Audio Engineering Society

Conference Paper

Presented at the Conference on
Semantic Audio
2017 June 22 – 24, Erlangen, Germany

A Multi-Profile Method for Key Estimation in EDM

Ángel Faraldo[1], Sergi Jordà[1], and Perfecto Herrera[1]

[1]*Music Technology Group, Universitat Pompeu Fabra, Barcelona*

Correspondence should be addressed to Ángel Faraldo (`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. In this paper, we present an algorithm that improves the performance of an existing method by introducing a system of multiple profiles, addressing difficult minor tracks as well as possibly amodal ones. After the explanation of our method, we use three independent datasets of electronic dance music to evaluate its performance, comparing it to other academic algorithms and commercially available solutions.

1 Introduction

Electronic dance music (EDM) is an umbrella term that generally refers to a number of subgenres originating in the 1980's and extending into the present, made almost solely with electronic equipment and mainly intended for dancing at nightclubs and raves [1]. Academic literature about EDM has significantly increased over the past years, even 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 when compared to other musical domains, as harmony and pitch are normally regarded as secondary aspects in this type of music, far behind rhythmic and timbral features. Despite this fact, we believe that automatic key estimation can assist EDM practitioners in classification and mixing endeavours, and in turn, shed light over some of the tonal practises present in this metagenre [2].

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

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 [3], recent years have seen publications focusing on rhythm similarity [4, 5], structure detection and segmentation (from short musical sections [6, 7] to complete DJ sets [8, 9]), as well as other typical MIR tasks such as genre identification [10, 11] or downbeat and tempo detection [12, 13].

Regarding key estimation, that is, the characterisation of a fragment of music as suggesting a certain central tone and mode, only the work by Sha'ath [14] and Faraldo et al. [15] address the problem specifically in

EDM, although this situation is likely to change due to the recent publication of new datasets [16].

One of the most common approaches to key estimation is to follow a template-matching method. Essentially, this technique relies in some kind of spectral transformation, in order to obtain a so-called *pitch-class profile* (PCP) from the audio signal [17]. A PCP is a vector of $12 \times i$ dimensions, where i represents the number of divisions per 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-class distribution characteristic of a given musical mode. Normally there is one template for each mode observed (typically two, major and minor), which are then cyclically shifted around the twelve tones to find the tonic of the key as the highest correlated profile.

Various types of profiles have been proposed in the literature upon different criteria. The pioneering Krumhansl-Schmuckler profiles were derived from research on tonality perception [18] whilst others are extracted from corpus analysis [19] or shaped based on music theoretical observation [20]. In the domain of EDM, both Sha'ath [14] and Faraldo et al. [15] use template-matching approaches. While Sha'ath made manual modifications over Krumhansl's profiles, Faraldo et al. created statistical profiles from a corpus of music.

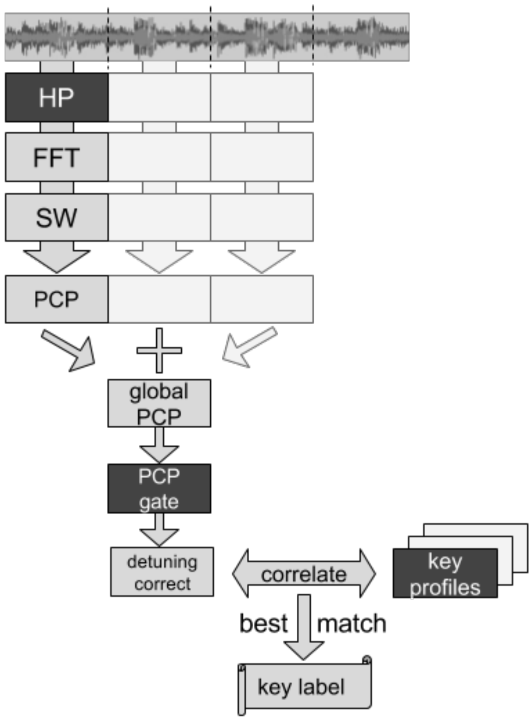

Fig. 1: Block diagram of our template-matching key estimation method. It is a variation of the one described by Faraldo et al. [15], to which we have added a high-pass filter, a PCP gate and new key profiles.

3 Method

In this paper, we propose a few variations over the algorithm described in Faraldo et al. [15].[2] That method is in turn based on the approach by Gómez [21], as implemented in *Essentia*, [3] a C++ library for audio information retrieval [22]. Our initial goal was to reduce the clear bias toward the minor mode present in the previous approach.

Figure 1 presents an overview of the method. The contributions of this paper are shown in dark, whereas the lighter areas are already described in [15]. A sound file is windowed in non-overlapping equally-sized frames (4096 samples), after what is high-pass filtered (HP) with a cut-off frequency of 200 Hz.[4] This stage provides cleaner pitch-class profiles, minimising low frequency noise likely from percussive instruments. After this process, we convert the signal to the frequency domain (FFT) and apply a spectral whitening (SW) function, in order to amplify the peaks present in the signal before calculating the PCP for each window. When the selected sound file is completely analysed, we aggregate all the PCP's and normalise the resulting vector. Bins with energy under a threshold of 0.2 are zeroed (PCP gate), as it will be explained later, and a simple detuning correction function is applied before correlating the output vector to a set of profiles extracted from a corpus of EDM, taking the best match as the estimated key.

[2] Available online at
`www.github.com/angelfaraldo/keyest-paper`
[3] `http://essentia.upf.edu/`

[4] We selected that frequency after informal experiments with various frequencies in the range 100-250 Hz.

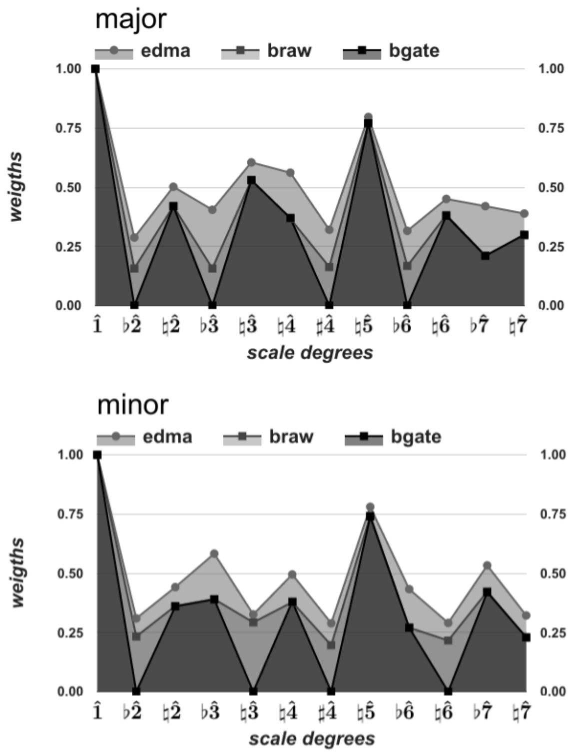

Fig. 2: Comparison of *edma* profiles [15] with the proposed profiles *braw* and *bgate*.

3.1 New Major and Minor Profiles

For this research, we curated a new dataset, manually annotated by two experts, with 1500 two-minute excerpts from *Beatport*,[5] an online music store for DJ's and producers. 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. From this collection, we created new major and minor key profiles based on a two subgroups of 300 tracks each, that were estimated correctly with three different templates (Faraldo et al. [15], Krumhansl [18] and Temperley [20]). The motivation to find a *consensus* between different profiles was to establish a sort of cross-stylistic baseline of modal categories.

The new key profiles, *braw* (beatport-raw), were obtained by calculating the median profile for each mode. An additional pair of profiles, *bgate* (beatport-gate), was obtained by zeroing the weights of the four lower elements on each vector. In the major profile, these corresponded to $\flat\hat{2}$, $\flat\hat{3}$, $\sharp\hat{4}$, $\flat\hat{6}$, all showing weights just under 0.2. The minor mode, however, presented a much flatter profile, only with $\sharp\hat{4}$ having energy below 0.2. In this case, we proceeded by analogy with the major mode, lowering $\flat\hat{2}$ and $\sharp\hat{4}$, as well as the $\natural\hat{3}$ and $\natural\hat{6}$, both indicators of major modalities.

Figure 2 shows a comparison of this profiles with *edma* profile [15]. Except for the first and fifth degrees, the weights of *edma* tend to be higher than the new profiles. Although the shapes are in general very similar, our new major profile seems to de-emphasise the $\flat\hat{7}$. It is also apparent how the effect of zeroing the theoretically less important degrees creates edgier profiles.

3.2 A PCP Gate

As mentioned above, we introduced a threshold function that automatically zeroes weights under 0.2. This operation is only performed once, after the global PCP is computed and normalised to 1. With this operation we intend to approximate the peakiness of the *bgate* profiles, reducing less modally relevant contributions.

3.3 Additional Profiles

We notice that the global PCP estimation from many minor tracks in the corpus showed equivalent energy for both third degrees in the pitch-class profile ($\flat\hat{3}$ and $\natural\hat{3}$, being normally mutually exclusive). As the reader might know, the third degree of the scale is one of the main identifiers of the modality of a piece of music: Depending on the interval it forms with the tonic, the piece will generally be considered as being in major or minor modality. The coexistence of both components, however, seems frequent in EDM tracks, leading 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 vaguer modality, compared to pop or euroclassical music [23], but mainly to the timbral qualities of most EDM. Our hypothesis is that synthesised sounds with very rich spectra, reinforce the 5^{th} harmonic of the tonic (i.e. $\natural\hat{3}$) even when the pitch relationships suggest a minor context, as it is illustrated in Figure 3.

In order to address this ambiguity, we used our *braw* profiles to obtain an additional profile as the median vector from a group of minor tracks estimated with the

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

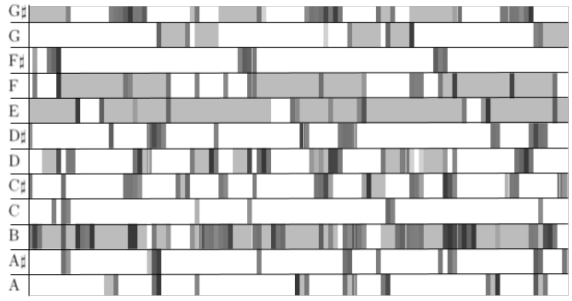

Fig. 3: Chromagram from 6 seconds of Williams Acidic Circutis' Remix of The Knife's *Silent Shout* (from the *GiantSteps* dataset). Note the presence of G♯ ($\natural\hat{3}$) over the E, on an otherwise phrygian (minor) fragment.

correct tonic but mistaken as major. Then, we incorporated this profile as a third choice into the system.

Furthermore, we included another profile trying to account for *amodal* tracks, (i.e. where the music has a clear tonic but no sense of modality), with all the energy concentrated on the tonic.

Figure 4 shows the two additional profiles. It is interesting to observe that *minor2* presents more energy on the major third degree than on the minor, contrary to music-theoretical intuitions. The *amodal* profile concentrates all the energy on the tonic, padding with zeroes all other values in the vector.

In summary, our method makes use of four profiles: one major, two minor and an extra one trying to account for possible cases of amodality). However, tracks estimated as amodal are reported as minor, given the bias of EDM toward the minor modality [15], and since all key datasets available provide a binary major/minor classification.

4 Results

In this section we introduce the results of our evaluations, including a study of the effect of different parameters in the method, as well as the different profiles proposed. All the experiments were computed on uncompressed mono audio files at a sampling rate of 44100 Hz. The files from Beatport, obtained at lower quality (mp3 at 96 kbps), were transcoded to WAV to meet this requirement.

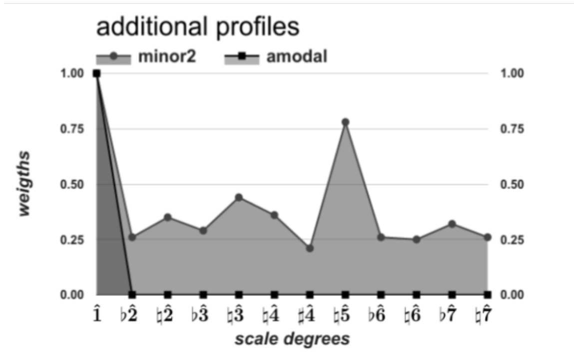

Fig. 4: Additional profiles addressing difficult minor tracks (*minor2*) and possibly amodal ones.

For the evaluation, we used three different datasets of EDM with a single estimation per track. At the moment of writing, only the so-called *GiantSteps* dataset is publicly available, comprising 604 two-minute excerpts from *Beatport* with ground-truth annotations extracted from user fora [16]. A second dataset of 1000 tracks, compiled by Sha'ath to improve his key estimation software *KeyFinder*, is not publicly available due to copyright issues. However, expert annotations are freely available online, together with his software.[6] Both datasets present a clear bias toward the minor modes (over 85% of the tracks are in minor in both datasets), characteristic of EDM [15]. Additionally, we used a sub-collection from our in-house dataset (referred to as *Beatport* in this paper), comprising 1160 tracks confidently annotated by two experts as being in one single key. The use of three independent datasets of EDM is justified to avoid possible overfitting, since we have extracted our profiles from a sub-group of 600 tracks from the same collection, as explained in Section 3.1.

Although we have conducted a more exhaustive analysis of errors, in this paper we report the most common ones as well as a weighted score following the MIREX convention (Music Information Retrieval Evaluation eXchange) for this task. However, the interested reader could access our code online,[7] containing the algorithms described in this text as well as evaluation tools with finer detail of analysis.

Figure 5 shows the proportion of correctly estimated tracks and the weighted score on the *GiantSteps* dataset

[6] http://www.ibrahimshaath.co.uk/keyfinder
[7] https://github.com/angelfaraldo/edmkey

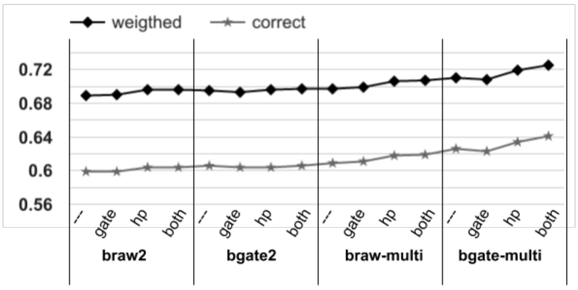

Fig. 5: MIREX weighted score and correct classification on the *GiantSteps* dataset showing the contribution of different components of the system: two and three profiles, high-pass filtering and PCP gating.

for various configurations of our algorithm. We compare a method with two-profiles to the new multi-profile system, both with the raw median profiles *braw* and the partially zeroed ones (*bgate*). We also show the effect of the high-pass filtering and the global PCP gating, alone and in combination. We observe improvement is small but mostly incremental steps from two to multiple profiles. Within each profile, high-pass filtering slightly improves the performance in all experiments whereas the effect of the PCP gate alone produces variable effects. In any case, the combination of the high-pass filter with the PCP gate provides the best results with all profile types.

Tables 1 and 2 compare the performance of our *multi-profile gated* method (bgate) with other algorithms' estimations. Table 1 presents the results of our system along the one described in [15] (*edma* and *edmm* variations). We also include the output of *Essentia*'s default key extractor [21], since we are using that framework as a basis to our analysis tools. In Table 2, we compare the output of our algorithm to dedicated software applications, used by DJ's and producers in real-life scenarios for key labelling and harmonic mixing. These include *KeyFinder* –the freely available piece of software by Sha'ath,– and a commercial product, *Mixed-In-Key*,[8] which is considered the state-of-the-art key detection algorithm among EDM producers.

5 Discussion

From the results shown in Table 1, we can observe that *Essentia*'s baseline algorithm scores much lower than any other result presented. The other two profiles (*edma, edmm*) were derived statistically from the *Shaath* dataset [15], what can be noted by observing that they perform better in 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. What becomes apparent from the results in Table 1, is that algorithms tailored for this specific style seem to outperform general purpose algorithms such as *Essentia*'s key extractor, something that could be indicative of the variety of distinct tonal practices across musical genres, and the need to acknowledge this difference in the design of algorithms.

The comparison with commercial applications in Table **??**, shows that *Mixed in Key* provides the best performance in all datasets, followed, also in all scenarios, by our solution. The application by Sha'ath[9]. shows greater variability across the different datasets, suggesting that there is a certain amount of overfitting when evaluated with his own dataset.

6 Conclusions

In this paper, we presented a modification of an existing algorithm for key detection in EDM. Its main contribution is a multi-profile system that detects difficult minor tracks, which contrary to musicological intuition present more energy on the major third than in the minor. With this approach, we have reduced the bias toward the minor mode present in an earlier method and improved the performance on three datasets, getting closer to state-of-the-art commercial software.

Naturally, there is room to improve the detection process and we have some preliminary evidence that a combined approach of this method with deep learning techniques could boost the performance of the algorithm, as evinced in recent publications [24, 25].

7 Acknowledgments

Special thanks to Eduard Mas for his contribution to the manual annotation process of 1500 excerpts and

[8] www.mixedinkey.com

[9] Results of *Key Finder* on the *GiantSteps* dataset differ from those shown in [16, 15] This is to be attributed to an update in Sha'ath's software between the experiments

	GiantSteps				***Beatport***				***Shaath***			
	essnt	*edma*	*edmm*	*bgate*	*essnt*	*edma*	*edmm*	*bgate*	*essnt*	*edma*	*edmm*	*bgate*
correct	.305	.581	**.642**	.641	.271	.517	.525	**.637**	.302	.598	**.701**	.663
fifth	.175	.101	.108	.094	.285	.116	.117	107	.262	.112	.105	.102
relative	.111	.066	.033	.088	.053	.058	.071	.051	.061	.035	.017	.063
parallel	.114	.106	.068	.051	.119	.219	.169	.141	.080	.118	.045	.041
other	.295	.146	.149	.126	.273	.090	.118	.065	.296	.137	.132	.132
mirex	*.448*	*.673*	*.720*	*.725*	*.453*	*.636*	*.638*	*.734*	*.467*	*.688*	*.767*	*.741*

Table 1: Typical errors and MIREX scores with various key profiles and for different evaluation collections: *Essentia*'s key extractor (*essnt*, *edma* and *edmm* [15] and the *bgate* multi-profile.

	GiantSteps			***Beatport***			***Shaath***		
	KF	*MIK*	*bgate*	*KF*	*MIK*	*bgate*	*KF*	*MIK*	*bgate*
correct	.604	**.672**	.641	.548	**.657**	.637	.674	**.720**	.663
fifth	.127	.093	.094	.155	.100	.107	.128	.094	.102
relative	.066	.056	.088	.060	.061	.051	.019	.021	.063
parallel	.056	.053	.051	.150	.118	.141	.041	.045	.041
other	.146	.126	.126	.087	.064	.064	.138	.121	.132
mirex	*.699*	*.742*	*.725*	*.673*	*.749*	*.741*	*.751*	*.782*	*.741*

Table 2: Typical errors and MIREX scores of our method (*bgate*) along two well-known software applications: *KeyFinder* (KF) and *Mixed in Key 7* (MIK).

to Sergio Latre for his valuable help annotating a sub-collection with difficult tracks.

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

References

[1] Doehring, A., "Andrés's 'New For U': New for Us. On Analysing Electronic Dance Music," in R. von Appen, A. Doehring, D. Hemls, and A. F. Moore, editors, *Song interpretation in 21st-Century Pop Music*, Ashgate Publishing, 2015.

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

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

[4] Rocha, B., Bogaards, N., and Honingh, A., "Segmentation and Timbre- and Rhythm-Similarity in Electronic Dance Music," in *Proceedings of the Sound and Music Computing Conference*, pp. 754–761, 2013.

[5] Panteli, M., Rocha, B., Bogaards, N., and Honingh, A., "Development of a Rhythm Similarity Model for Electronic Dance Music," in *Proceedings of the 53rd International Audio Engineering Society Conference: Semantic Audio*, pp. 537–542, 2014.

[6] Eigenfeldt, A. and Pasquier, P., "Evolving Structures for Electronic Dance Music," in *Proceeding of the 15th Genetic and Evolutionary Computation Conference*, 2013.

[7] Yadati, K., Larson, M., Liem, C. C. S., and Hanjalic, A., "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*, pp. 143–148, 2014.

[8] Scarfe, T., Koolen, W., and Kalnishkan, Y., "Segmentation of electronic dance music," Technical report, Royal Holloway, University of London, Surrey, 2014.

[9] Glazyrin, N., "Towards Automatic Content-Based Separation of Dj Mixes Into Single Tracks," in *Proceedings of the 15th International Society for Music Information Retrieval Conference*, pp. 149–154, 2014.

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

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

[12] Hockman, J., Davies, M., and Fujinaga, I., "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*, pp. 169—–174, 2014.

[13] Hörschlager, F., Vogl, R., Böck, S., and Knees, P., "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.

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

[15] Faraldo, Á., Gómez, E., Jordà, S., and Herrera, P., "Key Estimation in Electronic Dance Music," in *Proceedings of the 38th European Conference on Information Retrieval*, pp. 335–347, 2016.

[16] Knees, P., Faraldo, Á., Herrera, P., Vogl, R., Böck, S., Hörschläger, F., and Le Goff, M., "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*, pp. 364–370, Málaga, 2015.

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

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

[19] Noland, K., *Computational Tonality Estimation: Signal Processing and Hidden Markov Models*, Ph.D. thesis, Queen Mary, University of London, London, 2009.

[20] Temperley, D., "What's Key for Key? The Krumhansl-Schmuckler Key-Finding Algorithm Reconsidered," *Music Perception: An Interdisciplinary Journal*, 17(1), pp. 65–100, 1999.

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

[22] Bogdanov, D., Wack, N., Gómez, E., Gulati, S., Herrera, P., Mayor, O., Roma, G., Salamon, J., Zapata, J., and Serra, X., "ESSENTIA: an Audio Analysis Library for Music Information Retrieval," in *Proceedings of the 14th International Society for Music Information Retrieval Conference*, pp. 493–498, 2013.

[23] Tagg, P., *Everyday Tonality II - Towards a Tonal Theory of What Most People Hear*, The Mass Media Music Scholar's Press, New York & Huddersfield, 2014.

[24] S. Dieleman, P. B. and Schrauwen, B., "Audio-based music classification with a pretrained convolutional network," in *Proceedings of the 12th International Society for Music Information Retrieval Conference*, 2011.

[25] Melendez, B., *Studying the Application of Deep Learning to the Task of Key Estimation*, Master's thesis, Universitat Pompeu Fabra, Barcelona, 2016.