A New Method for Tracking Modulations in Tonal Music in Audio Data Format [1]

Hendrik Purwins[♮‡], Benjamin Blankertz[⊢], and Klaus Obermayer[‡]
[♮]CCRMA, Stanford
[‡]Technical University Berlin, FR 2-1, FB 13, Franklinstr. 28/29, 10587 Berlin, Germany,
{hendrik,oby}@cs.tu-berlin.de
[⊢]GMD FIRST, Rudower Chaussee 5, 12489 Berlin, Germany,
blanker@first.gmd.de

Abstract

Cq-profiles are 12-dimensional vectors, each component referring to a pitch class. They can be employed to represent keys. Cq-profiles are calculated with the constant Q filter bank [4]. They have the following advantages: (i) They correspond to probe tone ratings. (ii) Calculation is possible in real-time. (iii) Stability is obtained with respect to sound quality. (iv) They are transposable. By using the cq-profile technique as a simple auditory model in combination with the SOM [11] an arrangement of keys emerges, that resembles results from psychological experiments [13], and from music theory [1]. Cq-profiles are reliably applied to modulation tracking by introducing a special distance measure.

Introduction

The goal of this work is to derive an appropriate representation of tone centers based on the audio signal. To what degree does such a representation have some psychological plausibility? Such a method should be fast in calculation. It should be appliable for stylistic analysis and for tone center tracking. An interesting question is how far one can get just employing DSP, without deeper musical considerations.

The probe tone experiments were pursued by Carol Krumhansl and Roger Shepard [14]. Probe tone ratings are a quantitative description of a key, that creates the possibility of relating statistical or computational analysis of music to cognitive psychology. The probe tone experiment consists of two stages: establishment of a tonal context, and rating of the relation of a probe tone to that context. The tonal context is provided by examples, which are unambiguously written in a certain key. In our case the subjects listen to simple cadential chord progressions composed of Shepard tones [13]: IV-V-I, VI-V-I, II-V-I (Roman numerals indicating scale degrees of the root of the chords). Subsequently, a Shepard tone chosen randomly from the chromatic scale, the probe tone, is played. The subject is asked to judge, how well the note fits with the tonal context, provided by the cadential chord progression. The test subjects rate by a number from 1 (“fits poorly”) to 7 (“fits well”). After this procedure is repeated several times, with different chromatic notes, the average rating for each pitch class is calculated. The 12-dimensional vector containing the averaged answers for each pitch class is called the probe tone rating. There are two types of rating vectors, one for major and one for minor – depending on the mode of the contexts. Rating vectors of keys in the same mode but with different tonic keynotes are assumed to be related by a shift that compensates for the interval of transposition (cf. [13], p. 342).

context	c	c♯/d♭	d	d♯/e♭	e	f	f♯/g♭	g	g♯/a♭	a	a♯/b♭	b
c-major	6.35	2.23	3.48	2.33	4.38	4.09	2.52	5.19	2.39	3.66	2.29	2.88
c-minor	6.33	2.68	3.52	5.38	2.6	3.53	2.54	4.75	3.98	2.69	3.34	3.17

One observes, that the first scale degree is rated highest. The third and fifth scale degrees are also rated high. Diatonic notes are rated higher than non-diatonic notes. According to an observation reported in [12] (p. 66–76), each component in the probe tone rating vector corresponds to

[1] H. Purwins, B. Blankertz, and K. Obermayer. A new method for tracking modulations in tonal music in audio data format. In *International Joint Conference on Neural Network (IJCNN'00)* **6**:270-275. IEEE Computer Society, 2000.

the frequency and the overall duration of occurrence of the corresponding pitch class at metrically prominent positions in a tonal piece that is written in a given key. Key distances are calculated by comparing the corresponding probe tone ratings by correlation, Euclidean distance, etc.

Our goal is different from pitch recognition. We need not to know all exact pitches, just a profile which indicates the key, resp. the tone center.

To see how a piece is represented, we have to consider, how a note is represented. We will restrict us to a representation in a 12-dimensional vector. Each component in the vector corresponds to a pitch class in the well tempered chromatic scale.

There are some approaches for automatic tone center recognition. Gang and Berger [7] introduced a system, based on input in midi data format. Linking metrical and harmonic information a recurrent net learns to make harmonic predictions. Griffith [8] did tone center analysis on the simplest representation and referred to profiles that included interval use from each pitch class [5]. Fujishima [6] matches chords form the audio signal with some prototype chords based on Fourier techniques. Leman [15] did tone center analysis on the basis of references from Shepard tone cadential chord progressions, which where preprocessed by an auditory model. Izmirli and Bilgen [10] used the constant Q transformation [2] for tone center analysis in combination with a refined frequency estimation observing phase changes [3]. Context is integrated adaptively based on chord changes. By cancelling out harmonics of a detected fundamental, fundamentals of other tones are possibly cancelled out also. This method yields a quite reasonable, yet not perfect tone center analysis.

First we will introduce the constant Q profile technique, and we will indicate its correlation to the psychological probe tone data. Then we will show results in using the constant Q profile technique as a simple auditory model in combination with the Self Organizing Feature Map [11]. An arrangement of keys, which is perceptually relevant, evolves on the basis of the Préludes Op. 28 by Chopin. Then a special distance measure, the fuzzy distance, is introduced. It leads to good results in tracking of modulations across different tone centers.

Cq-Profiles

Cq-profiles are a new concept of key profiles. They unite features of Krumhansl's probe tone ratings [12] and Leman's correlograms [15]. Advantages comprise: (1) Each cq-profile has a simple interpretation, since it is a 12–dimensional vector like a probe tone rating. The value of each component corresponds to a pitch class. (2) A cq-profile can easily be calculated from an audio recording. Since no complex auditory model, or other time consuming method is used, the calculation is quick and can be done in real time. (3) The calculation of the cq-profiles is very stable with respect to sound quality. E.g. analyzing a recording of Alfred Cortot from 1933/34 works well.

Constant Q transform The calculation of the cq-profiles is based on the constant Q transform [2]. The letter 'Q' refers to the constant quotient of center frequency and bandwidth for each filter. The constant Q transform is useful in establishing a direct correspondence between filters and musical notes by identifying appropriate center frequencies. To minimize spectral leakage (cf. [9]), we use 36 filters per octave rather than 12.

Figure 1 shows how cq-profiles are calculated from the output of the transform. Hence, only every third filter output maps to a tone of the chromatic scale. Cq-profiles can be used to study pitch use in different composers and for modulation tracking. A cq-reference set is a sequence of 24 cq-profiles, one for each key. Every profile should reflect the tonal hierarchy that is characteristic for its key. Typically cq-reference sets are calculated from sampled cadential chord progressions or from small pieces of music.

Calculation of cq-transform Like the Fourier transform, a constant Q transform [2] is a bank of filters, but in contrast to the former it has geometrically spaced center frequencies $f_k = f_0 \cdot 2^{\frac{k}{b}}$ and a constant ratio of frequency to bandwidth $Q = \frac{f_k}{\Delta_k^{\text{cq}}} = (2^{\frac{1}{b}} - 1)^{-1}$ $(k = 0, \ldots)$, where b dictates the number of filters per octave.

This is achieved by choosing an appropriate window length N_k individually for each component of the constant Q transform (cq-bin). For integer values Q the k-th cq-bin is the Q-th DFT-bin with window length $Q\frac{f_s}{f_k}$. Calculation: First choose minimal frequency f_0 and the number of bins per octave b according to the requirements of the application and let[2]: $K := \lceil b \cdot \log_2(\frac{f_{\max}}{f_0}) \rceil$, $Q := (2^{\frac{1}{b}} - 1)^{-1}$, and $N_k := \lceil Q\frac{f_s}{f_k} \rceil$ (for $k < K$). Then the k–th cq-bin is equal to $N_k^{-1} \sum_{n<N_k} x[n]\ w_{N_k}[n]\ e^{-2\pi i n Q/N_k}$. Following [4] we use Hamming windows $\langle w_N[n] : n < N \rangle$. Using Parseval's rule a filter matrix is calculated in advance. Exploiting sparsity accelerates the calculation of the constant Q transform very much [4].

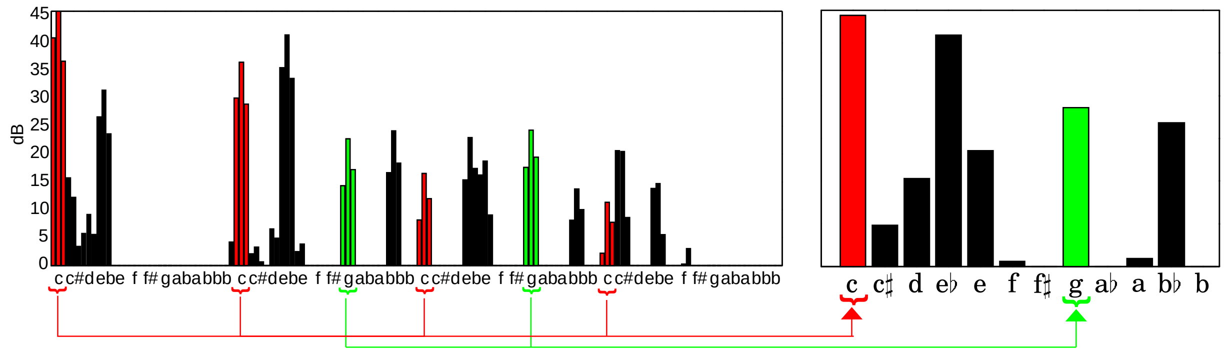


FIGURE 1: The constant Q transform is calculated from a minor third c – e♭ (played on piano) with three bins per half-tone (left figure). We yield the constant Q profile (right figure) by summing up bins for each tone over all octaves.

Applications

Derivation of a Toroidal Model of Inter-Key Relations (ToMIR) with cq-reference sets In an experiment with music in audio data format we combine the cq-transform and a toroidal Self Organizing Feature Map (SOM, [11]). After training the evolved configuration resembles ToMIR in music psychology ([12] p. 46) and music theory ([16] pp. 19). The circles of fifths wrap around the torus in three turns. On the configuration, keys share borders with the dominant, the subdominant, the parallel and the relative key, cf. figure 11 in [1].

Comparison of different hierarchies An important question that arises regarding cq-reference sets that are calculated from audio recordings is in what respect the results are affected by (1) musical interpretation, (2) the recorded instrument, and (3) the selected pieces of music.

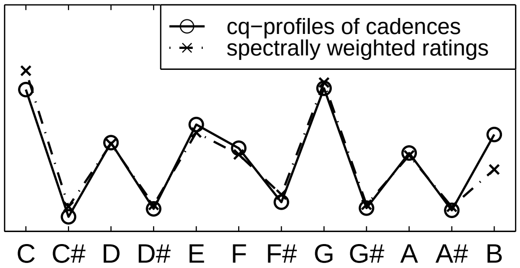
For the examination of (1) we compared radically different interpretations of the Chopin préludes and Bach WTC I preludia. The mean profiles showed a correlation of [3] 0.995/0.989.

For the investigation of (2) we compared the recordings of the preludia of Bach's WTC I preformed on modern pianos and on a (Pleyel) harpsichord: The correlation is 0.989/0.982.

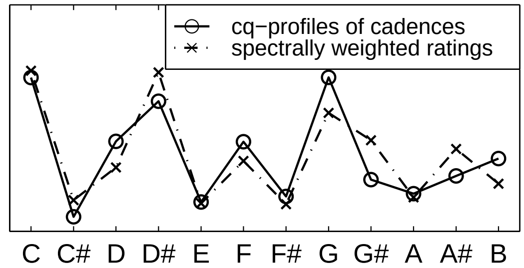
To study the impact of the selection of music (3) on the corresponding reference sets, we performed some inter and some across epoch comparisons. Group 1 consists of four reference sets calculated from the preludia/fugues cycles (separately) of both books of the well-tempered clavier (Glenn Gould's recording). Group 2 consists of two reference sets derived from Alfred Cortot's recording of the Chopin préludes op. 28 (finished 1839), and from Olli Mustonen's recording of Alkan's préludes op. 31 (finished 1847). Group 3 consists of a reference set based on Scriabin's preludes op. 11 (finished 1896) performed by Vladimir Sofronitsky, Heinrich Neuhaus, Vladimir Horowitz and the composer

[2] $\lceil x \rceil$ denotes the least integer greater than or equal to x.

[3] When writing correlation values in the form x/y we use the convention that x refers to major profiles and y to minor profiles.



(a) Major ratings



(b) Minor ratings

FIGURE 2: The cq-profiles of sampled piano cadences are compared with spectrally weighted ratings.

(reproduced from a Welte-Mignon piano roll). The inter-group correlations are 0.992/0.983 for the Bach reference sets (mean value) and 0.987/0.980 between Chopin's and Alkan's préludes. The mean across group correlations are 0.924/0.945 between groups 1 and 2, 0.935/0.949 between groups 1 and 3 and 0.984/0.952 between groups 2 and 3.

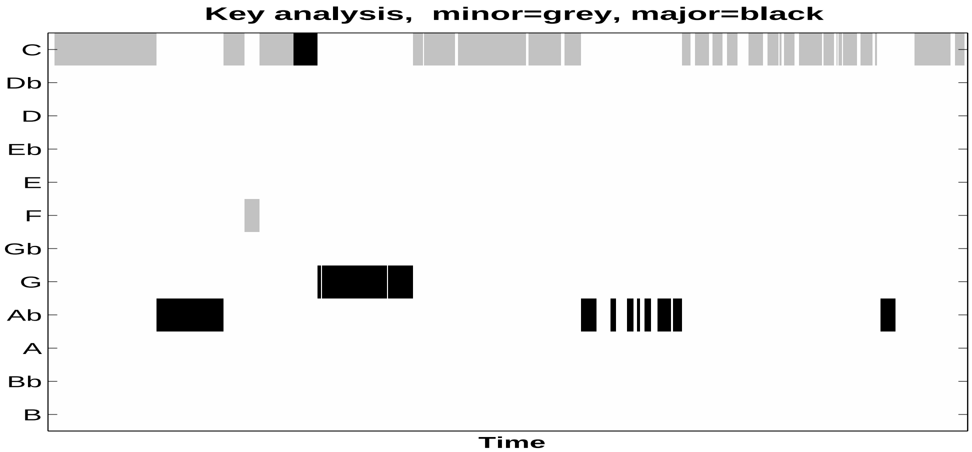
Relating cq-profiles to probe tone ratings Krumhansl observed in [12] a remarkable correlation between the probe tone ratings and the total occurrences of the twelve chromatic scale tones in musical compositions. In order to establish a direct correspondence between probe tone ratings and profiles of a cq-reference set, one fact has to be taken into consideration. In the cq-profiles not only the played tones are registrated, but all harmonics. For piano tones the strongest frequency contribution falls (modulo octaves) on the tonic and on the dominant keynote in an approximate average ratio 3:1. Hence cq-profiles should not be compared with the probe tone ratings, but with adapted ratings, in which the harmonic spectrum of the analyzed tones is accounted for. Such a spectrally weighted rating is calculated by adding to the rating value for each tone one third of the rating value for the tone seven chromatic steps above (modulo octave). Figure 2 shows the correlation of the cq-profiles of sampled piano cadences (I-IV-V^7-I and I-VI-V^7-I) with the spectrally weighted ratings.

Application in tone center tracking How can a piece be classified according to a cq-reference set? Generally we have the problem of matching a given cq-profile with a profile of the cq-reference set. A typical matching criteria is the closest fuzzy distance: Let y be a value subject to an uncertainty quantisized by a value σ (typically y is the mean and σ the standard deviation of some statistical data). The fuzzy distance of some value x to y regarding σ is defined by $d_\sigma(x,y) := |x-y| \cdot \left(1 - \frac{\sigma}{|x-y|+\sigma} e^{-\frac{|x-y|^2}{2\sigma^2}}\right)$. The fuzzy distance is similar to the Euclidean metric, but the greater the uncertainty the more relaxed is the metric. As an example, we present an analysis of Chopin's c-minor Prélude op. 28, No. 20. The reference vectors were calculated from all 24 Chopin Préludes in audio format.

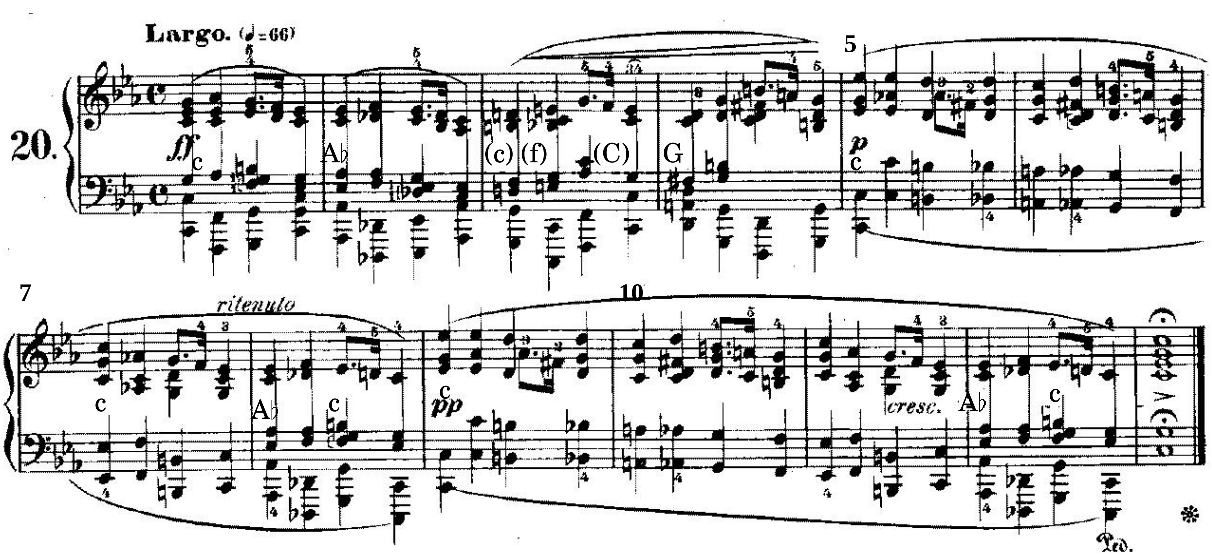
In the score (Figure 3 (b)) tone centers are marked. They were determined by a musical expert. Tone centers in parentheses indicate tonicizations on a very short time scale. Since the automatic tone center recognition (Figure 3 (a)) does not look ahead, there is a delay in recognizing tone centers. The program captures the prevailing key c–minor and the modulations: c–minor (1.measure), $A\flat$–major (2.measure). [4]

In measure 3, because of the interdominants G^7 (1.beat) and C^7 (2.beat) there is a short tonicization for c–minor and for f–minor. Then C–major is indicated (beat 4). Measure 4 shows G-major. In measures 5 and 6 a faux bordun in c–minor occurs, on a coarse time scale. Short tonicizations occur in measure 5, beat 4 (g–minor) and measure 6, beat 3 (G–major). In measure 7, there is a clear

[4] In measure 3 on beat 3 there is an e'. [17] points out this being a typo in Chopin's manuscript. He argues for replacing e' by $e\flat'$, because major and minor keys should alternate within the first 4 measures. We analyze an unrevised version.



(a) Result of automatic tone center analysis



(b) Score

FIGURE 3: Chopin's c-minor prélude, op. 28, No. 20. In (a) grey indicates minor, black indicates major. If there is neither black nor grey at a certain time, the significance of a particular key is below a given threshold. There is no distinction between enharmonic equivalent keys.

cadence in c–minor. In measure 8 beat 1 and 2 we have a flavour of $A\flat$–Major or $D\flat$–Major. The analysis indicates $A\flat$–Major. In measure 8 beat 4, the piece returns to c–minor. The analysis indicates this with a delay, because in the performance of Cortot $A\flat$–Major is heavily emphasized. Measures 9-12 are the same as 5-8, except the level is pp now. Therefore the analysis is more uncertain.

This result is astonishing. The only explicit musical knowledge utilized is the display of the signal in terms of pitch classes. The system receives musical knowledge only by choice of the music pieces, which lead to the reference vectors.

Discussion

The simple constant Q profile method incorporates context processing by averaging over the entire piece. Only very basic music theoretical assumptions like octave equivalence and the chromatic scale are used explicitly. However it can capture a large amount of harmonic structure including modulation and keys. Other musical knowledge is not explicitly used, like voice leading, harmony, metric, and rhythm. The constant Q profile method is a powerful tool that can be extended to different tunings, and to real time analysis. It could be improved by modeling masking phenomena.

Applications include automatic modulation tracking, and analysis of pitch use in different composers and epochs. A forthcoming paper will cover more details.

We are grateful to H. P. Reutter, and A. Budde. The first author was supported by "Studienstiftung des deutschen Volkes" and "Axel Springer Stiftung".

References

[1] B. Blankertz, H. Purwins, and K. Obermayer. Toroidal models of inter-key relations in tonal music. In *VI. International Conference on Systematic and Comparative Musicology*, 1999. submitted.

[2] J. Brown. Calculation of a constant Q spectral transform. *J. Acoust. Soc. Am.*, 89(1):425–434, 1991.

[3] J. C. Brown and M. S. Puckette. A high resolution fundamental frequency determination based on phase changes of the Fourier transform. *Journal of the Acoustical Society of America*, 1993.

[4] Judith C. Brown and Miller S. Puckette. An efficient algorithm for the calculation of a constant Q transform. *J. Acoust. Soc. Am.*, 92(5):2698–2701, 1992.

[5] R. Browne. Tonal implications in the diatonic set. *In Theory Only*, 5:3–21, 1981.

[6] T. Fujishima. Realtime chord recognition of musical sound: a system using Common Lisp Music. In *International Computer Music Conference*, pages 464–467. ICMA, 1999.

[7] D. Gang and J. Berger. A unified neurosymbolic model of the mutual influence of memory, context and prediction of time ordered sequential events during the audition of tonal music. In *Hybrid Systems and AI: Modeling, Analysis and Control of Discrete + Continuous Systems*. AAAI Technical Report SS-99-05, 1999.

[8] N. Griffith. Development of tonal centers and abstract pitch as categorizations of pitch use. In *Connection Science*, pages 155–176. MIT Press, Cambridge, 1994.

[9] F. J. Harris. On the use of windows for harmonic analysis with discrete fourier transform. In *Proc. IEEE*, volume 66, pages 51–83, 1978.

[10] Ö. Izmirli and S. Bilgen. A model for tonal context time course calculation from acoustical input. *Journal of New Music Research*, 25(3):276–288, 1996.

[11] T. Kohonen. Self-organized formation of topologically correct feature maps. *Biol. Cybern.*, 43:59–69, 1982.

[12] C. Krumhansl. *Cognitive Foundations of Musical Pitch*. Oxford University Press, Oxford, 1990.

[13] C. L. Krumhansl and E. J. Kessler. Tracing the dynamic changes in perceived tonal organization in a spatial representation of musical keys. *Psychological Review*, 89:334–68, 1982.

[14] C. L. Krumhansl and R. N. Shepard. Quantification of the hierarchy of tonal function with a diatonic context. *Journal of experimental psychology: Human Perception and Performance*, 1979.

[15] M. Leman. Schema-based tone center recognition of musical signals. *Journal of New Music Research*, 23:169–204, 1994.

[16] A. Schoenberg. *Structural functions of harmony*. Norton, 1969.

[17] E. Zimmermann. *Chopin Préludes op. 28: Kritischer Bericht*. Henle, 1969.