Detecting Harmonic Change In Musical Audio

Christopher Harte and Mark Sandler
Centre for Digital Music
Queen Mary, University of London
London, UK
christopher.harte@elec.qmul.ac.uk,
mark.sandler@elec.qmul.ac.uk

Martin Gasser
Austrian Research Institute for Artificial
Intelligence(ÖFAI)
Freyung 6/6, 1010 Vienna, Austria
martin.gasser@ofai.at

ABSTRACT

We propose a novel method for detecting changes in the harmonic content of musical audio signals.

Our method uses a new model for Equal Tempered Pitch Class Space. This model maps 12-bin chroma vectors to the interior space of a 6-D polytope; pitch classes are mapped onto the vertices of this polytope. Close harmonic relations such as fifths and thirds appear as small Euclidian distances.

We calculate the Euclidian distance between analysis frames $n+1$ and $n-1$ to develop a harmonic change measure for frame n. A peak in the detection function denotes a transition from one harmonically stable region to another. Initial experiments show that the algorithm can successfully detect harmonic changes such as chord boundaries in polyphonic audio recordings.

Categories and Subject Descriptors

J.0 [**Computer Applications**]: General

General Terms

Algorithms, Theory

Keywords

Pitch Space, Harmonic, Segmentation, Music, Audio

1. INTRODUCTION

In this paper we introduce a novel process for detecting changes in the harmonic content of audio signals. The Harmonic Change Detection Function (HCDF) combines a new theoretical model for equal tempered pitch space with a DSP front end to extract this information from digital audio recordings. The pitch space model projects collections of pitch classes as Tonal Centroid points in a 6-D space (section 2). Use of this 6-D space gives greatly improved results compared to using a straight forward 12-D pitch class distance measure.

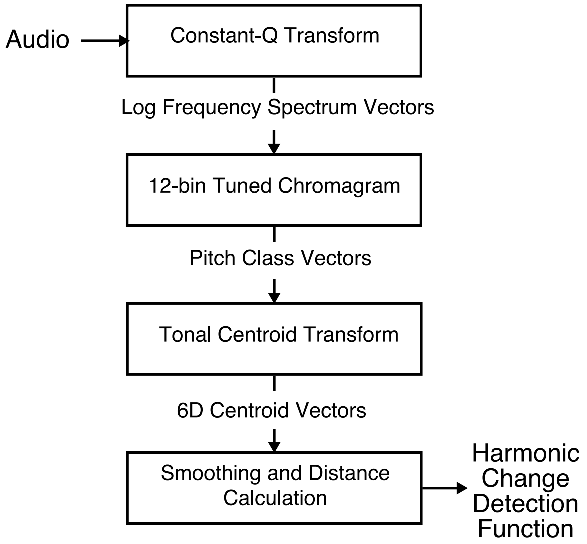

Figure 1: Flow Diagram of HCDF System.

Event-driven feature analysis has been shown to give more accurate musical feature extraction than more traditional approaches based on frames of equal length [1]. The HCDF has many potential applications in the segmentation of audio signals, particularly as a preprocessing stage for further harmonic recognition and classification algorithms. The primary motivation for this work is for use in chord recognition from audio. Used as a segmentation algorithm, this approach provides a good foundation for solving the general chord recognition problem of needing to know the positions of chord boundaries in the audio data before being able to successfully identify possible chord symbols as discussed in [16].

The HCDF system comprises several distinct elements. At the lowest level there is a Constant-Q spectral analysis followed by a 12-semitone Chromagram decomposition. A harmonic Centroid transform is then applied to the Chroma vectors which is then smoothed with a Gaussian filter before the distance measure is calculated (sections 3.1 and 3.2). Figure 1 shows the flow diagram for the system.

The results of our preliminary experiments are very promising with an f-measure value for overall chord change detection of 64.9% compared to a reference harmonic onset algorithm that scores 45.8% (section 4).

AMCMM'06, October 27, 2006, Santa Barbara, California, USA.

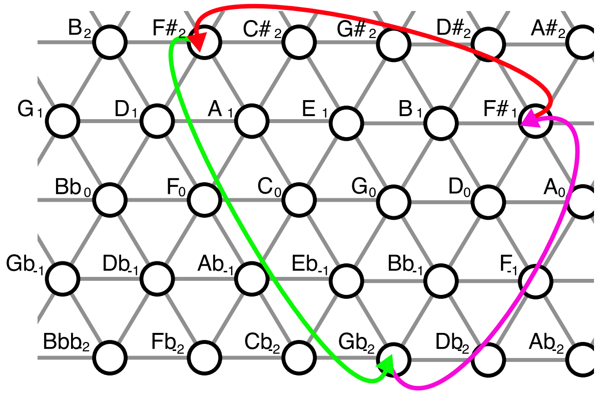

Figure 2: The Harmonic Network or *Tonnetz*. Arrows show the three circularities inherent in the network if enharmonic and octave equivalence are assumed.

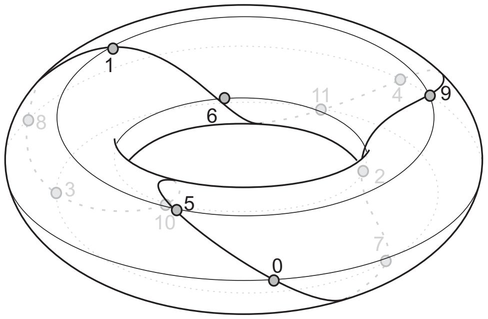

Figure 3: A projection showing how the Tonnetz wraps around the surface of a Hypertorus with the pitch classes following the spiral of fifths when enharmonic and octave equivalence are assumed.

2. MODELS FOR TONAL SPACE

The Harmonic Network or *Tonnetz* (shown in Figure 2) is a well known planar representation of pitch relations first attributed to Euler [6], later used extensively by 19th century music theorists such as Riemann and Oettingen and in recent years by Neo-Riemanninan Music Theorists [10, 6, 12]. Close harmonic relations are modelled by small distances on the plane. Lines of fifths travel from left to right, lines of major thirds travel from bottom left to top right and lines of minor thirds travel from top left to bottom right.

In Just Intonation, the Tonnetz is an infinite plane [13]. If it is assumed that a particular note spelling on one row is equivalent to the same note spelling on the next row (i.e. F$\sharp_1$ $\equiv$ F$\sharp_2$ etc. in Figure 2), the plane wraps up and forms a tube with the line of fifths becoming a helix on its surface. In the case where the helix is wrapped so that major third intervals are directly above each other on the surface of the tube this is Chew's Spiral Array [4]. Chew's model allows chords and keys to be projected as objects in a 3-D space on the interior of the tube and has been applied successfully to problems such as key finding and pitch spelling from symbolic data [5].

In the case of data derived from audio, it is very difficult to directly extract the correct enharmonic spelling of pitches. This is partly due to the fact that high resolution frequency analysis would be needed to resolve the small differences between them. On a more practical level, the majority of keyboard instruments are now tuned to twelve-tone equal temperament so the differences would not be present in the first place.

If enharmonic equivalence is assumed then instead of dealing with a theoretically infinite number of pitch names, there are now just the twelve different pitch *classes* (here we reference C as pitch class 0). In the Spiral Array model, this has the effect of joining the two ends of the tube together and the result is a hypertorus with the circle of fifths wrapping around its surface three times (see Figure 3). A form of this Hypertorus appears in many different areas of music research [10, 6, 11, 14].

We now propose a 6-dimensional interior space contained by the surface of the Hypertorus. This allows us to apply the same technique that Chew uses to develop the Centre of Effect in the Spiral Array to this equal tempered model for pitch space.

Since it is not possible to directly visualise 6-D space, it is helpful to imagine it as a projection onto the three circularities in the equal tempered Tonnetz: the circle of fifths, the circle of minor thirds and the circle of major thirds (Figure 4). Here, the six dimensions are viewed as three co-ordinate pairs $x1, y1$, $x2, y2$ and $x3, y3$. A collection of pitches (i.e. a chord) can be described as a single centroid point in the space. Chords with a tonal centre (such as the A major shown as point A in Figure 4) can be clearly assigned to a point in the circle of fifths. However, there are chords without defined tonal centres (e.g. diminished 7th and augmented chords). The centroid of each of these chords lies in the centre of the circle of fifths. On the circle of minor thirds, however, augmented chords can be unambiguously identified, while the circle of major thirds can uniquely depict diminished 7th chords.

3. ALGORITHM

The first stage of the system is the Constant-Q spectral analysis. This is a logarithmic frequency analysis based on the efficient algorithm described in [2]. We calculate a 36 bins-per-octave transform across five octaves between $f_{min} = 110$Hz (A2) and $f_{max} = 3520$Hz (A7) from a 11025Hz mono audio signal. To obtain this resolution at the lowest analysed frequencies requires a 743ms window length. This is a long analysis window in terms of musical signals so to improve time resolution we overlap analysis frames by $\frac{1}{8}$th of a window length giving the detection function a time resolution of 93ms per frame. A 12-bin tuned Chromagram is then calculated from the Constant-Q spectra using the method described in [9] giving a 12-dimensional chroma vector $\mathbf{c}$ for every frame.

3.1 Tonal Centroid Calculation

The six dimensional tonal centroid vector, ζ_n, for time frame n is given by the multiplication of the chroma vector, $\mathbf{c}$, and a transformation matrix Φ. 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 $\mathbf{c}$:

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

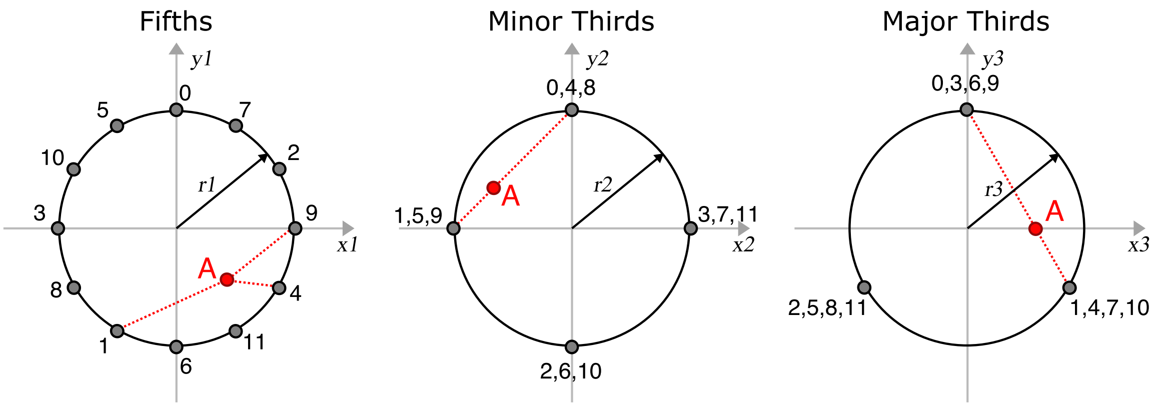

Figure 4: Visualising the the 6-D Tonal Space as three circles. Circles left to right: Fifths, Minor Thirds and Major Thirds. The Tonal Centroid for chord A Major (pitch classes 9,1 and 4) is shown at point A

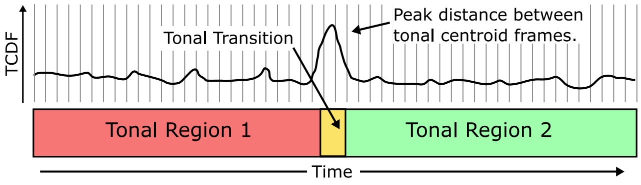

Figure 5: Harmonic Change Detection (Time frames are depicted by light vertical lines)

where l is the chroma vector pitch class index and d denotes which of the six dimensions of ζ_n is being evaluated. The transformation matrix Φ represents the basis of the 6-D space described in section 2 and is given as:

$$\Phi = [\phi_0, \phi_1 \ldots \phi_{11}] \qquad (2)$$

where

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

The values r_1, r_2 and r_3 are the radii of the three circles in Figure 4. To ensure that the distances between pitch classes in the 6-D space correspond to our perception of harmonic relations between pitches (i.e. that the fifth is the closest relation followed by the major third then the minor third and so on) we set the r_1, r_2 and r_3 to 1, 1 and 0.5 respectively. These values have been derived using the same approach that Chew uses to define the ratio of height to diameter for the Spiral Array [4].

3.2 Harmonic Change Detection Function

To reduce the effects of transient frames, the sequence of tonal centroid vectors is convolved with a 19-point Gaussian smoothing window with σ value of 4.034 in a row-by-row fashion (i.e. the individual dimensions are smoothed over time). We define the HCDF, ξ, as the overall rate of change of the smoothed tonal centroid signal. ξ_n is the Euclidian distance between the smoothed tonal centroid vectors $\hat{\zeta}_{n-1}$ and $\hat{\zeta}_{n+1}$ (equation 4) where ˆ denotes vectors from the Gaussian-smoothed signal. Peaks in this signal indicate transitions between regions that are harmonically stable (see Figure 5); an approach inspired by Chew's key modulation finding algorithm described in [5].

$$\xi_n = \sqrt{\sum_{d=0}^{5} \left[\hat{\zeta}_{n+1}(d) - \hat{\zeta}_{n-1}(d)\right]^2} \qquad (4)$$

We apply peak picking to the HCDF in order to identify harmonic transition times.

4. RESULTS

To test how the HCDF performed as a chord segmentation algorithm we analysed a set of sixteen Beatles songs (two from each of the first eight albums) for which we have chord transcription files. Harmonic change detection times are compared against the times of chord changes in the transcriptions.

Along with the results for the HCDF algorithm, we also give results for a harmonic onset detection algorithm by Hainsworth and Macleod [8] for comparison. This algorithm is a two-stage process in which peaks are detected in a spectral distance measure then the resulting segments are analysed for their harmonic content so that similar contiguous segments can be joined back together. The HCDF algorithm does not use such a second analysis step to obtain its results. The results for the experiment are shown in Table 1.

We defined a hit as a match within ±3 frames (278ms). The three performance measures used here are *Precision* (P), the ratio of Hits to Detected Changes and *Recall* (R), the ratio of hits to transcribed changes and the *f-measure* (F) which combines the two (see equation 5) [7].

$$F = \frac{2RP}{R+P} \qquad (5)$$

The f-measure scores show the HCDF to have the better overall performance of the two algorithms for chord boundary detection with a score of 64.9% compared to 45.8% for the Hainsworth algorithm. The Recall scores for both algorithms are fairly high with an average of 88% for Hainsworth's approach and 84% for the HCDF. However, the Precision scores for the two algorithms were much lower with averages

Table 1: Experimental results for HCDF peaks compared with hand labelled chord changes for 16 Beatles Songs (Songs arranged in chronological order of release date).

Song Title	TC	MN	MP	MR	MF	HN	HP	HR	HF
Please Please Me	78	267	27%	92%	41%	128	53%	87%	65.9%
Do You Want To Know A Secret	113	214	51%	97%	66.9%	126	72%	80%	75.8%
All My Loving	74	230	26%	81%	39.4%	129	50%	86%	63.2%
Till There Was You	93	265	33%	88%	48%	142	59%	90%	71.3%
A Hard Day's Night	101	343	28%	94%	43.1%	158	45%	70%	54.8%
If I Fell	81	237	32%	95%	47.8%	133	53%	87%	65.8%
Eight Days A Week	101	316	25%	80%	38%	169	49%	82%	61.3%
Every Little Thing	96	247	34%	84%	48.4%	133	49%	67%	56.6%
Help!	59	269	20%	84%	32.3%	152	29%	74%	41.7%
Yesterday	97	186	47%	83%	60%	132	64%	86%	73.4%
Drive My Car	84	328	24%	94%	38.2%	148	49%	86%	62.4%
Michelle	94	272	31%	90%	46.1%	160	53%	90%	66.7%
Eleanor Rigby	54	234	22%	96%	35.8%	128	34%	81%	47.9%
Here There And Everywhere	98	240	32%	73%	44.5%	127	65%	83%	72.9%
Lucy In The Sky With Diamonds	120	411	28%	97%	43.5%	213	50%	88%	63.8%
Being For The Benefit Of Mr Kite	113	255	38%	80%	51.5%	160	68%	95%	79.5%
Average over sixteen songs	**91**	**270**	**31%**	**88%**	**45.8%**	**146**	**53%**	**84%**	**64.9%**

Key to abbreviations

M	Denotes result for Hainsworth & Macleod's algorithm	**P**	Precision
H	Denotes result for HCDF algorithm	**R**	Recall
TC	Number of hand transcribed chord changes	**F**	f-measure
N	Number of detection function peaks		

of 31% for the Hainsworth algorithm and 53% for the HCDF. The significantly better Precision scores for the HCDF suggest that the new algorithm is better at discriminating important harmonic changes in the signal. This can be seen from the number of detection function peaks for each algorithm where the Hainsworth algorithm has an average of almost twice as many detections for each song as the HCDF. The HCDF was implemented in Matlab for these experiments. The low Precision scores for both algorithms can be explained by the fact that the transcription files only label chord changes. Both the detection functions here, however, pick up not only chord changes but also changes in harmonic content caused by strong melody or bass lines that include non-chord tones. Because of this a high number of false positives is to be expected for this experiment. Most misses in the HCDF are caused by transient signals introducing false peaks that mask the smaller peaks associated with the desired harmonic changes.

The songs that the HCDF algorithm performed best on were ones with fast harmonic rhythm such as *For The Benefit Of Mr Kite!*. The fast chord changes reduce the number of false positives and a strong organ part outlining the chords makes boundary detection easier. In contrast, songs such as *Every Little Thing* and *Help!* with slow harmonic rhythm and strong bass and melody lines producing false chord boundary detections scored less highly.

5. CONCLUSIONS AND FURTHER WORK

We have presented a novel feature detection function for audio data. A new model for equal tempered tonal space has been introduced on which the algorithm is based and the results of our preliminary experiments are encouraging. The algorithm has been implemented in Matlab and also in C++ as a visualisation plugin for the open source audio analysis tool Sonic Visualiser [3].

The results of our experiment show that the algorithm can successfully detect chord changes in polyphonic audio. However, other changes in harmonic content such as strong melody or bass line movement will also be detected. Developing a cleaner chromagram front end and the use of more sophisticated peak picking methods could improve the performance of the algorithm.

Applying adaptive thresholding in the peak picking stage of the algorithm may improve the detection of more important harmonic changes. Despite using the Gaussian function to smooth the detection function, strong transients in the audio signal still cause spurious peaks. Transient/Steady-State separation could be used to preprocess the audio in order to improve performance by reducing the effects of these transient components.

The HCDF has many potential applications in the segmentation of audio signals. It will be particularly useful as a preprocessing stage for many chord recognition and harmonic classification algorithms, especially those based on Hidden Markov Model techniques such as [1] and [15].

6. ACKNOWLEDGMENTS

This research was funded by EU-FP6-IST-507142 project SIMAC[1](acronym for Semantic Interaction with Music Audio Contents) and the WWTF project Interfaces to Music (I2M)[2].

[1] http://www.semanticaudio.org

[2] http://www.ofai.at/research/impml/projects/i2mproject

7. REFERENCES

[1] J. P. Bello and J. Pickens. A Robust Mid-level Representation for Harmonic Content in Musical Signals. In *Proceedings of the 6th International Conference on Music Information Retrieval, London*, 2005.

[2] 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.

[3] C. Cannam, C. Landone, M. Sandler, and J. Bello. The Sonic Visualiser: A Visualisation Platform For Semantic Descriptiors From Musical Signals. In *to appear in Proceedings of ISMIR*, Victoria, Canada, 2006.

[4] E. Chew. *Towards a Mathematical Model of Tonality.* PhD thesis, Operations Research Center, MIT. Cambridge, MA, 2000.

[5] E. Chew. The Spiral Array: An Algorithm For Determining Key Boundaries. *Proceedings of the Second International Conference, ICMAI 2002*, pages 18–31, 2002.

[6] R. Cohn. Introduction to Neo-Riemannian Theory: A Survey and a Historical Perspective. *The Journal of Music Theory*, 42(2), 1998.

[7] S. Dixon. Onset Detection Revisited. In *To appear in Proceedings of the 9th International Conference on Digital Audio Effects (DAFx06)*, Montreal, Canada, 2006.

[8] S. Hainsworth and M. Macleod. Onset Detection in Musical Audio Signals. In *Proceedings of ICMC*, Singapore, 2003.

[9] C. Harte and M. Sandler. Automatic Chord Identification Using a Quantised Chromagram. In *Proceedings of AES 118th Convention*, Barcelona, 2005.

[10] B. Hyer. Re-Imagining Riemann. *Journal of Music Theory*, 39(1):101–138, 1995.

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

[12] D. Lewin. *Generalized Musical Intervals and Transformations.* Yale University Press, New Haven, 1987.

[13] H. Longuet-Higgins. Second Letter to a Musical Friend. *The Music Review*, 23, 1962.

[14] H. Purwins. *Profiles of Pitch Classes Circularity of Relative Pitch and Key Experiments, Models, Computational Music Analysis, and Perspectives.* PhD thesis, Elektrotechnik und Informatik der Technischen Universitt Berlin, Berlin, 2005.

[15] A. Sheh and D. P. Ellis. Chord Segmentation and Recognition using EM-Trained Hidden Markov Models. *Proceedings of the ICMC 2003*, 2003.

[16] T. Yoshioka, T. Kitahara, K. Komatani, T. Ogata, and H. G. Okuno. Automatic Chord Transcription with Concurrent Recognition of Chord Symbols and Boundaries. In *Proceedings of ISMIR*, 2004.