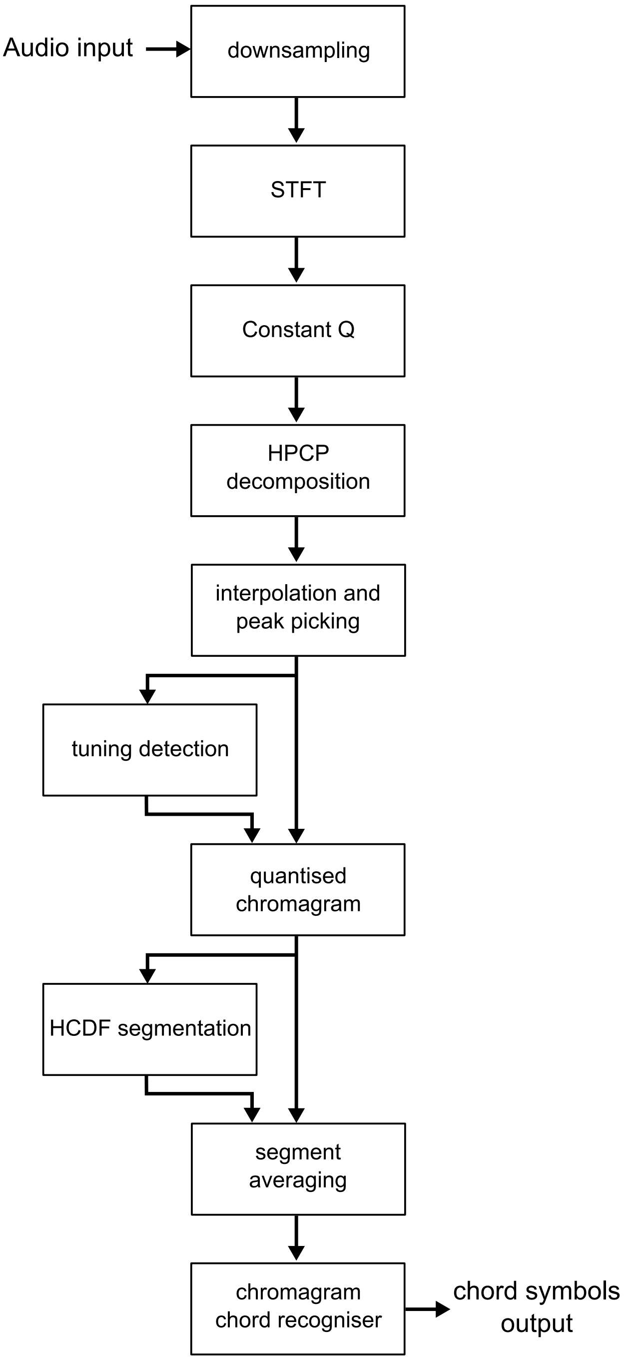

**Figure 1**. Flow diagram of chord recognition system

To deal with audio inputs where the tuning frequency deviates from A = 440Hz a tuning algorithm is used to identify the correct position of semitone centres in the HPCP. In order to accomplish this, each HPCP frame is processed using a simple peak picking algorithm and the position and magnitude of each peak is calulated using quadratic interpolation. To find the centre tuning frequency, we calculate the distribution of HPCP peaks across the width of a semitone (i.e. the modulus 3 value for each peak position).

By treating the peaks as vectors with a magnitude and angle (i.e. their position is the angle within one semitone with the centre of the semitone corresponding to A=440Hz), it is possible to sum them to arrive at a single tuning vector value for each frame. We find the tuning centre value for the whole piece by taking the mean value of all the frame tuning vectors.

With the tuning value known, we can now convert the peaks from the 36-bin HPCP into a 12-bin quantised chromagram. To reduce the effects of transients and other detuned signal components, we discard any HPCP peaks that lie outside a range of ±0.2 semitones relative to the tuning centre value.

Once the quantised chromagram has been calculated we calculate a harmonic change detection function (HCDF) in order to find possible chord boundaries. To calculate the HCDF, we first find the tonal centroid for each frame. A tonal centroid is a point in the six dimensional polytope that results from assuming enharmonic and octave equivalence in the Tonnetz map of tonal relations as described in [3]. For a 12-bin chroma vector $\mathbf{c}_n$ the six dimensional tonal centroid $\zeta_n$ for time frame $n$ is found by multiplying the chroma vector and a transformation matrix $\Phi$. To prevent numerical instability and ensure that the tonal centroid always lies within the 6-D polytope we divide the result by the $L_1$ norm of $\mathbf{c}$:

$$\zeta_n(d) = \frac{1}{||\mathbf{c}_n||_1} \sum_{l=0}^{11} \Phi(d,l)\mathbf{c}_n(l) \qquad \begin{array}{l} 0 \leq d \leq 5 \\ 0 \leq l \leq 11 \end{array} \tag{3}$$

where $l$ is the chroma vector pitch class index and $d$ denotes which of the six dimensions of $\zeta_n$ is being evaluated. The transformation matrix $\Phi$ represents the basis of the 6-D space and is given as:

$$\Phi = [\phi_0, \phi_1 \dots \phi_{11}] \tag{4}$$

where

$$\phi_l = \begin{bmatrix} \Phi(0,l) \\ \Phi(1,l) \\ \Phi(2,l) \\ \Phi(3,l) \\ \Phi(4,l) \\ \Phi(5,l) \end{bmatrix} = \begin{bmatrix} r_1 \sin l\frac{7\pi}{6} \\ r_1 \cos l\frac{7\pi}{6} \\ r_2 \sin l\frac{3\pi}{2} \\ r_2 \cos l\frac{3\pi}{2} \\ r_3 \sin l\frac{2\pi}{3} \\ r_3 \cos l\frac{2\pi}{3} \end{bmatrix} \qquad 0 \leq l \leq 11 \tag{5}$$

We set the $r_1$, $r_2$ and $r_3$ to 1, 1 and 0.5 respectively so that the distances between pitch classes in the 6-D space correspond to our perception of harmonic relations between pitches.

To reduce the effects of transients and noise, the sequence of tonal centroid vectors is convolved with a 19-point Gaussian smoothing window with $\sigma$ value of 4.034 in a row-by-row fashion. The HCDF, $\xi$, is defined as the overall rate of change of the smoothed tonal centroid signal. $\xi_n$ is the Euclidian distance between the smoothed tonal centroid vectors $\hat{\zeta}_{n-1}$ and $\hat{\zeta}_{n+1}$ (equation 6) where $\hat{}$ denotes vectors from the Gaussian-smoothed signal. The peaks in this signal indicate transitions between regions that are harmonically stable.

$$\xi_n = \sqrt{\sum_{d=0}^{5} \left[\hat{\zeta}_{n+1}(d) - \hat{\zeta}_{n-1}(d)\right]^2} \tag{6}$$

We apply peak picking to the HCDF in order to identify harmonic transitions which may be potential chord boundaries.

Using the the harmonic transition information from the HCDF stage, we can calculate an average chroma vector for each segment of the quantised chromagram.

We finally apply a chord recognition technique in order to estimate the chord symbol for each segment of the segmented chromagram. The chord recognition is achieved by simply multiplying the current chroma vector by a matrix of chroma chord templates. The templates used in this case are weighted bit patterns (weighted to sum to 1) corresponding to the chord types major, minor, augmented, diminished and ‘non-chord’. The major template is: [$\frac{1}{3}$,0,0,0,$\frac{1}{3}$,0,0,$\frac{1}{3}$,0,0,0,0]. We take the maximum element in the vector resulting from the multiplication to correspond to the chord which is currently being played.

## 2. RESULTS

In the initial MIREX results based on 206 song files, the algorithm scored 57% for the weighted average overlap score and 62.5% in the case that major and minor chords were merged in the test.

Evaluation of the algorithm using newer performance metrics of frame-based recall (using a specific chord dictionary) and a segmentation measure based on directional hamming distance [5] was also done. The new recall measure is based on a dictionary of allowable chord types. Those sections of the audio which are annotated with chord types that are not in the dictionary can be excluded from the evaluation. The segmentation measure is a combination of over-segmentation score $f$ and an under-segmentation score $m$ as described in [5]. The segmentation scores are not independent so to combine them we take the worst case to be our segmentation measure i.e. $1 - \max(m, f)$.

Evaluation results using these metrics were only available for the 180 Beatles songs at the time of writing. Of the 206 songs used for the initial evaluation results, 172 appeared in both sets. The average overlap score for those 172 songs was 58.9% whereas the new frame-based recall score was 59.8% with an average of 3% of frames excluded because of non-dictionary chords. The average segmentation score for the algorithm was 69.9%.

## 3. ANALYSIS

The system presented here did not perform as well as some of the other algorithms in the competition. However, the combination of the HCDF segmentation and simple chord recogniser has not been tuned fully yet. It is expected that some performance increases will be likely if the peak picking in the segmentation stage can be improved and also the chord templates that were used may also be enhanced by adding weightings for strong harmonics.

We believe that use of a segmentation measure in the evaluation of chord recognition algorithms is also important for judging performance. A highly fragmented chord recognition output is less useful in the real world than a more stable one even if the latter has a slightly lower recall value. The performance of the algorithm presented here is quite high in its segmentation accuracy compared to its recall measure and as such could compare more favourably with other algorithms in future evaluations where this factor is included as a performance metric.

## 4. REFERENCES

[1] J. C. Brown and M. S. Puckette. “An Efficient Algorithm for the Calculation of a Constant Q Transform”. *Journal of the Acoustical Society of America*, 92(5):2698–2701, 1992.

[2] C. Harte, M. Sandler: “Automatic chord recognition using a quantised chromagram” *Proceedings of the 118th Convention. of the AES*, pp. 66–71,Barcelona, Spain, 2005.

[3] C. Harte, M. Sandler, M. Gasser: “Detecting harmonic change in musical audio” *AMCMM’06: Proceedings of the 1st ACM workshop on Audio and music computing multimedia*, pp. 21–26,Santa Barbara, California, USA, 2006.

[4] C. Harte, M. Sandler, S. Abdallah and E.Gómez: “Symbolic representaion of musical chords: A proposed syntax for text annotations” *Proceedings of the 6th International Symposium on Music Information Retrieval (ISMIR05)*, pp. 66–71, September 2005.

[5] S. Abdallah, K. Noland, M. Sandler, M. Casey and C. Rhodes: “Theory and Evaluation of a Bayesian Music Structure Extractor” *Proceedings of the 6th International Symposium on Music Information Retrieval (ISMIR05)*, pp. 420-425), September 2005.