GLOBAL KEY EXTRACTION FROM CLASSICAL MUSIC AUDIO RECORDINGS BASED ON THE FINAL CHORD

Christof Weiß
Fraunhofer Institute for Digital Media Technology, Ilmenau, Germany
christof.weiss@idmt.fraunhofer.de

ABSTRACT

This paper presents a novel approach to global key extraction from audio recordings, restricted to the genre *Classical* only. Especially in this field of music, musical key is a significant information since many works include the key in their title. Our rule-based method relies on pre-extracted chroma features and puts special emphasis on the final chord of the piece to estimate the tonic note. To determine the mode, we analyze the chroma histogram over the complete piece and estimate the underlying diatonic scale. In both steps, we apply a multiplicative procedure to obtain high error robustness. This approach helps to minimize the amount of false tonic notes which is important for further key-related tonality analyses. The algorithm is evaluated on three different datasets containing mainly 18th and 19th century music for orchestra, piano, and mixed instruments. We reach accuracies up to 97 % for correct full key (correct tonic note and mode) classification and up to 100 % for correct tonic note classification.

1. INTRODUCTION

The key is an essential information about a musical work. Especially in Western art music, the usage of different keys shows some historical peculiarities that are connected to the evolution of the musical instruments and tuning schemes. Inspired by the ability to play all keys on keyboard instruments, J. S. Bach and several latter composers created series of works for every single key. In other works, musical keys obtain certain characteristics or special semantic meanings.

Therefore, automatized extraction of musical key is an important task in Music Information Retrieval (MIR). Besides applications for annotating classical music datasets, the key may also be necessary for further MIR tasks like genre classification or composer identification. For such scenarios, all key misclassifications constitute a problem. Especially, the system should avoid confusions of fifth-related keys that arise frequently in many common algorithms, e.g. in [1–4].

To this end, we first consider the special role of the final chord in this paper. For most pieces, the root of this chord (the first note in a cluster of ascending thirds) equals the tonic note of the written global key. We combine this information with a scale estimation of the complete piece. For this, we present the idea of multiplicative chroma processing to estimate a chord's root or a diatonic scale. We show that this reduces classification errors compared to template-based methods.

This rule-based approach is inspired by music theory and does not make use of machine learning techniques so far. We restrict ourselves to consider classical music only. For other genres like *Rock*, *Pop* or *Jazz*, such a method may not work since there may arise a considerable number of fade out endings or complex final jazz chords.

2. RELATED WORK

Since the concept of musical key is not defined precisely in many cases, automatic key extraction remains a challenging MIR task—also on classical music data. There are algorithms dealing with symbolic data only, as well as direct audio analysis methods on which we focus on in this paper. Recent overviews can be found in [2, 5], also comparing knowledge-based and data-driven algorithms—the two main approaches.

In general, the first step is an extraction of chroma features. Motivated by studies on human pitch perception [6, 7], many algorithms match the chroma statistics to pitch class profiles or use advancements of such approaches [1, 5, 8–10]. In the MIREX 2005 contest (1252 classical pieces synthesized from MIDI), the best results reached 87 % correctly identified keys [3].

Among the works concerning data-driven techniques, Hidden Markov Models (HMMs) are used most frequently [4, 11]. They also show promising results in localized tonality analysis and chord detection. Chai and Vercoe [4] combine HMMs with a two-step approach, considering diatonic scale and tonic note individually. Noland and Sandler [11] investigate the effect of the signal processing parameters and test their HMM-based approach on recordings of Bach's well-tempered piano, book 1 (48 tracks), yielding 98 % correct classification for the best parameter settings.

There are works considering special sections of the recordings: Izmirli [9] investigates the first seconds of 85 classical pieces by different composers (randomly

chosen from a NAXOS dataset) with up to 86 % success. Chuan and Chew [12] test their geometrical approach on the beginning of several Mozart symphonies yielding up to 96 % success rate. Extending these tests to a wide stylistic range, they reach 75 % correct accuracy [13]. Van de Par et al. [14] combine profile training with special weighting of the beginning and ending section. They evaluate on piano music [1] with high accuracies up to 98 %.

3. SYSTEM OVERVIEW

In the presented key detection system, we make use of the final chord's significance in Western classical music applying a two-step approach: First, we estimate the final chord's root and the complete piece's dominating diatonic scale individually. Then, we combine these informations to obtain the most probable full key consisting of the tonic note and the associated mode (major/minor). An overview is shown in Fig. 1.

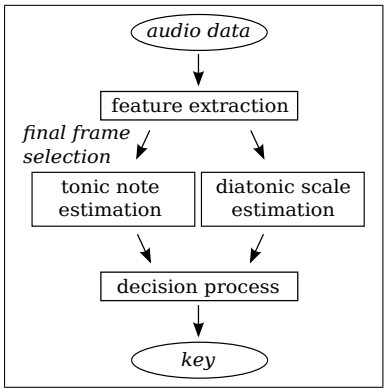

Figure 1. The key extraction process.

3.1 Feature Extraction

The system is based on chroma features which are commonly used to represent the harmonic content of music [15, 16]. We use the Chroma Toolbox of Müller and Ewert [17] to extract pitch and chroma features from the audio data in a preprocessing step. First, we calculate a pitch representation from the audio signals via a multi-rate filter bank, covering the pitch range of a grand piano (MIDI pitches Nos. 21–108). To account for the global tuning, we use the tuning estimation of this toolbox package and apply a shifted filter bank if the difference from a 440 Hz tuning exceeds 15 cent. We obtain a set of N_{tot} pitch feature vectors $\mathbf{p}$ (88-dim.), each covering a frame of length 50 ms:

$$\left(\mathbf{p}^1, \ldots, \mathbf{p}^i, \ldots, \mathbf{p}^{N_{\text{tot}}}\right) \tag{1}$$

To estimate the overall dynamics, we calculate the average L_1 norm of these vectors:

$$||\mathbf{p}||_{\text{mean}} = \frac{1}{N_{\text{tot}}} \sum_{i=1}^{N_{\text{tot}}} ||\mathbf{p}^i||_1 \tag{2}$$

[1] cf. Sec. 4.1 and [10] for details

Next, the energy of all pitch bands belonging to a pitch class (chroma) is summed up and normalized to obtain 12-dim. chroma vectors $\mathbf{c}$, where c_k^i is the k-th component of the i-th chroma vector:

$$\left(c_1^i, c_2^i, \ldots, c_{12}^i\right) \mathrel{\widehat{=}} (C, C\sharp, \ldots, B) \tag{3}$$

Then, we add up all vectors over the piece to obtain a normalized (using L_2 norm) chroma histogram $\mathbf{g}$:

$$\mathbf{g}' = \sum_{i=1}^{N_{\text{tot}}} \mathbf{c}^i \,, \quad \mathbf{g} = \frac{\mathbf{g}'}{||\mathbf{g}'||_2} \tag{4}$$

For all pitch- and chroma-related vectors, we identify flat and sharp notes ($C\sharp = D\flat$) and understand the indexing in a circular way ($k \to 1 + (k-1) \bmod 12$).

3.2 Tonic Note Estimation

3.2.1 Frame Selection

Starting from this feature set, we estimate the root of the piece's final chord. Since we do not want to consider frames containing silence, we take the last N final feature frames that exceed a defined energy threshold. To account for the overall loudness of the piece, we apply a dynamical adaption for the energy threshold. We calculate the L_1 norm of the pitch feature vectors [2] $\mathbf{p}^j$ and select only vectors fulfilling the condition

$$||\mathbf{p}^j||_1 > f_e \cdot ||\mathbf{p}||_{\text{mean}} \tag{5}$$

with a suitable factor f_e.

3.2.2 Chroma Processing

From the frame selection thus obtained (length N_{end}), we compute a 12-dim. chroma histogram $\mathbf{h}$:

$$\mathbf{h}' = \sum_{m=1}^{N_{\text{end}}} \mathbf{c}^m \,, \quad \mathbf{h} = \frac{\mathbf{h}'}{||\mathbf{h}'||_2} \tag{6}$$

Here, we are only interested in the root and not in the mode of the final chord, and thus ignore this chord's third. [3] To consider the tonal relationship between the chroma classes, we re-sort the entries of $\mathbf{h}$ according to a perfect fifth ordering by re-ordering the indices:

$$(1, 2, \ldots, 12) \to (2,\ 9,\ 4,\ 11,\ 6,\ 1,\ 8,\ 3,\ 10,\ 5,\ 12,\ 7)$$

$$\left(h_1^{\text{fifth}}, \ldots, h_{12}^{\text{fifth}}\right) \mathrel{\widehat{=}} (D\flat,\ A\flat, \ldots, C,\ G, \ldots, B,\ F\sharp)$$

We multiply these values for each two neighboring entries

$$h_k^{\text{prod}} = h_k^{\text{fifths}} \cdot h_{k+1}^{\text{fifths}} \tag{8}$$

to consider only such chroma peaks, where the respective upper fifths is also present. The principle is illustrated in Fig. 2.

[2] Since the chroma features are normalized, we compute the energy measure directly on the pitch features.

[3] In classical music, the final chord may not be representative for the overall mode of the piece: Many minor pieces end in the respective major chord ("Picardy third"), certain symphony movements show a development from minor to major, etc.

Since the majority of classical pieces' final chords—independently of their mode—contain strong energy in the root as well as in the fifth chroma, this procedure provides the final chord's root with a high reliability:

$$k^{\text{root}} = \arg\max_k h_k^{\text{prod}} \tag{9}$$

Also for third-less chords or even monophonic endings, this method works well, as the third partial of the root always produces some energy in the fifth chroma. To estimate the likelihoods, we calculate confidence measures P_k^{tonic} using the euclidean norm:

$$P_k^{\text{tonic}} = \frac{h_k^{\text{prod}}}{||\mathbf{h}^{\text{prod}}||_2} \tag{10}$$

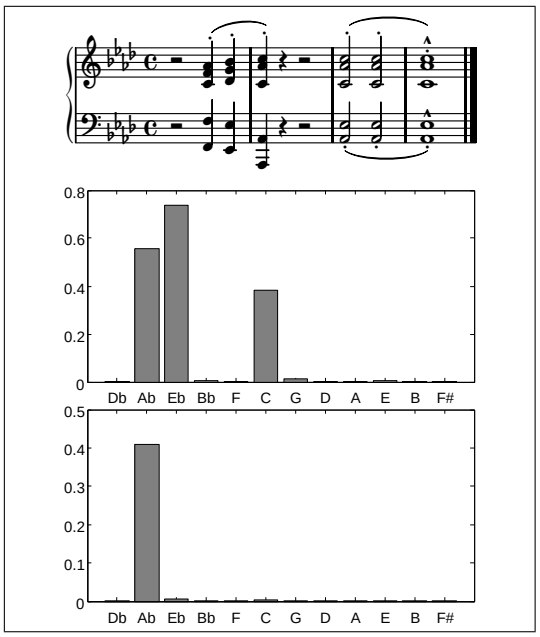

Figure 2. The final bars of Frédéric Chopin's Impromptu No. 1 for Piano, op. 29 in A♭ major. The upper plot shows the re-sorted chroma histogram $\mathbf{h}^{\text{fifths}}$ from the last $N = 30$ frames (cf. Eqs. 5–6), which results in the lower one $\mathbf{h}^{\text{prod}}$ after pairwise multiplication (cf. Eq. 8). From this, we identify the correct root A♭ even though the maximum value in the chroma histogram belongs to E♭.

3.3 Diatonic Scale Estimation

Since classical works or single movements may pass through certain tonal progressions, show parts in other keys, or even end in a key different from the global key [4], we consider the complete length data to identify the diatonic scale that corresponds to the global key's major or natural minor scale. To this end, we use the chroma histogram $\mathbf{g}$ from the preprocessing step (Sec. 3.1) and try to estimate the most probable diatonic scale or "tonal level". This concept, illustrated e.g. in [18], is suitable for various tonal analysis tasks. As an example, G major as well as E minor are denoted as "+1 level" (1♯), B♭ major and G minor as "-2 level" (2♭). As a diatonic scale consists of seven fifth-related notes (cf. Fig. 3), we again resort the histogram to a fifth ordering and compute a 12-dimensional vector by multiplying each seven fifth-related chroma energies corresponding to the respective diatonic scale (Indexing: $g_1^{\text{prod}} \rightarrow k = -5$ diatonic (= D♭ major scale), …, $g_{12}^{\text{prod}} \rightarrow k = +6$ diatonic):

$$g_k^{\text{prod}} = \prod_{l=1+k \bmod 12}^{1+(k+11) \bmod 12} \left(g_l^{\text{fifths}}\right)^{m_l^{(a)}} \tag{11}$$

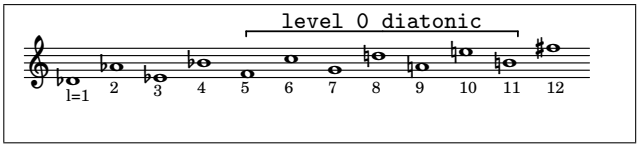

Figure 3. A diatonic scale (level 0) in a representation of neighboring fifths. The notes are signed with the indices of $\mathbf{m}^{(a)}$. The tonic note for the corresponding major scale is C ($l = 6$), for the minor scale A ($l = 9$).

To account for the individual relevance of the notes, we test a weighting [5] by five different templates of exponents $m_l^{(a)}$:

$$\mathbf{m}^{(1)} = (0\ 0\ 0\ 0\ 1\ 1\ 1\ 1\ 1\ 1\ 1\ 0)$$
$$\mathbf{m}^{(2)} = (0\ 0\ 0\ 0\ 1\ 3\ 2\ 1\ 2\ 3\ 1\ 0)$$
$$\mathbf{m}^{(3)} = (0\ 0\ 0\ 0\ 3.75\ 4.75\ 3.00\ 3.75\ 4.25\ 4.50\ 3.75\ 0)$$
$$\mathbf{m}^{(4)} = (0\ 0\ 0\ 0\ 4.04\ 5.87\ 4.27\ 3.51\ 5.00\ 4.57\ 3.20\ 0)$$

$\mathbf{m}^{(1)}$ corresponds to equal weighting. In $\mathbf{m}^{(2)}$, we emphasize the notes of the tonic chords (for the level 0 diatonic of Fig. 3, these are the C major and the A minor chord). $\mathbf{m}^{(3)}$ is computed from the templates of Temperley [7], where we summed up the major and the relative minor profile, multiplied with 0.5. $\mathbf{m}^{(4)}$ is the same for the Krumhansl templates [6]. The non-diatonic notes are exponentiated by 0 and thus not considered. Up to this, the scale estimation step basically equals a common template matching. [6] However, the multiplicative procedure leads to a high fidelity, since yet shifting by one fifth (i.e., one small g_l value in the product Eq. 11) leads to a small g_k^{prod} entry.

Again, we compute confidence measures for all 12 levels via

$$P_k^{\text{scale}} = \frac{g_k^{\text{prod}}}{||\mathbf{g}^{\text{prod}}||_2}. \tag{12}$$

3.4 Decision Process

To select the most probable key from the confidence measures computed before, we build a 24-dimensional

[4] Most frequently, this is the corresponding minor/major key; cf. remarks to Sec. 3.2

[5] Note that for a product calculation, weighting has to be done by exponentiation and not by multiplication.

[6] The fifth ordering is just for visualisation: In this representation, all diatonic scale notes are neighbors.

confidence vector, combining every tonic note confidence with the associated major and minor scale confidences, where the exponent s serves as a tuning parameter between root and scale influence:

$$P_k^{\text{major}} = \left(P_k^{\text{tonic}}\right)^s \cdot P_k^{\text{scale}} \qquad (13a)$$
$$P_k^{\text{minor}} = \left(P_k^{\text{tonic}}\right)^s \cdot P_k^{\text{mscale}}$$
$$\mathbf{P}^{\text{combined}} = \left(\mathbf{P}^{\text{major}}, \mathbf{P}^{\text{minor}}\right) \qquad (13b)$$

For the minor case, one has to shift the scale vector by three entries to associate the roots with the corresponding [7] minor scales:

$$\mathbf{P}^{\text{mscale}} = \left(P_{10}^{\text{scale}}, \ldots, P_{12}^{\text{scale}}, P_1^{\text{scale}}, \ldots, P_9^{\text{scale}}\right) \quad (14)$$

The highest P_k^{combined} provides the key:

$$k^* = \arg\max_k P_k^{\text{combined}} \qquad (15)$$

The normalized confidence vector for the full key is

$$P_k^{\text{key}} = \frac{P_k^{\text{combined}}}{||\mathbf{P}^{\text{combined}}||_2}. \qquad (16)$$

The confidence for the selected key is $P_{k^*}^{\text{key}}$.

4. EVALUATION

4.1 Description of the Datasets

To evaluate our algorithm, we consider three datasets of classical music audio recordings. The first one (*Symph*) contains classical and romantic symphonies (all movements) from 11 composers, 115 tracks in total (cf. Tab. 1), taken from a dataset of NAXOS recordings.

Composer	*Symphonies No.*
Beethoven, L. v.	2, 3, 8
Brahms, J.	2, 3
Bruckner, A.	3, 4, 8
Dvořak, A.	5, 7
Haydn, J.	22, 29, 60, 103
Mendelssohn-B., F.	3, 5
Mozart, W. A.	35, 39, 40, 41
Schubert, F.	2, 3, 8
Schumann, R.	2, 4
Sibelius, J.	3, 4
Tchaikovsky, P. I.	5, 6

Table 1. Contents of *Symph* dataset.

The second dataset (*SMD*) is a selection from Saarland Music Data Western Music, collected in a collaboration of Saarland University and MPI Informatik Saarbrücken with Hochschule für Musik Saar [19]. The dataset contains music for solo, voice and piano, as well as chamber and orchestral music. We annotated the key for the 126 tracks showing clear tonality. [8]

[7] We identify sharp and flat chromas (D♭ =C♯): E.g., the tonic confidence for C♯ is multiplied with the confidence of level −5 for the D♭ major likelihood, and with the confidence of level +4 for the C♯ minor case.

[8] To this end, we skipped works of Bellini, Berg, Debussy, Donizetti, Martin, Poulenc and Ravel as well as the first and second movement of Faure's op. 15. From Schumann's works, op. 15 and 48 have been removed, since they are work cycles and do not constitute separated pieces in some way. For detailed information, see http://www.mpi-inf.mpg.de/resources/SMD. The key annotations are also available on this website.

Third, we test our method on a dataset of piano music recordings (*Pno*). This data was used to investigate key determination in the publications [10] and [14] and thus, allows for a direct comparison. The set contains 237 piano pieces by Bach, Brahms, Chopin and Shostakovich which are explicitly dedicated to a special key, as in the "well-tempered piano". Detailed information about the recordings can be found in [10].

Dataset	*Symph*	*SMD*	*Pno*	**tot.**
major global key	70 %	57 %	49 %	**56 %**
minor global key	30 %	43 %	51 %	**44 %**
major final chord	72 %	55 %	70 %	**67 %**
minor final chord	12 %	20 %	14 %	**15 %**
third-less fin. chord	16 %	25 %	15 %	**18 %**
fin. chord ≙ gl. key	70 %	64 %	53 %	**60 %**
fin. root ≙ gl. tonic	99 %	98 %	98 %	**99 %**

Table 2. The datasets' properties with respect to global key and final chord.

Table 2 shows some properties of the datasets. Final chord and global key coinicide for only 60 % of the pieces. However, the final chord's root matches the global tonic note almost always. Most of the mode deviations are picardian thirds (20 %), where a minor piece ends in the relative major chord (The opposite case is rare). The rest is caused by third-less final chord (18 %) like empty fifths (1 %) or unisono endings (17 %). 71 % end in a full triad, 11 % end in a fifth-less chord.

4.2 Experimental Results

We investigate the influence of the system parameters in a large study (Tab. 3). First, we show selected results for different energy threshold factors f_e, where a value of $f_e = 0.15\,\%$ seems to separate best silence from music frames. In the test of the weight exponents $\mathbf{m}^{(a)}$, the emphasis of the chord notes in $\mathbf{m}^{(2)}$ and the template derived from Temperley $\mathbf{m}^{(3)}$ perform best. To estimate the individual influence of root and scale estimation, we also run the algorithm with different weight exponents s in the decision process, where a slight preference of the scale confidence yields best results. For the size of the final frame set, a value of $N = 40$ frames corresponding to 2 seconds performs best. This value seems to balance the requirements for short chords (no failures caused by previous chords) with a sufficiently high reliability. With the low dynamic threshold f_e, we are also including reverb to a certain extent. Because of this, and of the frequent occurence of final ritardando in classical music, we do not have to worry about choosing a fixed small number of final frames N independently of the tempo.

To check the influence of the individual steps, we perform single runs without the multiplicative procedure in the tonic note estimation and in the diatonic scale estimation, respectively (block (E) in Tab. 3). From this, we can see that the multiplication in the

diatonic scale does not improve much. However, the multiplication in the tonic note estimation leads to a clear advancement, even when combined with a basic template matching with the Krumhansl profile (E4).

Parameters	*Symph*	*SMD*	*Pno*
(A) $\mathbf{m} = \mathbf{m}^{(3)}$, $N = 40$, $s = 0.8$			
$f_e = 0.10\,\%$	92.2 %	94.4 %	96.2 %
$\boldsymbol{f_e = 0.15\,\%}$	**92.2 %**	**93.7 %**	**97.0 %**
$f_e = 0.25\,\%$	92.2 %	92.9 %	96.6 %
$f_e = 0.50\,\%$	92.2 %	92.1 %	94.9 %
(B) $f_e = 0.15\,\%$, $N = 35$, $s = 0.75$			
$\mathbf{m} = \mathbf{m}^{(1)}$	88.7 %	92.1 %	94.1 %
$\mathbf{m} = \mathbf{m}^{(2)}$	**93.0 %**	**95.2 %**	**95.8 %**
$\mathbf{m} = \mathbf{m}^{(3)}$	92.2 %	93.7 %	96.6 %
$\mathbf{m} = \mathbf{m}^{(4)}$	89.6 %	91.3 %	95.4 %
(C) $f_e = 0.15\,\%$, $N = 35$, $\mathbf{m} = \mathbf{m}^{(3)}$			
$s = 0.5$	89.6 %	91.3 %	95.8 %
$\boldsymbol{s = 0.8}$	**92.2 %**	**93.7 %**	**97.0 %**
$s = 1.0$	92.2 %	93.7 %	96.2 %
$s = 1.2$	92.2 %	92.9 %	96.2 %
(D) $f_e = 0.15\,\%$, $s = 0.8$, $\mathbf{m} = \mathbf{m}^{(3)}$			
$N = 10$	90.4 %	89.7 %	93.2 %
$N = 30$	92.2 %	93.7 %	96.6 %
$\boldsymbol{N = 40}$	**92.2 %**	**93.7 %**	**97.0 %**
$N = 60$	90.4 %	90.5 %	96.6 %
(E) $f_e = 0.15\,\%$, $s = 0.8$, $\mathbf{m} = \mathbf{m}^{(2)}$, $N = 35$			
(E1)	83.5 %	80.2 %	82.3 %
(E2)	91.3 %	92.0 %	92.0 %
(E3)	76.5 %	62.7 %	55.3 %
(E4)	90.4 %	91.3 %	93.7 %

Table 3. Correct full key classification results for different parameter sets. We test the influence of the energy threshold factor f_e (A), the weight exponent set $\mathbf{m}^{(a)}$ (B), the root–scale weight exponent s (C), and the size of the final frame set N (D). The best results for each parameter are printed bold. In (E1), the multiplication in the tonic note estimation Eq. 8 is replaced by a simple maximum-picking. In (E2), the product Eq. 11 is replaced by a weighted sum. (E3) considers both these changes at the same time. For (E4), we use the full combined (major + parallel minor) Krumhansl template (non-diatonic entries nonzero) and again calculate a sum instead of a product.

Most of the parameters discussed here show important impact especially on one of the databases. In our interpretation, this is caused by different acoustic behavior (orchestra vs. piano) as well as properties of the music (cf. Tab. 6) and its temporal dimensions (symphonic vs. solo/chamber music). Individual error rates for two of the best parameter sets are shown in Tab. 4. Hereby, we emphasize the small number of fifths errors that arise frequently in other approaches. Third errors include all tonic note relations of minor and major thirds, including the relative key. Especially on symphonic data, identification of the correct tonic note is clearly more reliable than full key detection.

For our best parameter sets, we reach results slightly below the state-of-the-art [11, 12]. Taking into account that these algorithms are evaluated on music

Dataset	*Symph*	*SMD*	*Pno*
Correct full key	**92.2** %	**93.7** %	**97.0** %
Correct tonic note	98.3 %	96.0 %	97.5 %
Fifth errors	0.9 %	2.3 %	0.8 %
Third errors	0.9 %	1.6 %	1.7 %
Ø confidence	96.5 %	96.5 %	98.2 %
Correct full key	**93.0** %	**95.2** %	**95.8** %
Correct tonic note	100 %	96.8 %	97.0 %
Fifth errors	0 %	1.6 %	0.4 %
Third errors	0 %	1.6 %	2.5 %
Ø confidence	96.1 %	96.2 %	97.1 %

Table 4. Key extraction results for $f_e = 0.15\,\%$, $N = 40$, $s = 0.8$, $\mathbf{m} = \mathbf{m}^{(3)}$ (upper block) and $f_e = 0.15\,\%$, $N = 35$, $s = 0.75$, $\mathbf{m} = \mathbf{m}^{(2)}$ (lower block).

Dataset	*Symph*	*SMD*	*Pno*
Correct full key	**73.0** %	**71.2** %	**62.9** %
Correct tonic note	78.4 %	71.2 %	62.9 %
Fifth errors	9.0 %	12.8 %	13.1 %
Third errors	12.6 %	14.4 %	20.2 %

Table 5. Results of the `MIRtoolbox` key algorithm.

from one composer for one type of orchestration, our results may be comparable, since we considered a wide range of styles and instrumentations. On Bach's well-tempered piano, we reach 100 % full key identification for the upper settings in Tab. 4. On the *Pno* set, we almost reach the 98 % accuracy presented in [14]. To compare to a public algorithm, we run the key detection algorithm of `MIRtoolbox` from Univ. Jyväskylä [20] on our data, a common chroma- and template-based approach. Looking at the results in Tab. 5, we see that our method performs clearly better for detection of the full key and especially of the tonic note .

Epoch	1)	2)	3)	4)
No. in *Symph*	0	46	26	43
No. in *SMD*	11	49	20	46
No. in *Pno*	144	0	0	93
total No.	**155**	**95**	**46**	**185**
Correct full key	**98** %	**96** %	**96** %	**92** %
Correct tonic note	99 %	98 %	100 %	96 %

Table 6. Results by epoch: Baroque (1), Classical (2), Early Romantic (3) and Late Romantic / Modern (4) music. Parameters like in Tab. 4, lower block.

Last, we show the results by musical epoch in Tab. 6. To this end, we clustered the results by composer and aggregate music by Bach (Baroque), Haydn, Mozart and Beethoven (Classical), Schubert, Schumann and Mendelssohn (Early Romantic), and the rest (Late Romantic and Modern). We see the accuracy decreasing with composition time as expected because of the increase of tonal complexity during the centuries.

5. CONCLUSIONS

We presented a new rule- and theory-based approach to extract the key from classical music audio recordings. The method puts special emphasis on the final chord of the piece. After extracting chroma features, a number of final frames exceeding a dynamic threshold is selected. From this, the final chord's root is determined via a pairwise multiplication of fifth-related chroma values. From a full-piece chroma statistics, the system estimates the underlying diatonic scale. Finally, combining these results by multiplying corresponding confidence measures provides the full key.

For the evaluation, we considered three datasets on symphonic, mixed and solo piano music containing 478 recordings in total. We performed a parameter study and reach an average success rate of 95.0 % for full key detection and 97.7 % for tonic note detection for the best parameter settings. Hence, our results are in the range of most state-of-the-art approaches for automatic key detection, specialized on classical music.

Since our method provides the final chord's root with a high reliability, our approach can be combined with other chroma processing as well as machine learning techniques. So, this may be a helpful tool to facilitate renaming, browsing, and analyzing of classical music.

Acknowledgments

The authors thank M. Müller and S. Ewert for publishing their Chroma Toolbox. The research presented in this paper is part of the SyncGlobal project. SyncGlobal is a 2-year collaborative research project between Piranha Womex AG, Bach Technology GmbH, 4FriendsOnly AG and the Fraunhofer IDMT in Ilmenau, Germany. The project is co-financed by the Germany Ministry of Education and Research in the framework of the SME innovation program (FKZ 01/S11007).

6. REFERENCES

[1] E. Gómez and P. Herrera, "Estimating the tonality of polyphonic audio files: Cognitive versus machine learning modelling strategies," in *Proc. 5th Int. Conf. Music Inf. Retr. (ISMIR)*, 2004.

[2] B. Schuller and B. Gollan, "Music theoretic and perception-based features for audio key determination," *J. New Music Research*, vol. 41, no. 2, pp. 175–193, 2012.

[3] Ö. Izmirli, "An algorithm for audio key finding," in *Proc. 1st Annual Music Inf. Retr. Evaluation eXchange (MIREX '05)*, 2005.

[4] W. Chai and B. Vercoe, "Detection of key change in classical piano music," in *Proc. 6th Int. Conf. Music Inf. Retr. (ISMIR)*, 2005.

[5] H. Papadopoulos and G. Peeters, "Local key estimation from an audio signal relying on harmonic and metrical structures," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 20, no. 4, pp. 1297–1312, 2012.

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

[7] D. Temperley, *The Cognition of Basic Musical Structures*. MIT Press, 2001.

[8] E. Gómez, "Key estimation from polyphonic audio," in *Proc. 1st Annual Music Inf. Retr. Evaluation eXchange (MIREX)*, 2005.

[9] Ö. Izmirli, "Template based key finding from audio," in *Proc. Int. Comput. Music Conf. (ICMC)*, 2005.

[10] S. Pauws, "Musical key extraction from audio," in *Proc. 5th Int. Conf. Music Inf. Retr. (ISMIR)*, 2004.

[11] K. Noland and M. Sandler, "Influences of signal processing, tone profiles, and chord progressions on a model for estimating the musical key from audio," *Comput. Music J.*, vol. 33, no. 1, pp. 42–56, 2009.

[12] C.-H. Chuan and E. Chew, "Polyphonic audio key finding using the spiral array CEG algorithm," in *2005 IEEE Int. Conf. Multim. and Expo*, 2005.

[13] ——, "Fuzzy analysis in pitch class determination for polyphonic audio key finding," in *Proc. 6th Int. Conf. Music Inf. Retr. (ISMIR)*, 2005.

[14] S. van de Par, M. F. McKinney, and A. Redert, "Musical key extraction from audio using profile training," in *Proc. 7th Int. Conf. Music Inf. Retr. (ISMIR)*, 2006.

[15] M. A. Bartsch and G. H. Wakefield, "To catch a chorus: Using chroma-based representations for audio thumbnailing," in *Proc. IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, 2001.

[16] M. Müller, F. Kurth, and M. Clausen, "Audio matching via chroma-based statistical features," in *Proc. 6th Int. Conf. Music Inf. Retr. (ISMIR)*, 2005.

[17] M. Müller and S. Ewert, "Chroma toolbox: Matlab implementations for extracting variants of chroma-based audio features," in *Proc. 12th Int. Conf. Music Inf. Retr. (ISMIR)*, 2011.

[18] Z. Gárdonyi and H. Nordhoff, *Harmonik*. Möseler, 1990. [in German]

[19] M. Müller, V. Konz, W. Bogler, and V. Arifi-Müller, "Saarland music data," in *Proc. 12th Int. Conf. Music Inf. Retr. (ISMIR)*, 2011.

[20] O. Lartillot and P. Toiviainen, "A toolbox for musical feature extraction from audio," in *Proc. 8th Int. Conf. Music Inf. Retr. (ISMIR)*, 2007.