CHROMA-BASED SCALE MATCHING FOR AUDIO TONALITY ANALYSIS

Christof Weiß[1], Julian Habryka[2]

[1] *Fraunhofer Institute for Digital Media Technology Ilmenau, Germany*

[2] *Hochschule für Musik Würzburg, Germany*

Correspondence should be addressed to: christof.weiss@idmt.fraunhofer.de

Abstract: **We present an algorithm for visualizing the tonal characteristics of classical music audio recordings. The method is inspired by music theory concepts on harmony and intended to assist musicological research. Our system bases on chroma features that are extracted from the audio data in order to represent the dominant local pitch classes. Hence, the approach is widely insensitive to the orchestration. We use a coarse time resolution to account for the overall local pitch content rather than for single melody notes.**

The first visualization type presented in this paper serves to display the temporal evolution of local keys within a movement. This method is inspired by Gárdonyi's analysis technique regarding diatonic key relationships. We calculate local estimates for the underlying diatonic scales. These scale estimates are arranged according to a perfect fifths series to account for tonal similarity. Visualizing the local results over time provides an overview of the modulation structure of a piece.

The second method refers to the general scale type and the symmetries of the local pitch content. This technique is related to scale-based theories of harmony such as the analysis methods by Gárdonyi and Lendvai or the Tonfeld concept by Simon. Scale models such as the whole tone scale, the octatonic scale or the acoustic scale play an important role in impressionistic music or in Messiaen's compositions, among others. With our method, we compute the maximal likelihood for all transpositions of a scale to measure the occurrence of the respective scale type. These estimates are displayed over the course of the piece to show the locally prominent scales. This allows for an analysis of the formal aspects of tonality.

1. INTRODUCTION

With the high efficency of today's personal computers and smartphones together with cheap storage media and fast internet connections, a large amount of music data has become accessible. On commercial download platforms and streaming services, thousands of music recordings are easily available. There are several aspects how musicological research could benefit from such audio collections. From a quantiative point of view, it could be interesting to test the validity of musicological theories for a wide variety of works. Consider the analysis of a particular chord change. Often, such musical observations are exemplified with a small amount of case studies. To search after such phenomena on a corpus of pieces could provide additional statistical information. This also applies to particularly long works such as operas or symphonies where the analysis of large-scale structures can be very costly.

The traditional object for musicological analysis is the written score since it contains that fraction of a piece that is predetermined by the composer. However, there are some cases where audio-based analysis methods could be beneficial. For less known works, a score is not always available. Since manual transcription is time-consuming and automatic systems often do not provide proper transcriptions, audio-based analysis could be helpful. Furthermore, such methods could provide possibilities to study performance-related aspects which are not written in the score. One example is the shaping of tempo or loudness that could be analysed comparing different interpretations of a work. In early music, we find even more phenomena "outside" the score, e. g., common deviations of the performed pitch alterations from the score ("Musica ficta"). In such cases, audio-based analysis ensures to cope with the "sounding reality" of a performance, whereas a score representation would have to be interpreted using sets of rules.

In this paper, we propose a novel method for the automatic audio analysis regarding the tonal content of the music. Without reliable systems for automatic transcription and separation of musical parts, a distinction between harmonic chord notes and melodic phenomena is challenging. Because of such problems, we do not estimate the underlying chords but perform a scale-based method which incorporates all sounding pitches at a time into the analysis. This approach links to several musicological theories. We present two specifications of our algorithm. The first one estimates the local diatonic scale in order to visualize modulation structures for tonal music. Related methods for local key analysis of audio data have been presented in [1, 2]. With the second technique, we calculate likelihoods for the general scale types of the local tonal content.

The paper is structured as follows: In Section 2, we motivate our method from a musicological point of view. The technical details of the algorithm are presented in Section 3. In Section 4, we show analysis examples and discuss their musicological interpretation.

2. MUSICOLOGICAL FOUNDATIONS

The analysis technique presented in this work is based on the local scale material used in a composition. In Western music theory from the 19th century on, there are two ways of dealing with scales and their relation to tonality: On the one hand, chord progressions or contrapunctual voice leading phenomena are emphasized without focussing on the pitch class content. Understanding harmony that way, a scale is the consequence of the used chords. This assumption is crucial for theories in the tradition of Riemann. On the other hand, musicologists such as Fétis stated a close but ambivalent relationship between scale and tonality. [3, 4]. Hungarian music theorists (Gárdonyi, Simon and Lendvai) particularly consider scales as the result of the chord progressions and [5–7]. The tonal material created by chord progressions constitutes a link between Lendvai's theory and the Neo-Riemannian theory [8] where the pitch classes are represented in a "Tonnetz". Opposed to this, theorists from the 17th century state an influence of the available tonal scale on the chord progressions. Heinichen distinguishes between progressions that are independent from the scale and progressions resulting from the scale. The tonal disposition of the music depends on chord progressions formed by a scale an leads to the "regola dell' ottava" [9].

Besides such local observations, our visualization method allows for analysing the formal aspects of tonality. In Schenkerian analysis, a piece of music constitutes a sequence of scale degrees ("Stufen"). Hereby, the term "scale degree" is understood in an extended and more abstract way. It denotes no longer a single note or triad but consumes several harmonies that can be observed as autonomous chords themselves. These scale degrees are prolongated and connected to formal concepts such as sonata form or fugue [10]. Other theories emphasizing the structural function of tonality have been proposed [11]. Further large scale analyses of tonality focus on the music drama of Richard Wagner such as the analyses performed by Lorenz [12] that relate to our visualization method of local keys. The general idea of aggregating pitches to tonal structures has been picked up in newer theories such as the pitch class set theory [13].

3. VISUALIZATION METHOD

3.1. *Basic Features*

For an automatic analysis of tonality, we first have to extract pitch information from the audio content. Because of the challenges with automatic transcription, we confine ourselves to extract pitch classes. A common way to obtain such a description are chroma

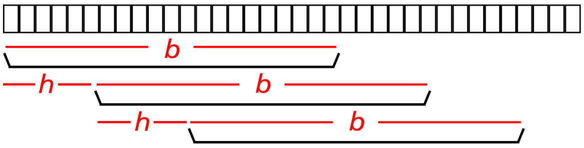

Figure 1: Segmentation of the chromagram with blocksize b and hopsize h. The boxes represent individual chroma vectors.

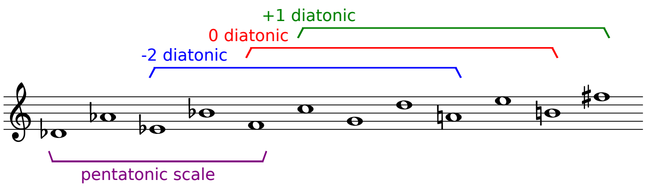

Figure 2: Fifth-relation of diatonic scales.

features. They have been shown to properly capture tonal characteristics of audio data for various applications such as audio thumbnailing [14], audio matching [15], or chord labelling [16, 17]. Using a public chroma extraction algorithm [18], we first derive frame-wise pitch features using a filter bank and then sum up all energies belonging to a pitch class. We obtain a chroma vector $\mathbf{c} = (c_0, c_1, \ldots, c_{11})^{\mathrm{T}} \in \mathbb{R}^N$ of dimension $N := 12$ for each time frame. The vector is normalized with respect to the ℓ_1 norm so that $||\mathbf{c}||_1 = 1$. In this basic representation, the chroma vector elements $c_0, c_1, \ldots, c_{11}$ describe the energy of the pitch classes C, C♯, …, B in the ordering of an ascending chromatic scale. To account for harmonic similarity of pitch classes, it turned out useful to re-order the chroma vector to a series of perfect fifths C, G, D …, F obtaining a vector $\mathbf{c}^{\text{fifth}}$ with

$$c_n^{\text{fifth}} = c_{(n \cdot 7 \mod 12)}. \qquad (1)$$

We compute the chroma vectors with a feature rate of $f = 10$ Hz. To analyse the local pitch content, we need larger analysis windows. Therefore, we group the chroma vectors to blocks of size b with a hopsize of h such as shown in Figure 1. A block of $b = 200$ feature frames corresponds to an analysis window of $b/f = 20$ s. For every block containing b chroma vectors $\mathbf{c}^1, \ldots, \mathbf{c}^b$, we compute a chroma histogram $\mathbf{g}$ by summing over all local vectors and normalize:

$$g_n = \sum_{i=1}^{b} c_n^i \;/\; ||\sum_{i=1}^{b} c_n^i||_1 \,. \qquad (2)$$

$\mathbf{g}$ is a vector of dimension $N = 12$ again. Equally, we obtain a fifth-ordered chroma histogram $\mathbf{g}^{\text{fifth}}$ for each block.

3.2. Diatonic Scale Estimation

The first analysis method proposed in this paper refers to the local tonality of the music. We therefore consider Gárdonyi's analysis method regarding the similarity of fifth-related keys [5]. When we re-order the chromatic scale to a series of perfect fifth related pitches, a diatonic scale corresponds to an excerpt of seven neighbours. In such a representation, two fifth-related diatonic scales such as the C major and the G major scale differ by only one note (in this example, F♯ instead of F). We use the nomenclature presented in [5] and denote the diatonic scales according to the number and type of accidentals necessary for notation. For example, a D major scale (2♯) is called "+2 diatonic", an A♭ major scale is called "-4 diatonic". In contrast, a pentatonic scale constitutes an excerpt of just five fifth-related pitches. In Figure 2, we show the systematics of such fifth relations.

To analyse modulations, we try to estimate the underlying diatonic scale of the local tonal content. For each analysis block, we multiply the entries of the chroma histogram $\mathbf{g}^{\text{fifth}}$ corresponding to the seven pitches of a diatonic scale. The absence of one or more scale notes results in a multiplication with a small number leading to a small likelihood for . This procedure is repeated 12 times calculating likelihoods D_k for all diatonic scales:

$$D_k = \prod_{i=0}^{11} \left(g_i^{\text{fifth}}\right)^{V_{(i+k) \bmod 12}}, \quad k \in \{-5, \ldots, 0, \ldots, +6\}. \qquad (3)$$

The index k in Equation 3 corresponds to the diatonic scale names introduced in Section 3.1. To account for the individual importance of the scale notes, the scale degrees are weighted with a set of exponents $\mathbf{V} = (V_0, V_1, \ldots, V_{11})$:

$$\mathbf{V} = (1.51 \;\; 2.97 \;\; 2.07 \;\; 1.38 \;\; 2.25 \;\; 2.64 \;\; 1.30 \;\; 0 \;\; 0 \;\; 0 \;\; 0 \;\; 0). \qquad (4)$$

The exponential weighting has shown to improve scale estimation in the context of global key detection [19]. The specific template $\mathbf{V}$ is derived from the Krumhansl tone profiles [20] combined with a weighting of the tonic triads (e. g., C major and A minor for the "0" diatonic) and has turned out to be useful in experiments. The off-scale notes are not considered and thus, exponentiated with zero. The proposed procedure corresponds to a multiplicative version of common template matching strategies. This has shown to be very useful for obtaining a robust scale estimation algorithm.

Finally, we normalize $\mathbf{D}$ with respect to the ℓ_2 norm to obtain the final diatonic scale likelihoods:

$$D_k^{\text{norm}} = D_k \;/\; ||\mathbf{D}||_2. \qquad (5)$$

With the normalization, we force the system to decide on the likeliest local diatonic scale (or combination of scales) even if all D_k are small. This showed to enhance the robustness of the system. As a drawback, the output for non-diatonic music is not always meaningful and thus, the preconditions for applying this analysis have to be considered carefully. For example, the presence of melodic or harmonic minor scales can produce misleading results. This will be discussed in Section 4.1.

3.3. Scale Type Estimation

To analyse whether the general scale type is diatonic or belongs to another scale model, we propose a second analysis method. Here, we do not estimate the transposition of one scale type but compare the likeliest transposition for different types to each other. To calculate the scale type estimates S_q, we replace the exponents $\mathbf{V}$ with binary templates $\mathbf{T}$:

$$S_q = \prod_{i=0}^{11} (g_i)^{T_{(i+q) \bmod 12}}, \quad q \in \{0, 1, \ldots, 11\}. \qquad (6)$$

Unlike Equation 3, we use the chromatic order of the chroma vector $\mathbf{g}$ here. The index q gives the transposition of the scale in semitones. The maximum likelihood S_q of all transpositions is used as scale type estimate:

$$S^{\max} = \max_q S_q. \qquad (7)$$

To investigate various concepts from music theory, we use templates for different scale models which are shown in Figure 3. Besides the diatonic and the pentatonic scale [1]

$$\mathbf{T}_{\text{Diatonic}} = (1\,0\,1\,0\,1\,1\,0\,1\,0\,1\,0\,1) \qquad (8)$$
$$\mathbf{T}_{\text{Pentatonic}} = (1\,0\,1\,0\,1\,0\,0\,1\,0\,1\,0\,0),$$

we use several symmetrical scale models that divide the octave into equal parts:

$$\mathbf{T}_{\text{Wholetone}} = (1\,0\,1\,0\,1\,0\,1\,0\,1\,0\,1\,0) \qquad (9)$$
$$\mathbf{T}_{\text{Octatonic}} = (1\,1\,0\,1\,1\,0\,1\,1\,0\,1\,1\,0)$$
$$\mathbf{T}_{\text{Hexatonic}} = (1\,1\,0\,0\,1\,1\,0\,0\,1\,1\,0\,0).$$

Additionally, we use templates for the so-called acoustic scale and for the complete chromatic scale:

$$\mathbf{T}_{\text{Acoustic}} = (1\,0\,1\,0\,1\,0\,1\,1\,0\,1\,1\,0) \qquad (10)$$
$$\mathbf{T}_{\text{Chromatic}} = (1\,1\,1\,1\,1\,1\,1\,1\,1\,1\,1\,1).$$

[1] Note that the entries of the template vectors $\mathbf{T}$ now refer to a chromatic pitch ordering.

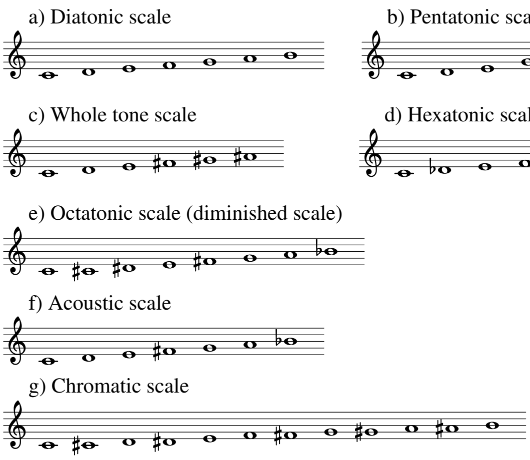

Figure 3: Scale models in basic form (without transposition).

For the symmetrical scales, several transposed versions are identical to each other. Since we pick the maximum likelihood of all transpositions, this does not constitute a problem. However, in order to compare the likelihoods for different scale types to each other, we have to account for the varying number of notes M in the scales:

$$M := \sum_{i=0}^{11} T_i. \qquad (11)$$

We therefore introduce a normalization factor dependending on the number of notes of a scale type and compute the final likelihoods as

$$S^{\text{norm}} = S^{\max}/\lambda \quad \text{with} \quad \lambda := (1/M)^M. \qquad (12)$$

Since the initial chroma histogram $\mathbf{g}$ is normalized to $||\mathbf{g}||_1 = 1$, we obtain a maximum value of $S^{\text{norm}} = 1$ if the scale notes have equal energy ($g_i = 1/M$) and the off-scale notes have zero values ($g_i = 0$). For a graphical visualization of these analyses, we show the scale likelihoods, indicated by the color, over a time axis. The results for each frame are displayed from the beginning of the analysis window until the beginning of the next window.

4. Examples

4.1. Analysis of Modulations

In this section, we show a number of different analyses and discuss the characteristics of our method on behalf of these plots. First, we show analyses of modulations using the diatonic scale estimation algorithm. Since this method is mainly inspired by Gárdonyi, we first look at J. S. Bach's Sinfonia in D major BWV 789 which is discussed in [5, p. 250]. Note that for such tonality analyses, the nomenclature of the diatonic scales is adapted to the global key (relative diatonic levels). For this example, the diatonic scale corresponding to D major (2♯) is denoted as "0 diatonic", the A major scale (3♯) as "+1" etc. Unlike Gárdonyi's approach, our method cannot discriminate between major and relative minor keys. The results are shown in Figure 4. In the analysis with fine time resolution (upper plot), we observe the general modulation structure with local keys at +1 in the beginning and -1 in the second half. At around 0:30 min, we see sudden jumps to the +2 level, in contrast to [5]. Here, a short modulation to F♯ minor is taking place (cadence in measure 14), introducing the pitches G♯ and D♯ (as part of the F♯ melodic minor scale). Using larger analysis windows (lower plot), these local alterations show less influence, leading to a sine-shaped structure very similar to [5]. From this observations, we see that the analysis results are meaningful, in general. Problems can arise from short-time local modulations as well as from non-diatonic scales such as the melodic minor scale. Hereby, the temporal resolution of the analysis windows plays a crucial role.

Next, we want to discuss visualizations composed in various musical styles. In Figure 5, we show an analysis of Palestrina's Kyrie from "Missa Papae Marcelli". The pitch classes used in this piece

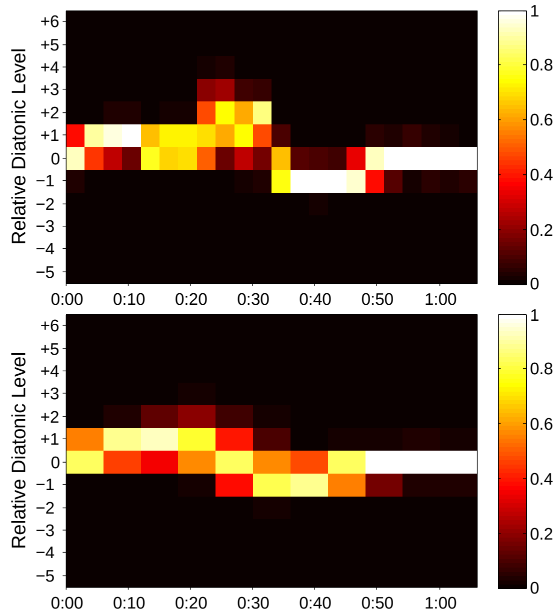

Figure 4: Diatonic scale visualization of J. S. Bach's Sinfonia No. 3 in D major, BWV 789 (0 $\widehat{=}$ 2♯). We compare two different time resolutions: blocksize 120, hopsize 30 frames (upper plot), blocksize $b = 240$ frames, hopsize $h = 60$ frames (lower plot).

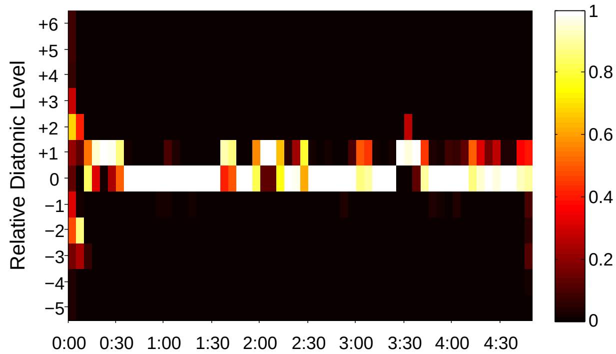

Figure 5: Diatonic scale visualization of the Kyrie from G. P. da Palestrina's "Missa Papae Marcelli" (0 $\widehat{=}$ no accidentals), $b = 100$, $h = 50$.

belong to one diatonic scale, to a great extent. Smaller deviations to the +1 level arise due to local voice leading phenomena, e. g., at 2:00 min where an F♯ is present. In contrast, the +1 scale detected at 3:30 min is caused by an ambiguity. Here, at the end of the "Christe eleison", a G major triad is sustained for a couple of seconds. This half-cadence is misinterpreted as a modulation to the +1 level. Further obscurities occur at the very beginning. After the initial silence, the voices come in gradually and thus, the full scale material is present after some seconds, firstly. On that account, scale detection is difficult here.

A contrasting example can be found in Figure 6 where we display the analysis of a piece by Lasso. Here, the preconditions of scale-based diatonic music are not fulfilled. Sometimes, we find a small number of chords belonging to one diatonic scale. However, most of the chord changes are based on chromatic movements of the voices such as the change from an F major to an A major chord on the first two syllables of "Militiae". On that account, the results at around 2:00 min show a rough chord estimation rather than a meaningful scale analysis. Overall, this example shows the limitations of this method for chromatic chord-based music.

In Figure 7, we show the analysis of a choral by J. S. Bach. The modulation to the +1 level in the repeated first phrase can

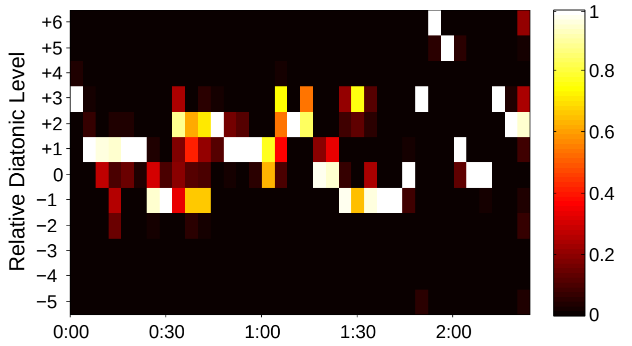

Figure 6: Diatonic scale visualization of No. 4 “Sibylla Cimmeria” from O. di Lasso’s “Prophetiae Sibyllarum” ($0 \mathrel{\widehat{=}} 1\flat$ according to the common notation, final chord is G major), $b = 80$, $h = 40$.

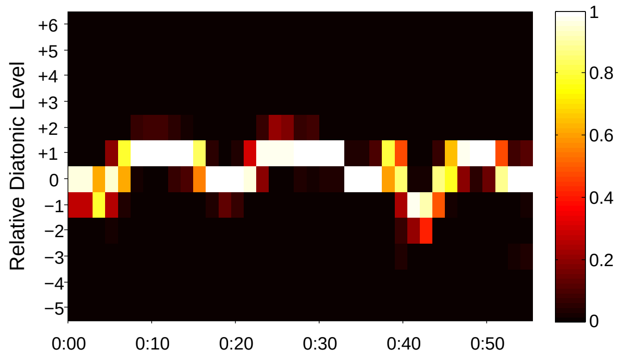

Figure 7: Diatonic scale visualization of the Choral No. 22 “Durch dein Gefängnis” from J. S. Bach’s Johannespassion BWV 245 ($0 \mathrel{\widehat{=}} 4\sharp$), $b + 42$, $h = 15$.

be recognized well. The deviation to the minus region at around 0:40 min may arise from the flat alterations at the chromatic elaboration of “unsere Knechtschaft”. The +1 level at 0:50 min is a misinterpretation of a long dominant chord, again.

Looking at Beethoven’s sonata Op. 14, No. 2 in G major (Figure 8), we observe the modulation shape of the classical sonata form, with some interesting details. In the modulation from the first to the second theme at 0:20 min and repeated at 2:00 min, we see even a small +2 area where we only expect +1. Indeed, the piece modulates to A major for a short time, indicated by the presence of the pitch G♯. In the development (3:30-5:00 min), we find keys in the minus region, in particular.

As the last example, we discuss R. Wagners overture from the opera “Die Meistersinger von Nürnberg” (Figure 9). Interestingly, we find a structure that roughly corresponds to the tonal shape of a sonata form. There are +1 regions in the first part, a highly modulation middle part, as well as an ending mainly based on the 0 level. The modulation path at around 4:00 min is remarkable, in particular. Here, our analysis indicates a modulation around the circle of fifths. After a short period at +4 and +3, the tonal structure slowly leads back to the tonic which is emphasized by a three minute coda based on the 0 level. For this particular example, the proposed method seems to provide an appropriate analysis. This has to be tested for other works by Wagner. For larger structures such as the tonal shape of a whole opera, it could be interesting to compare the output of our algorithm to the analyses in [12].

4.2. Analysis of Scale Types

Because of the limitations of the diatonic scale analysis for non-diatonic music, we now present examples for our second method dealing with the general scale type. Non-diatonic scale types such as symmetrical scales have become important from the late romantic period on. In particular, composers from the impressionistic period used pentatonic and whole tone scales, among others.

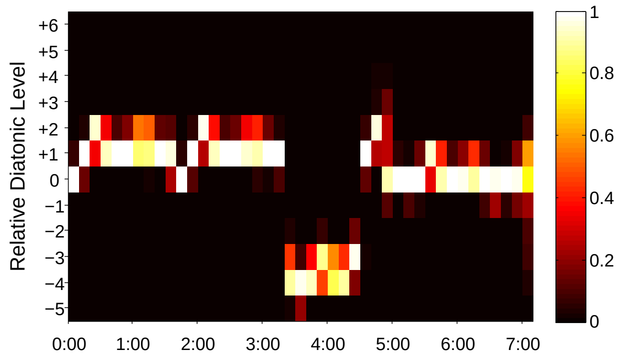

Figure 8: Diatonic scale visualization of L. v. Beethoven’s Sonata Op. 14, No. 2, 1st movement in G major ($0 \mathrel{\widehat{=}} 1\sharp$), $b = 150$, $h = 60$.

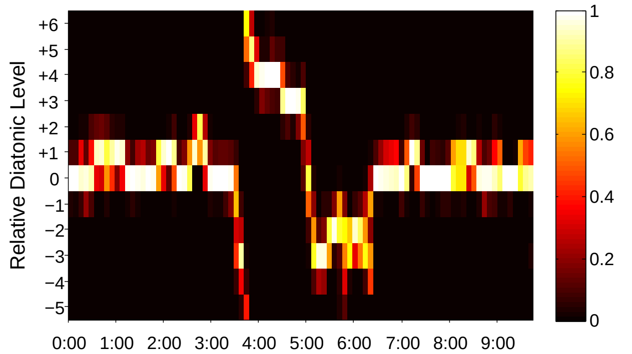

Figure 9: Diatonic scale visualization of the Overture from R. Wagner’s “Meistersinger von Nürnberg” in C major ($0 \mathrel{\widehat{=}}$ no accidentals), $b = 150$, $h = 65$.

In Figure 10, we show the analysis of the prelude “Voiles” by Debussy. The likelihoods for the different scale types are indicated by the color, with a logarithmic color axis. In the first part until 1:50 min, the whole tone scale is dominating. This corresponds to the score that contains only the pitches of such a scale for the first 41 measures. In contrast, the middle part is constructed from a pentatonic scale, while the ending returns to the whole tone scale.

In the parts with dominating whole tone scale, we see some contributions to the likelihood for the acoustic scale as well. This is not very surprising, since the acoustic scale contains five out of the six notes of a whole tone scale. Therefore, energies stemming from upper harmonics or effects such as resonances in the piano may lead to a non-zero likelihood for acoustic scale. A similar behaviour is observed comparing the pentatonic and the diatonic scales. Since the pentatonic scale pitches are a subset of the diatonic scale, small energy deviations in the silent chroma bands may produce a contribution to the diatonic scale likelihood, even if only the notes of a pentatonic scale are played.

Such effects may cause even more problems when dealing with complex orchestral music. We therefore show an analysis computed on a MIDI representation of Debussy’s orchestra piece “La Mer” and compare this analysis to the results of the audio-based method for the same piece (Figure 11). For the MIDI analysis, the pitches are weighted with their velocity values and aggregated to pitch classes in order to build chroma-like features. On these features, we perform our analysis as described in Section 3.3. Note that the time axes are not synchronized in a musically meaningful way so that the time positions are related only roughly.

Comparing the results for the two representations, we observe a very similar structure. Looking at the details, we find some smaller deviations. In the ending sections (8:00 min to 9:00 min), we find some “noisy” contributions to the likelihood of a chromatic scale for the audio analysis. In the beginning at around 0:30 min, we find more substantial differences. The reasons for the high likelihood *a)* MIDI-based

b) Audio-based

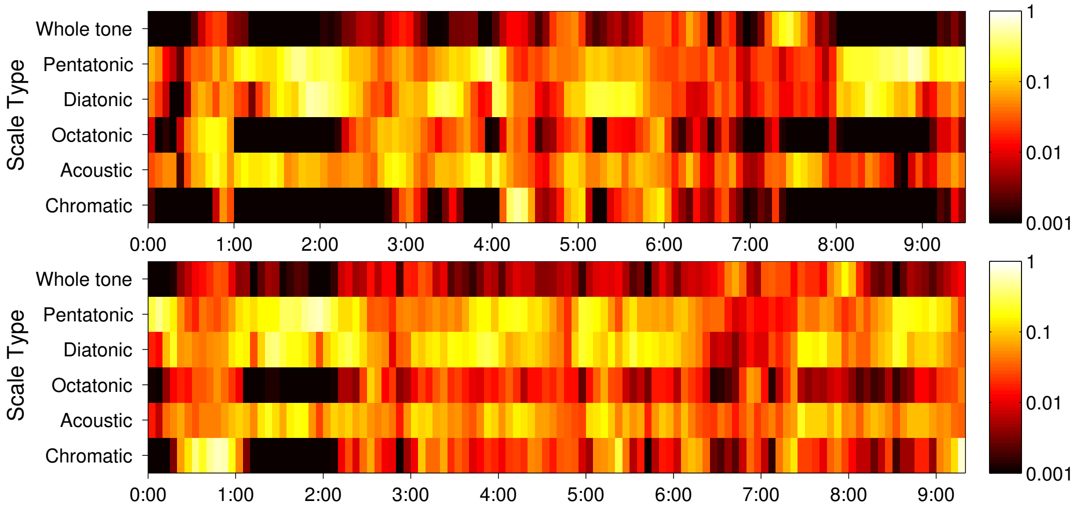

Figure 11: Scale type visualization of the first movement "De l'aube à midi sur la mer" from C. Debussy's "La Mer" from MIDI data (a) and audio data (b), $b = 200$, $h = 50$. The recording used for the audio analysis was played the Belgian Radio and Television Philharmonic Orchestra under A. Rahbari (Naxos 1997).

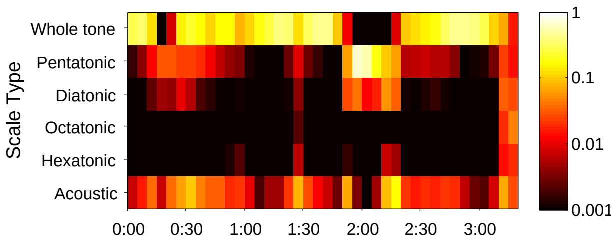

Figure 10: Scale type visualization of No. 2 "Voiles" from the first book of C. Debussy's "Préludes", $b = 100$, $h = 50$.

of the chromatic scale in the audio analysis are not very clear. However, the audio-based method may have advantages as well. In [5], the horn theme in octaves is proposed as an example for acoustic tonality. This motif comes first at around 1:45 (score letter 3) where it can be detected better in the audio-based analysis. The repetitions of that motif at around 3:00 min (letter 5) and around 4:00 min (short before letter 8) can be seen in the audio visualization as well but are more clear in the MIDI analysis. In general, we find a lot of pentatonic scales as well as some diatonic and acoustic scales. In contrast, there is almost no prominent whole tone scale. This may result from the fact that this scale is less used to generate simultaneous sounds than, for example, the pentatonic scale.

Next, we want to test our method on a piece containing atonal harmonies as well as parts dominated by percussion instruments. We therefore show an analysis of Stravinsky's famous "Le Sacre du Printemps" (Figure 12). As expected, we find high likelihoods for the chromatic scale in several parts. In particular, atonal and percussive phenomena may be present at the end of both parts. A contrasting section can be found at the begin of the "Spring Rounds" movement (between 8:00 min and 10:00 min in the first part). Here, we find a pitch class material related to the E♭ dorian scale. This is one of the few sections of the piece that is notated with a key signature (5♭). At some positions, acoustic tonality can be observed. A weak example for such an observation is in the first part at 6:30 min (score letter 32), in accordance to [5]. In the second part, there is a very clear indication for an acoustic scale on the beginning of the "Ritual Action of the Ancestors", at around 11:00 min (score letter 129). Here, we find a high likelihood for the acoustic scale, without ambiguities to other scales. A look into the score confirms this assumption. Another indication for an acoustic tonal structure is found in the second part at around 3:00 min (score letter 87). Analysing the score leads to a similar result: The pitch

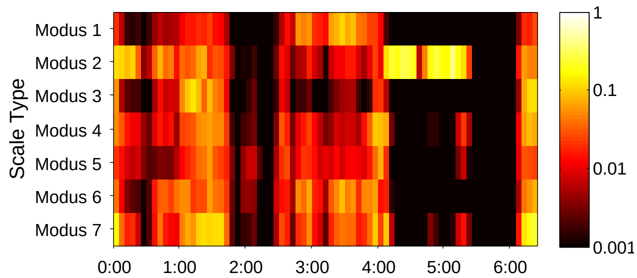

Figure 13: Visualization of Messiaen"s modes for No. 1 "La vierge et l'enfant" from O. Messiaen's "La Nativité du Seigneur" for organ, $b = 150$, $h = 50$.

classes of an acoustic scale on B♭ dominate this passage, with one additional pitch class (D♭). Altogether, we see that this method can be helpful to get an overview over the tonal structure of large pieces. For pieces where different concepts of tonality are combined, our approach can provide hints to particular tonal phenomena.

The scale type analysis method presented in this paper may be suitable to analyse the music of Messiaen since he proposes a set of symmetrical scales called "modes of limited transposition" to be crucial for his compositional approach [5, 21]. Some of the modes have been introduced yet. The first mode corresponds to the whole tone scale and the second mode is the octatonic scale. The third mode relates to the hexatonic scale since it shows a periodicity in major third distance. The other three modes are periodic with the tritone interval [21]. As an example, we perform an analysis of an organ piece from "La Nativité du Seigneur", shown in Figure 13. We find a clearly octatonic section in the last part of the piece between 4:00 min and 5:20 min (Modus 2). For the presence of other modes, we cannot see any clear indications. One reason for this may be the acoustic behaviour of the organ: In this recording, aliquot registers enhancing particular harmonics of the played pitches have a strong influence on the sound. This may lead to deviations of the chroma features from the notated pitch classes. At the end of the piece (between 5:30 min and 6:00 min), no scale type seems to be present. Here, only monophonic pitches are played after each other in a slow tempo. Overall, the analysis of scale types was not satisfying for this piece even though several modes are present in the score. To investigate the problem of such analyses, further studies including MIDI representations of the pieces should be conducted.

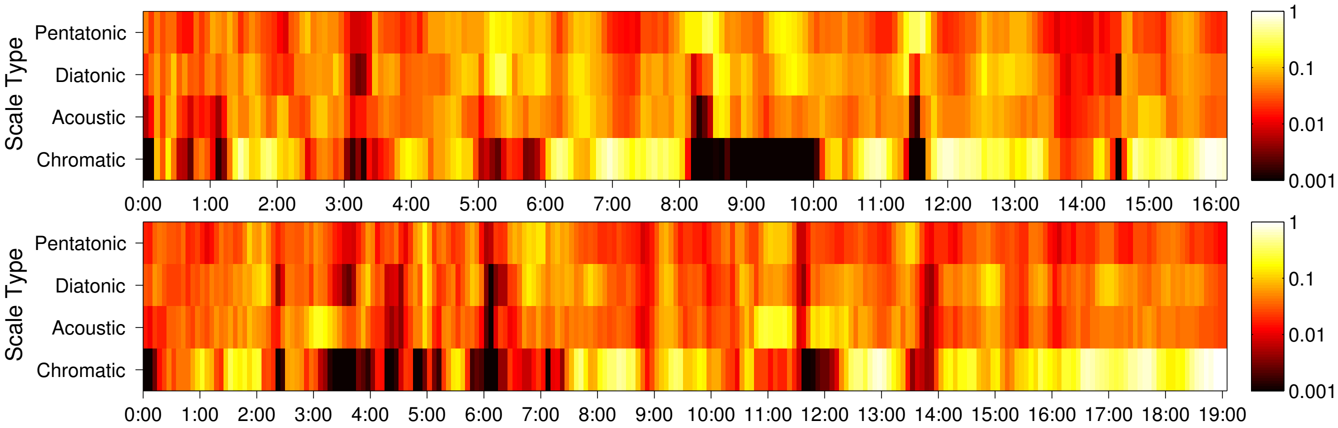

Figure 12: Scale type visualization of I. Stravinsky's "Le Sacre du Printemps" first part (upper plot) and second part (lower plot), $b = 200$, $h = 50$. We used a recording of the Belgian Radio and Television Philharmonic Orchestra, conducted by A. Rahbari (Naxos 1991).

5. Conclusion

We have presented a novel approach for automatic analysis of audio data with respect to the tonal and harmonic properties of the music. The method is based on a chroma feature representation of the audio data. The features are grouped into analysis windows of variable length. We have presented two post-processing methods inspired by several musicological theories. With the first method, the likelihoods for the twelve diatonic scales are estimated over time. This has been tested on music examples from several historical eras. Visualizing the results provides an overview of the modulation structure in a musically meaningful way, under the condition that the tonality of the music is dominated by diatonic scales. With the second analysis technique, we estimate the general scale type of the local tonal content. To do this, we match the chroma vectors to binary templates of the scale types and extract the maximum likelihood for all transpositions of a scale model. We have shown several examples from the 20th century where we identified fifth-based scale types (pentatonic, diatonic), symmetrical models (whole tone scale, octatonic), and acoustic tonality successfully. For atonal passages, an enhanced likelihood for the chromatic scale could be detected.

If only a fraction of the scale notes is presented locally, the proposed analysis method can lead to problems and ambiguities. On that account, the size and position and the analysis windows plays a crucial role. Information about the musical time from automatic beat tracking or a manual annotation of the measure positions could improve analysis quality in future work. This would also be helpful to link score positions to the analysis frames in an exact and reliable way. When comparing the audio-based analysis to results computed on a MIDI representation of the same piece, we have found only slight deviations pointing to a certain robustness against acoustical artifacts and noise. Altogether, both methods provide musically meaningful visualizations that can help to get an overview of the tonal structure of a musical piece.

References

[1] Ö. Izmirli: *Localized key finding from audio using nonnegative matrix factorization for segmentation*. In *Proceedings of the 8th International Conference on Music Information Retrieval Conference (ISMIR)*. 2007.

[2] C. S. Sapp: *Visual Hierarchical Key Analysis*. In *ACM Computers in Entertainment*, volume 3(4):1–19, 2005.

[3] C. Dahlhaus: *Untersuchungen über die Entstehung der harmonischen Tonalität*. Bärenreiter, Kassel, 1967.

[4] B. Hyer: *Tonality*. In *The Cambridge History of Western Music Theory*, pages 726–752. Cambridge University Press, Cambride, 2002.

[5] Z. Gárdonyi and H. Nordhoff: *Harmonik*. Möseler, Wolfenbüttel, 1990.

[6] B. Haas: *Die neue Tonalität von Schubert bis Webern: Hören und Analysieren nach Albert Simon*. Veröffentlichungen zur Musikforschung. F. Noetzel, Wilhelmshaven, 2004.

[7] E. Lendvai: *Symmetrien in der Musik: Einführung in die musikalische Semantik*. Universal Edition, Wien, 1995.

[8] R. L. Cohn: *Audacious Euphony*. Oxford University Press, Oxford, 2012.

[9] J. D. Heinichen: *Der General-Bass in der Composition*, volume 2. Dresden, 1728.

[10] H. Schenker: *Harmonielehre: Neue musikalische Theorien und Phantasien I*. Stuttgart and Berlin, 1906.

[11] D. W. Bernstein: *Nineteenth-century harmonic theory: the Austro-German legacy*. In *The Cambridge History of Western Music Theory*, pages 778–811. Cambridge University Press, Cambride, 2002.

[12] A. Lorenz: *Das Geheimnis der Form bei Richard Wagner: Der musikalische Aufbau des Bühnenfestspieles "Der Ring des Nibelungen"*. Berlin, 1924.

[13] A. Forte: *The Structure of Atonal Music*. Yale University Press, New Haven and London, 1973.

[14] M. A. Bartsch and G. H. Wakefield: *To catch a chorus: Using chroma-based representations for audio thumbnailing*. In *Proceedings Workshop on Applications of Signal Processing (WASPAA)*, pages 15–18. 2001.

[15] M. Müller, Frank Kurth, and Michael Clausen: *Chroma-Based Statistical Audio Features for Audio Matching*. In *Proceedings Workshop on Applications of Signal Processing (WASPAA)*, pages 275–278. 2005.

[16] T. Fujishima: *Realtime chord recognition of musical sound: A system using common lisp music*. In *Proceedings of the International Computer Music Conference 1999*, volume 1999, pages 464–467. 1999.

[17] A. Sheh and D. P. W. Ellis: *Chord Segmentation and Recognition using EM-Trained Hidden Markov Models*. In *Proceedings of the 4th International Conference on Music Information Retrieval (ISMIR)*. 2003.

[18] M. Müller and S. Ewert: *Chroma Toolbox: MATLAB implementations for extracting variants of chroma-based audio features*. In *Proceedings of the 12th International Society for Music Information Retrieval Conference (ISMIR)*, pages 215–220. 2011.

[19] C. Weiss: *Global Key Extraction from Classical Music Audio Recordings Based on the Final Chord*. In *Proceedings of the 10th Sound and Music Computing Conference (SMC)*. 2013.

[20] C. L. Krumhansl: *Cognitive Foundations of Musical Pitch*. Oxford psychology series. Oxford University Press, 1990.

[21] O. Messiaen: *The technique of my musical language*. Leduc, Paris, 1944.