FKEY-EDM: AN ALGORITHM FOR KEY ESTIMATION IN EDM

Ángel Faraldo, Sergi Jordà, Perfecto Herrera
Music Technology Group, Universitat Pompeu Fabra
angel.faraldo@upf.edu

ABSTRACT

Key detection in electronic dance music is important for producers and DJ's who want to mix their tracks harmonically or organise their music collection by tonal content. This extended abstract describes an algorithm for key estimation based on template matching, with a number of processing stages that improve the detection accuracy in electronic dance music over other academic algorithms.

1. INTRODUCTION

The notion of tonality is one of the most prominent concepts in Western music. In its broadest sense, it defines the systematic arrangements of pitch phenomena and the relations between them, specially in reference to a main pitch class. The idea of key conveys a similar meaning, but normally applied to a smaller temporal scope, being common to have several key changes along the same musical piece. Different periods and musical styles have developed different practices of tonality. For example, modulation (i.e. the process of digression from one local key to another according to tonality dynamics) seems to be one of the main ingredients of musical language in euroclassical [1] music, whereas pop music tends to remain in a single key for a whole song or perform key changes by different means.

We use the term *electronic dance music* (EDM) to refer to a number of subgenres originating in the 1980's and extending into the present, intended for dancing at nightclubs and raves, with a strong presence of percussion and a steady beat [2]. Some of such styles even seem to break up with notions such as chord and harmonic progression (two basic building blocks of tonality in the previously mentioned repertoires) and result in an interplay between pitch classes of a given key, but without a sense of tonal direction.

These differences in the musical function of pitch and harmony suggest that computational key estimation should take into account style-specific particularities and be tailored to specific genres rather than aiming at all-purpose solutions.

[1] We take this term from Tagg [17] to refer to European Classical Music of the so-called common practice repertoire, on which most treatises on harmony are based.

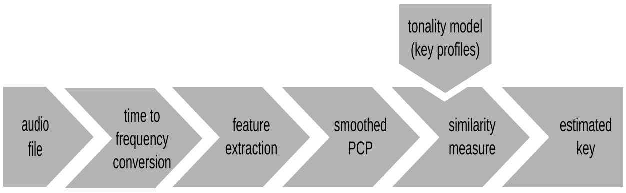

Figure 1. Basic template-based key estimation system.

1.1 Template-Based Key Estimation Methods

One of the most common approaches to key estimation is based on pitch-class profile extraction and template matching. Figure 1 shows the basic architecture of such key estimation system. Regular methodologies usually convert the audio signal to the frequency domain. The spectral representation is then folded into a so-called pitch class profile (PCP) or chromagram, a vector representing perceptually equal divisions of the musical octave, providing a measure of the intensity of each semitone of the chromatic scale per time frame. For improved results, a variety of preprocessing techniques such as tuning-frequency finding, transient removal or beat tracking can be applied. It is also common to smooth the results by weighting neighbouring vectors. Lastly, similarity measures serve to compare the averaged chromagram to a set of templates of tonality, and pick the best candidate as the key estimate. We refer the reader to [6], [14] for a detailed description of this method and its variations.

Regarding EDM, only the work by Sha'ath [16] and Faraldo et al. [7] address the problem specifically, although this situation is likely to change due to the recent publication of new datasets [11]. Both Sha'ath and Faraldo use template-matching approaches. Sha'ath makes manual modifications to the pioneering Krumhansl-Schmuckler key profiles [12], whereas Faraldo et al. derive statistical profiles from a corpus of EDM.

2. ALGORITHM

Our algorithm is based on a simple template-based method implemented in *Essentia*, [2] a C++ library for audio information retrieval [1], and builds upon previous work by Gómez [5, 6]. For a more detailed description, see [7].

[2] http://essentia.upf.edu/

2.1 High-Pass Filtering and Spectral Whitening

The input audio signal is processed with a 3^{rd} order high-pass filter prior to the spectral analysis. [3] We decided to set the cut-off frequency to 200 Hz after informal experimentation with various frequencies in the range 100-250 Hz. Besides, we flatten the spectrum according to its spectral envelope, based on a method by Röbel and Rodet [19]. The aim was to increase the weight of the predominant peaks, so that notes across the selected pitch range contribute equally to the final PCP. This technique has been previously used by Gómez [6], and other authors have proposed similar solutions (e.g. [10], [13]).

2.2 New Key Profiles

As explained above, one of the main ingredients in a template-based key estimator is the tonality model represented by the so-called key profile, a vector containing the relative weight of the different pitch classes for a given key. That is the reason to to submit two different algorithms to the context.

Fkey-edm (FJH3) incorporates new key profiles based on analysis of a balanced corpus with further manual adjustments. We obtained the basic major and minor profiles by calculating the median of a subgroup of around 300 tracks per mode, that were correctly estimated with different key profiles (Krumhansl [12], Temperley [18] and Faraldo et al. [7]). Then, we modified the profiles by minimising the weights of non-tonal degrees ($\flat$II, $\flat$III, $\sharp$IV, $\flat$VI in major; $\flat$II, $\sharp$IV, $\natural$VI in minor).

We also obtained a third profile from a group of minor tracks wrongly estimated as major with the newly created profiles. We incorporated this profile into the system, so that the template matching is performed against three different vectors (one major and two minor), trying to lower the parallel errors, frequent in this kind of music. This is explained in detail in [8].

The other algorithm (*fkey*, FJH2), incorporates all the processing stages present in FJH3, but uses the major and minor profiles proposed by Temperley in [18].

2.3 Detuning Correction

We noted that some of the estimations with the basic method produced tritone and semitone errors. Our hypothesis was that these could be due to possible de-tunings produced by record players with manual pitch/tempo corrections [16]. In order to tackle this, our algorithms use a PCP resolution of 3 bins per semitone, as it is usual in key detection algorithms [9], [15]. This allowed us to insert a post-processing stage that shifts the averaged PCP ± 33 cents, depending on the position of the maximum peak in the vector.

Various tuning-frequency estimation methods have been proposed, mostly based on statistics (cfr. [4]). Our approach is a simplification of that described in [9]. The algorithm finds the maximum value in the averaged chromagram and shifts the spectrum ± 1 bin, depending on this unique position. This shift is done only once per track, after all the PCP's are averaged together.

MIREX05				
	FHJ2	FHJ3	BD1	CN1
correct	.6342	.6070	.7260	.8267
fifth	.1829	.2404	.1406	.0599
relative	.1102	.0367	.0583	.0272
parallel	.0128	.0543	.0160	.0176
other	.0599	.0615	.0591	.0687
weighted	.7613	.7491	.8170	.8683

Table 1. Results on the MIREX05 Dataset.

GiantSteps				
	FHJ2	FHJ3	BD1	CN1
correct	.3411	.6209	.5530	.3974
fifth	.0712	.0613	.0662	.0480
relative	.1689	.0662	.0977	.1325
parallel	.0960	.0563	.0381	.0430
other	.3228	.1954	.2450	.3791
weighted	.4465	.6826	.6230	.4697

Table 2. Results on the GiantSteps Dataset.

3. RESULTS

Tables 1 and 2 show the results from the MIREX 2016 competition in the Audio Key Detection task. We can see how in the typical MIREX dataset for this task, consisting of euroclassical music *incipits*, our two algorithms perform notably below the algorithms by Bernardes and Davies (BD1); and Cannam and Noland (CN1). This was expected for FHJ3, an algorithm intended to detect keys in electronic dance music tracks. However, we see that *Essentia's* baseline algorithm with Temperley's profiles (FJH2) also scores below BD1 and CN1.

We observe, on the other hand, that the performance of all algorithms except for FJH3 drops dramatically in the *Giantsteps* dataset, a collection of 604 two-minute excerpts of EDM tracks [11]. This raises questions about the usability of key detection algorithms in style-agnostic applications, as well as to the validity of the evaluation dataset used so far in the MIREX evaluation context. More specifically, we see that:

- The highest number of errors of FJH3 in the MIREX05 dataset are fifth errors (24%). This could be explained as a clear divergence in the tonal language of euroclassical practice in comparison to popular music styles. [4]

[3] After informal testing, we chose the following settings: mix-down to mono; sampling rate: 44,100 Hz.; window size: 4,096 hanning; hop size: 4,096; frequency range: 25-3,500 Hz.; PCP size: 36 bins; weighting size: 1 semitone; similarity: cosine distance.

[4] In particular, we attribute it to the fact that V-I relationship so common to euroclassical music theory is somehow perceived as a I-IV relationship in pop-rock music (and extending into EDM), what would lead

- Despite being the lowest scoring algorithm in the MIREX05 dataset, FJH3 presents the smallest variance between the two datasets used for evaluation.
- In any case, the general results suggest that key detection algorithms should be evaluated according to the purpose they were designed for, and if not, at least evaluate them on different styles. In this regard, it would probably be informative to run the same algorithms on other available datasets containing different tonal languages (e.g. the beatles dataset).
- In this sense, the limitation of analysis to *incipits* of musical works based on the assumption of modulation is something that should not apply (at least systematically) to all repertoires.
- Results also confirm the intuition that different strategies should be applied to different musical styles, that tonality is a dynamic notion that evolves over history, and that even musical practices that are normally regarded as poorly interesting from a tonal viewpoint (such as EDM), pose research challenges both to Music Information Retrieval and Music Theory fields.
- Last, despite we still lack datasets with extended modal details, broadening the classification to other widespread modes would correspond better with the multiplicity of tonal practices, that in turn might be distinctive of different styles. [5]

4. ACKNOWLEDGEMENT

This research has been partially supported by the EU-funded *GiantSteps* project (FP7-ICT-2013-10. Grant agreement number 610591).

5. REFERENCES

[1] Bogdanov, D., Wack, N., Gómez, E., Gulati, S., Herrera, P., Mayor, O.: ESSENTIA: an open-source library for sound and music analysis. Proc. 21st ACM-ICM, 855–858 (2013)

[2] Dayal, G., Ferrigno, E.: Electronic Dance Music. Grove Music Online, Oxford University Press (2012)

[3] De Clerq, T., Temperley, D.: A Corpus Analysis of Rock Music Popular Music 30(1) 47–70 (2011)

[4] Dressler, K., Streich, S.: Tuning frequency estimation using circular statistics. Proc. 8th ISMIR, 2–5 (2007)

[5] Gómez, E.: Tonal description of polyphonic audio for music content processing. INFORMS Journal on Computing 18(3), 294–304 (2006)

to fifth errors in the current evaluation methodology. (cfr. De Clerq and Temperley [3] for an elaboration on the predominance of the I-IV harmonic relationship in rock music.)

[5] For example, a purely aeolian tune (pop, edm) differs significantly from a minor harmonic (euroclassical). Whilst in the current taxonomy they should both be classified as minor, truth is that they are representative of highly diverse notions of tonality.

[6] Gómez, E.: Tonal description of music audio signals. PhD Thesis, Universitat Pompeu Fabra, Barcelona (2006)

[7] Faraldo, Á., Gómez, E., Jordà, S., Herrera, P.: Key Estimation in Electronic Dance Music. Proc. 36th ECIR, (2016)

[8] Faraldo, Á., Jordà, S., Herrera, P.: Improving Key Estimation in Electronic Dance Music: A Multi-Profile Approach- *In preparation.*

[9] Harte., C.: Towards automatic extraction of harmony information from music signals. PhD Thesis, Queen Mary University of London (2010)

[10] Klapuri, A.: Multipitch analysis of polyphonic music and speech signals using an auditory model. IEEE Transactions on Audio, Speech, and Language Processing 16(2), 255—266 (2008)

[11] Knees, P., Faraldo, Á., Herrera, P., Vogl, R., Böck, S., Hörschläger, F., Le Goff, M.: Two Data Sets for Tempo Estimation and Key Detection in Electronic Dance Music Annotated from User Corrections. Proc. 16th ISMIR (2015)

[12] Krumhansl, C. L.: Cognitive foundations of musical pitch. Oxford Unversity Press, New York (1990)

[13] Müller, M., Ewert, S.: Towards timbre-invariant audio features for harmony-based music. IEEE Transactions of Audio, Speech & Language Processing 18(3), 649—662 (2010)

[14] Noland, K.: Computational Tonality Estimation: Signal Processing and Hidden Markov Models. PhD Thesis, Queen Mary University of London (2009)

[15] Noland, K., Sandler, M.: Signal Processing Parameters for Tonality Estimation. Proc. 122nd Convention Audio Engeneering Society (2007)

[16] Sha'ath., I.: Estimation of key in digital music recordings. Dep. Computer Science & Information Systems, Birkbeck College, University of London (2011)

[17] Tagg., P.: Everyday tonality II (Towards a tonal theory of what most people hear). The Mass Media Music Scholars' Press. New York and Huddersfield (2014)

[18] Temperley, D.: What's key for key? The Krumhansl-Schmuckler key-finding algorithm reconsidered. Music Perception: An Interdisciplinary Journal 17(1), 65–100 (1999)

[19] Röbel, A., Rodet, X.: Efficient spectral envelope estimation and its application to pitch shifting and envelope preservation. Proc. 8th DAFX (2005)