IMPROVING TONAL ESTIMATION IN ELECTRONIC DANCE MUSIC

First author
Affiliation1
author1@smcnetwork.org

Second author
Affiliation2
author2@smcnetwork.org

Third author
Affiliation3
author3@smcnetwork.org

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. Both research-oriented algorithms and commercially available applications provide relatively solid key estimation solutions. However, they are normally limited to a binary classification into major and minor keys, that does not reflect the variety of tonal practises in electronic dance music. In this paper, we propose some ideas that expand this dual classification, providing finer information about the modal and tonal characteristics of the tracks, setting the basis for future work on this area. We conclude our explanation with a comparison of our algorithm to other available solutions, and discuss about the suitability of this new perspective of tonality in electronic dance music.

1. INTRODUCTION

Academic literature about electronic dance music (EDM in the following) has significantly increased over the past years [1], leading to the appearance of peer-reviewed publications such as *Dancecult*. [1] This is probably due to a combination of musicological interest, the challenges it poses to the Music Information Retrieval community, and a myriad of real-world potential applications (from recommender systems to integration in music production software). However, the study of tonality in EDM is residual if compared to other musical domains, as harmony and pitch are normally regarded as secondary aspects in the production of electronic popular music.

Despite this fact, we believe that automatic key estimation does not only help EDM practitioners in classification and mixing endeavours, but could also shed light over some of the tonal practises present in this metagenre [2, 3]. These include typical modal variants such as *phrygian* or *mixolydian*, but also cases of *mono-tonal* excerpts (where the music has a clear tonic but no sense of modality), and *atonal* or *poly-modal* fragments.

Furthermore, it is common to find *un-pitched* passages, which although technically not belonging to actual tonal categories, would be significant to detect for the purpose of harmonic mixing, since they are highly neutral.

[1] http://dj.dancecult.net/

With this in mind, this paper presents a method addressing tonality estimation in EDM, with which we obtain an improvement over existing academic algorithms (see [3,4] for a recent evaluation of both academic and commercial algorithms).

Perhaps more importantly, we introduce new *experimental* categories, trying to expand the typical major/minor classification, and giving account of monotonic and atonal tracks as well as ambiguous or modally complex ones, thus setting the basis for further work on this area. Unfortunately, due to the lack of solid ground-truth containing the tonal nuances we are seeking to express, the output of our method still produces a binary classification (although with some extra verbose), allowing direct comparison with existing algorithms. Therefore, the experimental results regarding the new categories can only be taken as provisional, an issue that we will tackle in the near future.

The rest of the paper is organised as follows: Section 2 presents a short overview of MIR publications related to EDM, and contextualises the task of tonal induction. In Section 3 we elaborate on the issues we want to address and continue describing our methodology in Section 4. We follow with an evaluation and comparison to other available solutions in Section 5, before wrapping up and pointing at future development in Section 6.

2. RELATED WORK

As pointed above, there is an increasing attention in academia towards analysing EDM, especially addressing the domains of timbre, structure and rhythm. Since the pioneering study by Butler [5], recent years have seen publications focusing on rhythm similarity [6–8], structure detection and segmentation –from short musical sections [6,9,10] to complete DJ sets [11,12],– as well as other typical MIR tasks such as genre identification [13–15] or downbeat and tempo detection [16, 17].

Regarding tonality estimation, only the work by Sha'ath [18] and Faraldo et al. [3] address the problem specifically in EDM, although this situation is likely to change due to the recent publication of new datasets [4]. Both Sha'ath and Faraldo use template-matching approaches. Shaath makes manual modifications to the pioneering Krumhansl-Schmuckler key profiles [19], whereas Faraldo derives statistical profiles from a corpus of EDM.

Essentially, a template-based key estimation algorithm relies in some kind of spectral transformation, in order to obtain a so-called *pitch-class profile* (PCP) [20]. A PCP is a vector of $12n$ dimensions, where n stands for perceptually equal divisions of a semitone, representing a weighted distribution of all pitch-classes in a given time period. PCP's are then correlated with a number of key profiles, –an equivalent vectorial representation of the pitch-classes theoretically present in a given mode. Template-based approaches normally offer one template for each mode observed (typically two, major and minor), which are then cyclically shifted around the twelve possible tones to find the tonic of the key. We refer the reader to [21, 22] for a detailed explanation of standard key estimation methods.

3. PROBLEM STATEMENTS

3.1 Parallel Errors

One of the major identifiers of the modality of a piece of music lies on the third degree of the scale. Depending on the interval it forms with the tonic (either 2 or 1.5 tones), the modality of the piece will generally be regarded as major or minor. Major and minor thirds are normally mutually exclusive, since there lies too the difference between a major and a minor chord, (`C E G` vs. `C E♭ G`, for example). In EDM, however, wee see too often pitch-class distributions that show equivalent energy for both degrees, making difficult the modal estimation of the excerpt, what normally leads to parallel errors in the key estimation process (e.g. mistaking a track in `C minor` as being in `C major`). We think this is not only due to a perhaps "looser" modality, compared to pop or European classical music, but mainly to the timbral qualities of most EDM. Our hypothesis is that synthesiser sounds with extremely rich spectra, reinforce the 5^{th} harmonic of the tonic (that is, the major third) even when the pitch relationships are in a minor context. For us, it might be easy to discern between tonal and timbral levels. However, an algorithm –that is after all based on spectral analysis,– does not see the difference all that clear.

We envision a twofold solution to this problem: (a) creating new key profiles from a relatively balanced corpus of EDM, that reflect the inherent tonal ambiguity of this music, (b) before introducing a three-profile system, with an extra profile addressing specifically the parallel error issue.

3.2 Modality Expanded

Faraldo et al. [3] outline some tonal characteristics of EDM, including the presence of modal variants such as *mixolydian* (basically a major scale with $\flat\hat{7}$) or *phrygian* (minor, $\flat\hat{2}$). However, they do not incorporate any modal categories in their analysis.

Besides these considerations on modality *per se*, we are interested in other practices not reflected within a typical modal definition. In particular, we observed that many tracks offer a clear sense of tonal centre, but it is however difficult to infer a specific mode when only 1-to-3 pitch-classes are present, thus being quite ambivalent for harmonic mixing purposes. [2] Other tracks are almost atonal (in the sense they use almost all 12 pitch-classes, although the tonal centre is normally set clearly by the bass or the kick) and there are not few cases of *poly-modal* fragments, that arise by the superposition of loops in different keys or modes. On top of that, it is common to find *un-pitched* passages, just with percussion and/or spoken voice, which are harmonically neutral by nature.

In this work, we try to expand the typical binary classification addressing some of the variants just mentioned [3] with an array of dedicated key profiles. As we will show later, profile variants are prone to introduce new types of errors (especially fifth and relative errors). After all, modal differences arise from a circular shift of the intervalic pattern over the degrees of the scale, and this is also an essential part in the estimation chain to detect the tonic of the mode.

We minimize this problem by running the algorithm twice, combining a three-profile method with a multi-profile system. This way, we are also able to evaluate the system with available datasets (that so far, only offer a binary major/minor classification), while still providing an insight of other modal characteristics.

4. METHOD

This paper proposes a few additions to the algorithm described in [3], which is available online. [4] That method is in turn based of the approach by Gómez [21], as it is implemented in *Essentia*, [5] a C++ library for audio information retrieval [23]. One of our concerns is to diminish as much as possible the clear bias toward the minor mode present in that approach. Besides, we also attempt to expand the possible output of the system, in order to provide finer information about the modal qualities of a track, although this should be taken as preliminary work, due to the lack of available ground truth for evaluation.

4.1 Algorithm

As stated above, our key detector incorporates a few improvements and modifications to the one described in [3]:

4.1.1 High-Pass Filtering

First, we inserted a 3^{rd} order high-pass filter to the input audio file. We decided to set the cut-off frequency to 200 Hz after informal experimentation with various frequencies in the range 100-250 Hz. This pre-processing of the audio signal provides cleaner pitch-class profiles improving the results.

4.1.2 New Key Profiles

We created 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 [19], Temperley [24] and Faraldo et al. [3]). Then, we modified the profiles by minimising the weights

[2] The most extreme case of this would be a track where the only pitched sound is the tone of the bass drum.

[3] Specifically, in this paper we will not address the problems of poly-modality and pitch/un-pitch sound recognition.

[4] `www.github.com/angelfaraldo/keyest-paper`

[5] `http://essentia.upf.edu/`

	GiantSteps				*Beatport*				*Shaath*			
	essentia	*edma*	*edmm*	*new*	*essentia*	*edma*	*edmm*	*new*	*essentia*	*edma*	*edmm*	*new*
correct	**0.295**	**0.568**	**0.636**	**0.603**	**0.271**	**0.517**	**0.525**	**0.584**	**0.302**	**0.598**	**0.701**	**0.618**
fifth	0.235	0.106	0.108	0.098	0.285	0.116	0.117	0.116	0.262	0.112	0.105	0.110
relative	0.099	0.066	0.028	0.083	0.053	0.058	0.071	0.054	0.061	0.035	0.017	0.062
parallel	0.086	0.113	0.065	0.071	0.119	0.219	0.169	0.158	0.080	0.118	0.045	0.071
other	0.285	0.147	0.164	0.147	0.273	0.090	0.118	0.088	0.296	0.137	0.132	0.138
mirex	*0.459*	*0.663*	*0.711*	*0.691*	*0.453*	*0.636*	*0.638*	*0.690*	*0.467*	*0.688*	*0.767*	*0.706*

Table 1. Comparison of typical errors and MIREX scores using various key profiles. The profiles proposed in this paper are labeled as *new*.

of non-tonal degrees (♭II, ♭III, ♯IV, ♭VI in major; ♭II, ♯IV, ♮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 then incorporated this profile into the system, trying to lower the parallel errors, frequent in this kind of music.

4.1.3 Multi-Modal Profile System

In addition to the profiles just described, we incorporated an extended multi-modal profile system, with up to 6 simultaneous profiles accounting for modal variations. They were obtained by tweaking the profiles presented in section 4.1.2, based on music theoretical prescriptions. The six profiles include two major/minor alternatives and two other profiles accounting for *mono-tonic* and difficult, likely atonal, excerpts.

The modal variants comprise *ionian* and *mixolydian* profiles, both based on the major profile from the previous subsection, with a generous boost on the leading tone ($\natural\hat{7}$) and the $\flat\hat{7}$, respectively.

We also include *minor-harmonic* (again, with an increased weight on the $\natural\hat{7}$), and *phrygian* profiles, another minor mode with the energy shifted from $\hat{2}$ to $\flat\hat{2}$, both based in our basic minor mode.

Last, we incorporate a *monotonic* profile, with unit energy only in the first bin, followed by zeroes; and an almost flat profile (we did not use a completely flat profile because that would not produce any output), giving account of difficult, possibly atonal tracks. Although the level of detail provided by this approach has not been used in the evaluation process, for the reasons mentioned above, the system produces an output with modal information not present in other methods. `A major mixolydian`, `C# minor phrygian` or `Bb minor monotonic` are examples of the type of output produced by the system. Although on a very early stage of development, we have used a small number of decision rules to try improving the typical dual estimations with the multi-modal results.

4.2 Datasets

For this study, we have used three different datasets of EDM with a single estimation per track. At the moment of writing, only the so-called *GiantSteps* dataset [4] is publicly available, comprising 604 two-minute excerpts from *Beatport*, [6] an online music store for DJ's and producers.

A second dataset of 1000 tracks, compiled by Sha'ath to improve the key estimation software *KeyFinder*, is not publicly available due to copyright issues. However, the ground-truth annotations are freely available on his website. [7] We managed to obtain a subset of 925 tracks from this collection. As shown in [3], both datasets present a clear bias toward the minor modes (over 85% of the tracks are in minor in both datasets), so common of EDM.

Last, we curated an in-house dataset with 1500 excerpts from *Beatport*, manually annotated by two experts. The main purpose of this effort was to obtain a balanced collection in terms of major and minor tracks, with an even distribution across different subgenres and keys, something extremely difficult given the high prominence of minor tracks. For this research, we used a sub-collection of 1160 tracks that were confidently annotated as pertaining to a single key. Although still not ideal, we managed to make a dataset with 1/3 of the tracks in major.

5. RESULTS

In this section, we discuss the results of various experiments. They include an evaluation of different types and number of profiles, and a comparison to other available solutions. All the results shown were computed on uncompressed audio tracks at 44100 Hz, with a window size of 4096 points and a hop size of 16384 points. This provided good results, while keeping the computation time relatively fast.

As previously mentioned, we used three independent datasets of EDM, in order to detect possible overfitting (since we have extracted our profiles from a sub-group of our own *Beatport* dataset). Although we have conducted a more exhaustive analysis of errors, we report only the most common ones as well as a weighted score according to the MIREX standard (Music Information Retrieval Evaluation eXchange) for this task. However, the interested reader could access our code online, [8] containing estimation and evaluation tools for a finer detail of analysis.

5.1 Research Oriented Algorithms

Table 1 shows the results of various academic template-based methods. They are all based on dual profile correlation (major/minor). We include the output of *Essentia*'s default key estimation algorithm, since we are using that

[6] `https://pro.beatport.com/`

[7] `http://www.ibrahimshaath.co.uk/keyfinder`

[8] `http://www.hidden/for/review`

	GiantSteps			*Beatport*			*Shaath*		
	two	*three*	*many*	*two*	*three*	*many*	*two*	*three*	*many*
correct	**0.603**	**0.614**	**0.614**	**0.584**	**0.604**	**0.609**	**0.618**	**0.646**	**0.669**
fifth	0.098	0.098	0.099	0.116	0.114	0.110	0.110	0.112	0.117
relative	0.083	0.078	0.083	0.054	0.051	0.048	0.062	0.062	0.055
parallel	0.071	0.068	0.063	0.158	0.135	0.140	0.071	0.040	0.033
other	0.146	0.142	0.141	0.088	0.096	0.093	0.138	0.139	0.127
mirex	*0.691*	*0.700*	*0.701*	*0.690*	*0.703*	*0.706*	*0.706*	*0.729*	*0.750*

Table 2. Comparison of typical errors and MIREX scores of the proposed method with two, three and multiple profiles.

framework as a basis to our analysis tools, and which could be considered as the baseline for our experiments. We clearly see how this method performs far under any other result presented. The other two profiles (*edma, edmm*) are proposed by Faraldo et al. [3] and were derived statistically from the *shaath* dataset. We can infer that by observing how they perform notably better on this collection than in the other two. *Edmm* estimates any track as being in minor, based in the fact that most EDM is in minor. That is the reason why it offers the best performance on *giantSteps* and *shaath* datasets (both highly populated with minor tracks), but on a more modally balanced dataset as *beatport*, it scores behind our newly proposed profiles (labelled as *new* in the table). For this reason, we take *edma* as a fairer comparison candidate, against which our proposed profiles perform significantly better in all three datasets.

5.2 Two, Three and More Profiles

In Table 2 we present the results of our algorithm with the different profile approaches described in Section 4.1. For the sake of clarity, we replicate the results for the two-profiles just discussed (*new*, in Table1) in the first column of each dataset's results. Despite having used the *beaport* dataset for extracting our profiles (although we remind the reader that they were severely modified by hand afterwards), the table shows how these generalise well to the other two datasets (performing notably better in both of them).

Performance improves in all scenarios when switching to the three-profile method, there is a slight reduction of the parallel errors and the percentage of correct instances increases in all cases.

The increase, however, it is not significant as we move onto a multi-modal approach. As a matter of fact, the multi-modal algorithm is run in parallel with the three-profile method, providing additional labelling and only occasional influence on the final estimation, as explained in Section 4.1.3.

5.3 Final Evaluation

Let us finalise by comparing our algorithm to dedicated software applications, used by DJ's and producers in real-life scenarios for key labelling and harmonic mixing. In particular, Table 3 displays the results of our multi-modal algorithm (*new*) along with those of *KeyFinder* [9] –a freely available piece of software by Sha'ath,– and two commercial products, *Mixed-In-Key 7* [10] and *Rekordbox*. [11]

We can observe how *Mixed in Key* provides the best performance in all datasets, followed, also in all scenarios, by our multi-modal solution. *Rekordbox* is clearly the one offering lower results, whereas the application by Shaath shows greater variability across the different datasets, suggesting that there is a certain overfitting when evaluated with his own datasets.

6. CONCLUSIONS

In this paper, we presented our preliminary results towards a finer, musicologically motivated, key detection algorithm. In turn, we have made some significant improvements to template-matching methods, with the creation of new profiles and especially with the inclusion of more than two modal candidates, thus getting closer to the variety of tonal practices in EDM.

Although still in need of proper experimental validation due to the lack of detailed annotated corpora, overall, we think the multi-modal approach we described could give the electronic music producer a more meaningful description of the tracks she is working with, facilitating harmonic mixing. Simultaneously, the method described can be of help in a musicological study of some tonal characteristics of electronic dance music.

Naturally, there is still room to improve the detection process and we have some preliminary evidence that a combined approach of this method with supervised learning techniques could boost the performance of the algorithm. Similarly, as our annotated corpus grows in details, the modelling of the different modal profiles could be refined, hopefully leading to more precise estimations.

Another challenge for the future, at least with the practitioner in mind, is to encapsulate the knowledge of the algorithm in simple though meaningful ways, accounting for the variety of tonal practices and ambiguities in ways that empower the user instead of making her choose among an array of nomenclatures that, although informative for the musicologist, are of little significance for the creative user of such algorithm.

Acknowledgments

Special thanks to Name Surname for his contribution to the manual annotation process of 1500 tracks and Name

[9] www.ibrahimshaath.co.uk/keyfinder/

[10] www.mixedinkey.com

[11] http://rekordbox.com/

	GiantSteps				*Beatport*				*Shaath*			
	KF	*MIK*	*Rkbox*	*new*	*KF*	*MIK*	*Rkbox*	*new*	*KF*	*MIK*	*Rkbox*	*new*
correct	**0.604**	**0.668**	**0.526**	**0.614**	**0.548**	**0.657**	**0.480**	**0.609**	**0.674**	**0.720**	**-**	**0.669**
fifth	0.127	0.095	0.164	0.099	0.155	0.100	0.155	0.110	0.128	0.094	-	0.117
relative	0.066	0.055	0.053	0.083	0.060	0.061	0.048	0.048	0.019	0.021	-	0.055
parallel	0.056	0.050	0.079	0.063	0.150	0.118	0.158	0.140	0.041	0.045	-	0.033
other	0.146	0.133	0.177	0.141	0.087	0.064	0.158	0.093	0.138	0.121	-	0.127
mirex	*0.699*	*0.742*	*0.640*	*0.701*	*0.673*	*0.749*	*0.604*	*0.706*	*0.751*	*0.782*	-	*0.750*

Table 3. Comparison of typical errors and MIREX scores with three dedicated software applications: *KeyFinder* (KF), *Mixed in Key 7* (MIK) and *Rekordbox* (Rkbox). ***Due to logistic reasons we could not obtain results for MIK and Rekordbox in Shaath's dataset. We will incorporate them if the paper gets accepted.***

Surname for his valuable help annotating a sub-collection with atonal and monotonic excerpts.

This research has been partially supported by (omitted for blind review).

7. REFERENCES

[1] A. Doehring, "Andrés's 'new for u': New for us. on analysing electronic dance music," in *Song interpretation in 21st-Century Pop Music*, R. von Appen, A. Doehring, D. Hemls, and A. F. Moore, Eds. Ashgate Publishing, 2015.

[2] R. Wooller and A. Brown, "A framework for discussing tonality in electronic dance music," in *Proceedings Sound : Space - The Australasian Computer Music Conference*, 2008, pp. 91–95.

[3] Á. Faraldo, E. Gómez, S. Jordà, and P. Herrera, "Key estimation in electronic dance music," in *Proceedings of the 38th European Conference on Information Retrieval*, 2016, pp. 335–347.

[4] P. Knees, Á. Faraldo, P. Herrera, R. Vogl, S. Böck, F. Hörschläger, and M. Le Goff, "Two datasets for tempo estimation and key detection in electronic dance music annotated from user corrections," in *Proceedings of the 16th International Society for Music Information Retrieval Conference*, Málaga, 2015, pp. 364–370.

[5] M. Butler, "Unlocking the Groove: Rhythm, Meter and Musical Design in Electronic Dance Music," Ph.D. dissertation, Indiana University, 2003.

[6] B. Rocha, N. Bogaards, and A. Honingh, "Segmentation and timbre- and rhythm-similarity in electronic dance music," in *Proceedings of the Sound and Music Computing Conference*, 2013, pp. 754–761.

[7] M. Panteli, B. Rocha, N. Bogaards, and A. Honingh, "Development of a rhythm similarity model for electronic dance music," in *Proceedings of the 53rd International Audio Engineering Society Conference: Semantic Audio*, 2014, pp. 537–542.

[8] ——, "Modeling rhythm similarity for electronic dance music," in *Proceedings of the 15th International Society for Music Information Retrieval Conference*, 2014.

[9] A. Eigenfeldt and P. Pasquier, "Evolving structures for electronic dance music," in *Proceeding of the 15th Genetic and Evolutionary Computation Conference*, 2013.

[10] K. Yadati, M. Larson, C. C. S. Liem, and A. Hanjalic, "Detecting Drops in Electronic Dance Music: Content Based Approaches to a Socially Significant Music Event," in *Proceedings of the 15th International Society for Music Information Retrieval Conference*, 2014, pp. 143–148.

[11] T. Scarfe, W. Koolen, and Y. Kalnishkan, "Segmentation of electronic dance music," Royal Holloway, University of London, Surrey, Tech. Rep., 2014.

[12] N. Glazyrin, "Towards automatic content-based separation of dj mixes into single tracks," in *Proceedings of the 15th International Society for Music Information Retrieval Conference*, 2014, pp. 149–154.

[13] P. Kirss, "Audio Based Genre Classification of Electronic Music," Master's thesis, University of Jyväskylä, 2007.

[14] J. Sesmero Molina, "Electronic Dance Music Genre Classification," Master's thesis, Universitat Pompeu Fabra, Barcelona, 2008.

[15] N. Collins, "Influence in early electronic dance music: an audio content analysis investigation," in *Proceedings of the 13th International Society for Music Information Retrieval Conference*, 2012.

[16] J. Hockman, M. Davies, and I. Fujinaga, "One in the jungle: Downbeat detection in hardcore, jungle, and drum and bass," in *Proceedings of the 13th International Society for Music Information Retrieval Conference*, 2014, pp. 169—-174.

[17] F. Hörschlager, R. Vogl, S. Böck, and P. Knees, "Addressing tempo estimation octave errors in electronic music by incorporating style information extracted from wikipedia," in *Proceedings of the Sound and Music Computing Conference*, 2015.

[18] I. Sha'ath, "Estimation of key in digital music recordings," Master's thesis, Birbeck College, University of London, 2011.

[19] C. L. Krumhansl, *Cognitive foundations of musical pitch*. Oxford Unversity Press, 1990.

[20] T. Fujishima, "Realtime chord recognition of musical sound: A system using common lisp music," in *Proceedings of the International Computer Music Conference*, 1999, pp. 464–467.

[21] E. Gómez, "Tonal Description of Music Audio Signals," Ph.D. dissertation, Universitat Pompeu Fabra, Barcelona, 2006.

[22] K. Noland, "Computational Tonality Estimation : Signal Processing and Hidden Markov Models," Ph.D. dissertation.

[23] D. Bogdanov, N. Wack, E. Gómez, S. Gulati, P. Herrera, O. Mayor, G. Roma, J. Salamon, J. Zapata, and X. Serra, "Essentia: an audio analysis library for music information retrieval," in *Proceedings of the 14th International Society for Music Information Retrieval Conference*, 2013, pp. 493–498.

[24] D. Temperley, "What's Key for Key? The Krumhansl-Schmuckler Key-Finding Algorithm Reconsidered," *Music Perception: An Interdisciplinary Journal*, vol. 17, no. 1, pp. 65–100, 1999.