

TONALITY VISUALIZATION OF POLYPHONIC AUDIO

Emilia Gómez

Jordi Bonada

Music Technology Group, Institut Universitari de l'Audiovisual

Universitat Pompeu Fabra

{emilia.gomez,jordi.bonada}@iua.upf.es

<http://www.iua.upf.edu/mtg>

ABSTRACT

This paper presents a tool to visualize the tonal content of polyphonic audio signals. After a brief introduction to the problem of tonal analysis, we present different views that can help to analyze the tonal content of a piece of music in audio format and to investigate techniques for chord and key estimation and tonal similarity.

1. INTRODUCTION

A tonal music piece is usually described as being in a particular key, as for instance Mozart's Sonata in F major KV 533-494. This is usually true for classical music, where the key is included in the title of the piece or as editorial information. In other music genres, key information is usually not available or unknown.

Labelling a piece with a single key is often poor in terms of tonal description. A musical piece rarely maintains a single tonality throughout its duration. There are also some pieces where there is no clear tonality, and the tonal center is constantly moving. The instantaneous evolution of the tonality of a piece and its strength can give a more powerful tonal description of it. Applications of this description include structural description, genre classification and music similarity.

We present here a tool intended to visualize the tonal content of a piece of music by analyzing audio recordings. This system is inspired in the work by Craig Stuart Sapp on harmonic visualization [11] and other approaches for tonal visualization of MIDI representations [2,6,13]. We extend these ideas to the analysis of audio signals. Working directly with audio avoids the need of score transcriptions, being suitable for pieces where the score is unknown, as it often occurs. The tool introduces a set of additional visualizations which are specific to the analysis of audio.

This work has been carried out in the context of a system for automatic tonal description of audio [4,5]. Our main objective is to provide means of visually analyze audio features, tonality models, distance measures and analysis parameters (e.g. the length of the sliding window used for key tracking). Other intended applications for this tool include musical analysis and similarity. One example is to use tonal similarity to identify versions of the same song assuming that the tonal contour is kept.

After a brief introduction to the topic of tonal analysis and to our specific application context, we present the proposed tool as well as some examples of

the analysis obtained from different pieces which illustrate its utility.

2. TONAL ANALYSIS

2.1. Key estimation from MIDI

There has been much research on identifying the global tonality of a certain piece of music given its score (i.e. its MIDI representation). One of the most popular methods for determining the key in a region of music is the Krumhansl-Schmuckler key-finding algorithm based on probe-tone ratings generated from experimental results [7,11,12]. We refer to [1] for an extensive review of other approaches for key estimation and modelling.

Less research has been devoted to locate modulations. There have been some attempts, but it still remains a difficult task and quite hard to evaluate. The first problem to solve when trying to segment a piece according to its key is how to correctly identify regions of stable key centres and regions of modulations. Some approaches apply a sliding analysis window to the piece to generate a set of localized key estimations [3]. This set of measures gives a good description of the key evolution of the piece, but calls for the setting of a suitable window size, which normally depends on the tempo, musical style and the piece itself.

2.2. Key estimation from audio

Tonal description from audio becomes necessary when dealing with musical collections where the score is not available. When dealing with audio instead of MIDI, there is an additional difficulty: the pitches are unknown and it is quite difficult to estimate them. Given the current state of the art in automatic transcription, it is not achievable to extract a reliable score from an audio signal, especially for music with complex polyphony and percussive sounds.

There has been few research devoted to estimate the tonality from audio recordings, although it has become a very active topic within the last few years in the context of automatic audio description and music information retrieval [4,5,9,10,14]. Most of the approaches described in the literature are based on the extraction of a set of audio features and the employ of diverse tonal models to estimate the key of a piece from these audio descriptors. This scheme is represented in Figure 1.

The approach employed in this paper is based in this scheme and it is described in details in [4,5]. Some of

the aspects of the approach are illustrated in the following sections.

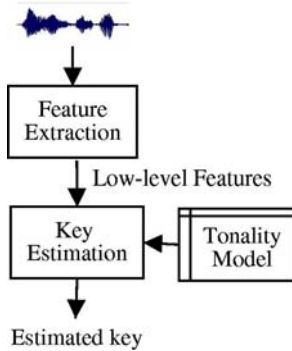


Figure 1. Overall diagram for key estimation from audio.

3. TONALITY VISUALIZATION

We present here a set of views which are related to the different blocks of the system presented in Figure 1. Color versions of this paper, as well as high-resolution color versions of the images and some animations are found at

<http://www.iaa.upf.es/~egomez/TonalDescription/GomezBonada-ICMC2005.html>

3.1. Feature Extraction

3.1.1. *TuningGram*

The first step of the feature extraction block in Figure 1 is to compute the reference frequency used to tune the piece. We obtain the evolution of this reference frequency, with respect to the standard A 440 Hz, as proposed in [14]. Figure 2 presents the evolution of the tuning frequency of an audio excerpt, measured in cents, and compared to its average over the excerpt (straight line). The frequency curve is smoothed using a moving average filter, which can be controlled through the user interface.

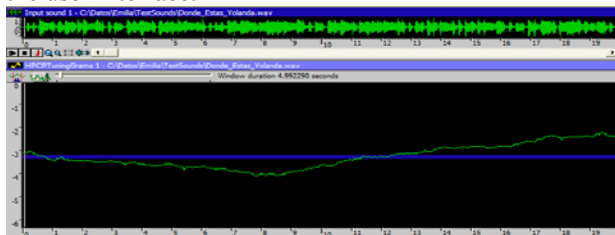


Figure 2. *TuningGram*.

3.1.2. *HPCPGram*

The second view we present is a display of the low-level audio features used for tonality estimation. The features used in our approach are called HPCP (*Harmonic Pitch Class Profile*) and represent the relative intensity of each bin in one octave (considering an equal-tempered scale tuned to the reference frequency computed before). 12 bins are equivalent to semitone resolution, 24 to quarter of tone, 36 to a third of semitone, etc. HPCP is computed using the magnitude of the spectral peaks that are located within a

certain frequency band, considered as the most significant frequencies carrying harmonic properties. A weight is introduced into the computation to get into account differences in tuning and inharmonicity. The HPCP vector is finally normalized for each analysis frame in order to discard energy information. Details are found in [3].

The *HPCPGram* presents the temporal evolution of the HPCP over the audio signal, and it is used to investigate complementary low-level instantaneous features which are relevant for tonal description of audio. An example is shown in Figure 3.

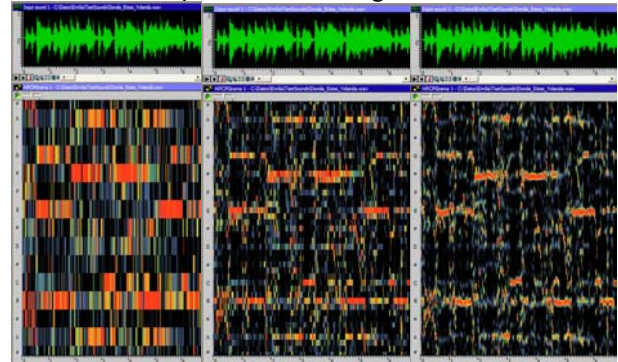


Figure 3. *HPCPGram* using 12 (left), 36 (center) and 120 bins (right). The vertical axe represents the pitch class (on a chromatic scale from A to G#).

3.2. Key Estimation

After computing a set of relevant low-level features from the audio signal, the system applies a tonal model to them in order to estimate the key. Several views are proposed in this second stage of the algorithm.

3.2.1. *Key Correlation*

This view displays the correlation of the average HPCP in a given temporal window with the possible minor and major keys, using the same frequency resolution than for the *HPCPGram*. This idea was considered by Sapp with a representation of the ‘clarity’ of the key [11]. However, it only represented the relationship within the second most correlated key without explicitly indicating both keys. In our approach, we consider the correlation of the average HPCP with a set of tonal profiles. These tonal profiles are derived from the probe-tone profiles proposed by Krumhansl-Schmuckler, which have been adapted to polyphonic audio (see [4] for a detailed explanation). This diagram compares this key estimation in a certain temporal window with the global key estimation. This is illustrated in Figure 4. The window size is an algorithm parameter which can be changed through the user interface. This view is used for comparing different tonal profiles (as [12]) and distance measures (e.g. correlation vs euclidean distance). It is also possible to order the keys, and hence the coloring scheme, according to the circle of fifths, as shown in Figure 4 (bottom).

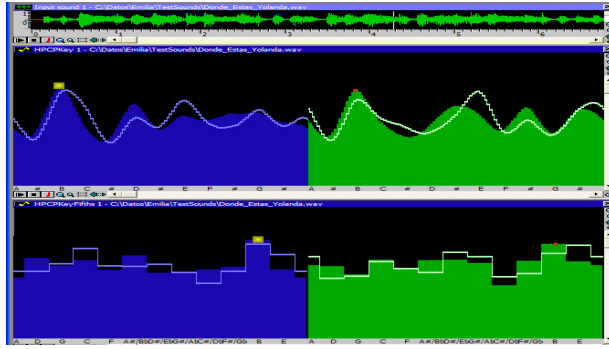


Figure 4. *Key Correlation* with major (left) and minor (right) keys. The horizontal axis represents the pitch class in chromatic scale from A to G# (top) and in the circle of fifths from A to E (bottom). Filled bars indicate the estimation on the current instant, while non filled bars represent the global estimation. The square represents the maximum correlation value, equivalent to the estimated key (B Major in this example).

We can also plot key correlation values using a rectangular representation, where which the points for the 24 major and minor keys are located on the surface of a torus (as in [7] pp. 46). Figure 5 shows an example.

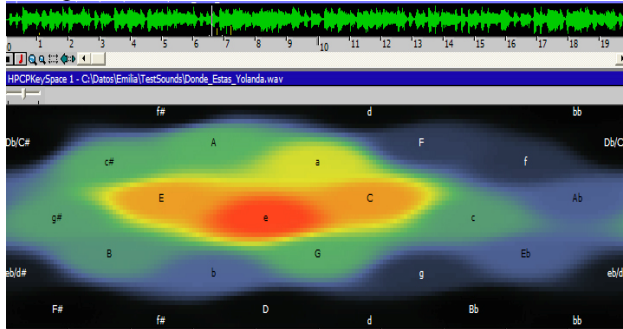


Figure 5. *Key Correlation* in the surface of a torus.

3.2.2. KeyGram

The previous visualization does not include information about the temporal evolution of the tonality, which is very relevant. The *KeyGram*, shown in Figure 6, represents the temporal evolution of the *KeyCorrelation*.

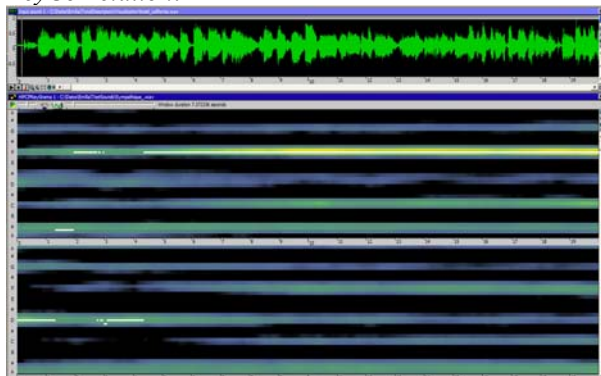


Figure 6. *KeyGram*. Instantaneous correlation with major (top) and minor (bottom) keys. White color indicates the highest correlation.

This evolution can also be displayed in the surface of a torus (as in Figure 5), defining a trajectory for tonality evolution. An example is showed in Figure 7. As for

KeyCorrelation, the size of the sliding window used for key estimation can be changed through the user interface, providing a way to navigate through different temporal scopes. This view is used to study the evolution of the key and its strength along the piece, to determine the most suited window size for each piece, and to investigate different distance measures and tonal profiles.

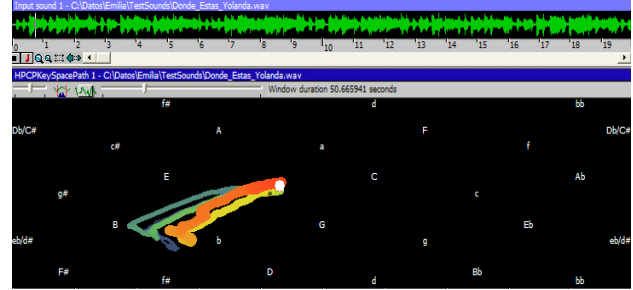


Figure 7. *KeyGram* in the surface of a torus.

3.2.3. KeyScope

This view introduces a multi-resolution representation of the estimated tonality, in order to visualize the evolution of the key (represented in the *KeyGram*) within different temporal scopes. To display tonal data in a compact visual manner, each key is mapped to a different color. We added to the original color representation in [11] the distinction between major and minor keys by assigning different brightness (brighter for minor keys). The colors are represented in Figure 8. It is also possible to order the keys, and hence the coloring scheme, according to the circle of fifths.

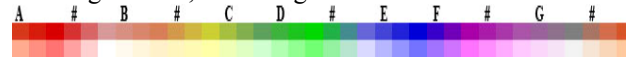


Figure 8. Colors used to represent Major (top) and minor (down) tonalities.

This diagram displays the tonality using different temporal scales. Each scale is related to the number of equal-duration segments in which the audio signal is divided to estimate its tonality. This way, the top of the diagram shows the overall tonality, the middle scales identify the main key areas present in the piece and the bottom scale displays chords. An example is presented in Figure 9.

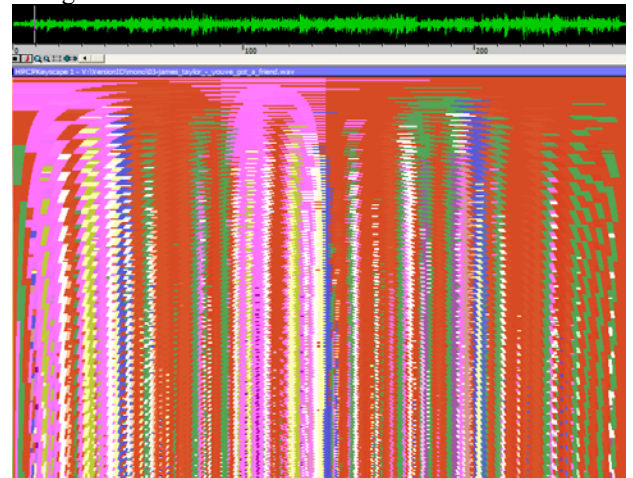


Figure 9. *KeyScope* from the song 'You've got a friend' by James Taylor. The global tonality is A major.

3.2.4. Tonal Contour

Pitch intervals are preferred to absolute pitch in melodic retrieval and similarity applications, given the assumption that melodic perception is invariant to transposition. We extend this idea to tonality, after observing that different versions of the same piece share the same tonal evolution but can be transposed. This is usually made to adapt the song to a singer or instrument tessitura. This view provides a relative representation of the key evolution. The distance between consecutives tonalities is measured in the circle of fifths: a transition from C major to F major is represented by -1, a transition from C major to A minor by 0, a transition from C major to D major by +2, etc. Figure 10 represents the tonal contour of the beginning of the song 'Imagine' by John Lennon, using a sliding window of 1 s. The estimated chord is moving from C major to F major, giving contour values equal to 0 and -1. A drawback of this contour representation is that relative major and minor keys are considered equivalent, which results in these modulations not being shown. This problem can be solved by using a representation of the torus *KeyGram* display shown in Figure 7 centered in the graphic.

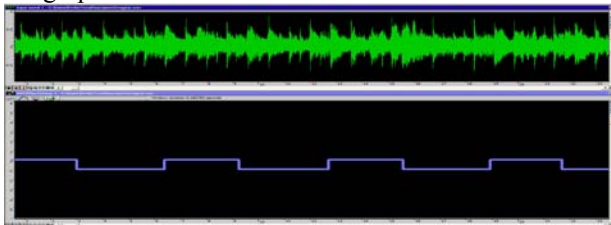


Figure 10. *Tonal Contour*

The size of the sliding window can be adjusted, providing a way to navigate through different temporal scopes. It is possible then to visualize, for instance, chord progressions or key progressions. The goal of this view is to analyze the relative evolution of the key as a valid descriptor for tonal similarity between pieces.

4. CONCLUSIONS AND FUTURE WORK

We have presented different ways of visualizing the tonal content of a piece of music by analyzing audio recordings. These diverse views can be combined in varied ways depending on the user needs. We offer also the possibility of visualizing two different pieces simultaneously, which can serve to study tonal similarity and define distances between pieces based on tonal features. One possible application that we plan to evaluate is the recognition of different versions of the same song assuming that the overall tonal progression is kept, using tonal contour description.

5. ACKNOWLEDGMENTS

This research has been partially funded by the EU-FP6-IST-507142 SIMAC project and the EU-FP6-eContent HARMOS project. The authors would like to thank Craig Stuart Sapp and the ICMC 2005 reviewers for their useful advices.

6. REFERENCES

- [1] Chew, E. 'Towards a Mathematical Model of Tonality', Ph.D. dissertation. Operations Research Center, MIT. Cambridge, MA, 2000.
- [2] Chew, E. and François, A. 'MuS.A.R.T - Music on the Spiral Array. Real-Time', *Proceedings of the ACM Multimedia Conference*, 2003.
- [3] Chew, E. 'Messiaen's Regard IV: Automatic Segmentation using the Spiral Array', *Proceedings of Sound and Music Computing Conference*, 2004.
- [4] Gómez, E. 'Tonal Description of polyphonic audio for music content processing', *INFORMS Journal on Computing, Special Cluster on Computation in Music*. Accepted for publication, 2004.
- [5] Gómez, E. and Herrera, P. 'Estimating The Tonality Of Polyphonic Audio Files: Cognitive Versus Machine Learning Modelling Strategies', *Proceedings of International Conference on Music Information Retrieval*, 2004.
- [6] Janata P., Birk, J., Van Horn, J. D., Leman, M., Tillmann, B. and Bharucha, J.J. 'The cortical topography of tonal structures underlying western music', *Science*, 298 pp. 2167-2170, 2002.
- [7] Krumhansl, C. L. 'Cognitive Foundations of Musical Pitch', *Oxford Psychology series*, Oxford University Press, New York, 1990.
- [8] Lindsay, A.T. and Herre, J. 'MPEG-7 and MPEG-7 Audio – An Overview', *Journal of the Audio Engineering Society*, 49 pp. 589-594, 2001.
- [9] Martens, G., De Meyer, H., De Baets, B., Leman, M., Lesaffre, M., Martens, J.P. and De Mulder, T. 'Distance-based versus Tree-based key recognition in musical audio', *Soft Computing*, 2004.
- [10] Pauws, S. 'Musical key extraction from audio', *Proceedings of International Conference on Music Information Retrieval*, 2004.
- [11] Sapp, C.S. 'Harmonic Visualizations of Tonal Music', *Proceedings of the International Computer Music Conference*, 2001.
- [12] Temperley, D. 'What's Key for Key? The Krumhansl-Schmuckler Key-Finding Algorithm Reconsidered', *Music Perception* 17(1), pp. 65-100, 1999.
- [13] Toiviainen, P. and Krumhansl, C. L. 'Measuring and modeling real-time responses to music: The dynamics of tonality induction', *Perception* 32(6), pp. 741 – 766, 2003.
- [14] Zhu, Y., Kankanhalli, M.S. Sheng Gao. 'Music Key Detection for Musical Audio', *Proceedings of the 11th International Multimedia Modelling Conference*, pp 30-37, 2005.